

Kernel density estimation on symmetric spaces of non-compact type

Dena Marie Asta

Department of Statistics, The Ohio State University, 1958 Neil Ave, Columbus, OH 43210, USA

ARTICLE INFO

Article history:

Received 31 December 2019

Received in revised form 1 September 2020

Accepted 2 September 2020

Available online 9 September 2020

AMS 2010 subject classifications:

primary 62G07

secondary 43-04

Keywords:

Harmonic analysis

Helgason–Fourier transform

Kernel density estimator

Non-Euclidean geometry

Non-parametric

ABSTRACT

We construct a kernel density estimator on symmetric spaces of non-compact type and establish an upper bound for its convergence rate, analogous to the minimax rate for classical kernel density estimators on Euclidean space. Symmetric spaces of non-compact type include hyperboloids of constant curvature -1 and spaces of symmetric positive definite matrices. This paper obtains a simplified formula in the special case when the symmetric space is the space of normal distributions, a 2-dimensional hyperboloid.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction

Data, while often expressed as collections of real numbers, are often more naturally regarded as points in non-Euclidean spaces. To take an example, radar systems can yield the data of bearings for planes and other flying objects; those bearings are naturally regarded as points on a sphere [13]. To take another example, diffusion tensor imaging (DTI) can yield information about how liquid flows through a region of the body being imaged; that three-dimensional movement can be expressed in the form of symmetric positive definite (3×3) -matrices [13]. To take yet another example, the nodes of certain hierarchical real-world networks can be regarded as having latent coordinates in a hyperboloid [1,8]. In all such examples, the spaces can be regarded as subsets of Euclidean space even though Euclidean distances do not reflect true distances between points. An ordinary kernel density estimator (KDE) applied to sample data generally will not be optimal in terms of the L_2 -risk with respect to the volume measure on the non-Euclidean manifold.

The idea of kernel density estimation is to smooth out, or convolve, an empirical estimator (an average of Dirac distributions centered at the data) with a smooth rapidly decaying kernel so as to obtain a smooth estimate of the true density. The literature offers some variants of kernel density estimation on non-Euclidean spaces. A simple variant, for compact manifolds [11] or more general compact subsets of manifolds [2,3], applies a Euclidean kernel having supports small enough to fit inside the charts. A more general version, defined on complete manifolds, generalizes the kernel to be defined on the tangent bundle — in effect, requiring a kernel for each point [7]. Minimax rates of convergence have been proven for all of these different variants in terms of a Hölder class exponent [7,11]. It is desirable to refine these convergence rates based on Sobolev constraints on the true densities.

E-mail address: dasta@stat.osu.edu.

On symmetric spaces like Euclidean space, a kernel need only be defined at one point and transported everywhere else. All symmetric spaces X can be decomposed into symmetric spaces of Euclidean, compact, and noncompact type in such a way that a KDE on X can be constructed from KDEs on the three types. Symmetric spaces of the first two types admit KDEs with minimax convergence rates. The goal of this paper is to begin to complete the picture for symmetric spaces of noncompact type, by proving the upper bound part of a conjectured minimax rate.

Kernel density estimation for random variables can be interpreted as an estimate of a Fourier transform of the density. Thus for manifolds on which Fourier analysis generalizes, there should exist some generalization of the KDE. One of the earliest such Fourier-based generalizations of the KDE is defined for compact manifolds and shown to be minimax [5]. Certain density estimators on the non-compact Poincaré halfplane [6] and the space of symmetric positive definite matrices, based on Helgason–Fourier Analysis, have been shown to be minimax for estimation from corrupted samples. In the special case where the noise is non-existent, these estimators can be regarded as special cases of a KDE where the kernel is a natural generalization of the sinc kernel. However, these estimators have not been shown to be minimax for estimation from uncorrupted samples. Moreover, the full of generality of Helgason–Fourier Analysis is not exploited in defining and analyzing this estimator.

The Helgason–Fourier transform, unlike the ordinary Fourier transform, sends functions on a symmetric space of noncompact type to functions on a different frequency space. The exact form of this frequency space, much less a usable formula for the transform, depends on a geometric understanding the original space. Countless symmetric spaces of interest in applications have well-understood geometries. When the original space is the Poincaré halfplane, for example, the frequency space is a cylinder. The Helgason–Fourier transform, like the ordinary Fourier transform, is an isometry on L_2 -function spaces and sends convolutions to products in a certain sense.

This paper uses the Helgason–Fourier transform to construct and analyze a version of a KDE on symmetric spaces of noncompact type. We define a kernel density estimator for symmetric spaces of non-compact type \mathbf{X} , for which Helgason–Fourier transforms are defined. Unlike the non-Fourier-based variant [11] for compact manifolds, this variant is differentiable everywhere and estimates densities with non-compact support. The analogue of a kernel is often just a density on a space \mathbf{G} of isometries invariant with respect to the subgroup \mathbf{H} of \mathbf{G} for which $\mathbf{X} = \mathbf{G}/\mathbf{H}$. An example is a Gaussian, a solution to the \mathbf{H} -invariant heat equation. We bound risk in terms of bandwidth h , the number n of sample points, a Sobolev parameter α , and the sum of the restricted roots of \mathbf{X} [Theorem 1]. Optimizing h in terms of n , we obtain an upper bound of

$$n^{-2\alpha/(2\alpha+\dim \mathbf{X})},$$

for the convergence rate, where α is a Sobolev parameter, under natural assumptions on the density space [Theorem 1]. We then obtain a simplified formula, that can be implemented on a computer, for the special case where \mathbf{X} is the 2-dimensional hyperboloid of constant curvature -1 . The proof for the upper bound of the convergence rates adapts techniques used in [6]. We conjecture that the same upper bound yields a lower bound and hence a minimax rate, and a proof is reserved for future work.

2. Preliminaries

We recall some preliminary constructions and results in this section. First, we recall the construction of a KDE, including a Fourier-based interpretation. We then recall some of the theory of symmetric spaces. Finally, we recall the theory of Helgason–Fourier Analysis.

2.1. Kernel density estimation

The kernel density estimator (KDE) $f_{(X_1, \dots, X_n)}^{(h)} : \mathbb{R} \rightarrow \mathbb{R}$, defined by

$$f_{(X_1, X_2, \dots, X_n)}^{(h)}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right), \quad (1)$$

estimates a density f on \mathbb{R} based on some observed points X_1, \dots, X_n identically and independently sampled from f , a tunable bandwidth $h > 0$, and a fixed choice of kernel function K , which in most cases amounts to a symmetric unimodal Lebesgue density on \mathbb{R} having mode 0.

A generalization of the KDE for certain Riemannian manifolds X is

$$f_{(X_1, X_2, \dots, X_n)}^{(h)}(x) = \frac{1}{nh^{\dim X}} \sum_{i=1}^n \theta_{X_i}(x)^{-1} K\left(\frac{\text{dist}(x - X_i)}{h}\right), \quad (2)$$

where θ_p denotes the density of the volume measure on the Riemannian manifold [11]. For Euclidean space, θ_p is just the constant 1 function and K can be taken to be a general kernel function. For compact manifolds, K is taken to have support $[-1, +1]$ and bandwidth h is bounded by the injectivity radius of the manifold. In these cases, (2) integrates to 1, defines a density when K is non-negative, and converges to f at a minimax rate in a suitable sense [11].

One way to think about kernel density estimation is that it is an estimation of the characteristic function ϕ_X of a random variable X , defined by

$$\phi_X(t) = \mathbb{E}[e^{itX}]. \quad (3)$$

An empirical characteristic function $\hat{\phi}_{x_1, \dots, x_n}$, defined by

$$\hat{\phi}_{x_1, \dots, x_n}(t) = \frac{1}{n} \sum_{i=1}^n e^{itx_i}, \quad (4)$$

estimates ϕ_X based on some observed points x_1, \dots, x_n sampled from X . While (3) often admits a convergent inverse Fourier transform, (4) does not. Therefore while the density for X can often be recovered by taking the inverse Fourier transform of (3), an estimated characteristic function for X does not analogously give an estimated density for X . Instead, (4) needs to be dampened by a rapidly decaying function $\mathcal{F}[K_h]$, after which it is in $L_2(\mathbb{R})$ and therefore has a well-defined inverse Fourier transform. If $\mathcal{F}[K]$ is fixed and $\mathcal{F}[K_h](s)$ is set to be $\mathcal{F}[K](hs)$, one then recovers the original definition (1) of a KDE. What is more, Fourier Analysis then makes it then possible to prove that the standard KDE achieves a minimax rate for densities in a Sobolev ball. Thus different types of Fourier transforms give different variants of the standard KDE. In this way, an alternative to (2) for certain compact spaces achieves a minimax rate for densities in a Sobolev ball.

2.2. Symmetric spaces

This paper assumes the basic definition of a smooth manifold. A Riemannian manifold is a smooth manifold M equipped with a Riemannian metric, a choice of inner product $g_x : T_x M \otimes T_x M \rightarrow T_x M$ on the real tangent vector spaces $T_x M$ smoothly varying in $x \in M$. The formula $\int_t \langle \gamma'(t), \gamma'(t) \rangle dt$ for arc lengths of curves $\gamma : [0, 1] \rightarrow \mathbb{R}^n$ straightforwardly generalizes to a definition of arc lengths of curves, and hence of geodesics and distances between points, in Riemannian manifolds. An isometry between Riemannian manifolds is a smooth map of manifolds whose differential defines an isometry of tangent spaces. An involution at a point x in a Riemannian manifold M is an isometry $\varphi : M \cong M$ fixing x such that the induced linear isometry $(\partial\varphi)_x : T_x M \rightarrow T_x M$ is defined by scalar multiplication by -1 . A Riemannian symmetric space is a Riemannian manifold admitting an involution at each of its points. For background on the basic theory of Riemannian symmetric spaces, the reader is referred to [9].

If a Riemannian symmetric space X decomposes as a product $X = Y \times Z$ of such spaces, then KDEs on the factors Y and Z can be combined to give a KDE on the entire space. If a Riemannian symmetric space X is a quotient E/G of a Riemannian symmetric space E by the action of a discrete group G , then a KDE on X can be defined as the restriction of a KDE on E by identifying all but a Borel measure 0 subspace of X with a subspace of E . Therefore it suffices to restrict our attention to irreducible simply connected Riemannian symmetric spaces, Riemannian symmetric spaces that do not admit decompositions into a product of non-trivial Riemannian symmetric spaces and do not arise as quotients E/G of Riemannian symmetric spaces E by non-trivial discrete groups G . The irreducible simply connected Riemannian symmetric spaces fall into three types: Euclidean, compact, and non-compact (and non-Euclidean). A minimax KDE on Euclidean space is classical. A minimax KDE on compact Riemannian manifolds already exists. This paper focuses on the (irreducible, simply connected, and) noncompact case.

Such spaces admit the following algebraic characterization. Recall that a Lie group is a smooth manifold G that is at once a group in such a way that the multiplication $G \times G \rightarrow G$ and inversion $G \rightarrow G$ are smooth maps. Let \mathbf{G} be a noncompact semisimple Lie group, a noncompact connected Lie group with finite center containing no non-trivial normal connected Abelian subgroups. Then \mathbf{G} admits an Iwasawa decomposition $\mathbf{G} = \mathbf{HAN}$, a decomposition of \mathbf{G} as the space \mathbf{HAN} of all triple products of elements from a maximal compact Lie subgroup $\mathbf{H} \leq \mathbf{G}$, an Abelian Lie subgroup $\mathbf{A} \leq \mathbf{G}$, and nilpotent Lie subgroup $\mathbf{N} \leq \mathbf{G}$. (Since the symbol K will assume its traditional role in statistics as denoting a kernel, the symbol \mathbf{H} is used here to denote a maximal compact subgroup of \mathbf{G} .) Nilpotency means that \mathbf{N}_k is the trivial group for large enough $k \gg 0$, where $\mathbf{N}_0 = \mathbf{N}$ and \mathbf{N}_{i+1} is the subgroup of \mathbf{N} generated by all elements of the form $xyx^{-1}y^{-1}$ for $x \in \mathbf{N}_i$ and $y \in \mathbf{N}$. In practice, \mathbf{G} is a suitable group of matrices under matrix multiplication, \mathbf{A} is a group of diagonal matrices and \mathbf{N} is a group of upper triangular matrices.

Example 1. The group SL_2 has Iwasawa decomposition

$$\mathrm{SL}_2 = \mathrm{SO}_2 \mathbb{R}_+ \mathbb{R},$$

where the positive reals \mathbb{R}_+ (under the operation of multiplication) is identified with the diagonal (2×2) -matrices in SL_2 with positive entries and the reals \mathbb{R} (under the operation of addition) is identified with the upper triangular (2×2) -matrices with 1's along the diagonal.

Let \mathbf{X} be the smooth manifold defined as the quotient space

$$\mathbf{X} = \mathbf{G}/\mathbf{H} \cong \mathbf{AN}.$$

The space \mathbf{X} is a smooth manifold with smooth structure characterized by the property that precomposition with the natural function $\mathbf{G} \rightarrow \mathbf{G}/\mathbf{H}$ bijectively identifies smooth functions $\mathbf{G}/\mathbf{H} \rightarrow \mathbb{R}$ with smooth functions $\phi : \mathbf{G} \rightarrow \mathbb{R}$ which are

right- \mathbf{H} -invariant ($\phi(gh) = \phi(g)$ for all $h \in \mathbf{H}$). A choice of a bi- \mathbf{H} -invariant, left \mathbf{G} -invariant inner product on \mathfrak{g} passes to a well-defined \mathbf{G} -invariant inner product on a tangent space of \mathbf{X} , which in turn uniquely extends to a Riemannian metric turning \mathbf{X} into a Riemannian symmetric space. Conversely, every Riemannian symmetric space of noncompact type arises in this manner.

Example 2. Continuing Example 1, we can give an algebraic construction of the Poincaré halfplane \mathbb{H}_2 , the 2-manifold defined as the subspace

$$\mathbb{H}_2 = \{z \in \mathbb{C} \mid \operatorname{Im}(z) > 0\},$$

of \mathbb{C} equipped with the Riemannian metric given by the arc length

$$ds^2 = (\operatorname{Im} z)^{-2}(d(\operatorname{Re} z)^2 + d(\operatorname{Im} z)^2).$$

The matrices in \mathbb{SL}_2 act on \mathbb{H}_2 as Möbius transformations:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}(z) = \frac{az + b}{cz + d}.$$

The matrices in \mathbb{SL}_2 fixing $i \in \mathbb{H}_2$ form the matrix subgroup \mathbb{SO}_2 . The action of \mathbb{SL}_2 on \mathbb{H}_2 implicitly gives a well-defined bijection $\mathbb{SL}_2/\mathbb{SO}_2 \cong \mathbb{H}_2$ sending an equivalence class of a matrix $m \in \mathbb{SL}_2$ to $m(i) \in \mathbb{H}_2$. This bijection $\mathbb{H}_2 \cong \mathbb{SL}_2/\mathbb{SO}_2$ defines an isometry for a suitable choice of bi- \mathbb{SO}_2 -invariant inner product on the Lie algebra \mathfrak{sl}_2 associated to \mathbb{SL}_2 [14, §3.1]. Thus \mathbb{H}_2 is a Riemannian symmetric space. This space can be interpreted as the information manifold of all univariate normal distributions, where the real coordinates describe the means, and the imaginary coordinates describe standard deviations, and the Riemannian metric is the Fisher metric. Alternatively, this space is a natural latent space for families of random graphs used to model real-world networks [8]. Alternatively, this space models electrical impedances on which certain circuit elements act as Möbius transformations [6].

From now on, fix a semisimple Lie group \mathbf{G} having finite center with Iwasawa decomposition $\mathbf{G} = \mathbf{HAN}$ and let $\mathbf{X} = \mathbf{G}/\mathbf{H}$. Let \mathfrak{g} , \mathfrak{a} , \mathfrak{n} denote the Lie algebras of the respective Lie groups \mathbf{G} , \mathbf{A} , \mathbf{N} , their tangent spaces over identities. Recall that the exponential map $\exp : \mathfrak{a} \rightarrow \mathbf{A}$ is defined as sending a tangent vector v to $\gamma_v(1) \in \mathbf{A}$ for γ_v the unique continuous group homomorphism $\gamma : \mathbb{R} \rightarrow \mathbf{A}$ with $\gamma'(0) = v$. For each $x \in \mathbf{X}$, there exist unique elements $a(x) \in \mathfrak{a}$ and $n(x) \in \mathbf{N}$ such that $x = \exp(a(x))n(x)$.

Example 3. Continuing Examples 1 and 2, for each $z \in \mathbb{H}_2$,

$$n(z) = \begin{pmatrix} 1 & \operatorname{Re} z \\ 0 & 1 \end{pmatrix}, \quad \exp(a(z)) = \begin{pmatrix} \sqrt{\operatorname{Im} z} & 0 \\ 0 & (\sqrt{\operatorname{Im} z})^{-1} \end{pmatrix}.$$

Let $[-, -] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ denote the Lie bracket operation, defined in the case \mathbf{G} is a group of $(n \times n)$ -invertible matrices and thus \mathfrak{g} is a vector space of $(n \times n)$ -matrices, by $[v, w] = vw - wv$. Recall that the Killing form on \mathfrak{g} is the symmetric bilinear map $\kappa : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathbb{R}$ sending a pair (v, w) of tangent vectors to the trace of the operator $[x, [y, -]]$ on \mathfrak{g} . The Killing form on \mathfrak{g} restricts to an inner product on \mathfrak{a} . In this manner, \mathfrak{a}^* will sometimes be naturally identified with \mathfrak{a} along the adjoint of this inner product, and \mathfrak{a} will often be identified with the Lebesgue measure space $\mathbb{R}^{\dim \mathfrak{a}}$ along an isometry between the two. Let \mathbf{M} be the set of all elements in \mathbf{H} commuting with all elements in \mathbf{A} . Then let $\mathbf{B} = \mathbf{H}/\mathbf{M}$. For each $\lambda \in \mathfrak{a}^*$, let \mathfrak{g}_λ denote the space of all $g \in \mathfrak{g}$ with $\lambda(x)g = xg - gx$ for all $x \in \mathfrak{a}$. The restricted root system Λ of \mathbf{G} is the set of all $\lambda \in \mathfrak{a}^*$ with $\dim \mathfrak{g}_\lambda > 0$. A positive restricted root system is a choice of subset $\Lambda_+ \subset \Lambda$ such that $\Lambda = \Lambda_+ \cup -\Lambda_+$ and $\Lambda_+ \cap -\Lambda_+ = \emptyset$. Set

$$\rho_{\mathbf{X}} = \#\{k\mathbf{M} \mid k\mathbf{A} = \mathbf{A}k, k \in \mathbf{H}\}, \quad \rho_{\mathbf{X}} = \frac{1}{2} \sum_{\lambda \in \Lambda_+} \dim \mathfrak{g}_\lambda.$$

The root systems Λ play a pivotal role in the classification of symmetric spaces. Note that the choice of Λ_+ , and hence also of $\rho_{\mathbf{X}}$, are only unique up to sign.

Example 4. For $\mathbb{H}_2 = \mathbb{SL}_2/\mathbb{SO}_2$, we have the following [14, §3.1]:

$$\rho_{\mathbb{H}_2} = 1/2, \quad w_{\mathbb{H}_2} = 1.$$

2.3. Helgason-Fourier analysis

Helgason-Fourier Analysis is an analogue of Fourier Analysis for symmetric spaces of noncompact type. The reader is referred to [12, Section 2] for a concise summary of the theory and [14] for details in the special case $\mathbf{X} = \mathbb{H}_2$. The Helgason-Fourier transform of $f \in L_1(\mathbf{X}, dx)$ at $\lambda \in \mathfrak{a}^*$ and $b = k\mathbf{M} \in \mathbf{B} = \mathbf{H}/\mathbf{M}$, when it exists, is

$$(\mathcal{H}f)(\lambda, b) = \int_{x=g\mathbf{H} \in \mathbf{X}=\mathbf{G}/\mathbf{H}} f(x) e^{-s(a(k^{-1}g))} dx,$$

where $s = \rho_{\mathbf{X}} - i\lambda$ lies in the complexification of \mathfrak{a}^* , and dx is the intrinsic volume measure on \mathbf{X} . The Helgason–Fourier transform defines a linear map

$$\mathcal{H} : L_2(\mathbf{X}, dx) \rightarrow L_2(\mathfrak{a}^* \times \mathbf{B}, |c|^{-2} d\lambda db).$$

such that we have the following Plancharel identity

$$\int_{\mathbf{X}} |f(x)|^2 dx = w_{\mathbf{X}}^{-1} \int_{\lambda \in \mathfrak{a}^*} \int_{b \in \mathbf{B}} |\mathcal{H}f(\lambda, b)|^2 |c(\lambda)|^{-2} d\lambda db. \quad (5)$$

Here $d\lambda$ is the Lebesgue measure under the natural identification $\mathfrak{a}^* \cong \mathbb{R}^r$, db is a suitably normalized K -invariant measure, and c is the Harish-Chandra c -function $\mathfrak{a}^* \rightarrow \mathbb{R}$ [4]; a proper treatment of the latter, while fundamental throughout geometric analysis, is beyond the scope of the paper. However, the following exact formula is used in the special case $\mathbf{X} = \mathbb{H}_2$.

Example 5. Continuing Examples 2 and 1,

$$c(\lambda) = \frac{1}{8\pi^2} \lambda \tanh(\pi\lambda).$$

The basic property of the Harish-Chandra c -function needed in the proofs is the following growth bound, noted in [10].

Lemma 1. For each $\lambda \in \mathfrak{a}^*$, $|c(\lambda)|^{-2} \leq (1 + |\lambda|)^{\dim \mathfrak{n}}$.

The Helgason–Fourier transform \mathcal{H} sends convolutions to products:

$$\mathcal{H}[f_{GX}] = \mathcal{H}[f_G] \mathcal{H}[f_X]$$

for each bi- \mathbf{H} -invariant density $f_G \in L_2(\mathbf{G})$ of a random quantity G on \mathbf{G} , identified with the induced left- \mathbf{H} -invariant density on \mathbf{X} , and density f_X of a random quantity X on \mathbf{X} , where f_{GX} is the density associated to GX .

The Sobolev ball $\mathcal{F}_\alpha(Q)$ in $L_2(\mathbf{X})$ can be defined as

$$\mathcal{F}_\alpha(Q) = \{f \in L_2(\mathbf{X}) \mid \|\Delta^{\alpha/2} f\|^2 \leq Q\},$$

just as in the Euclidean case, except that here $\Delta^{\alpha/2} f$ is defined by

$$\mathcal{H}[\Delta^{\alpha/2} f](\lambda, b) = (-\rho_{\mathbf{X}}^2 - \lambda^2)^{\alpha/2} \mathcal{H}[f_X](\lambda, b)$$

The inverse Helgason–Fourier transform, written \mathcal{H}^{-1} , is given by

$$\mathcal{H}^{-1} \psi(x) = w_{\mathbf{X}}^{-1} \int_{\mathfrak{a}^*} \int_{\mathbf{B}} \psi(\lambda, k\mathbf{M}) e^{(\rho_{\mathbf{X}} - i\lambda)(a(k^{-1}x))} |c|^{-2} d\lambda db,$$

We can recover $f \in C_c^\infty(\mathbf{X})$ from its transform by the inversion formula

$$f = \mathcal{H}^{-1} \mathcal{H} f.$$

2.4. \mathbf{G} -kernel density estimation

In an analogy with the classical definition, define the characteristic function

$$\phi_X : \mathfrak{a}^* \times \mathbf{B} \rightarrow \mathbb{C}$$

of a random quantity $g\mathbf{H}$ on the symmetric space $\mathbf{X} = \mathbf{G}/\mathbf{H}$ by

$$\phi_X(s, k\mathbf{M}) = \mathbb{E}[e^{\tilde{s}(-a(k^{-1}g))}].$$

An empirical characteristic function $\hat{\phi}_{x_1, \dots, x_n} : \mathfrak{a}^* \rightarrow \mathbb{C}$, defined by

$$\hat{\phi}_{g_1\mathbf{H}, \dots, g_n\mathbf{H}}(s, k\mathbf{M}) = \frac{1}{n} \sum_{i=1}^n e^{\tilde{s}(-a(k^{-1}g_i))}. \quad (6)$$

estimates ϕ_X based on some observed points $g_1\mathbf{H}, \dots, g_n\mathbf{H}$ sampled from a random quantity on \mathbf{X} . This function needs to be dampened by a kernel before taking \mathcal{H}^{-1} . A \mathbf{G} -kernel K on \mathbf{X} is a unimodal left- \mathbf{H} -invariant L_2 -function K on \mathbf{X} which integrates to 1.

A non-density example is an analogue K_{sinc} of the sinc kernel defined by

$$\hat{K}_{\text{sinc}}(s, b) \propto \begin{cases} 1 & |s| < 1, \\ 0 & |s| > 1. \end{cases}$$

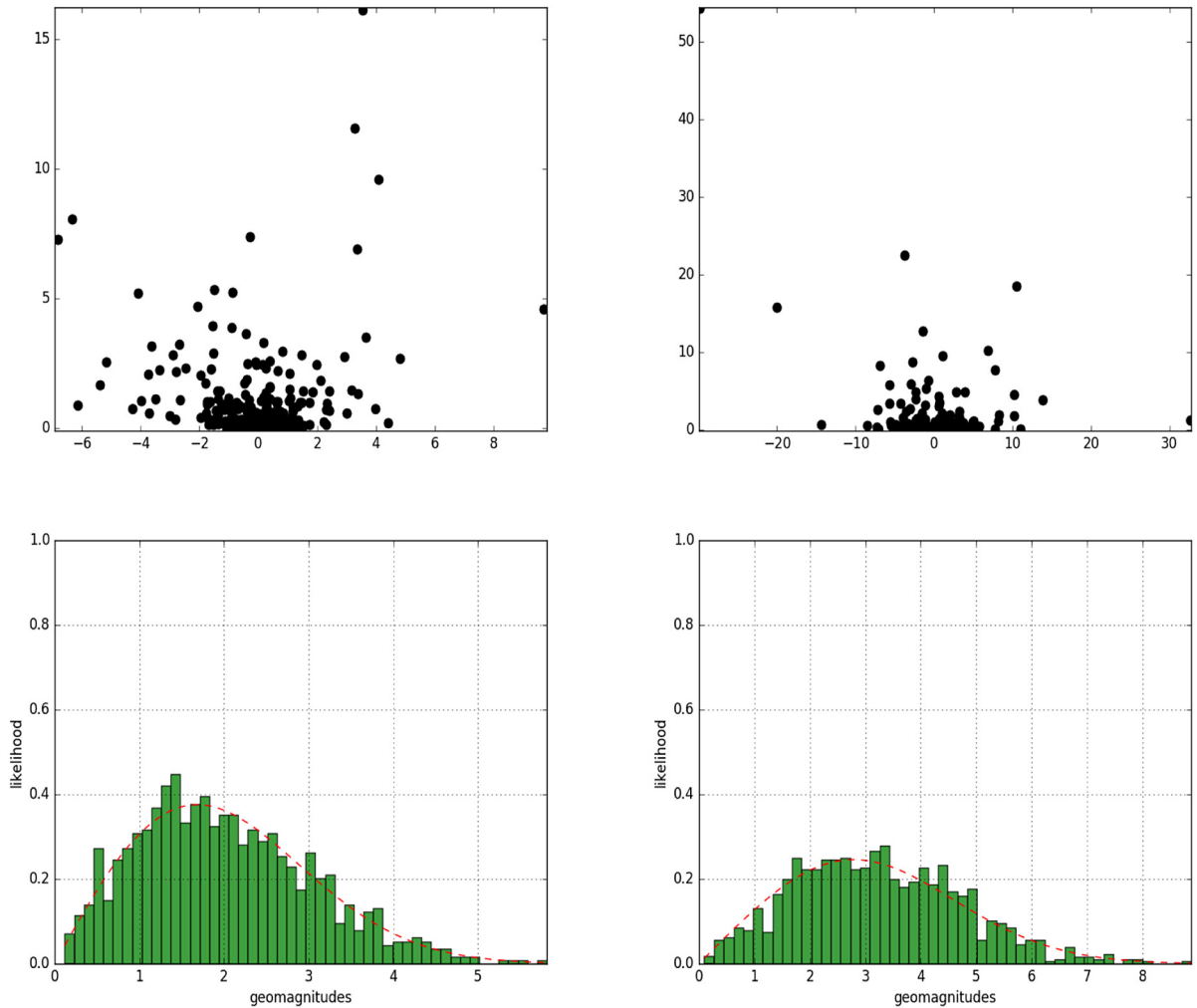


Fig. 1. Sampling 300 points from a Gaussian f_ζ on \mathbb{H}_2 , characterized by $\mathcal{H}[f_\zeta] = 2\pi e^{s(s-2\zeta)}$, with parameter $\zeta = 1$ (left) and $\zeta = 2$ (right). Sampling is implemented via Monte Carlo integration. The top row illustrates the covariates in \mathbb{H}_2 . The bottom row illustrates the associated geopotential magnitudes of the covariates.

A density example is a Gaussian K_{Gauss} defined by

$$\hat{K}_{\text{Gauss}}(s, b) \propto e^{s(s-2\zeta)},$$

for each choice of real parameter ζ , so named as it is a **H**-invariant solution to the heat equation (Fig. 1).

A **G**-kernel on **X** is (β, γ) -smooth if there exist $C_2, C_3 > 0$ such that

$$C_2 e^{-\frac{|s|^\beta}{\gamma}} \leq |\mathcal{H}K(s, b)| \leq C_3 e^{-\frac{|s|^\beta}{\gamma}}.$$

for all $s = \rho_{\mathbf{X}} + i\lambda$ and $b \in \mathbf{B}$. Examples include the previous two **G**-kernels.

For each **G**-kernel K on **X** and a bandwidth parameter $h > 0$, define K_h by

$$(\mathcal{H}K_h)(\lambda, b) = (\mathcal{H}K)(h\lambda, b).$$

The **G**-kernel density estimator $f_{x_1, \dots, x_n}^{(h)} : \mathbf{X} \rightarrow \mathbb{R}$ is

$$\hat{f}_{x_1, \dots, x_n}^{(h)} = \mathcal{H}^{-1} \left[\hat{\phi}_{x_1, \dots, x_n} \mathcal{H}[K_h] \right],$$

for each choice of **G**-kernel K , bandwidth $h > 0$, and $x_1, \dots, x_n \in \mathbf{X}$.

2.5. Main theorem

A proof of the following main result can be found in Section 4.

Theorem 1. For density $f_X \in \mathcal{F}_\alpha(Q)$ and (β, γ) -smooth kernel K on \mathbf{X} ,

$$\mathbb{E} \|f_{X_1, \dots, X_n}^{(h)} - f_X\|^2 \leq C_1 Q \rho_X^{2\alpha} h^{2\alpha} + K_1 Q T^{-2\alpha} + K_2 n^{-1} e^{-2(h\rho_X)^\beta / \gamma} T^{\dim \mathbf{X}}, \quad X_1, \dots, X_n \sim_{iid} f_X$$

where $C_1, K_1, K_2 > 0$ are constants not dependent on α, Q, n , for each choice of $T > 0$.

By choosing a smooth enough kernel K and optimal bandwidth h , we obtain the following upper bound.

Corollary 1. For density $f_X \in \mathcal{F}_\alpha(Q)$ and (β, γ) -smooth kernel K on \mathbf{X} ,

$$\mathbb{E} \|f_{X_1, \dots, X_n}^{(h)} - f_X\|^2 \leq C n^{-2\alpha/(2\alpha + \dim \mathbf{X})}, \quad X_1, \dots, X_n \sim_{iid} f_X,$$

for some constant $C > 0$ not dependent on α, Q, n and $h \in \mathcal{O}(n^{-1/(2\alpha + \dim \mathbf{X})})$.

3. Implementation

For $\mathbf{X} = \mathbb{H}_2$ and \mathbf{H} the Gaussian with $\zeta = 1$,

$$f_{Z_1, \dots, Z_n}^{(h)}(z) = \sum_{i=1}^n \int_{-\infty}^{+\infty} \int_0^{2\pi} \text{Im}(k_\theta(Z_i))^{\frac{1}{2}-i\lambda} e^{-(\frac{h^2}{4} + h^2 \lambda^2)} (\text{Im}(k_\theta(z)))^{\frac{1}{2}+i\lambda} d\mu,$$

where $d\mu = \frac{1}{8\pi^2 n} (\lambda \tanh(\pi\lambda)) d\theta d\lambda$ and k_θ denotes the rotation matrix associated to the angle $\theta \in [0, 2\pi)$. This double integral, computationally cumbersome, is currently the simplest form known for expressing the KDE – the usual simplifications that allow us to regard a KDE as a convolution of a kernel with an average of Dirac point-masses do not work for general \mathbf{X} . We leave for future work the task of optimizing the numerical approximation of the KDE, likely using analogues of discrete Fourier coefficients to construct discrete optimizations of the inverse Helgason–Fourier transform [12].

For the present paper, we simulate 1000 covariates from a Gaussian f_ζ with dispersion parameter $\zeta = 1$ on \mathbb{H}_2 and numerically compare KDEs defined on the first $n = 100, 200, \dots, 1000$ covariates with one another and with the true density. Every \mathbb{SO}_2 -invariant density is determined by its marginal on geopolar magnitudes. Given that f is \mathbb{SO}_2 -invariant, we simplify our computational task by only comparing the associated marginals on geopolar magnitudes. The associated marginal for f_ζ is

$$2\pi f_\zeta(e^{-r}i) = 2\pi(4\pi\zeta)^{-3/2} \sqrt{2} e^{-\zeta/4} \int_r^\infty \frac{b e^{-b^2/(4\zeta)}}{\sqrt{\cosh(b) - \cosh(r)}} db.$$

The associated marginal for the KDE $\hat{f}_{X_1, \dots, X_n}^h$ is given by

$$\hat{f}_{X_1, \dots, X_n}^h(e^{-r}i) = \sum_{i=1}^n \int_{-\infty}^{+\infty} (\text{Im} Z_i)^{\frac{1}{2}-i\lambda} e^{-(\frac{1}{4} + h^2 \lambda^2)} P_{-1/2+i\lambda}(\cosh r) d\mu.$$

where $d\mu = \frac{1}{8\pi^2 n} (\lambda \tanh(\pi\lambda)) d\theta d\lambda$ and $P_a(c)$ is the Legendre function defined by the integral

$$P_a(c) = \frac{1}{2\pi} \int_0^{2\pi} (c + \cos \theta \sqrt{c^2 - 1})^a d\theta.$$

Bandwidth selection is made by the rule-of-thumb

$$h = n^{-\alpha/(\alpha + \dim \mathbf{X})}$$

instead of a data-driven criterion like cross-validation, due to the computational burden of the formula for the KDE. An empirical calculation shows that the rule-of-thumb bandwidth closely tracks the ISE minimizer

$$h^* = \arg \min_h \|f - \hat{f}_{X_1, \dots, X_n}^h\|_2.$$

We plot some KDEs at $n = 300, 500, 700$ against the marginal of the true density f_ζ as functions on \mathbb{H}_2 (Fig. 2) and as associated Helgason–Fourier transforms on $\mathbb{R} \times \mathbb{SO}_2$ (Fig. 2). We then compare the rates at which the integrated square errors for the marginals of the KDEs and a Euclidean KDE on sample geopolar magnitudes decrease. Our bandwidth-selection criterion for the Euclidean KDE is Silverman's rule-of-thumb. We plot the Integrated Squared Error (ISE) for the marginal of the KDE and the Euclidean KDE on the space of geopolar magnitudes (Fig. 3).

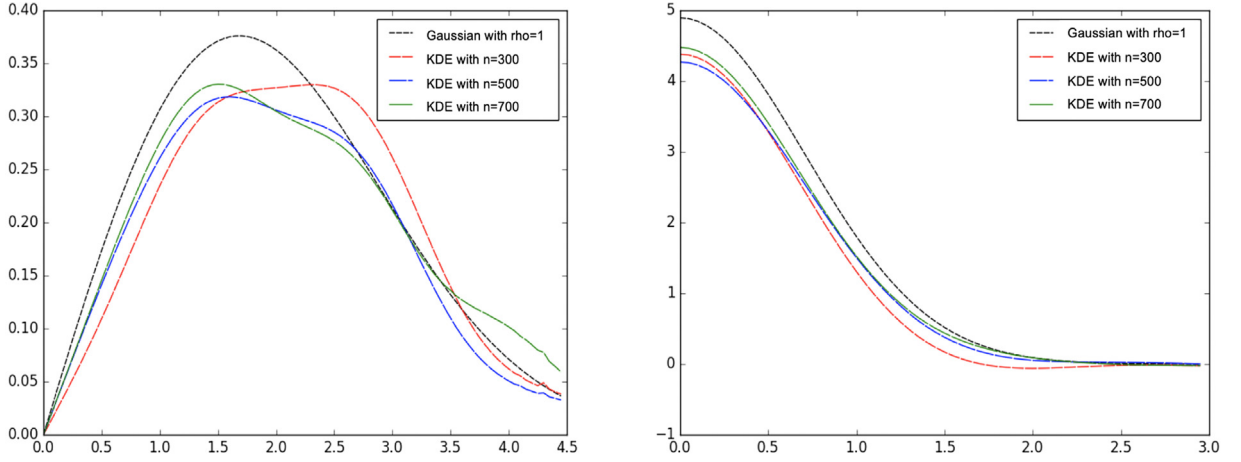


Fig. 2. Comparing KDEs. The true density of a Gaussian on \mathbb{H}_2 with dispersion parameter $\zeta = 1$ is compared with KDEs, whose bandwidths are chosen by the rule-of-thumb derived in the proof of the minimax rate. Illustrated above are true and estimated densities with respect to both Lebesgue measure (left) and the measure induced from the volume measure (right), on geopolar magnitudes conditioned on geopolar angles being 0. The densities on the right are used to calculate risk, which can be seen to be decreasing in the number of sample points.

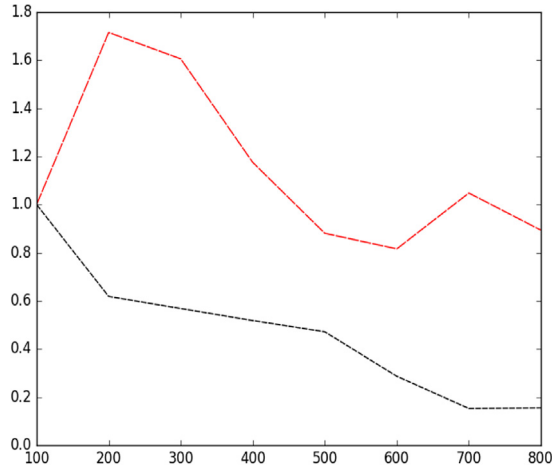


Fig. 3. Classical versus Non-classical KDEs. The integrated squared error (ISE) with respect to the volume measure on \mathbb{H}_2 is computed for a KDE on \mathbb{H}_2 (black) versus a Euclidean KDE (red) for $n = 300, 500, 700$ (x-axis) and normalized so that the ISEs for $n = 300$ are both 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Proofs

We use the mean integrated squared error to measure the performance of our generalized estimator. As previously noted, proofs here adapt techniques developed in [6]. We break the mean integrated squared error into two parts, variance and squared bias, and bound each part separately. Throughout, let s denote $\rho_X + i\lambda$, T denote a real number, and τ denote the spectral measure $|c(\lambda)|^{-2}d\lambda db$.

Proof of Main Theorem. Let $V = \mathbb{E} \left\| f_X^{(n,h)} - \mathbb{E} \left[f_X^{(n,h)} \right] \right\|^2$. To obtain the bound on the variance V , note that for each T ,

$$\begin{aligned} V &= \mathbb{E} \int_{\mathbf{X}} \left| f_X^{(n,h)} - \mathbb{E} \left[f_X^{(n,h)} \right] \right|^2 dx = w_{\mathbf{X}}^{-1} \mathbb{E} \int_{\mathbf{a}^*} \int_{\mathbf{B}} \left| \mathcal{H} \left[f_X^{(n,h)} \right] - \mathbb{E} \left[\mathcal{H} f_X^{(n,h)} \right] \right|^2 d\tau \\ &= w_{\mathbf{X}}^{-1} \mathbb{E} \int_{\mathbf{a}^*} \int_{\mathbf{B}} \left| \hat{\phi}_{X_1, \dots, X_n} \mathcal{H} K_h - \mathcal{H} f_X \mathcal{H} K_h \right|^2 d\tau = w_{\mathbf{X}}^{-1} \mathbb{E} \int_{\mathbf{a}^*} \int_{\mathbf{B}} |\mathcal{H} K_h|^2 \left| \hat{\phi}_{X_1, \dots, X_n} - \mathcal{H} f_X \right|^2 d\tau \\ &= w_{\mathbf{X}}^{-1} \int_{\mathbf{a}^*} \int_{\mathbf{B}} |\mathcal{H} K_h|^2 \mathbb{E} \left[|\hat{\phi}_{X_1, \dots, X_n}|^2 + |\mathcal{H} f_X|^2 - 2|\hat{\phi}_{X_1, \dots, X_n} \mathcal{H} f_X| \right] d\tau \end{aligned}$$

Let $I = \mathbb{E} \left[|\hat{\phi}_{X_1, \dots, X_n}|^2 + |\mathcal{H}f_X|^2 - 2|\hat{\phi}||\mathcal{H}f_X| \right]$. Observe:

$$\begin{aligned} I &= \left[|\mathcal{H}f_X|^2 + n^{-1} \left(|\mathbb{E}[e^{2\tilde{s}(a(X, b))}]| - |\mathcal{H}f_X|^2 \right) + |\mathcal{H}f_X|^2 - 2\mathcal{H}f_X \mathbb{E}[\hat{\phi}] \right] \\ &\leq \left[|\mathcal{H}f_X|^2 + n^{-1} \left(|\mathbb{E}[e^{2\rho_X(a(X, b))}]| - |\mathcal{H}f_X|^2 \right) + |\mathcal{H}f_X|^2 - 2\mathcal{H}f_X \mathbb{E}[\hat{\phi}] \right] \\ &= \left[2|\mathcal{H}f_X|^2 + n^{-1} \left(|\mathcal{H}f_X(2\rho_X, b)| - |\mathcal{H}f_X|^2 \right) - 2|\mathcal{H}f_X|^2 \right] = n^{-1} \left(|\mathcal{H}f_X(2\rho_X, b)| - |\mathcal{H}f_X|^2 \right) \\ &\leq n^{-1} |\mathcal{H}f_X(2\rho_X, b)|. \end{aligned}$$

It then follows that

$$\begin{aligned} V &\leq n^{-1} w_X^{-1} \int_{a^*} \int_{\mathbf{B}} |\mathcal{H}K_h|^2 |\mathcal{H}f_X(2\rho_X, b)| \, d\tau \\ &\leq n^{-1} w_X^{-1} \left(\int_{a^*} |\mathcal{H}K_h|^2 |c(\lambda)|^{-2} \, d\lambda \right) \left(\int_{\mathbf{B}} |\mathcal{H}f_X(2\rho_X, b)| \, db \right). \end{aligned}$$

The last term is an element inside

$$\mathcal{O} \left(\frac{1}{n} \left(\int_{|\lambda| < T} |\mathcal{H}K_h|^2 |c(\lambda)|^{-2} \, d\lambda + \int_{|\lambda| > T} |\mathcal{H}K_h|^2 |c(\lambda)|^{-2} \, d\lambda \right) \right),$$

therefore an element inside

$$\mathcal{O} \left(\frac{1}{n} \left(e^{-2(h\rho_X)^\beta/\gamma} (1+T)^{\dim a} T^{\dim a} + \int_{|\lambda| > T} e^{-2h^\beta |\lambda|^\beta/\gamma} (1+|\lambda|)^{\dim n} \, d\lambda \right) \right),$$

and hence an element inside

$$\mathcal{O} \left(\frac{1}{n} e^{-2(h\rho_X)^\beta/\gamma} T^{\dim \mathbf{X}} \right).$$

Let $B = \left\| \mathbb{E} f_{X_1, \dots, X_n}^{(h)} - f_X \right\|$. To obtain the bound on the squared bias B^2 , note that for each T ,

$$\begin{aligned} B^2 &= \int_{\mathbf{X}} \left| \mathbb{E} f_{X_1, \dots, X_n}^{(h)} - f_X \right|^2 \, dx = w_X^{-1} \int_{a^*} \int_{\mathbf{B}} \left| \mathbb{E} \mathcal{H}f_{X_1, \dots, X_n}^{(h)} - \mathcal{H}f_X \right|^2 \, d\tau = w_X^{-1} \int_{a^*} \int_{\mathbf{B}} |\mathcal{H}f_X \mathcal{H}K_h - \mathcal{H}f_X|^2 \, d\tau \\ &= w_X^{-1} \int_{a^*} \int_{\mathbf{B}} |\mathcal{H}f_X|^2 |\mathcal{H}K_h - 1|^2 \, d\tau = w_X^{-1} \int_{a^*} \int_{\mathbf{B}} \overline{s(s - 2\rho_X)}^{-\alpha} \overline{s(s - 2\rho_X)}^\alpha |\mathcal{H}f_X|^2 |\mathcal{H}K_h - 1|^2 \, d\tau \\ &\leq w_X^{-1} \int_{|\lambda'| < hT} \int_{\mathbf{B}} h^{2\alpha-1} (-(\lambda')^2 - \rho_X^2)^{-\alpha} |\mathcal{H}K_h - 1|^2 \, d\tau' + \mathcal{O} \left(\sup_{|\lambda| > T} |s|^{-2\alpha} |\mathcal{H}K_h - 1|^2 \right) \\ &\subset \mathcal{O} (C_1 h^{2\alpha} + K_1 T^{-2\alpha}) \end{aligned}$$

for constants $C_1, K_1 > 0$, where $\lambda' = h\lambda$ and $\tau' = h\tau$ above.

Proof of the Corollary. By [Theorem 1](#), for each $T > 0$

$$\mathbb{E} \|f_X^{(n, h)} - f_X\|^2 \leq C_1 h^{2\alpha} + K_1 T^{-2\alpha} + K_2 n^{-1} e^{-2(h\rho_X)^\beta/\gamma} T^{\dim \mathbf{X}}.$$

Setting $T^{-2\alpha} \propto n^{-1} e^{-2(h\rho_X)^\beta/\gamma} T^{\dim \mathbf{X}}$, we obtain

$$T \propto n^{1/(2\alpha + \dim \mathbf{X})}.$$

The upper bound converges at the fastest possible rate when

$$h^{2\alpha} \propto T^{-2\alpha} \propto n^{-2\alpha/(2\alpha + \dim \mathbf{X})},$$

or equivalently, when $h \propto n^{-\alpha/(\alpha + \dim \mathbf{X})}$. Thus

$$\mathbb{E} \|f_X^{(n, h)} - f_X\|^2 \in \mathcal{O}(n^{-2\alpha/(2\alpha + \dim \mathbf{X})}).$$

5. Conclusion

Until now, kernel density estimation in the non-compact setting has required either that one restricts to Euclidean space, requires that the densities have compact support, or requires that one specifies a kernel-like function for each point in a complete Riemannian manifold. We have introduced a new density estimator on a large class of non-compact symmetric spaces, sidestepping these various restrictions, and have proven an upper bound for the rate of convergence, conjecturally a minimax rate, identical to the minimax rate of convergence for a Euclidean kernel density estimator.

We then specialized our generalized kernel density estimator for the hyperboloid, motivated by applications to network inference. Future work will explore adaptivity, computational optimizations, and applications to symmetric spaces other than the hyperboloid.

Acknowledgments

This work was partially supported by an NSF Graduate Research Fellowship, USA under grant DGE-1252522. Also, the author is grateful to her advisor, Cosma Shalizi, for his invaluable guidance and feedback throughout this research. The author is also grateful to the anonymous referee for helpful comments, corrections, and suggestions.

References

- [1] D.M. Asta, C.R. Shalizi, Geometric network comparisons, in: M. Meila, T. Heskes (Eds.), *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence, UAI 2015, July 12–16, 2015, Amsterdam, The Netherlands (2015)* 102–110.
- [2] E. Chevallier, T. Forget, F. Barbaresco, J. Angulo, Kernel density estimation on the Siegel space with an application to radar processing, *Entropy* 18 (11) (2016) 347–364.
- [3] E. Chevallier, E. Kalunga, J. Angulo, Kernel density estimation on spaces of Gaussian distributions and symmetric positive definite matrices, *SIAM J. Imaging Sci.* 10 (1) (2017) 191–215.
- [4] S. Helgason, Harish-Chandra's c -function. A mathematical jewel, in: *Noncompact Lie Groups and Some of their Applications* (San Antonio, TX, 1993), NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci. 429 (1994) 55–67.
- [5] H. Hendriks, Nonparametric estimation of a probability density on a Riemannian manifold using Fourier expansions, *Ann. Statist.* 18 (2) (1990) 832–849.
- [6] S.F. Huckemann, P.T. Kim, J.-Y. Koo, A. Munk, Möbius deconvolution on the hyperbolic plane with application to impedance density estimation, *Ann. Statist.* 38 (4) (2010) 2465–2498.
- [7] Y.T. Kim, H.S. Park, Geometric structures arising from kernel density estimation on Riemannian manifolds, *J. Multivariate Anal.* 114 (2013) 112–126.
- [8] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, M. Boguñá, Hyperbolic geometry of complex networks, *Phys. Rev. E* (3) 82 (3) (2010) 036106, 18.
- [9] O. Loos, *Symmetric Spaces. I: General Theory*, W. A. Benjamin, Inc., New York-Amsterdam, 1969, p. viii+198.
- [10] E.K. Narayanan, S.K. Ray, The heat kernel and hardy's theorem on symmetric spaces of noncompact type, *Proc. Indian Acad. Sci. Math. Sci.* 112 (2) (2002) 321–330.
- [11] B. Pelletier, Kernel density estimation on Riemannian manifolds, *Statist. Probab. Lett.* 73 (3) (2005) 297–304.
- [12] I. Pesenson, A discrete helgason-fourier transform for sobolev and besov functions on noncompact symmetric spaces, in: *Radon Transforms, Geometry, and Wavelets*, vol. 464, *Contemp. Math.* (2008) 231–247.
- [13] I.U. Rahman, I. Drori, V.C. Stodden, D.L. Donoho, P. Schröder, Multiscale representations for manifold-valued data, *Multiscale Model. Simul.* 4 (4) (2005) 1201–1232.
- [14] A. Terras, *Harmonic analysis on symmetric spaces—Euclidean space, the sphere, and the Poincaré upper half-plane*, second ed., Springer, New York, 2013, p. xviii+413.