



HAL
open science

Invariant kernels on Riemannian symmetric spaces: a harmonic-analytic approach

Nathaël da Costa, Cyrus Mostajeran, Juan-Pablo Ortega, Salem Said

► **To cite this version:**

Nathaël da Costa, Cyrus Mostajeran, Juan-Pablo Ortega, Salem Said. Invariant kernels on Riemannian symmetric spaces: a harmonic-analytic approach. 2024. hal-04270515v2

HAL Id: hal-04270515

<https://hal.science/hal-04270515v2>

Preprint submitted on 9 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Invariant kernels on Riemannian symmetric spaces : a harmonic-analytic approach

Nathael Da Costa*, Cyrus Mostajeran*, Juan-Pablo Ortega*, and Salem Said†

Abstract. This work aims to prove that the classical Gaussian kernel, when defined on a non-Euclidean symmetric space, is never positive-definite for any choice of parameter. To achieve this goal, the paper develops new geometric and analytical arguments. These provide a rigorous characterization of the positive-definiteness of the Gaussian kernel, which is complete but for a limited number of scenarios in low dimensions that are treated by numerical computations. Chief among these results are the L^p -Godement theorems (where $p = 1, 2$), which provide verifiable necessary and sufficient conditions for a kernel defined on a symmetric space of non-compact type to be positive-definite. A celebrated theorem, sometimes called the Bochner-Godement theorem, already gives such conditions and is far more general in its scope but is especially hard to apply. Beyond the connection with the Gaussian kernel, the new results in this work lay out a blueprint for the study of invariant kernels on symmetric spaces, bringing forth specific harmonic analysis tools that suggest many future applications.

Key words. Riemannian manifold, symmetric space, harmonic analysis, Bochner's theorem, Godement theorem, positive-definite kernel

MSC codes. 43A35, 43A85, 43A90, 46E22, 53C35, 53Z50

1. Introduction. Positive-definite functions play a fundamental role in harmonic analysis and the theory of group representations, as well as in probability and statistical inference [30, 36, 53, 54]. In machine learning, these functions are of paramount importance in the context of kernel-based methods. Indeed, kernels are most often required to be positive-definite for Reproducing Kernel Hilbert Space (RKHS) methods and associated linear geometry and classification algorithms to apply [7, 48]. While many positive-definite kernels are known in Euclidean spaces, this is not yet true for non-Euclidean metric spaces, such as non-trivial Riemannian manifolds. Even when such kernels are known, they are far from easy to work with (in the first place, to evaluate). For example, just one evaluation of the heat kernel on a Riemannian symmetric space may require an elaborate Monte Carlo scheme [1, 2].

This being the case, one hopes to develop a tractable means of constructing closed-form positive-definite kernels on Riemannian manifolds or at least to develop an explicit criterion to verify whether a given closed-form expression yields a positive-definite kernel or not. This should apply, in particular, to the so-called distance kernels, which are of the form $k(x, y) = g(d(x, y))$, with g a suitable function and $d(x, y)$ the Riemannian distance.

The results developed in this work serve exactly this purpose. They are motivated by the study of the Gaussian kernel

$$(1.1) \quad k(x, y) = \exp(-\lambda d^2(x, y)) \quad \text{where } \lambda > 0.$$

In particular, they aim to show that this kernel is never positive-definite (meaning it is not

*Division of Mathematical Sciences, School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, 637371, Singapore

†Laboratoire Jean Kuntzman, Université Grenoble-Alpes, Grenoble, 38400, France

38 positive-definite for any value of λ) when defined on a non-Euclidean Riemannian symmetric
39 space M . In general, when M is a non-Euclidean Riemannian manifold, it is known that
40 there always exists some λ such that the Gaussian kernel is not positive-definite [34, 24, 23].
41 However, whether there exists any subset of values of the parameter λ for which the Gaussian
42 kernel is positive-definite is an open problem of theoretical and practical significance as noted
43 in the literature [41]. For instance, [50] shows that the Gaussian kernel may be positive-
44 definite for some non-trivial collection of parameters in certain non-Euclidean geometries.
45 Recently, it has been shown that if M is compact and not simply connected (*e.g.* a torus or
46 a real projective space), then the Gaussian kernel is never positive-definite [37]. While these
47 general results rely on somewhat complicated geometric arguments, they will be considerably
48 extended here, in the case of symmetric spaces, thanks to the introduction of new analytical
49 tools.

50 A typical result of choice in this circle of ideas is Bochner’s theorem, one of the most
51 famous theorems in harmonic analysis [25]. It states that a function defined on a locally
52 compact Abelian group is positive-definite if and only if it is the inverse Fourier transform
53 of some positive measure. For example, this theorem guarantees the existence of a spectral
54 power measure for any wide-sense stationary signal (Wiener-Khinchin theorem).

55 The generalization of Bochner’s theorem to symmetric spaces is due to Godement [27]. It
56 will be referred to as Godement’s theorem in this work, although it is sometimes called the
57 Bochner-Godement theorem elsewhere. This theorem is recalled in Section 2, where we explain
58 why it can be challenging to apply in practice. Focusing on non-compact symmetric spaces,
59 the L^2 -Godement theorem (Section 3) and the L^1 -Godement theorem (Appendix A), both
60 state that in the presence of additional integrability assumptions, the difficulties inherent in
61 its application are greatly alleviated: it turns out that a function is positive-definite if and only
62 if its spherical transform is positive and integrable (the definition of the spherical transform
63 is recalled in Section 3).

64 Basic examples of applications of the L^2 -Godement theorem are provided in Section 4.
65 Sections 5 and 6 then return to the Gaussian kernel (1.1). Section 5 proves that the Gaussian
66 kernel is never positive-definite on a compact symmetric space, expanding on [19] and [37],
67 which prove the corresponding result on the circle and non-simply connected closed Riemann-
68 ian manifolds, respectively. Section 6 takes up the case of non-compact symmetric spaces
69 (under the general assumptions of Sections 2 and 3, compact or non-compact have the same
70 meaning as “of compact or non-compact type” — no Euclidean spaces are allowed).

71 Section 6 provides a rigorous proof that deals with the case of non-compact symmetric
72 spaces in all but a few limited scenarios in low dimensions, which are investigated compu-
73 tationally instead. First, in the case of the hyperbolic plane, a numerical evaluation shows
74 that the positivity condition in the L^2 -Godement theorem does not hold, for any λ within
75 a certain continuous range, starting near $\lambda = 0$. However (Lemma 6.5), any non-compact
76 symmetric space contains an isometrically embedded hyperbolic plane, and this implies the
77 Gaussian kernel fails to be positive-definite on any non-compact symmetric space, for that
78 same range of values of λ . Section 6 also proves (with no recourse to numerical work) that
79 the Gaussian kernel is never positive-definite when defined on a three-dimensional hyperbolic
80 space. Again, this extends to any symmetric space that contains an isometrically embed-
81 ded hyperbolic space (*e.g.* the spaces of $n \times n$ real, complex, or quaternion, positive-definite

82 matrices, with the requirement that $n \geq 4$ for the real case).

83 Section 7 introduces an alternative to the Gaussian kernel (1.1) for hyperbolic spaces,
84 giving it the name of hyperbolic secant or Herschel-Maxwell kernel. Once more, this is a
85 distance kernel

$$86 \quad (1.2) \quad k(x, y) = (\cosh(d(x, y)))^{-a} \quad \text{where } a > 0.$$

87 This kernel is derived from first principles as a different generalization of the Gaussian kernel
88 on hyperbolic spaces, similar to the famous Herschel-Maxwell derivation [35]. In contrast with
89 the Gaussian kernel, it is shown to always be positive-definite on the hyperbolic plane and
90 positive-definite on the hyperbolic space for integer values of a .

91 The L^p -Godement theorems (where $p = 1, 2$) are most easily applied to distance kernels
92 (for example, to the Gaussian or Herschel-Maxwell kernels). A detailed examination of their
93 application to more general invariant kernels, which are not merely functions of distance, is
94 one of the main tasks to be carried out in ongoing and future work (see in particular [42]).

95 The present work focuses on symmetric spaces and does not discuss other, more general
96 classes of Riemannian manifolds. This restriction is rather essential if one wishes to pursue an
97 approach based on harmonic analysis, but it is still a restriction (Stiefel manifolds are relevant
98 to many applications, but are not symmetric spaces in general [13]). Fortunately, symmetric
99 spaces include various spaces of covariance matrices, in addition to Grassmannians, compact
100 Lie groups, and other kinds of manifolds, which arise across brain-computer interface analysis,
101 radar signal processing, computer vision, quantum information, *etc.* [17, 11, 6, 16]. Moreover,
102 symmetric spaces are also quite useful for graph-embedding applications [38]. Therefore, the
103 ongoing efforts directed at constructing positive-definite (as in [1, 2, 10, 42]) or positively
104 decomposable ([20]) kernels on symmetric spaces should lead to the emergence of new kernel-
105 based methods (for classification, hypothesis testing, *etc.*) applicable to the above-mentioned
106 fields.

107 The present work deals solely with kernels and does not make any mention of Gaussian
108 processes (Gaussian processes are considered in detail in [1, 2, 10]). This should come as no
109 surprise since the present approach is purely analytical rather than probabilistic. However, this
110 is not a limitation of the present work as positive-definite kernels and Gaussian processes are
111 equivalent. Given a Gaussian process, its covariance is a positive-definite kernel. The reverse
112 construction can be obtained by Kolmogorov's extension theorem [55]. Let us finally note that
113 the connection between invariant Gaussian processes and harmonic analysis is quite strong,
114 since these processes are typically obtained by solving invariant stochastic partial differential
115 equations [10, 43].

116 There is an additional direction where methods related to Godement's theorem can be
117 applied. The general setting of this theorem is that of Gelfand pairs, which goes beyond
118 symmetric spaces. In particular, Gelfand pairs include certain discrete spaces, such as spaces
119 of phylogenetic trees, very important in bio-informatics, linguistics, *etc.* [12]. A tentative
120 direction for future work is to use Godement's theorem to improve understanding of kernel
121 methods for phylogenetic trees.

122 **2. Godement's theorem.** Let M be a set and let $k : M \times M \rightarrow \mathbb{C}$. Assume that k satisfies
 123

$$124 \quad (2.1) \quad \sum_{i,j=1}^N k(x_i, x_j) c_i \bar{c}_j \geq 0$$

125 for all $x_i \in M$, $c_i \in \mathbb{C}$ and all $N \in \mathbb{N}$, where the bar denotes complex conjugation. Then, k is
 126 said to be a positive-definite kernel on M . Without loss of generality, it is enough to consider
 127 Hermitian kernels, that is, $k(x, y)$ is the conjugate of $k(y, x)$. In addition, assume that a group
 128 G acts transitively on M ; this action is denoted by $g \cdot x$, for $g \in G$ and $x \in M$. If k satisfies

$$129 \quad (2.2) \quad k(g \cdot x, g \cdot y) = k(x, y),$$

130 for all $g \in G$ and $x, y \in M$, then k is said to be a G -invariant (or just an invariant) positive-
 131 definite kernel. Let $o \in M$ and denote by H the stabiliser of o in G . A function $f : M \rightarrow \mathbb{C}$
 132 called H -invariant if $f(h \cdot x) = f(x)$ for all $h \in H$ and $x \in M$. Using such a function, we can
 133 define $k_f : M \times M \rightarrow \mathbb{C}$,

$$134 \quad (2.3) \quad k_f(x, y) = f(g_1^{-1} g_2 \cdot o) \text{ for } x = g_1 \cdot o \text{ and } y = g_2 \cdot o.$$

135 Indeed, H -invariance of f implies the right-hand side of (2.3) depends only on x and y (and not
 136 on the choice of g_1 and g_2). Now, k_f is G -invariant (it satisfies (2.2)). If k_f happens to be
 137 positive-definite (to satisfy (2.1)), then f is said to be a positive-definite function.

138 All invariant positive-definite kernels can be represented in this way, since (2.1) and (2.2)
 139 imply that $k = k_f$, where $f(x) = k(o, x)$ and f is H -invariant and positive-definite. In
 140 particular, if $k = k_f$, then k is positive-definite if and only if f is positive-definite.

141 Note that the setup that we just presented implies that the set M is necessarily isomorphic
 142 to the homogeneous manifold G/H , where G is the group that acts transitively on M and H
 143 the isotropy group of some element $o \in M$. Conversely, given any group G and an arbitrary
 144 subgroup H , we can cast the homogeneous space G/H in this framework by setting $M = G/H$,
 145 $o = eH$, and considering the transitive left action of G on G/H . Given these observations,
 146 in the sequel, M will be a Riemannian symmetric space and (G, H) a symmetric pair. G is
 147 assumed to be semisimple, of either compact or non-compact type, and with finite center [32].
 148 Godement's theorem generalizes Bochner's classical theorem to this setting [27]. Roughly
 149 speaking, it states that an H -invariant continuous function f is positive-definite if and only if
 150 it is a positive combination of positive-definite spherical functions. In [8], Godement's theorem
 151 is called the Bochner-Godement theorem.

152 Recall that a spherical function $\varphi : M \rightarrow \mathbb{C}$ is an H -invariant function, such that $\varphi(o) = 1$
 153 and φ is moreover a joint eigenfunction of all the G -invariant differential operators on M [30].
 154 The set Φ of positive-definite spherical functions on M is locally compact for the topology
 155 of uniform convergence on compact subsets of M [27]. Godement's theorem states that an
 156 H -invariant continuous f is positive-definite if and only if

$$157 \quad (2.4) \quad f(x) = \int_{\Phi} \varphi(x) \mu_f(d\varphi)$$

158 where μ_f is a unique finite positive measure on Φ .

159 It is complicated to use Godement's theorem to verify whether a given f is positive-
 160 definite. When M is non-compact, the set Φ may be challenging to determine [30] (Page 484),
 161 and there is no straightforward way of determining μ_f from f . The following section identifies
 162 a special case of Godement's theorem where these issues do not arise.

163 **3. L^2 -Godement theorem.** The idea is to combine Godement's theorem with the so-called
 164 spherical transform. Roughly speaking, the spherical transform can be seen as a radial Fourier
 165 transform for H -invariant functions [51, 52]. When $M = \mathbb{R}^2$, it boils down to a zero-order
 166 Hankel transform.

167 Assume that M is a non-compact symmetric space. Pick an Iwasawa decomposition of
 168 the Lie algebra \mathfrak{g} of G , say $\mathfrak{g} = \mathfrak{n} + \mathfrak{a} + \mathfrak{h}$ (where \mathfrak{n} is nilpotent, \mathfrak{a} is Abelian, and \mathfrak{h} is the Lie
 169 algebra of H [32]). For each $g \in G$, $g = n(g)\exp(a(g))h(g)$, where $n(g) \in N$, $a(g) \in \mathfrak{a}$ and
 170 $h(g) \in H$ are unique (N is the Lie subgroup of G with Lie algebra \mathfrak{n}).

171 For each $t \in \mathfrak{a}^*$ (the dual of \mathfrak{a} as a real vector space), there is a spherical function
 172 $\varphi_t : M \rightarrow \mathbb{C}$, given by the following integral formula [30],

$$173 \quad (3.1) \quad \varphi_t(g \cdot o) = \int_H \exp[(it + \rho)(a(hg))] dh \quad \text{for } g \in G.$$

174 Here, ρ is half the sum of the positive roots of \mathfrak{g} with respect to \mathfrak{a} , and dh denotes the
 175 normalized Haar measure on H (a gentle introduction to these concepts is provided in [51, 52]).
 176 Define the spherical transform of an H -invariant $f : M \rightarrow \mathbb{C}$ as $\hat{f} : \mathfrak{a}^* \rightarrow \mathbb{C}$,

$$177 \quad (3.2) \quad \hat{f}(t) = \int_M f(x)\varphi_{-t}(x)\text{vol}(dx)$$

178 where vol denotes the volume measure of M . This is well-defined when f is continuous and
 179 compactly supported, in which case it admits an inversion formula,

$$180 \quad (3.3) \quad f(x) = c_M \int_{\mathfrak{a}^*} \hat{f}(t)\varphi_t(x)|c(t)|^{-2} dt$$

181 where $c(t)$ is a certain function, called the Harish-Chandra function (for an overview, see [31]),
 182 dt denotes the Lebesgue measure, and c_M is a constant that is independent of the function f .

183 Defined as above, the spherical transform extends to the space $L^2_H(M)$ of H -invariant
 184 square-integrable functions (with respect to vol) and satisfies the Plancherel formula

$$185 \quad (3.4) \quad \int_M |f(x)|^2 \text{vol}(dx) = c_M \int_{\mathfrak{a}^*} |\hat{f}(t)|^2 |c(t)|^{-2} dt$$

186 for any $f \in L^2_H(M)$. Here, the following proposition will be called the L^2 -Godement theorem.
 187 It is a special case of Godement's theorem, which holds for square-integrable functions.

188 **Theorem 3.1 (L^2 -Godement Theorem).** *A continuous function $f \in L^2_H(M)$ is positive-
 189 definite if and only if $\hat{f}(t) \geq 0$ for almost all $t \in \mathfrak{a}^*$ and \hat{f} is integrable with respect to the
 190 measure $|c(t)|^{-2} dt$.*

191 The connection with Godement's theorem, from the previous Section 2, comes from the fact
 192 that all of the functions φ_t in (3.1) are positive-definite [30] (Page 484). The set of φ_t where
 193 $t \in \mathfrak{a}^*$ is therefore a subset of Φ (in fact, it is a proper subset). The proof of Theorem 3.1 is
 194 given in Section 8.1, where a second alternative proof is also discussed.

195 The following section spells out the spherical transform pair (3.2)–(3.3) in two basic cases,
 196 where M is a hyperbolic plane or a hyperbolic space of dimension equal to three, and provides
 197 two examples of applications of Theorem 3.1.

198 4. Basic examples.

199 **4.1. Hyperbolic plane.** M is a two-dimensional simply connected space with constant
 200 negative curvature equal to -1 . One may take $G = SL(2, \mathbb{R})$ and $H = SO(2)$ (in the usual
 201 notation, $SL(2, \mathbb{R}) = 2 \times 2$ real unit-determinant matrices, $SO(2) = 2 \times 2$ rotation matrices).
 202 A detailed exposition may be found in [51]. For any H -invariant function $f : M \rightarrow \mathbb{C}$ there
 203 exists $g : \mathbb{R}_+ \rightarrow \mathbb{C}$, with $f(x) = g(r)$ where $r = d(o, x)$, the distance in M between x and
 204 the “point of origin” o (the choice of this point is arbitrary). In particular, the spherical
 205 functions (3.1) are given in terms of Legendre functions [30, 51],

$$206 \quad (4.1) \quad \varphi_t(x) = P_{-\frac{1}{2}+it}(\cosh(r))$$

207 for $t \in \mathbb{R}$ (with regard to (3.1), \mathfrak{a} is one-dimensional, and therefore $\mathfrak{a} \simeq \mathfrak{a}^* \simeq \mathbb{R}$. Accordingly,
 208 $t \in \mathfrak{a}^*$ becomes $t \in \mathbb{R}$), and it should be noted that $\varphi_t = \varphi_{-t}$ (see [51] (Page 141)).

209 The spherical transform pair (3.2)–(3.3) can be expressed as follows (see [51] (Page 149))

$$210 \quad (4.2) \quad \hat{f}(t) = 2\pi \int_0^\infty g(r) P_{-\frac{1}{2}+it}(\cosh(r)) \sinh(r) dr,$$

$$211 \quad (4.3) \quad g(r) = \frac{1}{2\pi} \int_0^\infty \hat{f}(t) P_{-\frac{1}{2}+it}(\cosh(r)) \alpha(t) dt,$$

212 where $f(x) = g(r)$ as above, and $\alpha(t) = t \tanh(\pi t)$. As pointed out in [51], (4.2)–(4.3) is
 213 known as a Mehler-Fock transform pair in the mathematical physics literature. Note that
 214 (4.2) follows directly from (3.2) by using (4.1) and then expressing the volume integral in
 215 geodesic spherical coordinates, which leads to the appearance of $\sinh(r)$ under the integral
 216 (roughly, $\text{vol}(dx) = \sinh(r) dr d\theta$ where θ is a polar angle, which yields the pre-factor 2π).

217 **4.2. Hyperbolic space.** M is now of dimension 3 instead of 2. One may take $G = SL(2, \mathbb{C})$
 218 and $H = SU(2)$ (in other words, $SL(2, \mathbb{C}) = 2 \times 2$ complex unit determinant matrices, and
 219 $SU(2) =$ unitary matrices in $SL(2, \mathbb{C})$). As before, any H -invariant function f is of the form
 220 $f(x) = g(r)$. The spherical functions (3.1) are given by [30] (Page 432)

$$221 \quad (4.4) \quad \varphi_t(x) = \frac{\sin(tr)}{t \sinh(r)}.$$

222 Again, $\varphi_t = \varphi_{-t}$. The spherical transform pair (3.2)–(3.3) is essentially a sine transform pair,

$$223 \quad (4.5) \quad \hat{f}(t) = \frac{4\pi}{t} \int_0^\infty g(r) \sinh(r) \sin(tr) dr,$$

$$224 \quad (4.6) \quad g(r) \sinh(r) = \frac{1}{2\pi^2} \int_0^\infty t \hat{f}(t) \sin(tr) dt,$$

225 as can be seen using (4.4) and computing the integral (3.2) in geodesic spherical coordinates.

226 **4.3. Distance kernels.** For the hyperbolic plane and space that we studied in the previous
 227 examples, any H -invariant function f is of the form $f(x) = g(d(o, x))$. It follows from (2.3)
 228 that $k_f(x, y) = g(d(x, y))$. But, as stated in Section 2, any invariant kernel k is of the form
 229 $k = k_f$ and it is hence a distance kernel

$$230 \quad (4.7) \quad k(x, y) = g(d(x, y)),$$

231 for some $g : \mathbb{R}_+ \rightarrow \mathbb{C}$. The positive-definiteness of k implies that g is necessarily real-valued.
 232 In other words, any positive-definite kernel is a real-valued distance kernel. This statement
 233 holds whenever M is a rank-one symmetric space. This class of symmetric spaces includes
 234 spheres, hyperbolic spaces, projective and hyperbolic projective spaces of any dimension [32].

235 Intuitively, the rank of a symmetric space M is the (minimum) number of geometric
 236 invariants required to determine a pair of points (x, y) up to isometry. For a rank-one space,
 237 there is exactly one such invariant, the distance $d(x, y)$. Thus, if k is an invariant kernel,
 238 $k(x, y)$ depends only on $d(x, y)$.

4.3.1. Wishart kernel. Let M be the hyperbolic plane, and let $g(r) = \exp(-2a \cosh(r))$,
 where $a > 0$, in (4.7). It is well known that M can be identified with the space of 2×2
 unit-determinant positive-definite matrices [51]. Thinking of $x \in M$ as a matrix of this kind,
 and choosing o the identity matrix, $d(o, x) = r$ implies the trace of x is $\text{tr}(x) = 2 \cosh(r)$ [51]
 (Page 149. Exercice 20). Thus, by (4.7)

$$k(o, x) = \exp(-a \text{tr}(x)).$$

Recalling the action $g \cdot x = gxg^\dagger$ (a matrix product, with \dagger the transpose [51]) and $o = \text{identity}$,
 and setting $x = g_1 \cdot o$ and $y = g_2 \cdot o$, it follows from (2.2) that

$$k(x, y) = k(o, g_1^{-1} g_2 \cdot o) = \exp(-a \text{tr}(x^{-1} y)).$$

239 This is the unnormalized density of a Wishart distribution, and k is the Wishart kernel. Now,
 240 the L^2 -Godement theorem can be used to show that this kernel is not positive-definite, for any
 241 value of a . The spherical transform $\hat{f}(t)$ is given in [51] (Page 151),

$$242 \quad (4.8) \quad g(r) = \exp(-2a \cosh(r)) \text{ in (4.2)} \implies \hat{f}(t) = \left(\frac{4\pi}{a}\right)^{\frac{1}{2}} K_{it}(2a)$$

243 where K_{it} is a modified Bessel function of the second kind with imaginary order. For fixed
 244 a , $K_{it}(2a)$ has infinitely many simple zeros in t [3]. In particular, $\hat{f}(t)$ does not satisfy the
 245 positivity condition in the L^2 -Godement theorem (Theorem 3.1), so k is not positive-definite
 246 (the theorem can be applied because $f(x) = k(o, x)$ is continuous and in $L^2_H(M)$ for any a).

247 Furthermore, since the Wishart kernel is never positive-definite on the hyperbolic plane,
 248 and since a hyperbolic space contains infinitely many isometric copies of the hyperbolic plane,
 249 the Wishart kernel is never positive-definite on the hyperbolic space. This statement can be
 250 made rigorous using the embedding lemma below (Lemma 5.2).

4.3.2. Hyperbolic secant kernel. Let M be the hyperbolic plane and $g(r) = (\cosh(r))^{-a}$ where $a > 1/2$. Then, $f(x) = g(d(o, x))$ is clearly continuous, and $f \in L^2_H(M)$, because

$$\int_M |f(x)|^2 \text{vol}(dx) = 2\pi \int_0^\infty (\cosh(r))^{-2a} \sinh(r) dr = 2\pi \int_1^\infty u^{-2a} du < \infty$$

251 as follows by integrating in geodesic spherical coordinates. From [51] (Page 151)

$$252 \quad (4.9) \quad g(r) = (\cosh(r))^{-a} \text{ in (4.2)} \implies \hat{f}(t) = 2^{a-1} \frac{\sqrt{\pi}}{\Gamma(a)} |\Gamma((a - 1/2 + it)/2)|^2$$

which is positive for all t (Γ is the Gamma function). Now, by the L^2 -Godement theorem, the hyperbolic secant kernel

$$k(x, y) = (\cosh(d(x, y)))^{-a}$$

253 is positive-definite on the hyperbolic plane for any $a > 1/2$. This example of a positive-definite
254 kernel on the hyperbolic plane does not seem to be known in the literature. We believe it is
255 reasonable to name it the hyperbolic secant kernel.

256 The case where M is the hyperbolic space, rather than the hyperbolic plane, is investigated
257 in Appendix B. Positive-definiteness seems to be, once more, the general rule.

258 **5. Compact symmetric spaces.** This section considers the case in which M is a compact
259 symmetric space (for example, a sphere, a projective space, or a Grassmannian). The following
260 proposition will now be proved.

261 **Proposition 5.1.** *Let M be a compact symmetric space. The Gaussian kernel (1.1) is not*
262 *positive-definite on M , for any value of λ .*

263 In other words, the Gaussian kernel is never positive-definite on a compact symmetric space.
264 The case where M is not simply connected (real projective space or real Grassmannian, etc.)
265 is already taken care of in [37]. The case where M is simply connected will now be proved.
266 First, we prove the following result, referred to as the Embedding Lemma.

267 **Lemma 5.2 (Embedding Lemma).** *Assume M_o and M are metric spaces, with M_o isomet-*
268 *rically embedded in M . Let k be a distance kernel (i.e. a kernel of the form (4.7)). If k is*
269 *positive-definite on M , then it is positive-definite on M_o .*

270 Here, isometric embedding means that the distance function in M_o is the restriction of the
271 distance function in M (this is the isometric embedding of metric spaces, a stronger property
272 than the isometric embedding of Riemannian manifolds). Due to (4.7), this implies the kernel
273 k on M_o is just the restriction of the same kernel k from M . The fact that positive-definiteness
274 is hereditary from M to M_o is immediate from (2.1).

275 A corollary of the embedding lemma is that the Gaussian kernel (1.1) cannot be positive-
276 definite on M if it is not positive-definite on M_o .

277 **Lemma 5.3.** *Any compact simply connected symmetric space M contains an isometrically*
278 *embedded circle σ (the embedding being an isometric embedding of metric spaces).*

279 *Proof.* Assume that M is an irreducible symmetric space [32, Chapter VIII]. Denote the
280 maximum of the sectional curvatures of M by κ^2 . Then, M admits a simple periodic geodesic

281 σ of length $2\pi/\kappa$ [32] (Page 334). Let x be a point on σ and choose a parameterisation
 282 $\sigma : [0, 2\pi] \rightarrow M$ with $\sigma(0) = x$. Then, let $\sigma(\pi) = x'$.

283 It can be shown that σ is length-minimising from x to x' (when restricted to $[0, \pi]$) and
 284 also from x' to x (when restricted to $[\pi, 2\pi]$). Recall that the cut-locus of x in M is identical
 285 to its first conjugate locus [18]. By definition of κ , x has no conjugate points along σ before
 286 $\sigma(\pi) = x'$ [14, Morse-Schönberg Theorem, Page 86]. Thus σ is length-minimising from x to
 287 x' . An identical argument shows σ is also length-minimising from x' to x .

288 Denote $\sigma = \sigma([0, 2\pi])$ (the image of the geodesic curve σ in M). This is the isometrically
 289 embedded circle announced in Lemma 5.3. To see this, choose any $y \in \sigma$, $y = \sigma(t)$. If
 290 $t \in [0, \pi]$ the distance in M between x and y is $d(x, y) = t$. On the other hand, if $t \in [\pi, 2\pi]$,
 291 $d(x, y) = 2\pi - t$ (note both of these claims result from the length-minimising property of σ ,
 292 just obtained). In other words, the restriction of the distance function from M to σ is the same
 293 as the distance function of a circle (identified with the interval $[0, 2\pi]$ with glued endpoints).

294 The assumption that M is irreducible can now be removed since any simply connected
 295 symmetric space is a direct product of irreducible symmetric spaces [32, Page 381]. ■

296 We recall now (see [19]) that the Gaussian kernel is never positive-definite on the circle.
 297 With this in mind, Lemmas 5.2 and 5.3 yield Proposition 5.1.

298 Even though it should be possible to prove Proposition 5.1 using an analytic approach
 299 via Godement's theorem, the geometric approach used above is more elementary and easier
 300 to visualise.

301 **6. Non-compact symmetric spaces.** In this section, M will be a non-compact symmetric
 302 space, as in Section 3 (for example, a hyperbolic space, a space of positive-definite matrices,
 303 *etc.*). Using a mixture of numerical and analytical results, we show that the Gaussian kernel
 304 systematically fails to be positive-definite on M . First, the L^2 -Godement theorem is applied
 305 to the two basic cases where M is a hyperbolic plane or a three-dimensional hyperbolic space.
 306 Specifically, from (1.1), the Gaussian kernel corresponds to $g(r) = e^{-\lambda r^2}$, and upon replacing
 307 this into (4.2) and (4.5), the following observations are made:

- 308 • In both cases, the spherical transform $\hat{f}(t)$ is positive when t is small in comparison
 309 to λ (Proposition 6.1, and the remark after Formula (6.1)).
- 310 • In (4.2), a numerical evaluation shows that $\hat{f}(t)$ oscillates between positive and negative
 311 values, for any λ within a certain continuous range, starting near $\lambda = 0$ (Figure 1).
- 312 • In (4.5), $\hat{f}(t)$ has the same sign as $\sin(t/2\lambda)$. The Gaussian kernel is, therefore, never
 313 positive-definite on the hyperbolic space (Proposition 6.3).

314 Second, the general case uses the Embedding Lemma 5.2.

315 **6.1. Hyperbolic plane.** For the Gaussian kernel on the hyperbolic plane, the behavior of
 316 the spherical transform $\hat{f}(t)$, when t is small in comparison to λ , is the following.

317 **Proposition 6.1.** *Let $g(r) = e^{-\lambda r^2}$ in (4.2). For any $T > 0$, there exists $\lambda_T > 0$ such that*
 318 *$\lambda > \lambda_T$ implies $\hat{f}(t) \geq 0$ for all $t \leq T$.*

319 The proof of Proposition 6.1 is given in Section 8.2. The behavior of $\hat{f}(t)$ for larger values of
 320 t was investigated numerically, with the results reported in Figure 1. Note that, at the time
 321 of writing, we were not able to provide a full proof or counterexample to the negativity of the
 322 function $\hat{f}(t_\lambda)$ at some t_λ for all λ .

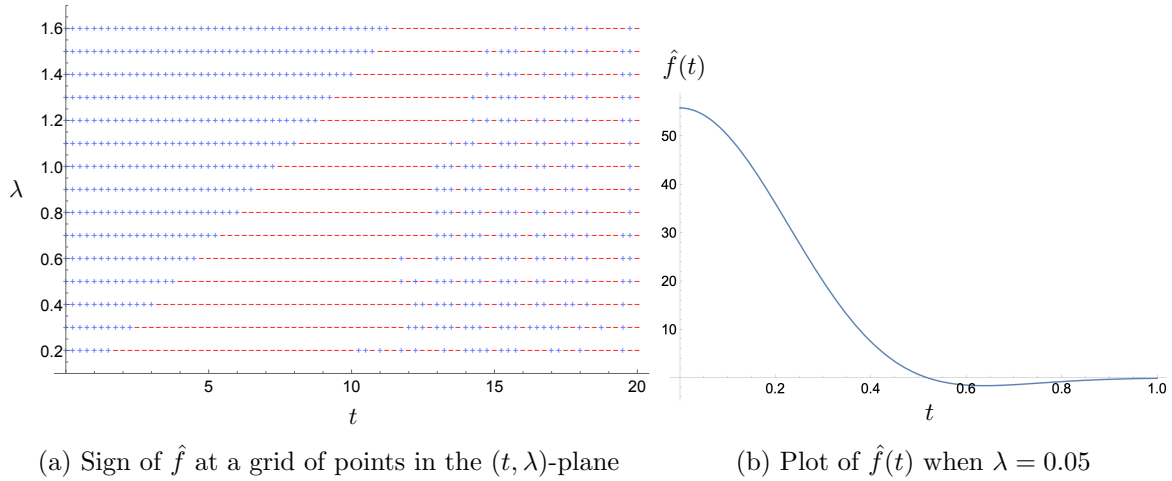


Figure 1: Spherical transform for the Gaussian kernel on the hyperbolic plane

323 Figure 1a shows the sign of \hat{f} in the (t, λ) -plane. The blue (positive) area on the left
 324 confirms Proposition 6.1. It is followed to the right by an alternation of red and blue areas,
 325 which shows that $\hat{f}(t)$ oscillates between positive and negative values (note that the λ axis
 326 begins at $\lambda = 0.2$, since the numerical evaluation of the integral (4.2) becomes unstable for
 327 much smaller values of λ). Figure 1b shows an individual plot of $\hat{f}(t)$, obtained for a fixed
 328 $\lambda = 0.05$.

329 It is clear from Figure 1a that the L^2 -Godement theorem implies the Gaussian kernel is
 330 not positive-definite on the hyperbolic plane, for any λ in the range of the λ -axis of this figure.

331 *Remark 6.2.* Proposition 6.1 continues to hold for a general non-compact M , as above.
 332 The proof of this general statement is not provided in order to maintain a reasonable length.

333 **6.2. Hyperbolic space.** In this case, $\hat{f}(t)$ can be found in closed form.

$$334 \quad (6.1) \quad g(r) = e^{-\lambda r^2} \text{ in (4.5)} \implies \hat{f}(t) = \left(\frac{2\pi}{t}\right) \left(\frac{\pi}{\lambda}\right)^{\frac{1}{2}} e^{(1-t^2)/4\lambda} \sin\left(\frac{t}{2\lambda}\right).$$

335 This shows that the sign of $\hat{f}(t)$ is determined by $\sin(t/2\lambda)$. This remains positive for $t \leq 2\pi\lambda$,
 336 and then oscillates indefinitely. From the L^2 -Godement theorem, the following is immediate.

337 **Proposition 6.3.** *The Gaussian kernel (1.1) is not positive-definite on the hyperbolic space*
 338 *for any value of λ .*

339 The expression (6.1) can be obtained from the following general lemma, proved in Section 8.3.

340 **Lemma 6.4.** *In (4.5), extend $g(r)$ to an even function of $r \in \mathbb{R}$. Then,*

$$341 \quad (6.2) \quad \hat{f}(t) = \frac{\pi}{t} \operatorname{Im} \{G(it + 1) - G(it - 1)\}$$

where Im denotes the imaginary part and G is the moment-generating function

$$G(s) = \int_{-\infty}^{+\infty} g(r) e^{sr} dr.$$

Indeed, (6.1) follows from (6.2) using the expression of the moment-generating function of a univariate normal distribution,

$$G(s) = \left(\frac{\pi}{\lambda}\right)^{\frac{1}{2}} e^{s^2/4\lambda}.$$

342 **6.3. The general case.** Let M be any non-compact symmetric space, as in Section 3
 343 (so the isometry group G of M is semisimple, of non-compact type, and has finite center [32]).
 344 The following lemma will be proved in Section 8.4.

345 **Lemma 6.5.** *Under the assumptions just stated, M contains an isometrically embedded*
 346 *hyperbolic plane.*

347 The numerical results of Paragraph 6.1 show that the Gaussian kernel is not positive-definite
 348 on the hyperbolic plane, at least for any λ in the range appearing in Figure 1a. The Embedding
 349 Lemma 5.2 and Lemma 6.5 therefore imply that the Gaussian kernel is not positive-definite
 350 on any non-compact Riemannian symmetric space, for λ in this same range (the additional
 351 assumptions of Lemma 6.5 merely serve to prevent M from being Euclidean).

352 On the other hand, Proposition 6.3 shows the Gaussian kernel is never positive-definite
 353 on the hyperbolic space. By the Embedding Lemma, it is therefore never positive-definite on
 354 symmetric spaces that contain an isometrically embedded hyperbolic space. Examples of such
 355 symmetric spaces include all spaces of $n \times n$ real, complex, or quaternion, positive-definite
 356 matrices (with the requirement that $n \geq 4$ for the real case) [38].

357 Hopefully, the numerical treatment of the hyperbolic plane case in Paragraph 6.1 will soon
 358 be replaced with a complete analytical treatment, similar to the one that we conducted for
 359 the hyperbolic space case in Paragraph 6.2. This is the only step still missing to obtain a
 360 fully analytical proof of the statement that the Gaussian kernel is never positive-definite on
 361 any non-Euclidean symmetric space.

362 **7. The hyperbolic secant revisited: non-Euclidean Herschel-Maxwell distributions.** Re-
 363 call the hyperbolic secant kernel on the hyperbolic plane from Section 4.3.2, which was shown
 364 to be positive-definite in contrast to the Gaussian kernel. Here, we will present a new inter-
 365 pretation of the hyperbolic secant kernel as the Herschel-Maxwell kernel on the hyperbolic
 366 plane. In Euclidean space, the Herschel-Maxwell theorem is an elegant characterization of the
 367 Gaussian distribution originally due to the astronomer J.F.W. Herschel, who was studying
 368 two-dimensional errors in astronomical observations. Ten years later, James Clerk Maxwell
 369 presented a three-dimensional version of the same argument to show that the stationary prob-
 370 ability distribution of the velocities of molecules in a gas follows a Gaussian distribution under
 371 the assumptions that (P1) the components of the random vector in an orthogonal coordinate
 372 system are independent, and (P2) the distribution only depends on the magnitude of the
 373 vector [29, 35, 39]. A modern statement of the theorem is that if the distribution of a random
 374 vector with independent components is invariant under rotations, then the components must
 375 be identically distributed as a Gaussian distribution.

376 Here, we will follow the Herschel-Maxwell derivation as it appears in Jaynes [35], which
 377 relies on the assumption of the existence of densities of the components of the random variables
 378 and solving a functional equation. This derivation closely follows Maxwell's own argument.
 379 Let (u, v) be a Cartesian coordinate system in the Euclidean plane. We seek to determine

380 a continuous joint probability distribution $p(u, v)dudv$ that satisfies assumptions P1 and P2.
 381 By assumption P1, we have

$$382 \quad (7.1) \quad p(u, v)dudv = h(u)du h(v)dv$$

383 for some function h . By assumption P2, we have

$$384 \quad (7.2) \quad p(u, v)dudv = g(r)rdrd\theta$$

385 as the density cannot depend on the angle θ when expressed in polar coordinates (r, θ) . (7.1)
 386 and (7.2) combine to give the functional equation

$$387 \quad (7.3) \quad h(u)h(v) = g\left(\sqrt{u^2 + v^2}\right),$$

388 which admits the general solution

$$389 \quad (7.4) \quad h(u) = \sqrt{\frac{\lambda}{\pi}}e^{-\lambda u^2}, \quad p(u, v) = \frac{\lambda}{\pi}e^{-\lambda(u^2+v^2)},$$

390 for any $\lambda > 0$ upon insisting that the distribution be normalized.

391 To replicate this derivation in the hyperbolic plane, we must first specify what we mean by
 392 an ‘orthogonal coordinate system’ in the hyperbolic plane. The most natural choice for such a
 393 coordinate system seems to be offered by Lobachevsky coordinates, determined by specifying
 394 a directed geodesic (u -axis) through the origin. The Lobachevsky coordinates of a point x
 395 then found by dropping a perpendicular to the u -axis. The signed distance from the foot of the
 396 perpendicular to the origin is the u -coordinate of the point, and the v -coordinate is the signed
 397 distance along the perpendicular to the u -axis (with the distance taken to be positive on one
 398 side of the u -axis and negative on the other). With respect to such a coordinate system, we
 399 seek a joint probability distribution $p(x)\text{vol}(dx) = p(u, v)\text{vol}(dx) = h(u)h(v)\text{vol}(dx)$, where
 400 $\text{vol}(dx)$ denotes the Riemannian volume form in the hyperbolic plane (P1). Moreover, (P2)
 401 generalizes to the assumption that $p(x)$ can only depend on the Riemannian distance to the
 402 origin $r = d(o, x)$, which once again yields the functional equation $h(u)h(v) = g(r)$. As this
 403 equation holds for all (u, v) , we can take any point on the u -axis to obtain $g(u) = h(u)h(0)$.
 404 That is, we have the functional equation

$$405 \quad (7.5) \quad h(u)h(v) = h(0)h(r)$$

406 where (u, v) and r are related by

$$407 \quad (7.6) \quad \cosh(r) = \cosh(u) \cosh(v)$$

408 according to the hyperbolic Pythagorean theorem. Setting $\tilde{h} = (h \circ \text{arcosh})/h(0)$, (7.5) be-
 409 comes $\tilde{h}(\cosh u)\tilde{h}(\cosh v) = \tilde{h}(\cosh u \cosh v)$, which admits the general solution $\tilde{h}(\cosh u) =$
 410 $(\cosh u)^{-a}$, for $a \in \mathbb{R}$ (i.e., \tilde{h} is a power function). Thus,

$$411 \quad h(r) = h(0)(\cosh(r))^{-a},$$

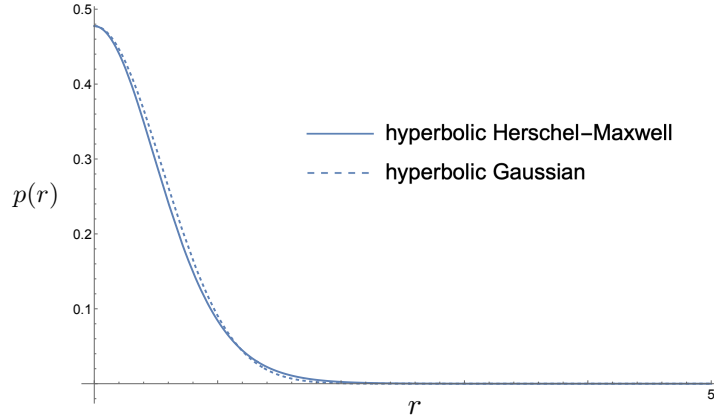


Figure 2: Plots of the normalized hyperbolic Herschel-Maxwell (solid curve, (7.7)) and Gaussian (dashed curve, (7.8)) densities as functions of $r = d(o, x)$ for parameters $a = 4$ and $\lambda = 1.66$.

412 which yields the probability density

$$413 \quad p(x) \text{vol}(dx) = h(0)^2 (\cosh(r))^{-a} \text{vol}(dx) = h(0)^2 (\cosh(r))^{-a} \sinh(r) dr d\theta,$$

414 in geodesic polar coordinates. Note that this density is normalizable only if $a > 1$. The
415 constant $h(0)$ can be determined through normalizing the distribution, which yields

$$416 \quad (7.7) \quad p(x) \text{vol}(dx) = \frac{1}{\pi} \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a-1}{2}\right)} (\cosh(d(o, x)))^{-a} \text{vol}(dx).$$

417 We call this distribution the Herschel-Maxwell distribution in the hyperbolic plane. It can
418 be viewed as one of several possible generalizations of the Gaussian distribution centered at
419 the point o to the hyperbolic plane alongside the distinct distributions derived from the heat
420 kernel interpretation of the Gaussian [28, 40] and

$$421 \quad (7.8) \quad p(x) \text{vol}(dx) = \frac{1}{\pi} \sqrt{\frac{\lambda}{\pi}} \frac{e^{-\frac{1}{4\lambda}}}{\text{erf}\left(\frac{1}{2\sqrt{\lambda}}\right)} e^{-\lambda d(o, x)^2} \text{vol}(dx),$$

422 which inherits an important characteristic property of Euclidean Gaussian distributions from
423 statistical inference: the maximum likelihood estimate of the mean reduces to a (Riemannian)
424 center of mass computation [44, 45, 47, 46, 15]. Figure 2 compares plots of the density functions
425 $p(r) \text{vol}(dx)$ in (7.7) and (7.8) as functions of $r = d(o, x)$ for parameters $a = 4$ and $\lambda = 1.66$.

426 Finally, we note that the hyperbolic Herschel-Maxwell distribution can be generalized
427 to n -dimensional hyperbolic space if we insist that the components of the point x in any
428 Lobachevsky coordinate system in *any* hyperbolic plane containing o and x are independent
429 and the distribution only depends on the geodesic distance $r = d(o, x)$. Reasoning as before,
430 we obtain

$$431 \quad (7.9) \quad p(x) \text{vol}(dx) = \frac{1}{\omega_{n-1}} \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{a+1-n}{2}\right) \Gamma\left(\frac{n}{2}\right)} (\cosh(d(o, x)))^{-a} \text{vol}(dx),$$

432 where ω_{n-1} denotes the area of the unit sphere in \mathbb{R}^n and the normalization factor in (7.9) is
 433 determined by integration with respect to the volume form $\text{vol}(dx)$ in n -dimensional hyperbolic
 434 space. The distribution is well-defined for $a > n - 1$.

435 *Remark 7.1.* We can repeat the derivation of the hyperbolic Herschel-Maxwell distribu-
 436 tion presented here in the case of spherical geometry, with the main difference being the
 437 replacement of the hyperbolic Pythagorean theorem with the spherical Pythagorean theorem
 438

$$439 \quad (7.10) \quad \cos(\theta) = \cos(u) \cos(v)$$

440 where o, x denote points on the 2-sphere S^2 , $\theta = d(o, x)$, and (u, v) is an orthogonal coordinate
 441 system centered on o , whereby the coordinates of a point x are obtained by dropping a
 442 perpendicular to a specified geodesic through o and measuring the signed lengths u along the
 443 geodesic from o to the foot of the perpendicular and v along the perpendicular to the geodesic
 444 from x . This yields the spherical Herschel-Maxwell kernel

$$445 \quad (7.11) \quad k(x, y) = |\cos(d(x, y))|^a,$$

446 where $a \geq 0$ to ensure the corresponding distribution is finite. Note that by representing
 447 $x, y \in S^2$ by unit vectors $w_x, w_y \in \mathbb{R}^3$, we have $\cos(d(x, y)) = \langle w_x, w_y \rangle$, and so (7.11) takes
 448 the form $k(x, y) = |\langle w_x, w_y \rangle|^a$. Thus, for non-negative even integer a , the spherical Herschel-
 449 Maxwell distribution can be identified with a homogeneous polynomial kernel $\langle x, y \rangle^n$, which
 450 is known to be positive-definite if and only if n is a non-negative integer [49].

451 **8. Proofs of the main results.**

452 **8.1. Proof of Theorem 3.1.**

453 **8.1.1. Notation.** The following notation will be used in the proof.

- 454 • Let W denote the Weyl group, which arises from the adjoint action of H on \mathfrak{a} [32].
 455 Also, let α denote the measure on \mathfrak{a}^* , which is given by $\alpha(dt) = c_M |c(t)|^{-2} dt$. Then,
 456 $L_W^2(\mathfrak{a}^*)$ stands for the space of W -invariant square-integrable functions (with respect
 457 to the measure α).
- 458 • Denote Φ_* the set of functions φ_t in (3.1) where $t \in \mathfrak{a}^*$. This is a proper subset of the
 459 set Φ of all positive-definite spherical functions, which appears in Godement's theorem
 460 (Section 2, in particular (2.4)).
- 461 • Functions or measures on \mathfrak{a}^* can be pulled back to Φ_* , assuming they are W -invariant.
 462 This is because $\varphi_{t'} = \varphi_t$ if and only if $t' = w \cdot t$ for some $w \in W$. Accordingly, the
 463 measure α on \mathfrak{a}^* can be identified with a measure on Φ_* , also denoted α . Similarly,
 464 if $\hat{f}(t)$ is a spherical transform, as in (3.2), $\hat{f}(\varphi)$ is just $\hat{f}(t)$ when $\varphi \in \Phi_*$ and $\varphi = \varphi_t$
 465 for some $t \in \mathfrak{a}^*$.
- 466 • Finally, $C_H^c(M)$ is the space of H -invariant continuous, compactly supported functions,
 467 with complex values, defined on M .

468 **8.1.2. The if part.** For square-integrable functions, the spherical transform is only defined
 469 in an abstract sense. It is a linear isometry between the Hilbert spaces $L_H^2(M)$ and $L_W^2(\mathfrak{a}^*)$,

470 $f \mapsto \hat{f}$ [30][52]. However, if \hat{f} is integrable, the inversion formula (3.3) holds

$$471 \quad (8.1) \quad f(x) = \int_{\mathfrak{a}^*} \hat{f}(t) \varphi_t(x) \alpha(dt) \quad \text{for all } x \in M$$

472 and this implies that f is continuous, since each φ_t is continuous and verifies $|\varphi_t(x)| \leq 1$,
473 being a positive-definite function with $\varphi_t(o) = 1$. The fact that f is positive-definite follows
474 from Godement's theorem. Indeed, (8.1) can be written

$$475 \quad (8.2) \quad f(x) = \int_{\Phi_*} \varphi(x) (\hat{f}\alpha)(d\varphi)$$

476 where $(\hat{f}\alpha)(d\varphi) = \hat{f}(\varphi) \alpha(d\varphi)$. Then, since $\hat{f}(\varphi) \geq 0$ for almost all $\varphi \in \Phi_*$ and \hat{f} is integrable,
477 with respect to α , $\hat{f}\alpha$ is a finite positive measure on Φ_* . Moreover, since Φ_* is a (measurable)
478 subset of Φ , (8.2) is exactly of the form (2.4), but with the measure μ_f supported on Φ_* ,
479 where it is equal to $\hat{f}\alpha$. This shows that f is positive-definite, and even determines the finite
480 positive measure μ_f (recall that μ_f is unique).

481 **8.1.3. The only-if part.** If f is H -invariant, continuous and positive-definite, then one
482 has

$$483 \quad (8.3) \quad \int_M (\psi * \psi^\dagger)(x) f(x) \text{vol}(dx) \geq 0$$

484 for any $\psi \in C_H^c(M)$, where $*$ denotes the convolution and $\psi^\dagger(g \cdot o) = \bar{\psi}(g^{-1} \cdot o)$ [22] (Page
485 286). This yields a scalar product on $C_H^c(M)$,

$$486 \quad (8.4) \quad \langle \psi_1, \psi_2 \rangle_f = \int_M (\psi_1 * \psi_2^\dagger)(x) f(x) \text{vol}(dx).$$

487 It will shortly be proved that this scalar product admits a spectral representation,

$$488 \quad (8.5) \quad \langle \psi_1, \psi_2 \rangle_f = \langle \hat{\psi}_1, \hat{\psi}_2 \rangle_{\hat{f}}$$

489 where the scalar product on the right-hand side is the usual “ L^2 ” scalar product, with respect
490 to the measure $\hat{f}\alpha$ on Φ_* . By [27] (Plancherel theorem, Page 101), this implies the measure
491 μ_f (in (2.4)) is supported on Φ_* , where it is equal to $\hat{f}\alpha$. Since μ_f is a finite positive measure,
492 $\hat{f}(t) \geq 0$ for almost all t and \hat{f} is integrable.

493 To prove (8.5), consider the left-hand side of (8.3). Both ψ and ψ^\dagger belong to $C_H^c(M)$.
494 The same is therefore true of their convolution. Since the spherical transform of ψ^\dagger is the
495 complex conjugate of the spherical transform of ψ , the convolution property of spherical
496 transforms (see [52] (Page 88)) implies that

$$497 \quad (8.6) \quad \widehat{\psi * \psi^\dagger}(t) = |\hat{\psi}(t)|^2.$$

498 Since both f and $\psi * \psi^\dagger$ are in $L_H^2(M)$, the Plancherel formula (3.4) yields

$$499 \quad (8.7) \quad \int_M (\psi * \psi^\dagger)(x) f(x) \text{vol}(dx) = \int_{\mathfrak{a}^*} |\hat{\psi}(t)|^2 \hat{f}(t) \alpha(dt) = \int_{\mathfrak{a}^*} |\hat{\psi}(t)|^2 (\hat{f}\alpha)(dt).$$

500 By (8.4), and by pulling back the integral of the right-hand side to Φ_* , this is

$$501 \quad (8.8) \quad \langle \psi, \psi \rangle_f = \int_{\Phi_*} |\hat{\psi}(\varphi)|^2 (\hat{f}\alpha)(d\varphi) = \langle \hat{\psi}, \hat{\psi} \rangle_{\hat{f}}$$

502 so that (8.5) follows immediately, using the polarization identity for scalar products. ■

503 **8.1.4. Alternative proof of Theorem 3.1.** The only-if part of the proposition admits at
 504 least one more proof. Indeed [22] (Page 302), for any continuous positive-definite $f \in L_H^2(M)$,
 505 it is possible to construct $\phi \in L_H^2(M)$ such that $f = \phi * \phi^\dagger$. If one can apply the convolution
 506 property as in (8.6), it would follow that $\hat{f}(t) = |\hat{\phi}(t)|^2$, so $\hat{f}(t)$ is positive for all t (not just
 507 for almost all t), and \hat{f} is integrable.

508 The convolution property does not apply directly because ϕ is not integrable. Letting
 509 $(\phi_n; n = 1, 2, \dots)$ belong to $C_H^c(M)$ and converge to ϕ in $L_H^2(M)$, the functions $f_n = \phi_n * \phi_n$
 510 belong to $C_H^c(M)$ and converge uniformly to f , as can be shown by using Young's convolution
 511 inequality [33] (Page 296).

512 Now, as in (8.6), one has $\hat{f}_n(t) = |\hat{\phi}_n(t)|^2$ for each n . However, since $\hat{\phi}_n$ converges to $\hat{\phi}$
 513 in $L_W^2(\mathfrak{a}^*)$, it follows that $|\hat{\phi}_n|^2$ converges to $|\hat{\phi}|^2$ in $L_W^1(\mathfrak{a}^*)$ (this is the space of W -invariant
 514 integrable functions) [9] (Page 298).

Therefore, a representation of f under the form (3.3) can be obtained by noting, for $x \in M$,

$$\lim f_n(x) = \lim \int_{\mathfrak{a}^*} \varphi_t(x) |\hat{\phi}_n(t)|^2 \alpha(dt) = \int_{\mathfrak{a}^*} \varphi_t(x) |\hat{\phi}(t)|^2 \alpha(dt)$$

515 where the first equality follows from (3.3) and $\hat{f}_n(t) = |\hat{\phi}_n(t)|^2$, and the second equality from
 516 $|\hat{\phi}_n|^2 \rightarrow |\hat{\phi}|^2$ in $L_W^1(\mathfrak{a}^*)$ and $|\varphi_t(x)| \leq 1$ (this is a consequence of φ_t being positive-definite).

However, the limit of $f_n(x)$ is just $f(x)$, since the f_n converge uniformly to f . Thus,

$$f(x) = \int_{\mathfrak{a}^*} \varphi_t(x) |\hat{\phi}(t)|^2 \alpha(dt).$$

517 Comparing to (3.3), it follows that $\hat{f}(t) = |\hat{\phi}(t)|^2$, as required. ■

518 **8.2. Proof of Proposition 6.1.** Let R be such that $r \leq R$ implies $|\cosh(r) - 1| < 1$.
 519 Then [51] (Page 142)

$$520 \quad (8.9) \quad P_{-\frac{1}{2}+it}(\cosh(r)) = 1 - (1/4 + t^2)(r^2/4) + \varepsilon(t, r) \quad \text{for } r \leq R$$

521 with remainder $|\varepsilon(t, r)| \leq C(T, R)r^4$ for all $t \leq T$ and $r \leq R$, where $C(T, R) > 0$ is a constant.
 522 Replacing this into (4.2), the following decomposition is obtained,

$$523 \quad (8.10) \quad \hat{f}(t) = \hat{f}_a(t) + \hat{f}_b(t) + \hat{f}_c(t)$$

524 in terms of the following integrals

$$525 \quad (8.11) \quad \hat{f}_a(t) = \int_0^R e^{-\lambda r^2} [1 - (1/4 + t^2)(r^2/4)] \sinh(r) dr,$$

$$526 \quad (8.12) \quad \hat{f}_b(t) = \int_0^R e^{-\lambda r^2} \varepsilon(t, r) \sinh(r) dr,$$

$$527 \quad (8.13) \quad \hat{f}_c(t) = \int_R^\infty e^{-\lambda r^2} P_{-\frac{1}{2}+it}(\cosh(r)) \sinh(r) dr.$$

528 Noting that $|P_{-\frac{1}{2}+it}(\cosh(r))| \leq 1$ [51] (Page 161), a simple calculation yields

$$529 \quad (8.14) \quad |\hat{f}_c(t)| \leq \int_R^\infty e^{-\lambda r^2} \sinh(r) dr = O(e^{-\lambda R^2}).$$

On the other hand,

$$(C(T, R))^{-1} |\hat{f}_b(t)| \leq \int_0^R e^{-\lambda r^2} r^4 \sinh(r) dr \leq \int_0^\infty e^{-\lambda r^2} r^4 \sinh(r) dr.$$

However, by performing an elementary change of variables and then assuming that $\lambda \geq 1$, this implies the following bound,

$$(C(T, R))^{-1} |\hat{f}_b(t)| \leq \lambda^{-\frac{5}{2}} \int_0^\infty e^{-u^2} u^4 \sinh(u) du$$

530 which means that

$$531 \quad (8.15) \quad |\hat{f}_b(t)| = O(\lambda^{-\frac{5}{2}}).$$

532 Finally, $\hat{f}_a(t)$ can be extended to an integral from 0 to ∞ , with an exponentially small error,

$$533 \quad (8.16) \quad \hat{f}_a(t) = Z(\lambda) - (1/4 + t^2) V(\lambda) + O(e^{-\lambda R^2})$$

where $Z(\lambda)$ and $V(\lambda)$ are the following integrals

$$Z(\lambda) = \int_0^\infty e^{-\lambda r^2} \sinh(r) dr, \quad V(\lambda) = \frac{1}{4} \int_0^\infty e^{-\lambda r^2} r^2 \sinh(r) dr.$$

534 By a direct evaluation

$$535 \quad (8.17) \quad Z(\lambda) = \frac{1}{2} \sqrt{\frac{\pi}{\lambda}} \exp\left(\frac{1}{4\lambda}\right) \operatorname{erf}\left(\frac{1}{2\sqrt{\lambda}}\right) = \frac{\lambda^{-1}}{2} + \frac{\lambda^{-2}}{6} + O(\lambda^{-3})$$

536 where erf denotes the error function [5] (Page 35), and then by noting that $4V(\lambda) = -Z'(\lambda)$
537 (the prime denotes derivation),

$$538 \quad (8.18) \quad V(\lambda) = \frac{\lambda^{-2}}{8} + O(\lambda^{-3}).$$

539 Therefore, replacing (8.17) and (8.18) into (8.16),

$$540 \quad (8.19) \quad \hat{f}_a(t) = \frac{\lambda^{-1}}{2} - \left(t^2 - \frac{13}{12}\right) \frac{\lambda^{-2}}{8} + O(\lambda^{-3}).$$

541 To conclude the proof, replace (8.14), (8.15) and (8.19) into (8.10). This yields

$$542 \quad (8.20) \quad \hat{f}(t) \sim \frac{\lambda^{-1}}{2} - \left(t^2 - \frac{13}{12}\right) \frac{\lambda^{-2}}{8} \quad \text{as } \lambda \rightarrow \infty$$

543 ($a \sim b$ means a/b converges to 1), where the convergence is uniform in $t \leq T$. It is now enough
544 to note that the right-hand side of (8.20) is positive for all $t \leq T$ when $\lambda \geq T^2/4$. ■

545 **8.3. Proof of Lemma 6.4.** Note that $\sinh(r) \sin(tr)$ is an even function of r . Moreover,
546 since $g(r)$ is also even, (4.5) can be written

$$547 \quad \hat{f}(t) = \frac{4\pi}{t} \int_0^\infty g(r) \sinh(r) \sin(tr) dr$$

$$548 \quad = \frac{2\pi}{t} \int_{-\infty}^{+\infty} g(r) \sinh(r) \sin(tr) dr$$

In other words,

$$\hat{f}(t) = \frac{2\pi}{t} \operatorname{Im} \left\{ \int_{-\infty}^{+\infty} g(r) \sinh(r) e^{itr} dr \right\}.$$

Using the definition of the hyperbolic sine, this becomes

$$\hat{f}(t) = \frac{\pi}{t} \operatorname{Im} \left\{ \int_{-\infty}^{+\infty} g(r) e^{(it+1)r} dr - \int_{-\infty}^{+\infty} g(r) e^{(it-1)r} dr \right\}$$

and then, by definition of the moment-generating function G ,

$$\hat{f}(t) = \frac{\pi}{t} \operatorname{Im} \{G(it+1) - G(it-1)\}$$

549 which is the same as (6.2). ■

550 **8.4. Proof of Lemma 6.5.** It will be enough to prove that there exists a totally geodesic
551 submanifold M_o of M , such that M_o is a hyperbolic plane. Indeed, M is a Hadamard manifold
552 (a simply connected complete Riemannian manifold with negative sectional curvatures [32]
553 (Chapter VI)). Therefore, if M_o is a totally geodesic submanifold, it is also an isometrically
554 embedded metric space. The proof will rely very heavily on the algebraic description of
555 symmetric spaces, detailed in [32] (in particular Chapters V and VI).

556 Recall that M is determined by a symmetric pair (G, H) . The Lie algebra \mathfrak{g} of G is a
557 real semisimple Lie algebra, and admits a Cartan involution θ with corresponding Cartan
558 decomposition $\mathfrak{g} = \mathfrak{h} + \mathfrak{p}$ (\mathfrak{h} is the Lie algebra of H). If B is the Killing form of \mathfrak{g} , then

$$559 \quad (8.21) \quad \langle \xi, \eta \rangle = -B(\xi, \theta(\eta)) \quad \xi, \eta \in \mathfrak{g}$$

560 defines a scalar product on \mathfrak{g} . Let \mathfrak{a} be a maximal Abelian subspace of \mathfrak{p} . The linear maps
 561 $\text{ad}_a : \mathfrak{g} \rightarrow \mathfrak{g}$ where $\text{ad}_a(\xi) = [a, \xi]$ for $a \in \mathfrak{a}$ are jointly diagonalisable, with real eigenvalues.
 562 These eigenvalues are linear forms $\lambda : \mathfrak{a} \rightarrow \mathbb{R}$, called the roots of \mathfrak{g} with respect to \mathfrak{a} .

563 Fix a non-zero root λ , and $a_\lambda \in \mathfrak{a}$ such that $\lambda(a) = B(a_\lambda, a)$ for $a \in \mathfrak{a}$. Then, let ξ_λ be a
 564 joint eigenvector, $\text{ad}_a(\xi_\lambda) = \lambda(a)\xi_\lambda$ for $a \in \mathfrak{a}$. Note that

$$565 \quad (8.22) \quad [a, \theta(\xi_\lambda)] = \theta([\theta(a), \xi_\lambda]) = -\theta([a, \xi_\lambda])$$

566 for any $a \in \mathfrak{a}$, where the first equality follows because θ is an involution, and the second
 567 because $a \in \mathfrak{a}$, a subspace of \mathfrak{p} , and \mathfrak{p} is by definition the (-1) -eigenspace of θ . However,

$$568 \quad (8.23) \quad \theta([a, \xi_\lambda]) = \lambda(a)\theta(\xi_\lambda)$$

569 by the definition of ξ_λ and linearity of θ . Thus, (8.22) and (8.23) imply that $\text{ad}_a(\theta(\xi_\lambda)) =$
 570 $-\lambda(a)\theta(\xi_\lambda)$ for any $a \in \mathfrak{a}$. Denoting $\theta(\xi_\lambda)$ by $\xi_{-\lambda}$ and recalling the definition of $a_\lambda \in \mathfrak{a}$, it is
 571 now seen that the following commutation relations hold

$$572 \quad (8.24) \quad [a_\lambda, \xi_\lambda] = \|a_\lambda\|^2 \xi_\lambda \quad [a_\lambda, \xi_{-\lambda}] = -\|a_\lambda\|^2 \xi_{-\lambda}$$

573 where $\|\cdot\|$ is the norm with respect to the scalar product (8.21). It will now be proved that

$$574 \quad (8.25) \quad [\xi_\lambda, \xi_{-\lambda}] = -\|\xi_\lambda\|^2 a_\lambda.$$

575 First, any $a \in \mathfrak{a}$ commutes with $[\xi_\lambda, \xi_{-\lambda}]$. Indeed,

$$576 \quad [a, [\xi_\lambda, \xi_{-\lambda}]] = [[a, \xi_\lambda], \xi_{-\lambda}] + [\xi_\lambda, [a, \xi_{-\lambda}]] \\ 577 \quad = \lambda(a)[\xi_\lambda, \xi_{-\lambda}] - \lambda(a)[\xi_\lambda, \xi_{-\lambda}]$$

578 where the first equality follows from Jacobi's identity for Lie brackets, and the second because
 579 ξ_λ and $\xi_{-\lambda}$ are joint eigenvectors with eigenvalues $\lambda(a)$ and $-\lambda(a)$. Second, $[\xi_\lambda, \xi_{-\lambda}] \in \mathfrak{p}$.
 580 Indeed, $\theta([\xi_\lambda, \xi_{-\lambda}]) = [\theta(\xi_\lambda), \theta(\xi_{-\lambda})]$, but $\theta(\xi_\lambda) = \xi_{-\lambda}$ by definition, and $\theta(\xi_{-\lambda}) = \xi_\lambda$ since θ
 581 is an involution. Therefore, $\theta([\xi_\lambda, \xi_{-\lambda}]) = -[\xi_\lambda, \xi_{-\lambda}]$ which means $[\xi_\lambda, \xi_{-\lambda}] \in \mathfrak{p}$. However, \mathfrak{a}
 582 is maximal Abelian in \mathfrak{p} , and this yields $[\xi_\lambda, \xi_{-\lambda}] \in \mathfrak{a}$.

Finally, to obtain (8.25), note that

$$B([\xi_\lambda, \xi_{-\lambda}], a) = B(\xi_{-\lambda}, [a, \xi_\lambda]) = B(\xi_{-\lambda}, \xi_\lambda)\lambda(a)$$

where the first equality holds since B is ad-invariant, and the second since ξ_λ is a joint
 eigenvector with eigenvalue $\lambda(a)$. Now, the fact that B is symmetric and the definition of a_λ
 imply $B([\xi_\lambda, \xi_{-\lambda}], a) = B(\xi_\lambda, \xi_{-\lambda})B(a_\lambda, a)$, so that

$$[\xi_\lambda, \xi_{-\lambda}] = B(\xi_\lambda, \xi_{-\lambda}) a_\lambda$$

583 and (8.25) follows after recalling (8.21) and $\xi_{-\lambda} = \theta(\xi_\lambda)$.

584 Moving on, let \mathfrak{g}_o denote the Lie subalgebra of \mathfrak{g} , generated by $\{\xi_{-\lambda}, a_\lambda, \xi_\lambda\}$. The Lie
 585 bracket of \mathfrak{g}_o is completely determined by (8.24) and (8.25). It is straightforward to see that
 586 $\mathfrak{g}_o \simeq \mathfrak{sl}(2, \mathbb{R})$ (the Lie algebra of 2×2 zero-trace real matrices). The isomorphism is given by

$$587 \quad (8.26) \quad \xi_{-\lambda} \longleftrightarrow \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix} ; \quad a_\lambda \longleftrightarrow \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} ; \quad \xi_\lambda \longleftrightarrow \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

588 after proper rescaling of $\{\xi_{-\lambda}, a_\lambda, x_{-\lambda}\}$ (this needs to ensure that $\|a_\lambda\|^2 = 2$ and $\|\xi_\lambda\|^2 = 1$).
 589 Let G_o be the connected subgroup of G with Lie algebra \mathfrak{g}_o and let M_o be the orbit of o
 590 under G_o (recall that $o \in M$ has stabiliser H in G and can be thought of as the “point of
 591 origin”). Then, M_o is a totally geodesic submanifold of M , because \mathfrak{g}_o is a Lie subalgebra of
 592 \mathfrak{g} [32] (Page 189). Moreover, M_o is a Riemannian homogeneous space under the action of G_o ,
 593 $M_o \simeq G_o/H_o$ where $H_o = G_o \cap H$ is the stabiliser of o in G_o . To see that M_o is a hyperbolic
 594 plane (up to an irrelevant scaling factor), note first that $G_o \simeq SL(2, \mathbb{R})$. It then remains to
 595 check that $H_o \simeq SO(2)$, since the Riemannian homogeneous space $SL(2, \mathbb{R})/SO(2)$ is just a
 596 hyperbolic plane [51]. To do so, let \mathfrak{h}_o be the Lie algebra of H_o . Then, $\mathfrak{h}_o = \mathfrak{g}_o \cap \mathfrak{h}$. However,
 597 since \mathfrak{h} is the (+1)-eigenspace of θ , it is easy to see that \mathfrak{h}_o is generated by $\tau = \xi_\lambda + \xi_{-\lambda}$.
 598 Since τ corresponds to the generator of $SO(2) \subset SL(2, \mathbb{R})$, under the isomorphism (8.26),
 599 $H_o \simeq SO(2)$, as required. ■

600 **Acknowledgments.** The authors acknowledge financial support from the School of Physi-
 601 cal and Mathematical Sciences, the Talent Recruitment and Career Support (TRACS) Office,
 602 and the Presidential Postdoctoral Programme at Nanyang Technological University (NTU).

603

REFERENCES

- 604 [1] A. AZANGULOV, A. SMOLENSKY, A. TERENIN, AND S. BOROVITSKY, *Stationary kernels and Gaussian*
 605 *processes on Lie groups and their homogeneous spaces I: the compact case*, Jan. 2023, [arxiv:2208.14960](https://arxiv.org/abs/2208.14960).
 606 [2] A. AZANGULOV, A. SMOLENSKY, A. TERENIN, AND S. BOROVITSKY, *Stationary kernels and Gaussian*
 607 *processes on Lie groups and their homogeneous spaces II: non-compact symmetric spaces*, Apr. 2023,
 608 [arxiv:2301.13088](https://arxiv.org/abs/2301.13088).
 609 [3] S. M. BAGIROVA AND A. K. KHANMAMEDOV, *On zeros of the modified Bessel function of the second kind*,
 610 *Computational Mathematics and Mathematical Physics*, 60 (2020), pp. 817–820.
 611 [4] W. D. BATEN, *The probability law for the sum of n independent variables, each subject to the law*
 612 *($1/(2h)$) sech($\pi x/(2h)$)*, *Bulletin of the American Mathematical Society*, 40 (1934), pp. 284 – 290.
 613 [5] R. BEALS AND R. WONG, *Special functions, a graduate text*, Cambridge University Press, Cambridge,
 614 UK, 2010.
 615 [6] T. BENDOKAT, R. ZIMMERMANN, AND P.-A. ABSIL, *A Grassmann manifold handbook: basic geometry*
 616 *and computational aspects*, *Advances in Computational Mathematics*, 50 (2024), pp. 1–51.
 617 [7] A. BERLINET AND C. THOMAS-AGNAN, *Reproducing-kernel Hilbert spaces in Probability and statistics*,
 618 Springer Science, New York, 2004.
 619 [8] N. H. BINGHAM AND T. L. SIMONS, *Probability, statistics, and planet earth, II: the Bochner-Godement*
 620 *theorem for symmetric spaces*, Nov. 2017, [arxiv:1707.05204](https://arxiv.org/abs/1707.05204).
 621 [9] V. I. BOGACHEV, *Measure theory, Vol. I*, Springer-Verlag, Berlin Heidelberg, 2007.
 622 [10] V. BOROVITSKIY, A. TERENIN, P. MOSTOWSKY, AND M. DEISENROTH, *Matérn Gaussian processes on*
 623 *Riemannian manifolds*, in *Advances in Neural Information Processing Systems*, 2020.
 624 [11] Y. CABANNES, *Multidimensional complex stationary centered Gaussian autoregressive time series machine*
 625 *learning in Poincaré and Siegel disks: application for audio and radar clutter classification*, Phd thesis,
 626 Université de Bordeaux, 2022. Available at <https://theses.hal.science/tel-03708515>.
 627 [12] T. CECCHERINI-SILBERSTEIN, F. SCARABOTTI, AND F. TOLLI, *Harmonic analysis on finite groups*, Cam-
 628 bridge University Press, Cambridge, 2008.
 629 [13] R. CHAKRABORTY AND B. C. VEMURI, *Statistics on the Stiefel manifold: theory and applications*, *The*
 630 *Annals of Statistics*, 47 (2019), pp. 415–438.
 631 [14] I. CHAVEL, *Riemannian geometry, a modern introduction (2nd Edition)*, Cambridge University Press,
 632 Cambridge, UK, 2006.
 633 [15] B. CHEN, C. MOSTAJERAN, AND S. SAID, *Geometric learning of hidden Markov models via a method of*
 634 *moments algorithm*, *Physical Sciences Forum*, 5 (2022), <https://doi.org/10.3390/psf2022005010>.

- 635 [16] G. S. CHIRIKJIAN, *Stochastic models, information theory, and Lie groups (vol. 2)*, Springer Science, New
636 York, 2012.
- 637 [17] M. CONGEDO, A. BARACHANT, AND R. BHATIA, *Riemannian geometry for EEG-based brain-computer*
638 *interfaces; a primer and a review*, *Brain-Computer Interfaces*, 4 (2017), pp. 155–174.
- 639 [18] R. CRITTENDEN, *Minimum and conjugate points in symmetric spaces*, *Canadian Journal of Mathematics*,
640 14 (1962), p. 320–328.
- 641 [19] N. DA COSTA, C. MOSTAJERAN, AND J.-P. ORTEGA, *The Gaussian kernel on the circle and spaces*
642 *that admit isometric embeddings of the circle*, in *Geometric Science of Information*, F. Nielsen and
643 F. Barbaresco, eds., Cham, 2023, Springer Nature Switzerland, pp. 426–435.
- 644 [20] N. DA COSTA, C. MOSTAJERAN, J.-P. ORTEGA, AND S. SAID, *Geometric learning with positively decom-*
645 *posable kernels*, Oct. 2023, [arxiv:2310.13821](https://arxiv.org/abs/2310.13821).
- 646 [21] E. B. DAVIES, *Heat kernels and spectral theory*, Cambridge University Press, Cambridge, UK, 1990.
- 647 [22] J. DIXMIER, *C^* -algebras*, North Holland Publishing Company, Amsterdam-New York, 1977.
- 648 [23] A. FERAGEN AND S. HAUBERG, *Open problem: Kernel methods on manifolds and metric spaces. what*
649 *is the probability of a positive definite geodesic exponential kernel?*, in *29th Annual Conference on*
650 *Learning Theory*, 2016.
- 651 [24] A. FERAGEN, F. LAUZE, AND S. HAUBERG, *Geodesic exponential kernels: When curvature and linear-*
652 *ity conflict*, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*
653 *(CVPR)*, June 2015.
- 654 [25] G. B. FOLLAND, *A course in abstract harmonic analysis*, CRC Press, 1995.
- 655 [26] R. GANGOLLI, *Asymptotic behaviour of the spectra of compact quotients of certain symmetric spaces*, *Acta*
656 *Mathematica*, 121 (1968), pp. 151 – 192.
- 657 [27] R. GODEMENT, *Introduction aux travaux de Selberg*, Séminaire Bourbaki, (no. 144), (1958), pp. 95–110.
- 658 [28] A. GRIGOR'YAN AND M. NOGUCHI, *The heat kernel on hyperbolic space*, *Bulletin of the London Mathemat-*
659 *ical Society*, 30 (1998), pp. 643–650, [https://doi.org/https://doi.org/10.1112/S0024609398004780](https://doi.org/10.1112/S0024609398004780).
- 660 [29] B. GYENIS, *Maxwell and the normal distribution: A colored story of probability, independence, and ten-*
661 *dency toward equilibrium*, *Studies in History and Philosophy of Science Part B: Studies in History*
662 *and Philosophy of Modern Physics*, 57 (2017), pp. 53–65, [https://doi.org/https://doi.org/10.1016/j.](https://doi.org/10.1016/j.shpsb.2017.01.001)
663 [shpsb.2017.01.001](https://www.sciencedirect.com/science/article/pii/S1355219817300102), <https://www.sciencedirect.com/science/article/pii/S1355219817300102>.
- 664 [30] S. HELGASON, *Groups and geometric analysis: integral geometry, invariant differential operators, and*
665 *spherical functions*, Academic Press, Orlando, Florida, 1984.
- 666 [31] S. HELGASON, *Harish-Chandra's c-function, a mathematical jewel*, in *Noncompact Lie Groups and some*
667 *of their applications*, E. A. Tanner and R. Wilson, eds., Springer Netherlands, 1994, pp. 55–67.
- 668 [32] S. HELGASON, *Differential geometry, Lie groups, and symmetric spaces*, Americal Mathematical Society,
669 Providence, RI, 2001.
- 670 [33] E. HEWITT AND K. A. ROSS, *Abstract harmonic analysis, Vol. I*, Springer-Verlag, Berlin Heidelberg,
671 1963.
- 672 [34] S. JAYASUMANA, R. HARTLEY, M. SALZMANN, H. LI, AND M. HARANDI, *Kernel methods on Riemann-*
673 *ian manifolds with Gaussian RBF kernels*, *IEEE Transactions on Pattern Analysis and Machine*
674 *Intelligence*, 37 (2015), pp. 2464–2477, <https://doi.org/10.1109/TPAMI.2015.2414422>.
- 675 [35] E. T. JAYNES, *Probability theory, the logic of science*, Cambridge University Press, Cambridge, UK, 2003.
- 676 [36] A. A. KIRILLOV, *Elements of the theory of representations*, Springer-Verlag, Berlin Heidelberg, 1976.
- 677 [37] S. LI, *Gaussian kernels on nonsimply connected closed Riemannian manifolds are never positive definite*,
678 *Bulletin of the London Mathematical Society*, (2023), <https://doi.org/10.1112/blms.12928>, <https://onlinelibrary.wiley.com/doi/abs/10.1112/blms.12928>.
- 679 [38] F. LOPEZ, B. POZZETTI, S. TRETTEL, M. STRUBE, AND A. WIENHARD, *Symmetric spaces for graph*
680 *embeddings: a Finsler-Riemannian approach*, in *Proceedings of the 38th International Conference on*
681 *Machine Learning*, 2021.
- 682 [39] J. MAXWELL, *Illustrations of the dynamical theory of gases*, *Philosophical Magazine*, 19 (1860), pp. 19–32.
- 683 [40] H. P. MCKEAN, *An upper bound to the spectrum of Δ on a manifold of negative curvature*, *Journal of*
684 *Differential Geometry*, 4 (1970), pp. 359 – 366, <https://doi.org/10.4310/jdg/1214429509>.
- 685 [41] H. Q. MINH AND V. MURINO, *Kernel Methods on Covariance Matrices*, Springer International Publish-
686 ing, Cham, 2018, pp. 55–64, https://doi.org/10.1007/978-3-031-01820-6_4, [https://doi.org/10.1007/](https://doi.org/10.1007/978-3-031-01820-6_4)
687 [978-3-031-01820-6_4](https://doi.org/10.1007/978-3-031-01820-6_4).
- 688

- 689 [42] C. MOSTAJERAN AND S. SAID, *Invariant kernels on the space of complex covariance matrices*, July 2024,
690 [arxiv:2404.02169](https://arxiv.org/abs/2404.02169).
- 691 [43] S. SAID, P. O. AMBLARD, AND J. H. MANTON, *Stationary random fields arising from second-order partial*
692 *differential equations on compact Lie groups*, IEEE Transactions on Information Theory, 59 (2013),
693 pp. 3497–3513.
- 694 [44] S. SAID, L. BOMBRUN, Y. BERTHOUMIEU, AND J. H. MANTON, *Riemannian Gaussian distributions*
695 *on the space of symmetric positive definite matrices*, IEEE Transactions on Information Theory, 63
696 (2017), pp. 2153–2170, <https://doi.org/10.1109/TIT.2017.2653803>.
- 697 [45] S. SAID, H. HAJRI, L. BOMBRUN, AND B. C. VEMURI, *Gaussian distributions on Riemannian symmetric*
698 *spaces: Statistical learning with structured covariance matrices*, IEEE Transactions on Information
699 Theory, 64 (2018), pp. 752–772, <https://doi.org/10.1109/TIT.2017.2713829>.
- 700 [46] S. SAID, S. HEUVELINE, AND C. MOSTAJERAN, *Riemannian statistics meets random matrix theory: To-*
701 *ward learning from high-dimensional covariance matrices*, IEEE Transactions on Information Theory,
702 69 (2023), pp. 472–481, <https://doi.org/10.1109/TIT.2022.3199479>.
- 703 [47] S. SAID, C. MOSTAJERAN, AND S. HEUVELINE, *Chapter 10 - Gaussian distributions on Riemannian*
704 *symmetric spaces of nonpositive curvature*, in Geometry and Statistics, F. Nielsen, A. S. Srinivasa
705 Rao, and C. Rao, eds., vol. 46 of Handbook of Statistics, Elsevier, 2022, pp. 357–400, [https://doi.](https://doi.org/10.1016/bs.host.2022.03.004)
706 [org/10.1016/bs.host.2022.03.004](https://doi.org/10.1016/bs.host.2022.03.004).
- 707 [48] B. SCHÖLKOPF AND A. J. SMOLA, *Learning with kernels: support-vector machine, regularization, opti-*
708 *mization, and beyond*, MIT Press, 2018.
- 709 [49] A. SMOLA, Z. ÓVÁRI, AND R. C. WILLIAMSON, *Regularization with dot-product kernels*, in Ad-
710 *vances in Neural Information Processing Systems*, T. Leen, T. Dietterich, and V. Tresp,
711 eds., vol. 13, MIT Press, 2000, [https://proceedings.neurips.cc/paper_files/paper/2000/file/](https://proceedings.neurips.cc/paper_files/paper/2000/file/d25414405eb37dae1c14b18d6a2cac34-Paper.pdf)
712 [d25414405eb37dae1c14b18d6a2cac34-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2000/file/d25414405eb37dae1c14b18d6a2cac34-Paper.pdf).
- 713 [50] S. SRA, *Positive definite matrices and the S-divergence*, Proceedings of the American Mathematical Soci-
714 ety, 144 (2016), pp. 2787–2797, <https://www.jstor.org/stable/procamermathsoci.144.7.2787> (accessed
715 2023-02-07).
- 716 [51] A. TERRAS, *Harmonic analysis on symmetric spaces and applications, Vol. I*, Springer-Verlag, New York,
717 1985.
- 718 [52] A. TERRAS, *Harmonic analysis on symmetric spaces and applications, Vol. II*, Springer-Verlag, New York,
719 1988.
- 720 [53] A. M. YAGLOM, *Second-order homogeneous random fields*, in Berkeley Symposium on Mathematical
721 Statistics and Probability, J. Neyman, ed., 1961, pp. 593–622.
- 722 [54] A. M. YAGLOM, *An introduction to the theory of stationary random functions*, Dover Publications, 1962.
- 723 [55] B. ØKSENDAL, *Stochastic Differential Equations*, Springer, Berlin, Heidelberg, 2003, [https://doi.org/10.](https://doi.org/10.1007/978-3-642-14394-6)
724 [1007/978-3-642-14394-6](https://doi.org/10.1007/978-3-642-14394-6).

725 **Appendix A. L¹-Godement theorem.** The L²-Godement theorem, stated back in Sec-
 726 tion 3, has the following variant, which will be called the L¹-Godement theorem. It seems
 727 useful to have this at hand, to deal with kernels k such that the function $f(x) = k(o, x)$
 728 is integrable but not square-integrable.

729 **Theorem A.1 (L¹-Godement Theorem).** *A continuous function $f \in L^1_H(M)$ is positive-*
 730 *definite if and only if $\hat{f}(t) \geq 0$ for all $t \in \mathfrak{a}^*$ and \hat{f} is integrable with respect to the measure*
 731 *$|c(t)|^{-2}dt$.*

732 Here, $L^1_H(M)$ denotes the space of H -invariant integrable functions (with respect to vol). The
 733 following proof is modeled on the one in Section 8.1. In particular, the notation is the same
 734 as in Paragraph 8.1.1.

735 **A.1. The if part.** It will be enough to prove that the inversion formula (3.3) holds

$$736 \quad (\text{A.1}) \quad f(x) = \int_{\mathfrak{a}^*} \varphi_t(x) \hat{f}(t) \alpha(dt) \quad \text{for all } x \in M.$$

737 The fact that f is continuous and positive-definite then follows exactly as in Paragraph 8.1.2.
 738 Denote by $G_s(x) = G_s(o, x)$ the heat kernel of M . For $s > 0$, G_s has spherical transform [26]

$$739 \quad (\text{A.2}) \quad \hat{G}_s(t) = (Z(s))^{-1} \exp[-s(\|t\|^2 + \|\rho\|^2)]$$

where $Z(s)$ is a normalizing constant, which guarantees that G_s integrates to 1 over M , and
 where ρ is half the sum of positive roots (as in Formula (3.1)). Recall that the heat semigroup
 is strongly continuous on $L^1_H(M)$ [21] (Page 148). This means the convolutions $f_s = G_s * f$
 converge to f in $L^1_H(M)$ as $s \rightarrow 0$. Of course, G_s satisfies the inversion formula (3.3),

$$G_s(x) = \int_{\mathfrak{a}^*} \varphi_t(x) \hat{G}_s(t) \alpha(dt).$$

740 Therefore, by taking the convolution under the integral

$$741 \quad (\text{A.3}) \quad f_s(x) = \int_{\mathfrak{a}^*} (f * \varphi_t)(x) \hat{G}_s(t) \alpha(dt) = \int_{\mathfrak{a}^*} \varphi_t(x) \hat{f}(t) \hat{G}_s(t) \alpha(dt)$$

where the second equality follows because φ_t is an eigenfunction of convolution operators [52]
 (Page 67). Now, from the assumption that \hat{f} is integrable (*i.e.* $\hat{f} \in L^1_W(\mathfrak{a}^*)$), and from (A.2),
 the dominated convergence theorem can be applied in (A.3), in order to obtain

$$\lim_{s \rightarrow 0} f_s(x) = \int_{\mathfrak{a}^*} \varphi_t(x) \hat{f}(t) \alpha(dt)$$

742 as $s \rightarrow 0$. But this limit is identical to $f(x)$, so (A.1) is proved.

743 **A.2. The only-if part.** As in Paragraph 8.1.3, the aim is to prove (8.7),

$$744 \quad (\text{A.4}) \quad \int_M (\psi * \psi^\dagger)(x) f(x) \text{vol}(dx) = \int_{\mathfrak{a}^*} |\hat{\psi}(t)|^2 \hat{f}(t) \alpha(dt)$$

which then implies the spectral representation (8.5). Since $\psi * \psi^\dagger \in C_H^c(M)$, it follows from (3.3) and (8.6) that

$$(\psi * \psi^\dagger)(x) = \int_{\mathfrak{a}^*} \varphi_t(x) |\hat{\psi}(t)|^2 \alpha(dt).$$

745 Replacing this under the integral on the left-hand side of (A.4), the right-hand side is obtained
746 by using the definition (3.2) of \hat{f} . Now, everything follows from (8.5) as in Paragraph 8.1.3.
747 ■

748 **Appendix B. The hyperbolic secant in three dimensions.** Recall the hyperbolic secant
749 kernel from 4.3.2. This is here considered in the case where M is a hyperbolic space. The aim
750 is to apply the L^2 -Godement theorem, with the help of the spherical transform pair (4.5).

Setting $g(r) = (\cosh(r))^{-a}$ and $f(x) = g(d(o, x))$, it follows that f is continuous and in $L_H^2(M)$ when $a > 1$. This is because,

$$\int_M |f(x)|^2 = 4\pi \int_0^\infty (\cosh(r))^{-2a} \sinh^2(r) dr = \int_1^\infty u^{-2a} (u^2 - 1)^{\frac{1}{2}} du < \infty$$

as follows by integrating in geodesic spherical coordinates. Thus, the L^2 -Godement theorem can be applied to f . The spherical transform \hat{f} can be found from Lemma 6.4, at least for integer values of a . Indeed, the moment-generating function G , required in the lemma, reads

$$G(s) = \int_{-\infty}^{+\infty} (\cosh(r))^{-a} e^{sr} dr.$$

751 When a is an (of course, positive) integer, this can be obtained using the residue theorem [4].
752 Precisely, if a is odd, $a = 2n + 1$,

$$753 \quad (\text{B.1}) \quad G(i\sigma) = \frac{\pi \operatorname{sech}(\pi\sigma/2)}{(2n)!} \prod_{m=0}^{n-1} [\sigma^2 + (2m+1)^2]$$

754 and if a is even, $a = 2n$,

$$755 \quad (\text{B.2}) \quad G(i\sigma) = \frac{\pi\sigma \operatorname{csch}(\pi\sigma/2)}{(2n-1)!} \prod_{m=1}^{n-1} [\sigma^2 + (2m)^2]$$

756 where $\operatorname{sech} = 1/\cosh$ and $\operatorname{csch} = 1/\sinh$ are the hyperbolic secant and cosecant, and where
757 the imaginary part of σ is ≥ -1 . With G given by (B.1) and (B.2), the spherical transform \hat{f}
758 can be computed using (6.2). Letting $H(\sigma) = G(i\sigma)$, this yields

$$759 \quad (\text{B.3}) \quad \hat{f}(t) = \frac{\pi}{t} \operatorname{Im} \{H(t-i) - H(t+i)\}.$$

760 **Lemma B.1.** *If $z \in \mathbb{C} \setminus \mathbb{Z}$ and n is a positive integer, then*

$$761 \quad \Gamma(n+z)\Gamma(n-z) = -\frac{\pi}{z \sin(\pi z)} \prod_{m=0}^{n-1} (m^2 - z^2),$$

$$762 \quad \Gamma\left(n + \frac{1}{2} + z\right) \Gamma\left(n + \frac{1}{2} - z\right) = \frac{\pi}{\cos(\pi z)} \prod_{m=0}^{n-1} \left[\left(m + \frac{1}{2}\right)^2 - z^2 \right].$$

763 *Proof.* The proof of the first equality follows by induction using Euler's reflection formula
 764 $\Gamma(z)\Gamma(1-z) = \pi/\sin(\pi z)$ for $z \notin \mathbb{Z}$ and the identity $\Gamma(z+1) = z\Gamma(z)$. The second equality
 765 follows in the same manner using the formula $\Gamma(1/2+z)\Gamma(1/2-z) = \pi/\cos(\pi z)$, which itself
 766 follows from replacing z with $z+1/2$ in Euler's reflection formula. ■

767 **Proposition B.2.** *The hyperbolic secant kernel is positive-definite in three-dimensional hy-*
 768 *perbolic space whenever a is a positive integer.*

769 *Proof.* If a is odd with $a = 2n + 1$, then

$$770 \quad H(\sigma) = G(i\sigma) = \frac{4^n \pi \operatorname{sech}(\pi\sigma/2)}{(2n)!} \prod_{m=0}^{n-1} \left[\left(m + \frac{1}{2}\right)^2 - \left(\frac{i\sigma}{2}\right)^2 \right]$$

$$771 \quad = \frac{4^n}{(2n)!} \Gamma\left(n + \frac{1}{2} + i\frac{\sigma}{2}\right) \Gamma\left(n + \frac{1}{2} - i\frac{\sigma}{2}\right)$$

772 using Lemma B.1 with $z = i\sigma/2$. Thus,

$$773 \quad H(t-i) - H(t+i) = \frac{4^n}{(2n)!} \left[\Gamma\left(n+1 + \frac{it}{2}\right) \Gamma\left(n - \frac{it}{2}\right) - \Gamma\left(n + \frac{it}{2}\right) \Gamma\left(n+1 - \frac{it}{2}\right) \right]$$

$$774 \quad = \frac{4^n}{(2n)!} \left[\left(n + \frac{it}{2}\right) \Gamma\left(n + \frac{it}{2}\right) \Gamma\left(n - \frac{it}{2}\right) - \left(n - \frac{it}{2}\right) \Gamma\left(n + \frac{it}{2}\right) \Gamma\left(n - \frac{it}{2}\right) \right]$$

$$775 \quad = i \frac{4^n t}{(2n)!} \Gamma\left(n + \frac{it}{2}\right) \Gamma\left(n - \frac{it}{2}\right) = i \frac{4^n t}{(2n)!} \left| \Gamma\left(n + \frac{it}{2}\right) \right|^2$$

776 where we have used the fact that $t \in \mathbb{R}$ and $\overline{\Gamma(z)} = \Gamma(\bar{z})$ for $z \in \mathbb{C}$. Thus, (B.3) reduces to

$$777 \quad \hat{f}(t) = \frac{4^n \pi}{(2n)!} \left| \Gamma\left(n + \frac{it}{2}\right) \right|^2 \geq 0$$

778 for any $t \in \mathbb{R}$. If a is even with $a = 2n$, a similar calculation using Lemma B.1 yields

$$779 \quad \hat{f}(t) = \frac{4^n n \pi}{(2n)!} \left| \Gamma\left(n - \frac{1}{2} + \frac{it}{2}\right) \right|^2 \geq 0$$

780 for any $t \in \mathbb{R}$, which completes the proof. ■