



HAL
open science

Equivariant log-extrinsic means on irreducible symmetric cones

Emmanuel Chevallier, Frank Nielsen

► **To cite this version:**

Emmanuel Chevallier, Frank Nielsen. Equivariant log-extrinsic means on irreducible symmetric cones. 2024. hal-04642282v1

HAL Id: hal-04642282

<https://hal.science/hal-04642282v1>

Preprint submitted on 11 Jul 2024 (v1), last revised 25 Jul 2024 (v2)

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.

Equivariant log-extrinsic means on irreducible symmetric cones

Emmanuel Chevallier

Aix Marseille University, CNRS, Centrale Marseille, Institut Fresnel

Frank Nielsen

Sony Computer Science Laboratories Inc

Abstract

In this paper, we introduce a notion of mean on irreducible symmetric cones, based on the product decomposition between the determinant one hypersurface and the determinant. Irreducible symmetric cones and their determinant on surfaces form an important class of spaces for statistics and data science, since they encompass positive definite self-adjoint operators as well as Lorentz cones and hyperbolic spaces. By construction, log-extrinsic means have similar equivariance properties as those of the Fréchet means. Moreover, the two means coincide under some symmetry assumption on the distribution. However, the log-extrinsic mean admits an explicit expression and is much simpler to compute. Numerical experiments show that the log-extrinsic means are a relevant alternative to log-Euclidean means. Furthermore, along with the log-extrinsic mean, we introduce a corresponding notion of Gaussian distributions, called log-extrinsic Gaussians. A classification experiment on stereo audio signals demonstrates the practical interest of the log-extrinsic framework.

Keywords: Symmetric cone; Jordan Euclidean algebra; Fréchet mean; log-Euclidean mean; equivariance; Gaussian-like distributions

1 Introduction

When considering statistics on a homogeneous space Arvanitogeorgos (2003) G/H , the equivariance of statistical objects with respect to the group G is a desirable property. However, in order to be used in applications, statistical objects should be computed in a reasonable time. These two considerations often results in a trade-off between these two properties. For instance, the Fréchet mean on a Riemannian homogeneous space is equivariant with respect to the isometry group, but it is sometimes replaced by non-equivariant approximations that are easier to compute, see Arsigny et al. (2006) (log-euclidean mean).

A symmetric cone \mathcal{C} is a self-dual convex open subset of an Euclidean vector space V whose group of linear automorphism $G(\mathcal{C})$ is transitive (see section 2). Symmetric cones are therefore homogeneous spaces. The most classical examples are the cones of positive definite symmetric or hermitian matrices and the Lorentz cone.

In this paper we propose a notion of mean, called *log-extrinsic mean*, on symmetric cones which both respects equivariance properties and is easy to compute. Note that the arithmetic mean on \mathcal{C} induced by the vector structure is both equivariant with respect to $G(\mathcal{C})$ and easy to compute. The automorphism group $G(\mathcal{C})$ can be decomposed as a product group of linear maps of determinant ± 1 and positive scaling factors $\mathbb{R}_{>0}$. Unlike the arithmetic mean, the Fréchet mean computed from the Riemannian structure of \mathcal{C} is not only equivariant with respect to $G(\mathcal{C})$, but also respects the orbits of the two factors of $G(\mathcal{C})$. This results in a deeper equivariance property, at the cost of more involved computations.

Given a set of samples $x_1, \dots, x_n \in \mathcal{C}$, the proposed log-extrinsic mean is defined as,

$$L(x_1, \dots, x_n) = e^{\frac{1}{rn} \sum_i \log \det x_i} \frac{\sum_i \frac{x_i}{\det(x_i)^{1/r}}}{\det \left(\sum_i \frac{x_i}{\det(x_i)^{1/r}} \right)^{1/r}}, \quad (1)$$

where r is the rank of the symmetric cone. Log-extrinsic means have properties similar to those of Fréchet means but ensures a low computational cost. In particular, it is shown that they coincide under some symmetry assumptions on the probability distribution. From a practical standpoint, the log-extrinsic means can be seen as an equivariant alternative to log-Euclidean means.

The paper is organized as follows:

Section 2 starts by a brief introduction to symmetric cones. We then describe the product structure of its automorphism groups and the corresponding orbital decomposition of the cone, as done in Massam (1994). The section ends with Eq.4, defining the log-extrinsic mean for arbitrary probability measures.

In section 3, we describe the Riemannian structure of symmetric cones and the corresponding Fréchet mean. We describe an invariant metric, making explicit the product structure on the orbital decomposition. We recall the definition of the Fréchet mean on a Riemannian manifold and show that all the invariant metrics on a symmetric cones lead to the same Fréchet mean.

In section 4, we study some properties of the log-extrinsic mean. We show that under a symmetry assumption on the probability distribution the log-extrinsic mean and the Fréchet mean are equal. We also provide a central limit theorem for log-extrinsic mean.

In section 5, we propose a type of Gaussian distributions associated with log-extrinsic means and show some of their properties. These distributions are defined as a product between a Wishart density conditioned on the hyper-surface of determinant 1 and a log-normal Gaussian.

In section 6, we show the practical relevance of the log-extrinsic means and corresponding Gaussians. We provide first convergence curves as well as computation times of empirical log-extrinsic means, Fréchet means, and log-Euclidean means. We describe then a classification experiment on the cone of 2 by 2 covariance matrices and compare the classification results using these different approaches.

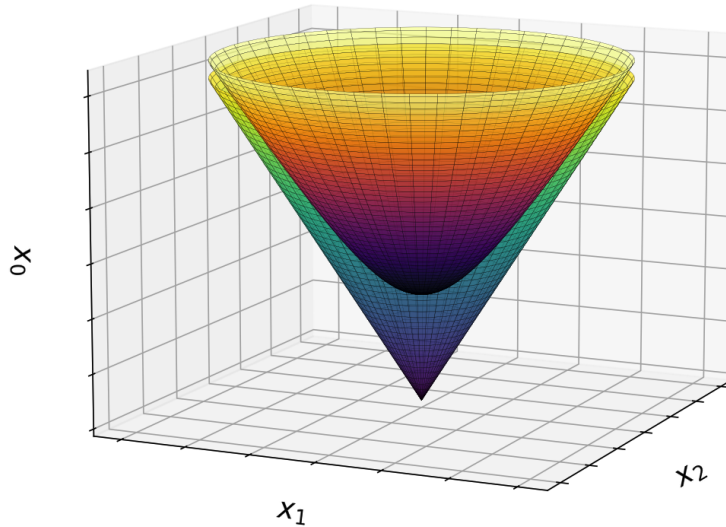


Figure 1: The Lorentz cone $L(3)$ and its determinant one surface \mathcal{S} .

2 Log-extrinsic means

This section assumes some knowledge on symmetric cone. The necessary facts about symmetric cones are summarized in the appendix 8.1. The log-extrinsic means described at the end of this section can be defined on arbitrary symmetric cones. However, most of their properties studied in this paper hold only for irreducible symmetric cones. For the sake of simplicity, we restrict the scope of this paper to irreducible symmetric cones. Hence in the rest of the paper, the letter V refers to a finite dimensional real vector space endowed with an inner product. Unless stated otherwise, the letter \mathcal{C} refers to an irreducible symmetric cone in V . The most important irreducible symmetric cones for applications in statistics and data sciences are the cones $\mathcal{P}(n, \mathbb{R})$ and $\mathcal{P}(n, \mathbb{C})$ of symmetric and Hermitian positive definite matrices of size $n \times n$ on \mathbb{R} and \mathbb{C} as well as Lorentz cones $L(n)$ for $n \geq 3$:

$$L(n) = \left\{ (x_0, \dots, x_{n-1}) \in \mathbb{R}_{>0} \times \mathbb{R}^{n-1}, \quad x_0^2 > \sum_{i=1}^{n-1} x_i^2 \right\}.$$

Lorentz cones are also called second-order cones, quadratic cones, or icecream cones in the literature, see Fig.1.

Before defining the log-extrinsic mean, let us introduce the so-called orbital parametrization of \mathcal{C} . Recall that the Jordan product associated with a symmetric cone leads to a notion of rank r of a symmetric cone, and to a determinant function. For matrix cones $\mathcal{P}(n, \mathbb{R})$ or $\mathcal{P}(n, \mathbb{C})$, $r = n$, while for Lorentz cones $L(n)$ the rank r is always 2. The determinant function obeys the following scaling rule: $\det(\lambda x) = \lambda^r \det(x)$ for any $\lambda \in \mathbb{R}$ and $x \in V$, and

is always real and strictly positive on \mathcal{C} .

Let $\mathcal{S} = \{x \in \mathcal{C} : \det x = 1\}$ be the codimension one sub-manifold of determinant one, and note π the projection on \mathcal{S} defined as

$$\pi(x) = \frac{x}{\det(x)^{\frac{1}{r}}}. \quad (2)$$

The map $\tilde{\phi} : \mathcal{C} \rightarrow \mathcal{S} \times \mathbb{R}_{>0}$ defined by

$$\tilde{\phi}(x) = (\pi(x), \det(x))$$

is a diffeomorphism. Constant coordinate sets of $\tilde{\phi}$ are shown on Fig.2 on a vertical section of $L(3)$. Let $G(\mathcal{C})$ be the Lie group of automorphisms of V which leaves \mathcal{C} invariant,

$$G(\mathcal{C}) = \{g \in \text{GL}(V), \quad g(\mathcal{C}) = \mathcal{C}\}.$$

Write $\text{Det}(g)$ the usual determinant of an element $g \in G(\mathcal{C})$. The subgroup $G(\mathcal{S})$ of $G(\mathcal{C})$ preserving \mathcal{S} is given by elements $g \in G(\mathcal{C})$ such that $|\text{Det}(g)| = 1$. The group $G(\mathcal{C})$ is isomorphic to $G(\mathcal{S}) \times \mathbb{R}_{>0}$, and the action of $G(\mathcal{C})$ on \mathcal{C} can be decomposed using $\tilde{\phi}$ as the product between an action of $G(\mathcal{S})$ on \mathcal{S} and an action of $\mathbb{R}_{>0}$ onto itself. These actions are given by

$$\forall (g, \alpha) \in G(\mathcal{S}) \times \mathbb{R}_{>0}, \forall (x, d) \in \mathcal{S} \times \mathbb{R}_{>0}, \quad (g, \alpha)(x, d) = (gx, \alpha^r d).$$

The submanifold \mathcal{S} is an orbit of the action of $G(\mathcal{S})$, and fixing the first or the second coordinate in the product $\mathcal{S} \times \mathbb{R}_{>0}$ describes respectively an orbit of the action of the $\mathbb{R}_{>0}$ component or of the $G(\mathcal{S})$ component. Hence the decomposition $\tilde{\phi} : \mathcal{C} \rightarrow \mathcal{S} \times \mathbb{R}_{>0}$ is called the orbital parametrization in Massam (1994).

The log-extrinsic mean is defined from the Cartesian product of an equivariant mean on \mathcal{S} with an equivariant mean on $\mathbb{R}_{>0}$. This results in a mean equivariant with respect to the action of $G(\mathcal{C})$ which respects the orbits of the actions of $\mathbb{R}_{>0}$ and $G(\mathcal{S})$.

In the rest of the paper, let \mathbb{E} denote the arithmetic mean or expectation. For a probability distribution μ on \mathcal{C} ,

$$\mathbb{E}(\mu) = \int_{\mathcal{C}} x \, d\mu.$$

Given a probability distribution μ on \mathcal{S} , let $E(\mu)$ be the extrinsic mean on \mathcal{S} defined as

$$E(\mu) = \pi(\mathbb{E}(\mu)). \quad (3)$$

Linear maps commute with \mathbb{E} , and it is not difficult to see that the action of $G(\mathcal{S})$ commutes with π . Hence E is equivariant under the action of $G(\mathcal{S})$. This type of extrinsic mean have been considered in the context of Lorentz cones in Galperin (1993).

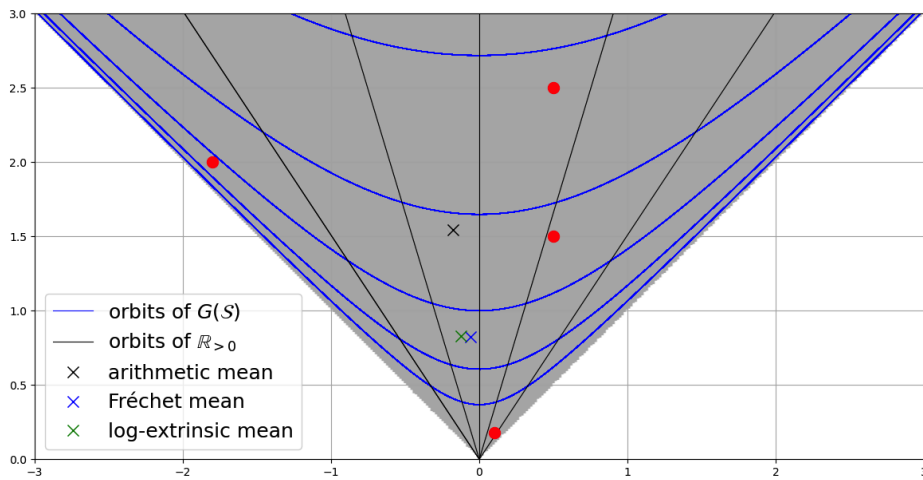


Figure 2: The grey area is the intersection between a Lorentz cone and a plane containing the coordinate x_0 . The black, blue and green crosses are respectively the arithmetic mean, the Fréchet mean, see section 3, and the log-extrinsic mean, see Eq.4, of the four red points. Recall that the log-Euclidean mean is obtained by linearizing the space using a Riemannian logarithm. It can be shown that every such plane of a Lorentz cone is a flat manifold for the invariant Riemannian metrics introduced in section 3, hence the log-Euclidean mean and the Fréchet mean coincide.

Recall that a scaling α acts on a determinant d as $\alpha.d = \alpha^r d$. Hence, there exists a variety of candidates for the mean on determinants, such as so-called power means M_p which include the geometric mean M_0 , respectively defined in the finite case by

$$M_p(d_1, \dots, d_n) = \left(\frac{1}{n} \sum_i d_i^p \right)^{1/p} \quad \text{and} \quad M_0(d_1, \dots, d_n) = e^{\frac{1}{n} \sum_i \log(d_i)},$$

with $p > 0$. Due to its link with the Riemannian structures which will be described in section 3, and its superiority in some classification experiment, see section 6, we focus on the geometric mean on determinants.

Let $H : \mathcal{C} \rightarrow \mathbb{R}$ be the map defined by

$$H(x) = \log(\det(x)),$$

and let $\phi : \mathcal{C} \rightarrow \mathcal{S} \times \mathbb{R}$ be the diffeomorphism defined by $\phi = (\pi, H)$.

Write $\pi_*\mu$ and $H_*\mu$, the pushforwards of μ by π and H respectively. The log-extrinsic mean on \mathcal{C} is then defined by

$$L(\mu) = \phi^{-1}(E(\pi_*\mu), \mathbb{E}(H_*\mu)) = e^{\frac{1}{r}\mathbb{E}(H_*\mu)} E(\pi_*\mu). \quad (4)$$

Like the arithmetic mean on \mathcal{C} , it is easy to check that L is equivariant with respect to the action of $G(\mathcal{C})$. Indeed, for $g \in G(\mathcal{S})$ and $\alpha \in \mathbb{R}_{>0}$, the equivariance on both factors of $\mathcal{C} \sim \mathcal{S} \times \mathbb{R}$ enables to write

$$L((\alpha g)_*\mu) = \phi^{-1}(gE(\pi_*(\mu)), r \log(\alpha) + \mathbb{E}(H_*\mu)) = \alpha gL(\mu).$$

However, unlike the arithmetic mean, L descends to equivariant means on $\mathcal{C}/\mathbb{R}_{>0}$ and $\mathcal{C}/G(\mathcal{S})$. Indeed, by construction of L , we have the following commutation diagrams,

$$\begin{array}{ccc} \mathbb{P}(\mathcal{C}) & \xrightarrow{L} & \mathcal{C} \\ \pi_* \downarrow & & \downarrow \pi \\ \mathbb{P}(\mathcal{S}) & \xrightarrow{E} & \mathcal{S} \sim \mathcal{C}/\mathbb{R}_{>0} \end{array} \quad , \quad \begin{array}{ccc} \mathbb{P}(\mathcal{C}) & \xrightarrow{L} & \mathcal{C} \\ H_* \downarrow & & \downarrow H \\ \mathbb{P}(\mathbb{R}) & \xrightarrow{\mathbb{E}} & \mathbb{R} \sim \mathcal{C}/G(\mathcal{S}) \end{array} ,$$

where $\mathbb{P}(X)$ is the set of probability measures of the measurable space X .

Let us make three remarks. First the Jordan product on an irreducible symmetric cone is unique only up to a scaling factor, and consequently, so is the determinant function. It can be checked that the definition of log-extrinsic means is independent of this scaling factor.

Second, any section \mathcal{S}' of \mathcal{C} could serve as representatives of the quotient $\mathcal{C}/\mathbb{R}_{>0}$, in order to define an extrinsic mean on $\mathcal{C}/\mathbb{R}_{>0}$. Another natural candidate is the section given by

$\text{trace}(x) = 1$. Though, it can be checked that the extrinsic mean defined in this way on $\mathcal{C}/\mathbb{R}_{>0}$ is not equivariant with respect to the action of linear maps. We conjecture that the only sections leading to an extrinsic mean which is equivariant with respect to the action of linear maps, are proportional to \mathcal{S} . They correspond to different choice of scaling factors in the Jordan product and lead to the same log-extrinsic mean.

Third, on cones of positive definite matrices, a notion of geometric mean has been formalised in Ando et al. (2004) as those verifying a set of axioms, known as the ALM-axioms. Our mean does not verify the first axiom (consistency with scalars), and is therefore not a geometric mean, unlike the Fréchet mean introduced in the next section, see Yamazaki (2013).

3 Riemannian metrics and Fréchet means

The groups $G(\mathcal{C})$ and $G(\mathcal{S})$ are transitive on \mathcal{C} and \mathcal{S} respectively. The cone \mathcal{C} and the section \mathcal{S} are hence homogeneous manifolds. Beyond homogeneity, it can be proved that they are simply connected complete Riemannian symmetric spaces of non-positive curvature Sasaki (1980); Rothaus (1960). Recall that on such spaces, the Riemannian exponentials at each point are bijections and the Riemannian logarithm is well defined.

Let us start by describing the invariant metrics. The characteristic function $\varphi : \mathcal{C} \rightarrow \mathbb{R}$ of a homogeneous cone is defined by

$$\varphi(x) = \int_{\mathcal{C}} e^{-\langle x, y \rangle} dy$$

where dy is the Lebesgue volume element in V . For $g \in G(\mathcal{C})$, we have

$$\varphi(gx) = |\det(g)^{-1}| \int_{\mathcal{C}} e^{-\langle x, y \rangle} dy.$$

Hence φ is invariant under $G(\mathcal{S})$. A calculation shows that the Hessian of $\log(\varphi)$ is positive definite, see Faraut and Korányi (1994) Proposition I.3.3. Hence its restriction to the submanifold \mathcal{S} defines an invariant Riemannian metric on \mathcal{S} . Denote $d_{\mathcal{S}}^{\varphi}$ the associated distance function. The following theorem describes the invariant distances on \mathcal{C} .

Theorem 1. *Let \mathcal{C} be an irreducible symmetric cone. Any invariant Riemannian distance on \mathcal{S} is of the form $d_{\mathcal{S}} = \alpha d_{\mathcal{S}}^{\varphi}$ for some $\alpha \in \mathbb{R}_{>0}$. Furthermore, any invariant Riemannian distance $d_{\mathcal{C}}$ on \mathcal{C} can be decomposed as a product metric*

$$d_{\mathcal{C}}(x, y)^2 = \alpha d_{\mathcal{S}}^{\varphi}(\pi(x), \pi(y))^2 + \beta(H(x) - H(y))^2, \quad (5)$$

where $\alpha, \beta \in \mathbb{R}_{>0}$.

Though this result should be known, we are not aware of a reference mentioning it and we provide a proof in appendix 8.2.

Note that the choice of an arithmetic mean on the logarithm of the determinants in the previous section, is compatible with the logarithmic behavior of the distance on determinants.

Let us recall the definition of the Fréchet mean on an arbitrary Riemannian manifold (\mathcal{M}, d) . Let μ be a probability distribution on \mathcal{M} and $\mathcal{F}_\mu : \mathcal{M} \rightarrow \mathbb{R} \cup \{\infty\}$ be the function defined by

$$\mathcal{F}_\mu(x) = \int_{\mathcal{M}} d(x, y)^2 d\mu(y).$$

Definition 1. Let F be the functional defined by

$$F(\mu) = \operatorname{argmin}_{x \in \mathcal{M}} \mathcal{F}_\mu(x).$$

When $F(\mu)$ contains a single element m , we abuse notation and write $F(\mu) = m$. $F(\mu)$ is then called the Fréchet mean of μ .

Let s be an isometry of \mathcal{M} . Clearly, the Fréchet mean commutes with s ,

$$s(F(\mu)) = F(s_*\mu).$$

As we will see in the next section, when the distribution μ has enough symmetries, the above commutation relation enables to determine its mean.

The uniqueness of the mean strongly depends on the geometry of the manifold. When the curvature is non-positive, the function \mathcal{F} has remarkable properties leading to the following result.

Theorem 2. Let (\mathcal{M}, d) be $(\mathcal{S}, d_{\mathcal{S}})$ or $(\mathcal{C}, d_{\mathcal{C}})$. Let μ be a probability measure on (\mathcal{M}, d) . If $\mathcal{F}_\mu(x)$ has a finite value for some x in \mathcal{M} , then it is uniformly convex and has a unique minimum. Since \mathcal{F}_μ is differentiable, the Fréchet mean $F(\mu)$ is the only solution of

$$\nabla(\mathcal{F}_\mu)_x = 2 \int_{\mathcal{M}} \log_x(y) d\mu(y) = 0.$$

The proof of the uniqueness of the Fréchet mean as well as the definition of uniform convexity can be found in Sturm (2003) (proposition 4.3) in the more general setting of metric spaces. It shows that when the Fréchet mean exists it can be obtained by a gradient descent. Recall that the Riemannian logarithms on \mathcal{S} and \mathcal{C} are defined globally, and that the gradient of the squared distance function is given by $\nabla_x d(x, y)^2 = 2 \log_x(y)$. The expression of $\nabla(\mathcal{F}_\mu)_x$ is obtained by interchanging derivation and integral.

According to theorem 1, the function \mathcal{F}_μ can be expressed as the following sum,

$$\mathcal{F}_\mu(x) = \alpha \int_{\mathcal{S}} d_{\mathcal{S}}^{\mathcal{C}}(\pi(x), \pi(y))^2 d\pi_*\mu(y) + \beta \int_{\mathbb{R}} (H(x) - y)^2 dH_*\mu(y). \quad (6)$$

This decomposition of \mathcal{F}_μ enables to show that given a probability distribution μ supported on \mathcal{S} , the Fréchet mean computed in \mathcal{S} or in \mathcal{C} coincide,

$$\operatorname{argmin}_{x \in \mathcal{S}} \mathcal{F}_\mu(x) = \operatorname{argmin}_{x \in \mathcal{C}} \mathcal{F}_\mu(x).$$

More generally, theorem 1 implies that \mathcal{S} is a totally geodesic submanifold of \mathcal{C} . Another consequence inferred from Eq.6 is that the Fréchet mean for a measure μ on \mathcal{C} is independent of α and β , and hence of the choice of the invariant Riemannian metric. Finally, it can be checked using Eq.6 that similarly to the log-extrinsic mean, the Fréchet mean can be expressed as

$$F(\mu) = \phi^{-1}(F(\pi_*\mu), \mathbb{E}(H_*\mu)) = e^{\frac{1}{r}\mathbb{E}(H_*\mu)} F(\pi_*\mu). \quad (7)$$

4 Properties of log-extrinsic means

4.1 Extrinsic means on \mathcal{S}

4.1.1 Extrinsic means and Fréchet means

Note K_x the subgroup of the identity component of $G(\mathcal{C})$ that leave the point $x \in \mathcal{C}$ fixed. The following important result is a direct consequence of proposition III.4.1 of Faraut and Korányi (1994).

Theorem 3. *Let \mathcal{C} be an irreducible symmetric cone. For all $x, y \in \mathcal{C}$, if y is fixed by K_x , then there exists $\alpha > 0$ with $y = \alpha x$.*

Proof. The proposition III.4.1 of Faraut and Korányi (1994) states a similar result in the context of simple Euclidean Jordan algebras, which are in direct correspondence with irreducible symmetric cones. \square

As a corollary, we get the following results relating the extrinsic mean and the Fréchet mean.

Corollary 1. *Let \mathcal{C} be an irreducible symmetric cone, and μ be a K_x -invariant measure for some $x \in \mathcal{S}$, i.e.,*

$$\forall k \in K_x, k_*\mu = \mu.$$

When the means $E(\mu)$ and $F(\mu)$ exist

$$E(\mu) = F(\mu) = x.$$

Proof. Being at the same time linear maps and Riemannian isometries, the elements of K_x commute with the log-extrinsic mean as well as the Fréchet mean. Hence, theorem 3 states that $E(\mu)$ and $F(\mu)$ are proportional to x . Since the three elements are of determinant 1, they must be equal. \square

It is known that an homogeneous space G/H , with G a locally compact group and H a compact subgroup of G , admits a unique G -invariant measure up to a scaling factor (see Wijsman (1990), Chapter 7, Corollary 7.4.4). It follows that a symmetric cone \mathcal{C} , (resp. its section \mathcal{S}), admits a unique $G(\mathcal{C})$ -invariant measure (resp. $G(\mathcal{S})$ -invariant measure) up to a scaling factor.

The invariant measure on \mathcal{S} is easy to construct. Given a measurable set $A \subset \mathcal{S}$, note $\tilde{A} = \{tx, t \in]0, 1], x \in A\}$. Since the elements of $G(\mathcal{S})$ are of determinant ± 1 , the Lebesgue measure in V is $G(\mathcal{S})$ -invariant and $v_{\mathcal{S}}(A) = \text{Leb}(\tilde{A})$ is an invariant measure. In the rest of the paper, densities on \mathcal{S} are always taken with respect to $v_{\mathcal{S}}$.

Corollary 2. *Let \mathcal{C} be an irreducible symmetric cone. Let μ be a probability distribution on \mathcal{S} with a density f such that there exists a function $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and a point x in \mathcal{S} with*

$$f(y) = h(d_{\mathcal{S}}(x, y)).$$

When the means exist,

$$E(\mu) = F(\mu) = x.$$

Proof. As a function of y , $d_{\mathcal{S}}(x, y)$ is K_x -invariant. Hence the result is implied by corollary 1. \square

These results are similar to theorem 3.3 of Bhattacharya and Patrangenaru (2003), where authors consider extrinsic means on a Riemannian manifold embedded in \mathbb{R}^n , as the orthogonal projection for the Euclidean distance, of the arithmetic mean on the embedding. Although the projection π on \mathcal{S} is not the orthogonal projection for the Euclidean distance, the result is obtained using similar arguments.

To analyse the symmetries of probability distributions in concrete examples, it is important to have characterisations of the automorphism groups for each cone. Let us describe $G(\mathcal{C})$ and K_x for $L(n)$, $\mathcal{P}(n, \mathbb{R})$ and $\mathcal{P}(n, \mathbb{C})$.

The automorphism group $G(L(n))$ of $L(n)$ is isomorphic to $O^+(1, n-1) \times \mathbb{R}_{>0}$ where $O^+(1, n-1)$ is the group preserving the Minkowski metric of signature $(+ - \dots -)$ and the sign of the first coordinate, and $\mathbb{R}_{>0}$ is the multiplicative group acting as homothety. Denote $o = (1, 0, \dots, 0)^T \in L(n)$. It is easy to check that K_o is the group of matrices of the form

$$M = \begin{pmatrix} 1 & 0 \\ 0 & R \end{pmatrix}$$

where $R \in \text{SO}(n-1)$. At an arbitrary point $x \in L(n)$, K_x is obtained by a conjugation of K_o , $K_x = gK_o g^{-1}$, with any $g \in G(\mathcal{C})$ such that $x = go$.

For families $\mathcal{P}(n, \mathbb{R})$ and $\mathcal{P}(n, \mathbb{C})$, the automorphism groups are fully described by congruence operations. For $\mathbb{K} = \mathbb{R}$ or \mathbb{C} , let Γ be the map defined as

$$\begin{aligned} \Gamma : \text{GL}(n, \mathbb{K}) &\rightarrow G(\mathcal{P}(n, \mathbb{K})) \\ A &\mapsto (X \mapsto AXA^*) \end{aligned}$$

where $*$ is the transpose or the conjugate transpose. When $\mathbb{K} = \mathbb{R}$, $G(\mathcal{P}(n, \mathbb{R}))$ is the image of Γ , see Dolcetti and Pertici (2018); Orlitzky (2024). When $\mathbb{K} = \mathbb{C}$, $G(\mathcal{P}(n, \mathbb{C}))$ is the group generated by the image of Γ and the conjugation map, see Schneider (1965); Orlitzky (2024). At the identity matrix I , it is possible to check that K_I is the image by Γ of $\text{SO}(n)$ in the real case, and of $\text{SU}(n)$ in complex cases. At an arbitrary X , K_X is the image by Γ of $\text{ASU}(n)A^{-1}$, for any A invertible with $X = AA^*$. Since X admits a symmetric square root we can choose $A = X^{\frac{1}{2}}$.

In the particular case where $\mathcal{C} = L(n)$, the section

$$\mathcal{S} = \left\{ (x_0, \dots, x_{n-1}) \in \mathbb{R}_{>0} \times \mathbb{R}^{n-1}, x_0^2 - \sum_{i=1}^{n-1} x_i^2 = 1 \right\}$$

is a hyperbolic space of dimension n (hyperboloid surface). Hyperbolic spaces are symmetric space with a maximal amount of symmetries. This enables to restate the previous corollaries with a weaker symmetry hypothesis on the measure μ . For $x \in \mathcal{S}$, define s_x as a conjugation of the Jordan inverse by an element $g \in G(\mathcal{S})$ such that $gx = o$,

$$gs_x(y) = (gy)^{-1},$$

It can be checked that the Jordan inverse preserves \mathcal{S} , hence s_x preserves \mathcal{S} . Furthermore, it can be proved that x is the only fixed point of s_x and that the definition does not depend on the choice of g . This involution is often called the “geodesic symmetry” at x because it reverses the time parameter of the Riemannian geodesics emanating from x ,

$$s_x(\gamma(t)) = \gamma(-t),$$

where γ is a geodesic curve with $\gamma(0) = x$.

Theorem 4. *Let \mathcal{C} be a Lorentz cone. Assume that a distribution μ on \mathcal{S} is s_x -invariant, for some $x \in \mathcal{S}$. When the means exist,*

$$E(\mu) = F(\mu) = x.$$

Proof. On a Lorentz cone, the Jordan inverse is given by $x^{-1} = \det(x)^{-1}(x_0, -x_1, \dots, -x_{n-1})$. Hence on \mathcal{S} , the Jordan inverse is the restriction of the linear map $k((x_0, \dots, x_{n-1})) = (x_0, -x_1, \dots, -x_{n-1})$. Using the definition of the determinant given in appendix 8.1, it is clear that $k \in G(\mathcal{S})$. Since $s_x = g^{-1}kg$ for any g with $gx = o$, $s_x \in G(\mathcal{S})$. Hence s_x is an isometry and commutes both with the log-extrinsic mean as well as the Fréchet mean. Since x is the only fix point of s_x , the result holds. \square

Fig.3 gives an illustration of the result for the average of two points. Note that this result does not hold on arbitrary cones because the geodesic symmetries on \mathcal{S} are usually

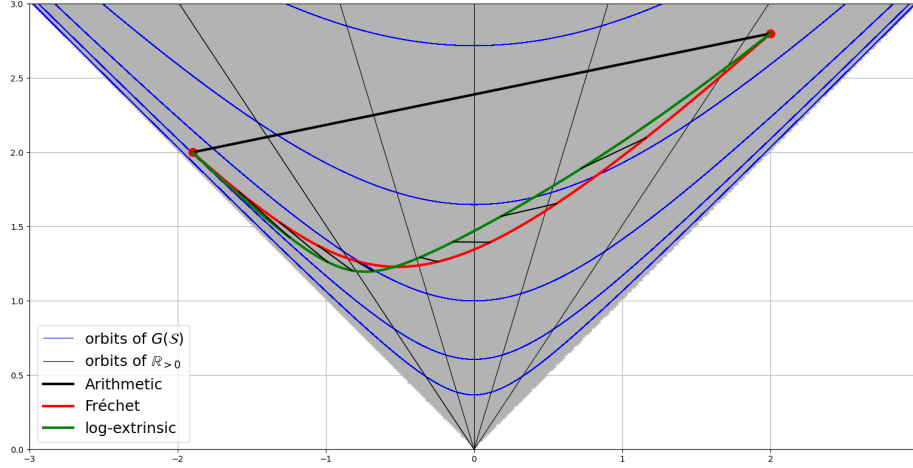


Figure 3: As in Fig.1, consider a vertical section of a Lorentz cone. The black, red and green curves are weighted averages of the two red points, with weights varying between 0 and 1, for the arithmetic mean, the Fréchet mean and the log-extrinsic mean respectively. The red curve described by the Fréchet mean is the Riemannian geodesic. The black segments show the correspondences between the Fréchet and the log-extrinsic means for identical weights. The two curves intersect for weights equal to 0.5. In that case, the Fréchet mean is the mid-point of the Riemannian geodesic, and the distribution is invariant under the geodesic symmetry at the Fréchet mean. By Theorem 4, the log-extrinsic mean and the Fréchet mean coincide.

not restrictions of linear maps. This enable to relax the hypothesis of corollary 2, to densities with an anisotropy encoded by a quadratic form in the tangent space at the mean. Recall that hyperbolic spaces have a well defined logarithm attached at each point, defined as the inverse of the exponential map.

Corollary 3. *Let \mathcal{C} be a Lorentz cone. Let $x \in \mathcal{S}$ and Q be a positive definite quadratic form in $T_x\mathcal{S}$. Let f be a probability density on \mathcal{S} of the form*

$$f(y) = h(Q(\log_x(y))),$$

where h is a function $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$. When the means exist,

$$E(\mu) = F(\mu) = x.$$

Proof. Since s_x is a geodesic symmetry, $\log_x(s_x(y)) = -\log_x(y)$. Hence the density f is s_x -invariant, and the result is a particular case of theorem 4. \square

4.1.2 A central limit theorem

The next theorem is a central limit theorem (CLT) for a sequence of i.i.d. random variables X_i with values in \mathcal{S} , without symmetry assumptions about their distribution. The CLT for random variables valued in a vector space involves a homogeneous dilation of the space by a factor \sqrt{n} , n being the number of random variables. However on manifolds, homogeneous dilations have no intrinsic definition. That is why, CLT on manifolds are usually expressed in a coordinate system Bhattacharya and Patrangenaru (2003, 2005).

Since $\det tx = t^r \det x$ for all $x \in \mathcal{C}$ and $t > 0$, we have $V = T_x \mathcal{S} \oplus \mathbb{R}.x$ for all $x \in \mathcal{S}$. Denote $P_x : V \rightarrow T_x \mathcal{S}$ the projection along $\mathbb{R}.x$ onto $T_x \mathcal{S}$. To express the CLT, we project \mathcal{S} to the tangent space attached to the extrinsic mean $E(X_i)$.

Theorem 5. *Let $(X_i)_{i \in \mathbb{N}}$ be a sequence random variables i.i.d. with values in \mathcal{S} of law μ and with extrinsic mean ϵ . Suppose that the Euclidean norms of the variable, $\|X_i\|$, are square integrable and consider the covariance Σ :*

$$\Sigma = \int_{\mathcal{S}} (x - \mathbb{E}(\mu)) \otimes (x - \mathbb{E}(\mu)) d\mu(x).$$

Denote $\Sigma_{|T}$ the marginalisation of the covariance Σ on $T_\epsilon \mathcal{S}$ with respect to the decomposition $V = T_\epsilon \mathcal{S} \oplus \mathbb{R}\epsilon$. Let $\Delta_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ be the empirical measure and let E_n be the empirical extrinsic mean,

$$E_n = E(\Delta_n).$$

Then we have,

$$\sqrt{n} P_\epsilon(E_n) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\Sigma_{|T}}{\det(\mathbb{E}(\mu))^{2/r}} \right).$$

where P_ϵ is the projection onto $T_\epsilon \mathcal{S}$ parallel to $\mathbb{R}\epsilon$.

The proof is given in appendix 8.3.

4.2 Log-extrinsic means on symmetric cones

As seen in section 2, the log-extrinsic mean on \mathcal{C} is constructed from the extrinsic mean on \mathcal{S} and an arithmetic mean on the logarithm of determinants. We do not express explicitly the CLT of log-extrinsic means on \mathcal{C} since it is readily obtained from the CLT for extrinsic means on \mathcal{S} and the CLT for arithmetic means on \mathbb{R} . The following points are reformulations of the results of section 4.1 for the log-extrinsic means. Points 2-3-4) rely on the following fact. For $p, q \in \mathbb{R}$, let $s_p^{\det}(q) = p + (p - q)$ be the symmetry with respect to p . For μ a distribution on \mathcal{C} , if $H_* \mu$ is s_p^{\det} -invariant, then when it exists, $\mathbb{E}(H_* \mu) = p$.

- 1) Let μ be a K_x -invariant measure on a irreducible symmetric cone \mathcal{C} , for some $x \in \mathcal{C}$,

$$\forall k \in K_x, \quad k_* \mu = \mu.$$

When the means exists,

$$L(\mu) = F(\mu) = e^{\frac{1}{r}\mathbb{E}(H_*\mu)}\pi(x). \quad (8)$$

Eq.8 can be derived from Eq.4, Eq.7 and corollary 1, after noting that $K_x = K_{\pi(x)}$. Remark : the symmetry under K_x does not impose constraints on the distribution of determinants. Hence the determinant of the means can be different from the determinant of x .

- 2) Note λ the canonical Lebesgue measure on \mathbb{R} and $v_{\mathcal{C}} = \phi_*^{-1}(v_{\mathcal{S}} \otimes \lambda)$. It can be checked that $v_{\mathcal{C}}$ is an invariant measure on \mathcal{C} . Let μ be a probability distribution on \mathcal{C} , with a density f of the form

$$f(y) = h(d(x, y)),$$

where $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, d is an invariant Riemannian distance on \mathcal{C} , and $x \in \mathcal{C}$. When the means exist, Eq. 8 holds since f is K_x -invariant. Using the product decomposition of invariant Riemannian distances given in Theorem 1, it is possible to check that $H_*\mu$ is $s_{H(x)}^{\det}$ -invariant. Hence $\mathbb{E}(H_*\mu) = H(x)$ and

$$L(\mu) = F(\mu) = x.$$

- 3) Let \mathcal{C} be a Lorentz cone. For x in \mathcal{C} , the geodesic symmetry $s_x^{\mathcal{C}}$ of the Riemannian symmetric space \mathcal{C} is given by $s_x^{\mathcal{C}} = \phi^{-1} \circ (s_{\pi(x)}, s_{H(x)}^{\det}) \circ \phi$, where $s_{\pi(x)}$ is the geodesic symmetry on \mathcal{S} at $\pi(x)$. Assume that a distribution μ on \mathcal{C} is $s_x^{\mathcal{C}}$ -invariant, for some $x \in \mathcal{C}$. We have that $H_*\mu$ is $s_{H(x)}^{\det}$ invariant and that $\pi_*\mu$ is $s_{\pi(x)}$ -invariant. By theorem 4 and the fact that $\mathbb{E}(H_*\mu) = H(x)$, when the means exist,

$$L(\mu) = F(\mu) = x.$$

- 4) Let \mathcal{C} be a Lorentz cone, let $x \in \mathcal{C}$ and let Q be a positive definite quadratic form on $T_x\mathcal{C} = V$. Let f be a probability density on \mathcal{C} of the form,

$$f(y) = h(Q(\log_x(y))).$$

where h is a function $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and \log_x is the Riemannian logarithm at x . Recall that $\log_x(s_x^{\mathcal{C}}(y)) = -\log_x(y)$, hence f is $s_x^{\mathcal{C}}$ -invariant and when the means exist,

$$L(\mu) = F(\mu) = x.$$

5 Gaussian like distributions

Beyond the CLT, there exists several characterizations of Gaussian distributions on the real line that are relevant for statistical inference. We state informally two of them:

- i)* Given the mean and the variance, the Gaussian distribution is the unique distribution maximizing the differential entropy: The set of Gaussians forms an exponential family Efron (2022).
- ii)* At fixed variance, Gaussian distributions form a location family parametrized by the mean. It is the only location family such that the maximum likelihood estimation of the location parameter from i.i.d. samples is the empirical mean, see for instance Jaynes (2003).

Most of the characterizations of Gaussian distributions rely on a notion of mean. We propose here a family of log-extrinsic Gaussian distributions compatible with *i)* and *ii)* for the log-extrinsic means. We first define extrinsic Gaussian distributions on the section \mathcal{S} , and extend them to the cone by taking a product with a log-Gaussian distribution on determinants.

5.1 Extrinsic Gaussians on \mathcal{S}

Let $\theta \in V$. When it exists, define $Z(\theta)$ as

$$Z(\theta) = \int_{\mathcal{S}} e^{-\langle \theta, x \rangle} dv_{\mathcal{S}}(x).$$

Using the definition of $v_{\mathcal{S}}$ given in section 4.1.1, we have that

$$Z(\theta) = \int_{\mathcal{S}} e^{-\langle \theta, x \rangle} dv_{\mathcal{S}}(x) \leq \int_{\cup_{0 \leq t \leq 1} t\mathcal{S}} e^{\langle \theta, x \rangle} dLeb(x) \leq \int_{\mathcal{C}} e^{-\langle \theta, x \rangle} dLeb(x).$$

Since symmetric cones are self dual, it is possible to check that $Z(\theta)$ is properly defined if and only if $\theta \in \mathcal{C}$. For $\theta \in \mathcal{C}$, let $f_{\theta}(x) : \mathcal{S} \rightarrow \mathbb{R}_+$ be the probability density on \mathcal{S} defined as

$$f_{\theta}(x) = \frac{1}{Z(\theta)} e^{-\langle \theta, x \rangle}. \tag{9}$$

Denote $\mathcal{M}_{\mathcal{S}} = \{f_{\theta}, \theta \in \mathcal{C}\}$ the corresponding exponential family. These distributions have been considered in Massam (1994), as Wishart distributions conditioned by the value of the determinant with $\det(X) = 1$. Note that Wishart distributions are traditionally defined on $\mathcal{P}(n, \mathbb{R})$ or $\mathcal{P}(n, \mathbb{C})$, but they can be defined more generally on arbitrary symmetric cones, see for instance Faraut and Korányi (1994); Massam (1994); Casalis and Letac (1996). When \mathcal{C} is a Lorentz cone, \mathcal{S} is a connected component of the unit sphere for the pseudo inner

product of signature $(+ - \dots -)$. Hence these are the analogues of the von Mises-Fisher distributions for the pseudo-inner product of signature $(+ - \dots -)$ and have been studied in Barndorff-Nielsen (1978); Jensen (1981); Barndorff-Nielsen et al. (1982); Massam (1994).

Denote \mathbb{E}_θ the usual barycenter

$$\mathbb{E}_\theta = \int x f_\theta(x) dv_{\mathcal{S}}.$$

and define

$$\epsilon = E(f_\theta) = \pi(\mathbb{E}_\theta) \quad \text{and} \quad s = \det(\mathbb{E}_\theta)^{\frac{1}{r}} - 1, \quad (10)$$

where $r = \text{rank } \mathcal{C}$. We have then $\mathbb{E}_\theta = (s+1)\epsilon$. The family of densities f_θ form an exponential family. It follows that the densities f_θ can also be parametrized by their expectations, and hence by ϵ and s .

For $g \in G(\mathcal{S})$, note g_*f the pushforward of the density f by g . Note that since g is $v_{\mathcal{S}}$ -preserving, the pushforward of functions and density functions are identical, i.e., $g_*f = f \circ g^{-1}$.

Theorem 6. *We have:*

1. For all $g \in G(\mathcal{S})$

$$g_*f_\theta = f_{g^{-1}\theta} \quad \text{and} \quad g_*f_{\epsilon,s} = f_{g\epsilon,s}.$$

2. The map $\theta \mapsto (\epsilon, s)$ is a bijection from \mathcal{C} to $\mathcal{S} \times \mathbb{R}_{>0}$ with

$$\epsilon = \pi(\theta^{-1}) \quad \text{and} \quad s = \alpha(\det(\theta)^{\frac{1}{r}}) - 1,$$

where $\alpha : \mathbb{R}_{>0} \rightarrow]1, \infty]$ is a bijective decreasing function.

The proof is given in appendix 8.4. On the one hand, when $s \rightarrow \infty$ it is possible to check that the volume $\int_A f_{\epsilon,s} dv_{\mathcal{S}}$ of any compact set $A \subset \mathcal{S}$ tends to 0. On the other hand, when $s \rightarrow 0$, the density $f_{\epsilon,s}$ concentrates around ϵ . Hence ϵ and s play the role of a location and a scale parameter.

By deriving the integral defining $Z(\theta)$, it is easy to check that the expectation $(s+1)\epsilon$ and θ are related by

$$(s+1)\epsilon = -\nabla \log(Z(\theta)).$$

The expression of the normalizing factor $Z(\theta)$ provided in Massam (1994) for an arbitrary symmetric cone involves a generalized hypergeometric function. In general, deriving or approximating the function $(\epsilon, s) \mapsto \theta$ might not be a trivial task. It is nonetheless the case when $\mathcal{C} = \mathcal{P}(2, \mathbb{R}) \sim L(3)$, where $(\epsilon, s) \mapsto \theta$ admits a remarkably simple expression.

Theorem 7. *When $\mathcal{C} = \mathcal{P}(2, \mathbb{R}) \sim L(3)$,*

$$Z(\theta) = \det(\theta)^{-1} e^{-\det(\theta)} \quad \text{and} \quad \theta = (s\epsilon)^{-1}$$

Proof. The expression of $Z(\theta)$ can be found in Nielsen and Okamura (2023) and the relation $\theta = \frac{1}{s}\epsilon^{-1}$ is derived in appendix 8.5. \square

The differential entropy of a density f on \mathcal{S} is defined as

$$S(f) = \int f(x) \log(f(x)) dv_{\mathcal{S}}(x).$$

The following theorem shows that the family of densities $f_{\epsilon,s}$ admits characterizations analogous to *i*).

Theorem 8 (Maximum entropy density). *Given a location parameter $\epsilon \in \mathcal{S}$ and a scale parameter $s \in \mathbb{R}_{>0}$, the density $f_{\epsilon,s}$ is the unique maximum of S among probability densities f on \mathcal{S} verifying the constraint*

$$(s + 1)\epsilon = \int x f(x) dv_{\mathcal{S}}.$$

This is a general result on exponential families, see for instance Efron (2022). Recall that in exponential families, the maximum likelihood estimator of the expectation parameter, given independent samples, is the empirical vector mean. Hence the maximum likelihood estimators of ϵ and s are readily derived from Eq.10. The characterization *ii*) of Gaussians does not involve the full location-scale model of Gaussians but only location models at fixed scale. For $s > 0$, consider the location model $\mathcal{M}_{\mathcal{S}}^s = \{f_{\epsilon,s}, \epsilon \in \mathcal{S}\}$. The next theorem relates such models and the characterization *ii*) of Gaussians. Note $\hat{\epsilon}_{MLE}$ the maximum likelihood estimator (MLE), given independent samples, of the location parameter ϵ .

Theorem 9 (Location family and MLE). *Let $x_1, \dots, x_n \in \mathcal{S}$ and $\Delta_n = \frac{1}{n} \sum_{i=0}^n \delta_{x_i}$ be the corresponding empirical measure. The MLE of ϵ in the model $\mathcal{M}_{\mathcal{S}}$ is the extrinsic mean,*

$$\hat{\epsilon}_{MLE}(x_1, \dots, x_n) = E(\Delta_n).$$

Call location model a set of densities f_p parametrized by a point $p \in \mathcal{S}$ such that for all $g \in G(\mathcal{S})$, $f_{gp} = g_ f_p$. When \mathcal{C} is a Lorentz cone, the models $\mathcal{M}_{\mathcal{S}}^s$ for $s > 0$ are the only location models with continuously differentiable density functions such that the MLE coincides with the extrinsic mean.*

The proof is given in appendix 8.6. We are currently unaware whether or not this is a characterizing property of $\mathcal{M}_{\mathcal{S}}$ when the cone is not Lorentzian. The proof of uniqueness on Lorentz cones is an adaptation of the Gauss derivation of Gaussians provided in Jaynes (2003). By turning hyperbolic trigonometric functions into usual trigonometric functions, it is possible to check that the characterisation result also holds for von Mises-Fisher distributions on spheres. We are not aware of an existing reference showing this characterising property, though it is probably not new due to the importance of von Mises-Fisher distributions for directional statistics.

5.2 Log-extrinsic Gaussians on symmetric cones

Let $g_{m,\sigma}$ be the Gaussian density on \mathbb{R} ,

$$g_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Building on the exponential family $\mathcal{M}_{\mathcal{S}}$, define $f_{(\epsilon,s,m,\sigma)}$ on \mathcal{C} as a product between $f_{(\epsilon,s)}$ and a Gaussian on the logarithm of the determinant : $f_{(\epsilon,s,m,\sigma)} = \phi_*^{-1}(f_{(\epsilon,s)} \otimes g_{(m,\sigma)})$,

$$f_{(\epsilon,s,m,\sigma)}(x) = f_{(\epsilon,s)}(\pi(x)) g_{m,\sigma}(H(x)). \quad (11)$$

Defined in this way, $f_{(\epsilon,s,m,\sigma)}$ is a probability density with respect to $\nu_{\mathcal{C}}$, defined in section 4.2. Without loss of information, the locations parameters ϵ and m can be merged in a global location parameter $l = e^{\frac{m}{r}} \epsilon$. It is easy to check that l is the log-extrinsic mean of $f_{(l,s,\sigma)}$,

$$L(f_{(l,s,\sigma)}) = l.$$

Since the maximum entropy and the MLE properties holds for both $f_{(\epsilon,s)}$ and $g_{(m,\sigma)}$, theorems 8 and 9 have their counter parts on \mathcal{C} .

- i)* Given l , s and σ , $f_{(l,s,\sigma)}$ is the maximum entropy distribution on \mathcal{C} .
- ii)* Consider the location model $\mathcal{M}_{(s,\sigma)} = \{f_{(l,s,\sigma)}, l \in \mathcal{C}, s, \sigma \in \mathbb{R}_{>0}\}$. Let X_1, \dots, X_n , be i.i.d. random variables valued on \mathcal{C} , and let $\Delta_n = \sum_{i=0}^n \delta_{X_i}$ be the corresponding empirical measure. The MLE estimator of l in $\mathcal{M}_{s,\sigma}$ is the extrinsic mean,

$$\widehat{l}_{MLE}(X_1, \dots, X_n) = L(\Delta_n).$$

When \mathcal{C} is a Lorentz cone, $\mathcal{M}_{s,\sigma}$ is the only location model which verifies this property.

6 Numerical experiments

6.1 Empirical means convergence curves

We show here convergence curves for the empirical log-extrinsic, Fréchet and log-Euclidean means in several scenarios. Recall that the log-Euclidean mean of a distribution μ on $\mathcal{P}(n, \mathbb{R})$ is defined as

$$LEc(\mu) = \exp \left(\int_{\mathcal{P}(n, \mathbb{R})} \log(X) d\mu(X) \right), \quad (12)$$

where \exp and \log are the matrix exponential and logarithm. Since the Riemannian logarithm at the identity and the matrix logarithm coincide, the log-Euclidean mean can be seen

as a linear approximation of the Fréchet mean in the tangent space at the identity matrix. The log-Euclidean mean is popular for practical applications due to its low computational complexity.

We focus on two different types of distributions: Wishart distributions and exponential-wrapped normals on $\mathcal{P}(n, \mathbb{R})$, and set the dimension n to 2 and 20.

Define first Wishart distributions on $\mathcal{P}(n, \mathbb{R})$ and compute their means. Let $(Y_i)_{i \in \mathbb{N}}$ be i.i.d. random vectors on \mathbb{R}^n following a normal distribution of mean 0 and covariance Σ . The Wishart distribution $\mathcal{W}(\Sigma, k)$ is defined as the law of the empirical second order moment

$$X = \sum_{i=1}^k Y_i Y_i^T.$$

When $k \geq n$, $\mathcal{W}(\Sigma, k)$ is supported on $\mathcal{P}(n, \mathbb{R})$. Recall that the group K_I fixing the identity matrix is given by $\Gamma(SO(n))$ where $\Gamma(A)(X) = AXA^T$. Hence the group K_Σ fixing Σ is given by $K_\Sigma = \Gamma(\Sigma^{1/2}SO(n)\Sigma^{-1/2})$. Let $A \in GL(n, \mathbb{R})$, it is easy to check that

$$AXA^T \sim \mathcal{W}(k, A\Sigma A^T), \quad (13)$$

which shows that K_Σ leaves $\mathcal{W}(\Sigma, k)$ fixed. Hence the log-extrinsic and the Fréchet mean can be computed using Eq.8:

$$L(\mathcal{W}(\Sigma, k)) = F(\mathcal{W}(\Sigma, k)) = e^{\frac{1}{n}\mathbb{E}(H_*\mathcal{W}(\Sigma, k))}\pi(\Sigma).$$

By differentiating the log partition of the Wishart density, a standard calculation gives

$$\mathbb{E}(H_*\mathcal{W}(\Sigma, k)) = n \ln(2) + \ln(\det(\Sigma)) + \sum_{i=1}^n \psi\left(\frac{k+1-i}{2}\right),$$

where ψ is the diagama function, i.e. the logarithmic derivative of the gamma function. Note $\psi_n(\frac{k}{2}) = \sum_{i=1}^n \psi\left(\frac{k+1-i}{2}\right)$ the multivariate diagama function. We obtain

$$L(\mathcal{W}(\Sigma, k)) = F(\mathcal{W}(\Sigma, k)) = 2 \det(\Sigma)^{1-\frac{1}{n}} e^{\psi_n(\frac{k}{2})} \Sigma.$$

Defined now exponential-wrapped normal distributions on $\mathcal{P}(n, \mathbb{R})$ as done for instance in Chevallier et al. (2022). In the rest of the paper, let d be the following invariant Riemannian distance on $\mathcal{P}(n, \mathbb{R})$:

$$d(A, B) = \|\log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})\|, \quad (14)$$

and denote g_Σ the local metric. Let U be a random vector following a normal distribution on the tangent space $T_\Sigma\mathcal{P}(n, \mathbb{R})$, of mean 0 and of covariance Cov . The exponential-wrapped normal distribution $WN(\Sigma, Cov)$ is defined as the law of

$$X = \exp_\Sigma(U),$$

where \exp_Σ is the Riemannian exponential at Σ . By theorem 2, since $\mathbb{E}(\log_\Sigma(X)) = 0$, the Fréchet mean of $WN(\Sigma, Cov)$ is Σ .

Assume now that the covariance is a multiple of the identity matrix in a g_Σ -orthonormal basis of $T_\Sigma\mathcal{P}(n, \mathbb{R})$: $Cov = \frac{v}{\dim} I_{\dim \times \dim}$ with $v > 0$ and $\dim = \frac{n(n+1)}{2}$. It is possible to show that K_Σ leaves $WN(\Sigma, Cov)$ stable, and Eq.8 enables to relate the log-extrinsic mean to the Fréchet mean:

$$L(WN(\Sigma, v)) = F(WN(\Sigma, v)) = \Sigma.$$

Denote M the Fréchet or log-extrinsic mean of either Wishart or wrapped normal distributions. Let $Emp(p)$ refer to an empirical mean of p i.i.d. samples, and note $RRMSE(p)$ the Riemannian root mean squared error between the empirical mean and M ,

$$RRMSE(p) = \sqrt{\mathbb{E}(d(M, Emp(p))^2)}.$$

Let $D(n)$ denote the diagonal matrix with n eigenvalues linearly spaced between 1 and 10. We estimated the convergence curves of the empirical means for the Wishart distributions $\mathcal{W}(D(2), 4)$, $\mathcal{W}(D(20), 40)$. We also considered wrapped normal distributions in dimension 2 and 20, whose means M coincide with those of the Wishart distributions, and whose covariance matrices are the identity matrix in an orthonormal basis of $T_M\mathcal{P}(n, \mathbb{R})$.

Due to the invariance of the metric, and the equivariance of the log-extrinsic mean and of the Fréchet mean, the convergence curves of the empirical means do not depend on the mean M . The log-Euclidean mean however is not equivariant under $G(\mathcal{C})$. Hence, the curve shown in Fig.4 would differ for different choices of M : the closer M is of the identity matrix, the better the log-Euclidean approximation is.

For each distribution, the curves $RRMSE(p)$ are estimated using Monte-Carlo sampling using 500 samples and shown on Fig.4. Fig.4 also show the average computing time of $Emp(p)$ from the i.i.d. samples in each scenario. These curves illustrate the relevance of the log-extrinsic mean as an alternative to Fréchet means in practical applications.

Fréchet means are computed using the python package Geomstats with NumPy as the execution backend. In order to take advantage of the symmetric structure of matrices, the matrix logarithms of the log-Euclidean means are computed from the eigenvalues. In dimension 2 the eigenvalues are obtained from their close form while in dimension 20, we used the NumPy diagonalization algorithm.

6.2 Classification experiment

We performed classification experiments on stereo sound signals. A set of 13 stereo recordings of the first movement of the moonlight sonata of Beethoven has been gathered. For each recording, a set of covariance descriptors in $\mathcal{P}(2, \mathbb{R})$ is extracted. The task is then to classify a new descriptor in order to correctly retrieve the corresponding recording. The classification is performed using $G(\mathcal{P}(2, \mathbb{R}))$ -equivariant statistical tools.

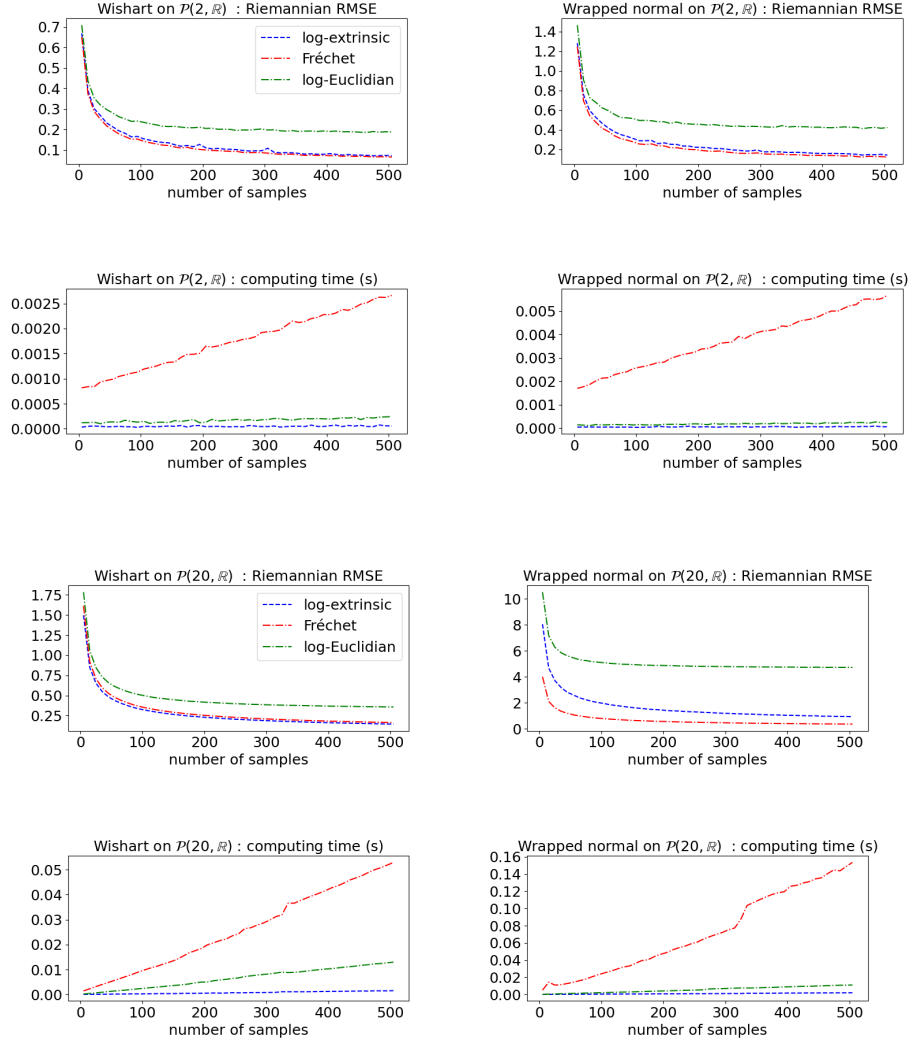


Figure 4: These plots show the convergence of empirical means toward the true mean for four different distributions, see text. On each plot, the red, blue and green curves respectively correspond to the Fréchet mean, the log-extrinsic mean and the log-Euclidean mean. The computations are made on a cpu Intel i9-11900K, and the computing time is averaged over the Monte-Carlo trials.

It has been mentioned in section 2 that several possible equivariant mean can be considered on determinants. For this classification experiment we consider only two frameworks: the geometric mean on determinants, which lead to log-extrinsic means of Eq4, and the arithmetic mean on determinants, which lead to the following formula,

$$A(\mu) = \left(\int \det(X) d\mu(X) \right)^{\frac{1}{r}} E(\pi_*\mu),$$

and which we will call an arithmetic-extrinsic mean.

The experiment indicate a superiority of the log-extrinsic framework over the arithmetic-extrinsic framework.

6.2.1 Dataset

Each recording is represented by a 2-dimensional array whose first index refers to the channel and whose second index refers to time. Given an arbitrary 2-dimensional array s , let \widehat{s} be the array of same dimensions obtained taking the discrete Fourier transform channel-wise:

$$\widehat{s}(c, \cdot) = \text{DTF}(s(c, \cdot)), \quad c \in \{1, 2\}.$$

For each frequency k , $\widehat{s}(\cdot, k) \in \mathbb{C}^2$ is a vector of Fourier coefficients. The considered descriptor is the second order moments of these vectors. At time t , let s_t be the slice of signal comprised between time t and t plus one second. The array s_t is then of dimension $2 \times N$, with N depending of the sample rate. The local covariance descriptor is obtained by

$$\Sigma_t = \frac{1}{N} \sum_{k=0}^{N-1} \widehat{s}_t(\cdot, k) \widehat{s}_t(\cdot, k)^T.$$

Since $\widehat{s}_t(\cdot, k) \in \mathbb{C}^2$, Σ_t is a priori a complex matrix. However, $\widehat{s}_t(\cdot, 0) = \overline{\widehat{s}_t(\cdot, 0)}$ and when $1 \leq k \leq N$, $\widehat{s}_t(\cdot, k) = \overline{\widehat{s}_t(\cdot, N - k)}$, hence Σ_t is a real matrix. It can be shown that provided that the two microphones have different spatial locations, the matrix Σ_t is definite. Hence $\Sigma_t \in \mathcal{P}(2, \mathbb{R})$. This covariance descriptor is inspired by polarization matrices in optics.

For each recording, 100 time indices are uniformly sampled to build a train set of covariance descriptors and 100 time indices to build a test set. The dataset of covariance descriptors is available on the webpage of the first author.

6.2.2 Classification methods and results

Let Δ_i be the empirical distribution of the i -th class, $\Delta_i = \sum_j \delta_{x_j^i}$. We consider several classification framework. A first group of classifiers is defined by computing centroids for each class and classifying a new data $x \in \mathcal{P}(2, \mathbb{R})$ according to the nearest centroid. We consider the rules

$$\begin{aligned}
(\text{R}) & : c(x) = \arg \min_i (d(\mathcal{F}(\Delta_i), x)) \\
(\text{LEx}) & : c(x) = \arg \min_i (d(L(\Delta_i), x)) \\
(\text{AEx}) & : c(x) = \arg \min_i (d(A(\Delta_i), x))
\end{aligned}$$

where LEx, AEx refer respectively to log-extrinsic and arithmetic extrinsic, and where d is the invariant metric defined in Eq.14. Note that using an invariant distance on \mathcal{S} combined with an Euclidean distance on determinants would not lead to an invariant distance on \mathcal{C} . Hence classification in the AEx framework is also performed using d .

In the second set of classifiers, the training sets are modeled by Gaussian distributions. Data are then classified according to the most likely class. Since the classes are equally represented, it coincides with the maximum a posteriori estimator. Recall that every invariant Riemannian distances can be decomposed as a product between \mathcal{S} and \mathbb{R} , see Theorem Eq.1. Since nothing relates apriori the scales of the two factors, there is no intrinsic notion of isotropic Gaussian on symmetric cones. Hence, the considered Gaussians on $\mathcal{P}(2, \mathbb{R})$ are parametrized by their mean and two scale parameters for their marginals on \mathcal{S} and \mathbb{R} . We considered several types of Gaussian distributions.

- ($W\mathcal{N}$). Exponential-wrapped Gaussians as defined in section 6.1, parametrized by their mean, a variance on \mathcal{S} and a variance on determinants. Parameters are estimated by empirical means and variances.
- ($\text{LEx-}\mathcal{N}$). Log-extrinsic Gaussians defined in Eq.11. Parameters are estimated by maximum likelihood.
- ($\text{AEx-}\mathcal{N}$). Gaussians corresponding to arithmetic means are the traditional Gaussian distributions. Hence arithmetic-extrinsic Gaussian can be defined as a products between an extrinsic Gaussian on \mathcal{S} , and a traditional Gaussian on determinants. Parameters are estimated by maximum likelihood. Note that this model is stable under the action of $G(\mathcal{P}(2, \mathbb{R}))$ and does not respect the positive constraint on the determinant. We nonetheless show the corresponding classification results.
- ($\text{AEx-}\Gamma$). Since $\text{AEx-}\mathcal{N}$ distributions are not supported on $\mathcal{P}(2, \mathbb{R})$, we also consider modeling the distribution of determinants with a Gamma distribution. Parameters are fitted by maximum likelihood using the stats package of the SciPy library. Unlike $\text{AEx-}\mathcal{N}$ distributions, this model is supported on $\mathcal{P}(2, \mathbb{R})$ and is stable under the action of $G(\mathcal{P}(2, \mathbb{R}))$.

Each classification experiment is repeated 100 times. The following table reports the average classification results \pm a standard deviation. All the codes have been written in Python and are available on the webpage of the first author.

Method	Accuracy (%)	Method	Accuracy (%)
(R)	35.9 ± 1.4	($W\mathcal{N}$)	43.6 ± 1.2
(LEx)	35.9 ± 1.4	(LEx- \mathcal{N})	43.6 ± 1.2
(AEx)	25.5 ± 1.7	(AEx- Γ)	34.2 ± 1.0
		(AEx- \mathcal{N})	33.2 ± 1.7

These results suggest that the arithmetic-extrinsic framework is less suited to describe experimental data than its logarithmic counterpart. They also show very similar performances for the Riemannian and log-extrinsic framework.

7 Discussion and open problems

The experimental section confirms the relevance of log-extrinsic means for practical applications. They maintain the equivariance properties of the Fréchet mean, while being more computationally efficient than log-Euclidean means. Future efforts will focus on the computation of the normalizing factor of log-extrinsic Gaussians, as well as the relation between expected and natural parameters. The difficulty of the computations is expected to depend on the type of the irreducible symmetric cone, the simplest cases being Lorentz cones. In the numerical experiment presented in section 6, the distance based classification is achieved using an invariant Riemannian distance. Future works will investigate the possibility of using non-Riemannian invariant distances based on the log-extrinsic framework, in order to reduce the computational complexity of the classification.

From a theoretical standpoint, two questions remain open. First, as mentioned at the end of section 4.2, it would be interesting to determine whether or not the hypersurface \mathcal{S} is the only one which lead to an equivariant mean on $\mathcal{C}/\mathbb{R}_{>0}$. Second, a perhaps more difficult question is to determine whether the characterizing property of extrinsic Gaussians given in Theorem 9 for Lorentz cones, also holds for other irreducible cones.

Finally, in the spirit of Pálfi (2012), it could be interesting to link the curves formed by weighted averages between two points, as shown in Fig.3, with an adapted notion of geodesic.

8 Appendix

8.1 Symmetric cones and Jordan algebras

Let V be a finite dimensional real vector space equipped with a scalar product $\langle x, y \rangle$. Let \mathcal{C} be an open convex cone in V . Call $G(\mathcal{C})$ the set of linear isomorphism $u : V \rightarrow V$ such that $u(\mathcal{C}) = \mathcal{C}$. The cone \mathcal{C} is symmetric if both $G(\mathcal{C})$ acts transitively on \mathcal{C} and if $\mathcal{C} = \mathcal{C}^*$ where

$$\mathcal{C}^* = \{y \in V : \forall x \in \bar{\mathcal{C}} \setminus \{0\}, \langle x, y \rangle > 0\}.$$

In that case, the identity component of $G(\mathcal{C})$, noted $G(\mathcal{C})_0$, is also transitive on \mathcal{C} .

A bilinear product on V verifying the following properties,

$$xy = yx \text{ and } x((xx)y) = (xx)(xy),$$

is called a Jordan product, and makes V a Jordan algebra. For $x \in V$, note l_x the left multiplication by x : $l_x(y) = xy$. The Jordan algebra V is called Euclidean if there is an inner product $\langle \cdot, \cdot \rangle$ such that all the l_x are self-adjoint. If \mathcal{C} is a symmetric cone, then $(V, \langle \cdot, \cdot \rangle)$ can be endowed with a Jordan product such that

- $(V, \langle \cdot, \cdot \rangle)$ is a Euclidean Jordan algebra with a unit element, noted o
- $\mathcal{C} = \{x \in V : l_x \text{ is symmetric positive definite}\}$.

When there are no non-trivial subspaces $V_1 \oplus V_2 = V$ and symmetric cones $\mathcal{C}_1 \subset V_1$ and $\mathcal{C}_2 \subset V_2$ such that $\mathcal{C} = \mathcal{C}_1 + \mathcal{C}_2$, the cone \mathcal{C} is said irreducible. Irreducible symmetric cones can be classified into four families plus an exceptional cone. Denote by $\mathcal{P}(n, \mathbb{K})$ positive definite self-adjoint matrices with coefficients in $\mathbb{K} = \mathbb{R}, \mathbb{C}, \mathbb{H}$ (quaternions) or \mathbb{O} (octonions). Every irreducible symmetric cone is isometric to a cone belonging to the following families

- Lorentz cones for $n > 2$: $L(n) = \{(x_0, \dots, x_{n-1}) \in \mathbb{R}_{>0} \times \mathbb{R}^{n-1}, x_0^2 > \sum_{i=1}^{n-1} x_i^2\}$
- $\mathcal{P}(n, \mathbb{R}), \mathcal{P}(n, \mathbb{C}), \mathcal{P}(n, \mathbb{H})$, with $n \geq 1$
- $\mathcal{P}(3, \mathbb{O})$.

On Lorentz cones $L(n)$, the product on \mathbb{R}^n is usually set as

$$\begin{pmatrix} x_0 \\ x_1 \\ \cdot \\ \cdot \\ x_{n-1} \end{pmatrix} \circ \begin{pmatrix} y_0 \\ y_1 \\ \cdot \\ \cdot \\ y_{n-1} \end{pmatrix} = \begin{pmatrix} \sum_{i=0}^{n-1} x_i y_i \\ x_0 y_1 + y_0 x_1 \\ \cdot \\ \cdot \\ x_0 y_{n-1} + y_0 x_{n-1} \end{pmatrix}$$

with identity element $o = (1, 0, \dots, 0)^T$. On self adjoint matrices, it is given by

$$A \circ B = \frac{1}{2}(AB + BA),$$

and the identity element is given by the identity matrix: $o = I$. Let A be an element of the symmetric cone $\mathcal{P}(\mathbb{K}, n)$. Since l_A is positive definite, it is invertible. Since $l_A(A^{-1}) = I$ where A^{-1} is the matrix inverse, the Jordan inverse coincide with the matrix inverse.

Given an element $x \in V$, let $m(x)$ be the maximal value of k such that the o, x, x^2, \dots, x^k form an independent family. Let V_x be

$$V_x = \text{span}\{o, x, x^2, \dots, x^{m(x)}\}.$$

The rank r of V is defined as

$$r = \max_{x \in V} m(x).$$

The rank of a symmetric cone is defined as the rank of the underlying Jordan algebra. Elements $x \in \mathcal{V}$ such that $m(x) = r$ are called regular elements and are dense in V . For $x \in V$ a regular element, let \tilde{l}_x be the restriction of l_x to V_x . The trace and determinant of x are then defined as

$$\det(x) = \text{Det}(\tilde{l}_x) \text{ and } \text{trace}(x) = \text{Trace}(\tilde{l}_x),$$

where Det and Trace refers to the usual determinant and trace. The definitions are extended by continuity to all elements of V . It can be checked that $\det(o) = 1$ and $\text{trace}(o) = r$. For matrix cones $\mathcal{P}(n, \mathbb{K})$, the Jordan determinant, trace and inverse coincide with the usual ones. The determinant verifies the following formula,

$$\forall g \in G(\mathcal{C}), \forall x \in V, \det(g(x)) = |\text{Det}(g)|^{\frac{r}{\dim(V)}} \det(x), \quad (15)$$

see Faraut and Korányi (1994) proposition.III.4.3., where the proof can easily be generalised to $g \in G(\mathcal{C})$. Hence the subgroup of $G(\mathcal{C})$ preserving \mathcal{S} , the surface of determinant 1, noted $G(\mathcal{S})$, is the set of elements $g \in G(\mathcal{C})$ with $\text{Det}(g) = \pm 1$. Recall that the action of $G(\mathcal{C})_0$ is transitive on \mathcal{C} . Eq.15 shows that the subsets of constant determinant are the orbits of $G(\mathcal{S})$. Since the homotheties with strictly positive scaling factor preserve \mathcal{C} , it is clear that $G(\mathcal{C})$ is isomorphic to $G(\mathcal{S}) \times \mathbb{R}_{>0}$.

On Lorentz cones trace, determinant and inverse are given by

- trace: $\text{trace}(x) = 2x_0$
- determinant: $\det(x) = x_0^2 - x_1^2 - \dots - x_n^2$
- when $\det(x) \neq 0$, the inverse for the Jordan product is given by: $x^{-1} = \frac{(x_0, -x_1, \dots, -x_n)}{\det(x)}$.

It can be checked that the maps

$$\begin{pmatrix} x_0 \\ x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_0 + x_1 & x_2 \\ x_2 & x_0 - x_1 \end{pmatrix} \text{ and } \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto \begin{pmatrix} x_0 + x_1 & x_2 + ix_3 \\ x_2 - ix_3 & x_0 - x_1 \end{pmatrix}$$

are respectively Jordan algebra isomorphisms mapping $L(3)$ to $\mathcal{P}(\mathbb{R}, 2)$ and mapping $L(4)$ to $\mathcal{P}(\mathbb{C}, 2)$.

8.2 Proof of theorem 1

Let \mathcal{C} be an irreducible symmetric cone and note $G(S)^0$ the identity component of $G(S)$. Recall that K_o is defined as the identity component of the subgroup of $G(\mathcal{C})$ fixing o . It is easy to check that $K_o \subset G(S)^0$. We have the following lemma.

Lemma 1. *The homogeneous space $G(S)^0/K_o$ is an irreducible symmetric space.*

Proof. The list of groups $G(S)^0$ and K_o for different irreducible symmetric cones can be found in Orlitzky (2024). It can be checked in Sasaki (1980) and Lopez et al. (2021) that each homogeneous spaces $G(S)^0/K_o$ corresponds to a an irreducible symmetric spaces. \square

Let $g^{\mathcal{C}}$ be an invariant Riemannian metric on \mathcal{C} . At e , the tangent space can be decomposed as

$$T_e\mathcal{C} = \ker(d\det_e) \oplus \ker(d\pi_e) = T_e\mathcal{S} \oplus \mathbb{R}.e.$$

Since $G(S)_0$ is transitive on \mathcal{S} , \mathcal{S} is naturally identified with the quotient $G(S)_0/K_e$, which shows that \mathcal{S} is an irreducible symmetric space. On irreducible symmetric space, the isotropy representation is irreducible: K_e acts irreducibly on $T_e\mathcal{S}$. Then, it can be proved that the irreducibility of the action implies that there exists only one inner product $g_e^{\mathcal{S}}$ on $T_e\mathcal{S}$ invariant under K_e , up to a constant.

Let W be the orthogonal complement of $T_e\mathcal{S}$ with respect to $g_e^{\mathcal{C}}$. We want to show that $W = \mathbb{R}.e$. Since all $g \in K_e$ are isometries with respect to $g_e^{\mathcal{C}}$, the one dimensional subspace W is stable. Since $V = T_e\mathcal{S} \oplus W$, there exists $u \in T_e\mathcal{S}$ and $w \in W$ unique such that $e = u + w$. Therefore, for all $g \in K_e$, we have $e = ge = gu + gw$ with $gu \in T_e\mathcal{S}$ and $gw \in W$. The uniqueness of the decomposition implies that $gu = u$ for all $g \in K_e$ which in turn implies that $u = 0$ because the action of K_e is irreducible. Therefore $e \in W$ and $W = \mathbb{R}.e$. Hence the orthogonal complement of $T_e\mathcal{S}$ is $\mathbb{R}.e$, and the inner product $g_e^{\mathcal{C}}$ is of the form $g_e^{\mathcal{C}} = \alpha g_e^{\mathcal{S}} + \beta g_{\mathbb{R}}$, where $\alpha, \beta > 0$, and $g_e^{\mathcal{S}}$ and $g_{\mathbb{R}}$ are inner products on $T_e\mathcal{S}$ and $\mathbb{R}.e$ respectively.

For $\alpha, \beta > 0$, let $d_{\alpha, \beta}(x, y) = \alpha d_{\mathcal{S}}(\pi(x), \pi(y))^2 + \beta(H(x) - H(y))^2$. It is possible to check that $d_{\alpha, \beta}$ is an invariant Riemannian distance whose inner product at identity is $\alpha g_e^{\mathcal{S}} + \beta g_{\mathbb{R}}$. Since invariant Riemannian distance are determined by their inner product at e , every Riemannian distance is of the form $d_{\alpha, \beta}$ for some $\alpha, \beta > 0$.

8.3 Proof of theorem 5

We shall need two Lemmas. The first is a general result about convergence of probability measures.

For ω in vector space F and $t \in \mathbb{R}$, denote $h_{\omega, t} = h_{\omega, t}^F : F \rightarrow F$ the homothety with center ω and ratio t .

Lemma 2. *Let V and W be two finite dimensional vector spaces, let $U \subset V$ be an open set, let $g : U \rightarrow W$ be a map and let $\omega \in U$. Suppose that g is continuous, $g(\omega) = 0$, g is*

differentiable at ω and that $\mathrm{d}g_\omega(\omega) = 0$. Let $(\mu_n)_n$ be a sequence of probability measures in V such that $\mu_n(U) = 1$ for all n and suppose that the sequence $(h_{\omega, \sqrt{n}, * }^V \mu_n)_n$ converges in law to a probability measure μ . Then the sequence $(h_{0, \sqrt{n}, * }^W g_* \mu_n)_n$ converges in law to $(\mathrm{d}g_\omega)_* \mu$.

Proof. It is enough to prove that for any continuous function $\phi : W \rightarrow \mathbb{R}$ with compact support, we have

$$\lim_{n \rightarrow \infty} \int \phi(\sqrt{n}g(x)) \mathrm{d}\mu_n(x) = \int \phi(\mathrm{d}g_\omega(x)) \mathrm{d}\mu(x).$$

First observe that using the convergence in law of the sequence $(h_{\omega, \sqrt{n}, * }^V \mu_n)_n$ with the bounded continuous function $\phi \circ \mathrm{d}g_\omega$, we have

$$\int \phi(\mathrm{d}g_\omega(x)) \mathrm{d}\mu(x) = \lim_{n \rightarrow \infty} \int \phi(\mathrm{d}g_\omega(\omega + \sqrt{n}(x - \omega))) \mathrm{d}\mu_n(x) = \lim_{n \rightarrow \infty} \int \phi(\sqrt{n} \mathrm{d}g_\omega(x)) \mathrm{d}\mu_n(x).$$

Let $\varepsilon > 0$. Let $R > 0$ be such that $\mu(B_V(\omega, R-1)) > 1 - \varepsilon$. Since ϕ is uniformly continuous we can find $\delta > 0$ such that $\mathrm{d}_W(x, y) \leq \delta \implies |\phi(x) - \phi(y)| \leq \varepsilon$. By definition of the differential, and since $g(\omega) = \mathrm{d}g_\omega(\omega) = 0$, there exists $r > 0$ such that $\mathrm{d}_W(g(x), \mathrm{d}g_\omega(x)) < \frac{\delta}{R} \mathrm{d}_V(x, \omega)$ for all $x \in B_V(\omega, r)$. Let n be any integer such that $R/\sqrt{n} \leq r$. On the one hand, if $x \in B_V(\omega, R/\sqrt{n})$ then

$$\mathrm{d}_W(\sqrt{n}g(x), \sqrt{n} \mathrm{d}g_\omega(x)) = \sqrt{n} \mathrm{d}_W(g(x), \mathrm{d}g_\omega(x)) \leq \delta,$$

therefore

$$\left| \int_{B_V(\omega, R/\sqrt{n})} (\phi(\sqrt{n}g(x)) - \phi(\sqrt{n} \mathrm{d}g_\omega(x))) \mathrm{d}\mu_n(x) \right| \leq \varepsilon.$$

On the other hand, since $h_{\omega, \sqrt{n}, * }^V \mu_n \rightarrow \mu$ in law,

$$\begin{aligned} 1 - \varepsilon &\leq \mu(B_V(\omega, R-1)) \leq \liminf_{n \rightarrow \infty} \mu_n((h_{\omega, \sqrt{n}}^V)^{-1}(B_V(\omega, R))) \\ &= \liminf_{n \rightarrow \infty} \mu_n(B_V(\omega, R/\sqrt{n})), \end{aligned}$$

so that $\mu_n(B_V(\omega, R/\sqrt{n})) \geq 1 - 2\varepsilon$ for all n large enough. Therefore,

$$\left| \int_{U \setminus B_V(\omega, R/\sqrt{n})} (\phi(\sqrt{n}g(x)) - \phi(\sqrt{n} \mathrm{d}g_\omega(x))) \mathrm{d}\mu_n(x) \right| \leq 4\varepsilon \|\phi\|_\infty.$$

and

$$\left| \int_U (\phi(\sqrt{n}g(x)) - \phi(\sqrt{n} \mathrm{d}g_\omega(x))) \mathrm{d}\mu_n(x) \right| \leq \varepsilon(1 + 4\|\phi\|_\infty).$$

□

Lemma 3. Let $\varepsilon \in \mathcal{S}$. $\mathrm{d}\pi_\varepsilon = P_\varepsilon$ where P_ε is the projection onto $T_\varepsilon \mathcal{S}$ parallel to $\mathbb{R}\varepsilon$.

Proof. For any $u \in V$ we have,

$$\pi(\epsilon + u) = \frac{\epsilon + u}{\det(\epsilon + u)^{\frac{1}{r}}} = (\epsilon + u)(1 - d(\det^{\frac{1}{r}})_\epsilon(u) + o(\|u\|)).$$

By definition, $u \in T_\epsilon \mathcal{S}$ implies $d(\det)_\epsilon(u) = 0$ and $d(\det^{\frac{1}{r}})_\epsilon(u) = 0$. Hence $d\pi_\epsilon(u) = u$ for all $u \in T_\epsilon \mathcal{S}$. On the other hand, for all $\lambda > 0$, $\pi(\lambda\epsilon) = \pi(\epsilon) = \epsilon$, hence $d\pi_\epsilon(\epsilon) = 0$. It follows that $d\pi_\epsilon = P_\epsilon$. \square

Let $(X_i)_{i \in \mathbb{N}}$ be a sequence of random variables i.i.d. on \mathcal{S} of extrinsic mean ϵ and covariance Σ :

$$\Sigma = \int_{\mathcal{S}} (x - \mathbb{E}(\mu)) \otimes (x - \mathbb{E}(\mu)) d\mu < \infty.$$

The empirical extrinsic mean is

$$E_n = \pi(X_1 + \dots + X_n)$$

Consider the random variables $Y_i = \frac{X_i}{\det(\mathbb{E}(\mu))^{1/r}}$ and

$$S_n = \frac{Y_1 + \dots + Y_n}{n}.$$

Clearly $\pi(S_n) = E_n$ and S_n is a random variable with values in \mathcal{C} . Note μ_n the distribution of S_n . Since $\epsilon = \frac{\mathbb{E}(\mu)}{\det \mathbb{E}(\mu)^{1/r}}$ is the extrinsic mean of the common law μ of the variables X_i , we have $\mathbb{E}(S_n) = \epsilon$.

The classical central limit theorem gives that $h_{\epsilon, \sqrt{n}, *}\mu_n$ converges in law towards $\mathcal{N}(\epsilon, \frac{\Sigma}{\det(\mathbb{E}(\mu))^{2/r}})$.

By the second lemma, we know that $d\pi_\epsilon = P_\epsilon$. Since P_ϵ is a projection $d(P_\epsilon \circ \pi)_\epsilon(u) = dP_\epsilon \circ (d\pi_\epsilon(u)) = P_\epsilon(P_\epsilon(u)) = P_\epsilon(u)$. Hence P_ϵ and $P_\epsilon \circ \pi$ both send ϵ to 0 and have the same differential at ϵ .

We now use the first lemma about the convergence in law with $V, W = T_\epsilon \mathcal{S}, U = \mathcal{C}, \omega = \epsilon$ and $g = P_\epsilon \circ \pi$. We obtain that $(h_{\epsilon, \sqrt{n}, *}(P_\epsilon \circ \pi)_*\mu_n)_n$ converges in law to $P_{\epsilon*} \mathcal{N}(\epsilon, \frac{\Sigma}{\det(\mathbb{E}(\mu))^{2/r}}) = \mathcal{N}(0, \frac{\Sigma|_T}{\det(\mathbb{E}(\mu))^{2/r}})$, where $\Sigma|_T$ is the marginalisation of Σ on $T_\epsilon \mathcal{S}$ with respect to $\mathbb{R}\epsilon$. Expressed with random variables, we obtain

$$\sqrt{n}P_\epsilon(\pi(S_n)) = \sqrt{n}P_\epsilon(E_n) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\Sigma|_T}{\det(\mathbb{E}(\mu))^{2/r}}\right).$$

8.4 Proof of theorem 6

1. Let us describe now the action of $G(\mathcal{S})$ on the density f_θ . The invariance of the measure $\nu_{\mathcal{S}}$ shows that for any $g \in G(\mathcal{S})$,

$$Z(\theta) = \int e^{-\langle \theta, x \rangle} d\nu_{\mathcal{S}}(x) = \int e^{-\langle \theta, g \cdot x \rangle} d\nu_{\mathcal{S}}(x) = \int e^{-\langle g^* \cdot \theta, x \rangle} d\nu_{\mathcal{S}}(x).$$

Hence, $Z(\theta) = Z(g^*.\theta)$. Since the cone \mathcal{C} is self dual, $g \in G(\mathcal{S})$ iff $g^* \in G(\mathcal{S})$ and therefore, $Z(\theta) = Z(g.\theta)$ for all $g \in G(\mathcal{S})$. Furthermore, the pushforward of f_θ by g is given by $g_*f_\theta = f_\theta \circ g^{-1}$ since $v_{\mathcal{S}}$ is g -invariant. We have then

$$(g_*f_\theta)(x) = \frac{1}{Z(\theta)} e^{-\langle \theta, g^{-1}.x \rangle} = \frac{1}{Z(g^{-1*}.\theta)} e^{-\langle g^{-1*}.\theta, x \rangle} = f_{g^{-1*}.\theta}(x).$$

Similarly,

$$g.\mathbb{E}_\theta = \int g.x f_\theta(x) dv_{\mathcal{S}} = \int x f_\theta(g^{-1}.x) dv_{\mathcal{S}} = \int x f_{g^{-1*}.\theta}(x) dv_{\mathcal{S}} = \mathbb{E}_{g^{-1*}.\theta}.$$

Since $\mathbb{E}_\theta = (s+1)\epsilon$, $g.\mathbb{E}_\theta = (s+1)g\epsilon$ and with the two above relations we obtain

$$g_*f_{\epsilon, s} = f_{g\epsilon, s}.$$

2. In order to prove the 2nd point of theorem 6, we start by assuming that θ is of the form to , with $t > 0$, and prove that there is a bijection between t and s . Using the equivariance properties shown above we can then derive the links between θ , ϵ and s for arbitrary θ . Recall that $o \in \mathcal{S}$ and that $K_o \subset O(V)$, see Faraut and Korányi (1994) proposition I.1.9 and Theorem III.3.1. Since for $g \in K_o$, $g = g^{-1*}$, \mathbb{E}_o is fixed by K_o . By Theorem 3, there is a function α valued in $\mathbb{R}_{>0}$ such that for $t > 0$,

$$\mathbb{E}_{to} = \frac{1}{Z(to)} \int_{\mathcal{S}} x e^{-t\langle o, x \rangle} dv_{\mathcal{S}} = \alpha(t)o. \quad (16)$$

By differentiating with respect to t we obtain,

$$(\mathbb{E}_{to})' = \alpha'(t)o = -\frac{Z'(to)}{Z(to)^2} \int_{\mathcal{S}} x e^{-t\langle o, x \rangle} dv_{\mathcal{S}} - \frac{1}{Z(to)} \int_{\mathcal{S}} \langle o, x \rangle x e^{-t\langle o, x \rangle} dv_{\mathcal{S}}$$

Both integrals are elements of \mathcal{C} invariant by the action of K_o , hence by Theorem 3 they are strictly positive multiples of o . Since $Z'(to) = -\int_{\mathcal{S}} \langle o, x \rangle e^{-t\langle o, x \rangle} dv_{\mathcal{S}} < 0$, the function $\alpha'(t)$ is strictly negative and α is strictly decreasing.

According to Proposition III.4.3 and Proposition I.3.3 in Faraut and Korányi (1994), the function $H : x \in \mathcal{C} \mapsto \log \det x$ is strictly concave. Since the probability measure with a density f_θ is not supported by a single point, it follows from the Jensen inequality that

$$H(\mathbb{E}_\theta) = H\left(\int_{\mathcal{S}} x f_\theta(x) dv_{\mathcal{S}}(x)\right) > \int_{\mathcal{S}} H(x) f_\theta(x) dv_{\mathcal{S}}(x) = 0,$$

where the last equality follows from the fact that $\det x = 1$ on \mathcal{S} . Hence $\det \mathbb{E}_\theta > 1$ and $\alpha(t) > 1$.

Our aim is to show that the map $\alpha : \mathbb{R}_{>0} \rightarrow]1, +\infty]$ is a bijection. Given $1 < t_1 < t_2$, denote $A_1 = \{x \in \mathcal{S}, \langle x, o \rangle < t_1\}$, $A_2 = \{x \in \mathcal{S}, t_1 \leq \langle x, o \rangle < t_2\}$ and $A_3 = \{x \in \mathcal{S}, t_2 \leq \langle x, o \rangle\}$. It can be checked that $v_{\mathcal{S}}(A_i) > 0$, $i = 1, 2, 3$. Denote

$$m_i(t) = \int_{A_i} f_{t,o} dv_{\mathcal{S}} \text{ and } f_{i,t,o} = \frac{1}{m_i(t)} f_{t,o} 1_{A_i},$$

where 1_{A_i} is the indicator function of A_i . The function $f_{i,t,o} : \mathcal{S} \rightarrow \mathbb{R}$ is the density $f_{t,o}$ conditioned on A_i . Denote also $\mathbb{E}_{i,t,o}$ the expectation of the conditional density. Since $k(A_i) = A_i$ for all $k \in K_o$, Theorem 3 implies that $\mathbb{E}_{i,t,o} = \alpha_i(t)o$ with $\alpha_i(t) > 0$.

Let us show that $\lim_{t \rightarrow \infty} \alpha(t) = 1$. Let $1 < t_1 < t_2$. Note first that $\alpha(t) = m_1(t)\alpha_1(t) + m_2(t)\alpha_2(t) + m_3(t)\alpha_3(t)$. It is easy to check that $\alpha_1(t), \alpha_2(t) \leq t_2$ hence

$$\alpha(t) \leq (m_1(t) + m_2(t))t_2 + m_3(t)\alpha_3(t).$$

Next we prove that $\lim_{t \rightarrow \infty} m_3(t) = 0$. We have,

$$\int_{A_1} e^{-\langle t_o, x \rangle} dv_{\mathcal{S}} \geq e^{-tt_1} v_{\mathcal{S}}(A_1) \text{ and } \int_{A_3} e^{-\langle t_o, x \rangle} dv_{\mathcal{S}} \leq e^{-tt_2} v_{\mathcal{S}}(A_3).$$

Hence,

$$\frac{m_3(t)}{m_1(t)} \leq e^{-t(t_2-t_1)} \frac{v_{\mathcal{S}}(A_3)}{v_{\mathcal{S}}(A_1)}.$$

Since $m_1(t) \leq 1$, $\lim_{t \rightarrow \infty} m_3(t) = 0$. The same reasoning as previously tells us that the $\alpha_i(t)$ are decreasing functions of t . Hence $\lim_{t \rightarrow \infty} m_3(t)\alpha_3(t) = 0$. At last, since $m_1(t) + m_2(t) < 1$, we have for all $t > 0$,

$$\alpha(t) \leq t_2 + m_3(t)\alpha_3(t).$$

Since t_2 can be chosen arbitrarily close to 1, $\lim_{t \rightarrow \infty} \alpha(t) = 1$.

Let us now show that $\lim_{t \rightarrow 0} \alpha(t) = \infty$. Note first that

$$\lim_{t \rightarrow 0} \int_{A_1 \cup A_2} e^{-\langle t_o, x \rangle} dv_{\mathcal{S}} < v_{\mathcal{S}}(A_1) + v_{\mathcal{S}}(A_2) \text{ and } \lim_{t \rightarrow 0} \int_{A_3} e^{-\langle t_o, x \rangle} dv_{\mathcal{S}} = \infty.$$

Hence $\lim_{t \rightarrow 0} \frac{m_1(t) + m_2(t)}{m_3(t)} = 0$, and since $m_1(t) + m_2(t) + m_3(t) = 1$, $\lim_{t \rightarrow 0} m_3(t) = 1$. It is easy to check that $\alpha_3(t) \geq t_2$. Since t_2 can be chosen arbitrarily, the inequality

$$\alpha(t) = m_1(t)\alpha_1(t) + m_2(t)\alpha_2(t) + m_3(t)\alpha_3(t) \geq m_3(t)\alpha_3(t) \geq m_3(t)t_2$$

implies that $\lim_{t \rightarrow 0} \alpha(t) = \infty$. Hence α is a decreasing bijective function from $\mathbb{R}_{>0}$ to $]1, +\infty[$.

Let $g \in G(\mathcal{S})$. Theorem III.5.3 of Faraut and Korányi (1994) states that for $x \in \mathcal{C}$, $(gx)^{-1} = g^{-1*}x^{-1}$. Hence,

$$\mathbb{E}_{g,t,o} = g^{-1*}.\mathbb{E}_{t,o} = \alpha(t)g^{-1*}.o = \alpha(t)(g.o)^{-1}.$$

Note that $o = \pi(to)$ and recall that π commutes with $g \in G(\mathcal{S})$, hence $g.o = \pi(g.to)$. Since any $\theta \in \mathcal{C}$ can be written as $g.to$ with $g \in G(\mathcal{S})$ and $t \in \mathbb{R}_{>0}$, we have

$$\mathbb{E}_\theta = \alpha(\det(\theta)^{\frac{1}{r}})\pi(\theta)^{-1}.$$

Let us show that π commutes with the inverse map, $x \in \mathcal{C} \rightarrow x^{-1}$. If $x = gto$ then

$$\pi(x^{-1}) = \pi(g^{-1*}(to)^{-1}) = g^{-1*}\pi((to)^{-1}) = g^{-1*}o = (go)^{-1} = (\pi(gto))^{-1} = (\pi(x))^{-1}.$$

Hence

$$\mathbb{E}_\theta = \alpha(\det(\theta)^{\frac{1}{r}})\pi(\theta)^{-1} = \alpha(\det(\theta)^{\frac{1}{r}})\pi(\theta^{-1}).$$

It follows that $\epsilon(\theta) = \pi(\theta)^{-1}$ and $s(\theta) = \alpha(\det(\theta)^{1/r}) - 1$.

Finally, consider the two maps

$$\begin{array}{ccc} \mathcal{C} & \longrightarrow & \mathcal{S} \times \mathbb{R}_{>0} \\ \theta & \longrightarrow & (\pi(\theta), \det(\theta)^{\frac{1}{r}}) \end{array} \qquad \begin{array}{ccc} \mathcal{S} \times \mathbb{R}_{>0} & \longrightarrow & \mathcal{S} \times \mathbb{R}_{>0} \\ (x, t) & \longrightarrow & (x^{-1}, \alpha(t) - 1) \end{array}.$$

They are bijective, the first being a parametrization of \mathcal{C} , and the second a product of two bijections. Therefore, their composition

$$\begin{array}{ccc} \mathcal{C} & \longrightarrow & \mathcal{S} \times \mathbb{R}_{>0} \\ \theta & \longrightarrow & (\pi(\theta)^{-1}, \alpha(\det(\theta)^{\frac{1}{r}}) - 1) \end{array}$$

is a bijection.

8.5 Proof of theorem 7

Let $\mathcal{C} = L(3) \sim \mathcal{P}(2, \mathbb{R})$ and let $\theta \in \mathcal{C}$. Let $|\theta| = \sqrt{\det \theta} = \sqrt{\theta_0^2 - \theta_1^2 - \theta_2^2}$ be the pseudo norm of $\theta = (\theta_0, \theta_1, \theta_2)$ and $\bar{\theta} = (\theta_0, -\theta_1, -\theta_2)$. Note that $|\theta| = |\bar{\theta}|$ and recall that $\theta^{-1} = |\theta|^{-2}\hat{\theta}$ in Lorentz cones. When $\mathcal{C} = L(3)$, the distribution f_θ takes the form

$$f_\theta(x) = c|\theta|e^{|\theta|}e^{-\langle \theta, x \rangle},$$

see for instance Nielsen and Okamura (2023) Eq.(12). Hence $Z(\theta) = c^{-1}|\theta|^{-1}e^{-|\theta|}$ and $\ln(Z(\theta)) = -\ln(|\theta|) - |\theta| - \ln(c)$. We have

$$\nabla|\theta| = \nabla\sqrt{\theta_0^2 - \theta_1^2 - \theta_2^2} = \frac{1}{2\sqrt{\theta_0^2 - \theta_1^2 - \theta_2^2}}\nabla(\theta_0^2 - \theta_1^2 - \theta_2^2) = \frac{1}{2\sqrt{\theta_0^2 - \theta_1^2 - \theta_2^2}}2\bar{\theta} = \frac{\bar{\theta}}{|\theta|},$$

and

$$\nabla \ln(Z) = -\frac{\bar{\theta}}{|\bar{\theta}|^2} - \frac{\bar{\theta}}{|\bar{\theta}|} = -\bar{\theta}\frac{1 + |\bar{\theta}|}{|\bar{\theta}|^2}.$$

Recall that $\mathbb{E}_\theta = \epsilon(s + 1) = -\nabla \ln(Z)$. Hence,

$$\theta^{-1}(1 + |\theta|) = \epsilon(s + 1)$$

Since $\epsilon = \pi(\theta^{-1})$ and that the rank of $L(3)$ is 2, $\epsilon = |\theta|\theta^{-1}$. Hence,

$$1 + |\theta| = |\theta|(s + 1).$$

and $s = \frac{1}{|\theta|}$, and $\theta = (s\epsilon)^{-1}$.

8.6 Proof of theorem 9

1. We shall use the following lemma.

Lemma 4. *For all $x \in \mathcal{C} \cap (o + o^\perp)$, $\det x < 1$ unless $x = o$.*

Proof. Suppose on the contrary that $\det x \geq 1$ for some $x \in \mathcal{C} \cap (o + o^\perp)$, $x \neq o$. Since $H(x) = \log \det x$ is strictly concave on \mathcal{C} , the derivative of the function

$$h : t \in [0, 1] \rightarrow H((1 - t)x + to)$$

is decreasing. Since $h(0) = H(x) \geq 0$ and $h(1) = 0$, we have $h'(1) < 0$ which contradicts that $(o + o^\perp)$ is tangent to \mathcal{S} at o by Propostion III.4.2 of Faraut and Korányi (1994). \square

Let $x_1, \dots, x_n \in \mathcal{S}$. Note

$$\mathcal{L}(\theta) = \prod_{1 \leq i \leq n} \log(f_\theta(x_i)) = -n \log(Z(\theta)) - \sum_{1 \leq i \leq n} \langle \theta, x_i \rangle = -n \log(Z(\theta)) - \langle \theta, \sum_{1 \leq i \leq n} x_i \rangle.$$

We want to show that the function $\theta \mapsto \mathcal{L}(\theta)$ has a maximum under the constraint $\det(\mathbb{E}_\theta)^{1/r} - 1 = s$ and that this maximum is attained in the unique point θ such that $\pi(\mathbb{E}_\theta) = \pi(\sum_i x_i)$. By Theorem 6, the constraint $\det(\mathbb{E}_\theta)^{1/r} - 1 = s$ can be expressed as $\alpha(\det(\theta)^{\frac{1}{r}}) - 1 = s$ and is hence a level set of the determinant $\theta \rightarrow \det \theta$. Let $t > 0$ be such that $\det(\mathbb{E}_{to})^{1/r} - 1 = s$, and denote $\mathcal{S}_t = t\mathcal{S}$ the set of parameters θ verifying the constraint.

Suppose first that $\sum_i x_i = \lambda o$ where $\lambda > 0$. Let us show that

$$\operatorname{argmax}_{\theta \in \mathcal{S}_t} \mathcal{L}(\theta) = to.$$

Recall that the level-sets of the determinant are the orbits of $G(\mathcal{S})$. Hence $\mathcal{S}_t = \{g(to) : g \in G(\mathcal{S})\}$. We showed in 8.4 that for any $g \in G(\mathcal{S})$, $Z(to) = Z(g(to))$. Hence we are reduced to show that

$$\operatorname{argmin}_{\theta \in \mathcal{S}_t} \langle \theta, \lambda o \rangle = to$$

which is equivalent to $\operatorname{argmin}_{x \in \mathcal{S}} \langle x, o \rangle = o$. Let $x \in \mathcal{S}$, $x \neq o$. There exists $\beta > 0$ such that $y = \beta x \in (o + o^\perp)$. By the above lemma $\det y < 1$, hence $\beta < 1$ which in turn implies that $\langle x, o \rangle = (1/\beta) \langle y, o \rangle > 1$. Therefore $\operatorname{argmin}_{x \in \mathcal{S}} \langle x, o \rangle = o$.

Recall that $\mathbb{E}_{to} = \alpha(t)o$, see appendix 8.4 Eq (16), hence $\widehat{\epsilon}_{MLE}(x_1, \dots, x_n) = \pi(\mathbb{E}_{to}) = o = \pi(\lambda o) = \pi(\sum_i x_i)$.

Suppose now that $\sum_i x_i = \lambda g(o)$ where $\lambda > 0$ and $g \in G(\mathcal{S})$. The sample $x'_i = g^{-1}x_1, \dots, x'_n = g^{-1}x_n$ is such that $\sum_i x'_i = \lambda o$. By the equivariance properties of Theorem 6, we have $(g^{-1})_* f_\theta = f_{g^*\theta}$. Therefore,

$$\begin{aligned} \mathcal{L}(g^*\theta, g^{-1}x_1, \dots, g^{-1}x_n) &= \prod_{1 \leq i \leq n} \log f_{g^*\theta}(g^{-1}x_i) \\ &= \prod_{1 \leq i \leq n} \log((g^{-1})_* f_\theta(g^{-1}x_i)) \\ &= \prod_{1 \leq i \leq n} \log f_\theta(gg^{-1}x_i), \end{aligned}$$

and $\mathcal{L}(\theta, x_1, \dots, x_n) = \mathcal{L}(g^*\theta, g^{-1}x_1, \dots, g^{-1}x_n)$. Hence

$$\operatorname{argmax}_{\theta \in \mathcal{S}_t} \mathcal{L}(\theta, x_1, \dots, x_n) = g^{-1*}(to),$$

and

$$\widehat{\epsilon}_{MLE}(x_1, \dots, x_n) = \pi(\mathbb{E}_{g^{-1*}(to)}) = g(o) = \pi(\lambda g(o)) = \pi\left(\sum_i x_i\right).$$

2. We will now prove that on Lorentz cones this property characterizes the model $\mathcal{M}_\mathcal{S}$ among location models. Let \mathcal{C} be a Lorentz cone. Let $\{\tilde{f}_p\}$ be a set of densities on \mathcal{S} parametrized by $p \in \mathcal{S}$ such that for all $g \in G(\mathcal{S})$, $g.\tilde{f}_p = \tilde{f}_{g.p}$.

Let $o = (1, 0, \dots, 0)^T \in \mathcal{S}$. Consider the curve $p(t) = (\sqrt{1+t^2}, t, 0, \dots, 0)^T \in \mathcal{S}$ with $t \in \mathbb{R}$ and let $q(x) = (\sqrt{1+x^2}, x, 0, \dots, 0)^T \in \mathcal{S}$ with $x \in \mathbb{R}$. Note $g_t \in G(\mathcal{S})$ the linear map with matrix

$$\begin{pmatrix} \sqrt{1+t^2} & -t & 0 \\ -t & \sqrt{1+t^2} & 0 \\ 0 & 0 & I \end{pmatrix}$$

where the zeros are block matrices and I is the identity matrix. Clearly, $g_t(p(t)) = o$, hence $\tilde{f}_o(g_t(q(x))) = g_t^{-1}.\tilde{f}_o(q(x)) = \tilde{f}_{p(t)}(q(x))$. Since $g_t(q(x))$ is in \mathcal{S} and since only the first two coordinates of $g_t(q(x))$ are nonzero, $g_t(q(x))$ is a function of its second coordinate: $-t\sqrt{1+x^2} + x\sqrt{1+t^2}$. Therefore there exist two functions h and g such that $\ln(\tilde{f}_{p(t)}(q(x))) = h(-t\sqrt{1+x^2} + x\sqrt{1+t^2}) = g(t, x)$.

Consider the sample of size $n+1$ formed by n times $q(-x)$ and one time $q(nx)$ for some $x \in \mathbb{R}$: $(q(-x), \dots, q(-x), q(nx)) \in \mathcal{S}^{n+1}$. Since the usual empirical mean of the second coordinates of the sample is 0, the empirical log-extrinsic mean of the sample is o . A necessary condition for o to maximize the likelihood is

$$n \frac{\partial g}{\partial t}(0, -x) + \frac{\partial g}{\partial t}(0, nx) = 0$$

because the derivative with respect to t of the log-likelihood at o must be 0. We have

$$\frac{\partial g}{\partial t}(t, x) = \left(-\sqrt{1+x^2} + \frac{xt}{\sqrt{1+x^2}} \right) h' \left(-t\sqrt{1+x^2} + x\sqrt{1+t^2} \right),$$

hence $\frac{\partial g}{\partial t}(0, x) = -\sqrt{1+x^2}h'(x)$ and thus the above necessary condition becomes

$$n\sqrt{1+x^2}h'(-x) + \sqrt{1+(nx)^2}h'(nx) = 0.$$

Choosing $n = 1$, we obtain $h'(-x) = -h'(x)$. Therefore

$$\forall x \in \mathbb{R}, \forall n \in \mathbb{N} \setminus \{0\}, n\sqrt{1+x^2}h'(x) = \sqrt{1+(nx)^2}h'(nx).$$

Since the function $\phi(x) = \sqrt{1+x^2}h'(x)$ is continuous, the condition $\phi(nx) = n\phi(x)$ for all x and n implies that ϕ is a linear function: there is $a \in \mathbb{R}$ such that $\sqrt{1+x^2}h'(x) = ax$. With an integration, we obtain $h(x) = a\sqrt{1+x^2} + b$ for some $b \in \mathbb{R}$, and $g(0, x) = h(x) = a\sqrt{1+x^2} + b$. It follows that

$$\tilde{f}_o(q(x)) = e^{a\sqrt{1+x^2}+b} = e^{\langle ao, q(x) \rangle + b}. \quad (17)$$

For any g in the stabilizer $G(\mathcal{S})_o$ of o , $\tilde{f}_o = \tilde{f}_{g_o} = g \cdot \tilde{f}_o = \tilde{f}_o \circ g^{-1}$. Furthermore, it is possible to check that any $q \in \mathcal{S}$ is of the form $g^{-1}q(x)$ for some $g \in K_o$ and $x \in \mathbb{R}$. Since $K_o \subset G(\mathcal{S})_o \subset O(V)$, Eq.17 is valid for any $q \in \mathcal{S}$:

$$\tilde{f}_o(q) = e^{\langle ao, q \rangle + b}.$$

Let $s = \alpha(a) - 1$, where α is the function defined in theorem 6. We have then

$$\tilde{f}_o = f_{o,s}.$$

Since a location model is entirely determined by its density at any location, both model coincide.

References

- Tsuyoshi Ando, Chi-Kwong Li, and Roy Mathias. Geometric means. *Linear algebra and its applications*, 385:305–334, 2004.
- Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 56(2): 411–421, 2006.
- Andreas Arvanitogeorgos. *An introduction to Lie groups and the geometry of homogeneous spaces*, volume 22. American Mathematical Soc., 2003.

- Barndorff-Nielsen. Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of Statistics*, 5(3):151–157, 1978.
- O Barndorff-Nielsen, Preben Blaesild, J Ledet Jensen, and B Jørgensen. Exponential transformation models. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 379(1776):41–65, 1982.
- Rabi Bhattacharya and Vic Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds. *The Annals of Statistics*, 31(1):1–29, 2003.
- Rabi Bhattacharya and Vic Patrangenaru. Large sample theory of intrinsic and extrinsic sample means on manifolds—ii. 2005.
- Muriel Casalis and Gérard Letac. The Lukacs-Olkin-Rubin characterization of Wishart distributions on symmetric cones. *The Annals of Statistics*, 24(2):763–786, 1996.
- Emmanuel Chevallier, Didong Li, Yulong Lu, and David Dunson. Exponential-wrapped distributions on symmetric spaces. *SIAM Journal on Mathematics of Data Science*, 4(4):1347–1368, 2022.
- Alberto Dolcetti and Donato Pertici. Differential properties of spaces of symmetric real matrices. *arXiv preprint arXiv:1807.01113*, 2018.
- Bradley Efron. *Exponential families in theory and practice*. Cambridge University Press, 2022.
- Jacques Faraut and Adam Korányi. *Analysis on symmetric cones*. Oxford university press, 1994.
- G.A. Galperin. A concept of the mass center of a system of material points in the constant curvature spaces. *Communications in Mathematical Physics*, 154:63–84, 1993.
- Edwin T Jaynes. *Probability theory: The logic of science*. Cambridge university press, 2003.
- Jens Ledet Jensen. On the hyperboloid distribution. *Scandinavian Journal of Statistics*, 8(4):193–206, 1981.
- Federico Lopez, Beatrice Pozzetti, Steve Trettel, Michael Strube, and Anna Wienhard. Symmetric spaces for graph embeddings: A finsler-riemannian approach. In *International Conference on Machine Learning*, pages 7090–7101. PMLR, 2021.
- Hélène Massam. An exact decomposition theorem and a unified view of some related distributions for a class of exponential transformation models on symmetric cones. *The Annals of Statistics*, pages 369–394, 1994.

- Frank Nielsen and Kazuki Okamura. On the f -divergences between hyperboloid and poincaré distributions. In *International Conference on Geometric Science of Information*, pages 176–185. Springer, 2023.
- Michael Orlitzky. Jordan automorphisms and derivatives of symmetric cones. *Linear Algebra and its Applications*, 2024.
- Miklós Pálfi. Semigroups of operator means and generalized karcher equations. *arXiv preprint arXiv:1208.5603*, 2012.
- Oscar S Rothaus. Domains of positivity. In *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, volume 24, pages 189–235. Springer, 1960.
- Takeshi Sasaki. Hyperbolic affine hyperspheres. *Nagoya Mathematical Journal*, 77:107–123, 1980.
- Hans Schneider. Positive operators and an inertia theorem. *Numerische Mathematik*, 7(1): 11–17, 1965.
- Karl-Theodor Sturm. Probability measures on metric spaces of nonpositive. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces: Lecture Notes from a Quarter Program on Heat Kernels, Random Walks, and Analysis on Manifolds and Graphs: April 16-July 13, 2002, Emile Borel Centre of the Henri Poincaré Institute, Paris, France*, 338: 357, 2003.
- Robert A. Wijsman. *Invariant Measures on Groups and their Use in Statistics*. Institute of Mathematical Statistics Lecture Notes-Monograph Series, 1990.
- Takeaki Yamazaki. A brief introduction of the Karcher mean (Research on structures of operators via methods in geometry and probability theory). *Mathematics Research Institute Kokyuroku*, 1839:31–39, 2013.