On the f-Divergences Between Hyperboloid and Poincaré Distributions

Frank Nielsen[1](✉) and Kazuki Okamura[2]

[1] Sony Computer Science Laboratories Inc., Tokyo, Japan
Frank.Nielsen@acm.org

[2] Department of Mathematics, Faculty of Science, Shizuoka University, Shizuoka, Japan
okamura.kazuki@shizuoka.ac.jp

Abstract. Hyperbolic geometry has become popular in machine learning due to its capacity to embed discrete hierarchical graph structures with low distortions into continuous spaces for further downstream processing. It is thus becoming important to consider statistical models and inference methods for data sets grounded in hyperbolic spaces. In this work, we study the statistical f-divergences between two kinds of hyperbolic distributions: The Poincaré distributions and the related hyperboloid distributions. By exhibiting maximal invariants of group actions, we show how these f-divergences can be expressed as functions of canonical terms.

Keywords: exponential family · group action · maximal invariant · Csiszár's f-divergence · hyperbolic distributions

1 Introduction

Hyperbolic geometry[1] [2] is very well suited for embedding tree graphs with low distortions [20] as hyperbolic Delaunay subgraphs of embedded tree nodes. So a recent trend in machine learning and data science is to embed *discrete* hierarchical graphs into *continuous* spaces with low distortions for further downstream processing. There exists many models of hyperbolic geometry [2] like the Poincaré disk or upper-half plane conformal models, the Klein non-conformal disk model, the Beltrami hemisphere model, the Minkowski or Lorentz hyperboloid model, etc. We can transform one model of hyperbolic geometry to another model by a bijective mapping yielding a corresponding isometric embedding [11]. As a byproduct of the low-distortion hyperbolic embeddings of hierarchical graphs, many embedded data sets are nowadays available in hyperbolic model spaces, and those data sets need to be further processed. Thus it is important to build *statistical models* and *inference methods* for these hyperbolic data

[1] Hyperbolic geometry has constant negative curvature and the volume of hyperbolic balls increases exponentially with respect to their radii rather than polynomially as in Euclidean spaces.

F. Nielsen and F. Barbaresco (Eds.): GSI 2023, LNCS 14071, pp. 176–185, 2023.
https://doi.org/10.1007/978-3-031-38271-0_18

sets using probability distributions with support hyperbolic model spaces, and to consider statistical mixtures in those spaces for modeling arbitrary smooth densities.

Let us quickly review some of the various families of probability distributions defined in hyperbolic models as follows: One of the very first proposed family of such "hyperbolic distributions" was proposed in 1981 [16] and are nowadays commonly called the *hyperboloid distributions* in reference to their support: The hyperboloid distributions are defined on the Minkowski upper sheet hyperboloid by analogy to the von-Mises Fisher distributions [3] which are defined on the sphere. Another work by Barbaresco [4] defined the so-called Souriau-Gibbs distributions (2019) in the Poincaré disk (Eq. 57 of [4], a natural exponential family) with its Fisher information metric coinciding with the Poincaré hyperbolic Riemannian metric (the Poincaré unit disk is a homogeneous space where the Lie group $\mathrm{SU}(1,1)$ acts transitively).

In this paper, we focus on Ali-Silvey-Csiszár's f-divergences between hyperbolic distributions [1,14]. In Sect. 2, we prove using Eaton's method of group action maximal invariants [15,19] that all f-divergences (including the Kullback-Leibler divergence) between Poincaré distributions [21] can be expressed canonically as functions of three terms (Proposition 1 and Theorem 1). Then, we deal with the hyperboloid distributions in dimension 2 in §3. We also consider q-deformed family of these distributions [23]. We exhibit a correspondence in §4 between the upper-half plane and the Minkowski hyperboloid 2D sheet. The f-divergences between the hyperboloid distributions are in spirit very geometric because it exhibits a beautiful and clear maximal invariant which has connections with the side-angle-side congruence criteria for triangles in hyperbolic geometry. This paper summarizes the preprint [18] with some proofs omitted: We refer the reader to the preprint for more details and other topics than f-divergences.

2 The Poincaré Distributions

Tojo and Yoshino [21–23] described a versatile method to build exponential families of distributions on homogeneous spaces which are invariant under the action of a Lie group G generalizing the construction in [13]. They exemplify their so-called "G/H-method" on the upper-half plane $\mathbb{H} := \{(x,y) \in \mathbb{R}^2 : \ y > 0\}$ by constructing an exponential family with probability density functions invariant under the action of Lie group $G = \mathrm{SL}(2,\mathbb{R})$, the set of invertible matrices with unit determinant. We call these distributions the Poincaré distributions, since their sample space $\mathcal{X} = G/H \simeq \mathbb{H}$, and we study this set of distributions as an exponential family [8]: The probability density function (pdf) of a Poincaré distribution [21] expressed using a 3D vector parameter $\theta = (a,b,c) \in \mathbb{R}^3$ is given by

$$p_\theta(x,y) := \frac{\sqrt{ac-b^2}\exp(2\sqrt{ac-b^2})}{\pi} \exp\left(-\frac{a(x^2+y^2)+2bx+c}{y}\right)\frac{1}{y^2}, \quad (1)$$

where θ belongs to the parameter space

$$\Theta := \{(a, b, c) \in \mathbb{R}^3 \ : a > 0, c > 0, \ ac - b^2 > 0\}.$$

The set Θ forms an open 3D convex cone. Thus the Poincaré distribution family has a 3D parameter cone space and the sample space is the hyperbolic upper plane. We can also use a matrix form to express the pdf. Indeed, we can naturally identify Θ with the set of real symmetric positive-definite matrices $\mathrm{Sym}^+(2, \mathbb{R})$ by the mapping $(a, b, c) \mapsto \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. Hereafter, we denote the determinant of θ by $|\theta| := ac - b^2 > 0$ and the trace of θ by $\mathrm{tr}(\theta) = a + c$ for $\theta = (a, b, c) \simeq \begin{bmatrix} a & b \\ b & c \end{bmatrix}$.

The f-divergence [1,14] induced by a convex generator $f : (0, \infty) \to \mathbb{R}$ between two pdfs $p(x, y)$ and $q(x, y)$ defined on the support $\mathbb{H}$ is defined by

$$D_f[p : q] := \int_{\mathbb{H}} p(x, y)\, f\left(\frac{q(x, y)}{p(x, y)}\right) \mathrm{d}x\, \mathrm{d}y. \tag{2}$$

Since $D_f[p : q] \geq f(1)$, we consider convex generators $f(u)$ such that $f(1) = 0$. Moreover, in order to satisfy the law of the indiscernibles (i.e., $D_f[p : q] = 0$ iff $p(x, y) = q(x, y)$), we require f to be strictly convex at 1. The class of f-divergences includes the total variation distance ($f(u) = |u - 1|$), the Kullback-Leibler divergence ($f(u) = -\log(u)$, and its two common symmetrizations, namely, the Jeffreys divergence and the Jensen-Shannon divergence), the squared Hellinger divergence, the Pearson and Neyman sided χ^2-divergences, the α-divergences, etc.

We state the notion of maximal invariant by following [15]: Let G be a group acting on a set X. We denote the group action by $(g, x) \mapsto gx$.

Definition 1 (Maximal invariant). *We say that a map φ from X to a set Y is maximal invariant if it is invariant, specifically, $\varphi(gx) = \varphi(x)$ for every $g \in G$ and $x \in X$, and furthermore, whenever $\varphi(x_1) = \varphi(x_2)$ there exists $g \in G$ such that $x_2 = gx_1$.*

It can be shown that every invariant map is a function of a maximal invariant [15]. Specifically, if a map ψ from X to a set Z is invariant, then, there exists a unique map Φ from $\varphi(X)$ to Z such that $\Phi \circ \varphi = \psi$.

These invariant/maximal invariant concepts can be understood using group orbits: For each $x \in X$, we may consider its orbit $O_x := \{gx \in X : \ g \in G\}$. A map is invariant when it is constant on orbits and maximal invariant when orbits have distinct map values.

We denote by $A^\top$e the transpose of a square matrix A and $A^{-\top}$ the transpose of the inverse matrix A^{-1} of a regular matrix A. It holds that $A^{-\top} = (A^\top)^{-1}$.

Let $\mathrm{SL}(2, \mathbb{R})$ be the group of 2×2 real matrices with unit determinant.

Proposition 1. *Define a group action of* $\mathrm{SL}(2, \mathbb{R})$ *to* $\mathrm{Sym}^+(2, \mathbb{R})^2$ *by*

$$(g, (\theta, \theta')) \mapsto (g^{-\top}\theta g^{-1}, g^{-\top}\theta' g^{-1}). \tag{3}$$

Define a map $S : \mathrm{Sym}^+(2,\mathbb{R})^2 \to (\mathbb{R}_{>0})^2 \times \mathbb{R}$ *by*

$$S(\theta,\theta') := \left(|\theta|, |\theta'|, \mathrm{tr}(\theta'\theta^{-1})\right). \tag{4}$$

Then, the map S *is maximal invariant of the group action.*

Proof. Observe that S is invariant with respect to the group action: $S(\theta,\theta') = S(g.\theta, g.\theta')$. Assume that $S\left(\theta^{(1)},\theta^{(2)}\right) = S\left(\widetilde{\theta^{(1)}},\widetilde{\theta^{(2)}}\right)$. We see that there exists $g_{\theta^{(1)}} \in \mathrm{SL}(2,\mathbb{R})$ such that $g_{\theta^{(1)}}.\theta^{(1)} = g_{\theta^{(1)}}^{-\top}\theta^{(1)}g_{\theta^{(1)}}^{-1} = \sqrt{|\theta^{(1)}|}I_2$, where I_2 denotes the 2×2 identity matrix. Then, $\theta^{(1)} = \sqrt{|\theta^{(1)}|}g_{\theta^{(1)}}^{\top}g_{\theta^{(1)}}$. Let $\theta^{(3)} := g_{\theta^{(1)}}.\theta^{(2)} = g_{\theta^{(1)}}^{-\top}\theta^{(2)}g_{\theta^{(1)}}^{-1}$. Then $\mathrm{tr}\left(\theta^{(3)}\right) = \mathrm{tr}\left(\theta^{(2)}g_{\theta^{(1)}}^{-1}g_{\theta^{(1)}}^{-\top}\right) = \sqrt{|\theta^{(1)}|}\,\mathrm{tr}\left(\theta^{(2)}(\theta^{(1)})^{-1}\right)$. We define $g_{\widetilde{\theta^{(1)}}}$ and $\widetilde{\theta^{(3)}}$ in the same manner. Then, $\mathrm{tr}\left(\theta^{(3)}\right) = \mathrm{tr}\left(\widetilde{\theta^{(3)}}\right)$ and $|\theta^{(3)}| = \left|\widetilde{\theta^{(3)}}\right|$. Hence the set of eigenvalues of $\theta^{(3)}$ and $\widetilde{\theta^{(3)}}$ are identical with each other. By this and $\theta^{(3)}, \widetilde{\theta^{(3)}} \in \mathrm{Sym}(2,\mathbb{R})$, there exists $h \in \mathrm{SO}(2)$ such that $h.\theta^{(3)} = \widetilde{\theta^{(3)}}$. Hence $(hg_{\theta^{(1)}}).\theta^{(2)} = g_{\widetilde{\theta^{(1)}}}\widetilde{\theta^{(2)}}$. We also see that

$$(hg_{\theta^{(1)}}).\theta^{(1)} = g_{\theta^{(1)}}.\theta^{(1)} = \sqrt{\left|\theta^{(1)}\right|}\,I_2 = \sqrt{\left|\widetilde{\theta^{(1)}}\right|}\,I_2 = g_{\widetilde{\theta^{(1)}}}.\widetilde{\theta^{(1)}}.$$

Thus we have $\left(\widetilde{\theta^{(1)}},\widetilde{\theta^{(2)}}\right) = (g_{\widetilde{\theta^{(1)}}}^{-1}hg_{\theta^{(1)}}).(\theta^{(1)},\theta^{(2)})$.

Remark 1 (This is pointed by an anonymous referee.). We can consider an extension of Proposition 1 to a case of higher degree of matrices. Let $n \geq 2$ and assume that $\theta,\theta' \in \mathrm{Sym}(n,\mathbb{R})$. Let $P_{\theta,\theta'}(t) := |(1-t)\theta + t\theta'|$ for $t \in \mathbb{R}$. where $|A|$ denotes the determinant of a square matrix A. This is a polynomial in t with degree n. Assume that $P_{\theta_1,\theta_1'} = P_{\theta_2,\theta_2'}$ for $\theta_1,\theta_1',\theta_2,\theta_2' \in \mathrm{Sym}(n,\mathbb{R})$. We can factor θ_i as $\theta_i = L_i^\top L_i$ for some L_i, $i = 1,2$. Let I_n be the identity matrix of degree n. Then, $P_{\theta_i,\theta_i'}(t) = |\theta_i||I_n + t(L_i^{-\top}\theta_i'L_i^{-1} - I_n)|$, $i = 1,2$. Since $L_i^{-\top}\theta_i'L_i^{-1} \in \mathrm{Sym}(n,\mathbb{R})$, the set of eigenvalues of $L_1^{-\top}\theta_i'L_1^{-1}$ and $L_2^{-\top}\theta_i'L_2^{-1}$ is identical with each other. Hence there exists an orthogonal matrix Q such that $L_2^{-\top}\theta_i'L_2^{-1} = Q^\top L_1^{-\top}\theta_i'L_1^{-1}Q$. Let $G := L_1^{-1}QL_2$. Then, $\theta_2 = G^\top\theta_1 G$ and $\theta_2' = G^\top\theta_1'G$. We finally remark that $P_{\theta_1,\theta_1'} = P_{\theta_2,\theta_2'}$ holds if and only if $P_{\theta_1,\theta_1'}(t) = P_{\theta_2,\theta_2'}(t)$ for $n+1$ different values of t.

If $n = 2$, then,

$$P_{\theta,\theta'}(t) = (1-t)^2|\theta|^2 + t^2|\theta'|^2 + t(1-t)\,|\theta|\,\mathrm{tr}(\theta'\theta^{-1}). \tag{5}$$

Hence the arguments above give an alternative proof of Proposition 1.

Proposition 2 (Invariance of f-divergences under group action).

$$D_f\left[p_\theta : p_{\theta'}\right] = D_f\left[p_{g^{-\top}\theta g^{-1}} : p_{g^{-\top}\theta' g^{-1}}\right].$$

For $g \in \mathrm{SL}(2,\mathbb{R})$, we denote the pushforward measure of a measure ν on $\mathbb{H}$ by the map $z \mapsto g.z$ on $\mathbb{H}$ by $\nu \circ g^{-1}$.

The latter part of the following proof utilizes the method used in the proof of [21, Proposition 1].

Proof. We first see that for $g \in \mathrm{SL}(2,\mathbb{R})$,

$$D_f\left[p_\theta : p_{\theta'}\right] = D_f\left[p_\theta \circ g^{-1} : p_{\theta'} \circ g^{-1}\right]. \quad (6)$$

Let $\mu(\mathrm{d}x\mathrm{d}y) := \mathrm{d}x\mathrm{d}y/y^2$. Then it is well-known that μ is invariant with respect to the action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{H}$, that is, $\mu = \mu \circ g^{-1}$ for $g \in \mathrm{SL}(2,\mathbb{R})$.

Define a map $\varphi : \Theta \times \mathbb{H} \to \mathbb{R}_{>0}$ by

$$\varphi(\theta, x+yi) := \frac{a(x^2+y^2)+2bx+c}{y}, \ \theta = \begin{bmatrix} a & b \\ b & c \end{bmatrix}.$$

Then, $\varphi(\theta, z) = \varphi(g.\theta, g.z)$ for $g \in \mathrm{SL}(2,\mathbb{R})$.

Since

$$p_\theta(x,y)\mathrm{d}x\mathrm{d}y = \frac{\sqrt{|\theta|}\exp(2\sqrt{|\theta|})}{\pi}\exp(-\varphi(\theta, x+yi))\mu(\mathrm{d}x\mathrm{d}y),$$

we have $p_\theta \circ g^{-1} = p_{g.\theta}$. Hence,

$$D_f\left[p_\theta \circ g^{-1} : p_{\theta'} \circ g^{-1}\right] = D_f\left[p_{g.\theta} : p_{g.\theta'}\right]. \quad (7)$$

The assertion follows from (6) and (7).

By Propositions 1 and 2, we get

Theorem 1. *Every f-divergence between two Poincaré distributions p_θ and $p_{\theta'}$ is a function of $\left(|\theta|, |\theta'|, \mathrm{tr}\left(\theta'\theta^{-1}\right)\right)$ and invariant with respect to the $\mathrm{SL}(2,\mathbb{R})$-action.*

We obtained exact formulae for the Kullback-Leibler divergence, the squared Hellinger divergence, and the Neyman chi-squared divergence.

Proposition 3. *We have the following results for two Poincaré distributions p_θ and $p_{\theta'}$.*

(i) (Kullback-Leibler divergence) Let $f(u) = -\log u$. Then,

$$D_f\left[p_\theta : p_{\theta'}\right] = \frac{1}{2}\log\frac{|\theta|}{|\theta'|} + 2\left(\sqrt{|\theta|} - \sqrt{|\theta'|}\right) + \left(\frac{1}{2} + \sqrt{|\theta|}\right)\left(\mathrm{tr}(\theta'\theta^{-1}) - 2\right). \quad (8)$$

(ii) (squared Hellinger divergence) Let $f(u) = (\sqrt{u}-1)^2/2$. Then,

$$D_f[p_\theta : p_{\theta'}] = 1 - \frac{2|\theta|^{1/4}|\theta'|^{1/4}\exp\left(|\theta|^{1/2} + |\theta'|^{1/2}\right)}{|\theta+\theta'|^{1/2}\exp\left(|\theta+\theta'|^{1/2}\right)}. \quad (9)$$

(iii) (Neyman chi-squared divergence) Let $f(u) := (u-1)^2$. Assume that $2\theta' - \theta \in \Theta$. Then,

$$D_f[p_\theta : p_{\theta'}] = \frac{|\theta'|\exp(4|\theta'|^{1/2})}{|\theta|^{1/2}|2\theta'-\theta|^{1/2}\exp\left(2(|\theta|^{1/2} + |2\theta'-\theta|^{1/2})\right)} - 1. \quad (10)$$

We remark that $|\theta + \theta'|$ and $|2\theta' - \theta|$ can be expressed by using $|\theta|, |\theta'|$, and $\mathrm{tr}(\theta'\theta^{-1})$. Indeed, we have

$$|\theta + \theta'| = |\theta| + |\theta'| + |\theta| \,\mathrm{tr}(\theta'\theta^{-1}),$$
$$|2\theta' - \theta| = 4|\theta'| + |\theta| - 2|\theta| \,\mathrm{tr}(\theta'\theta^{-1}).$$

Thus the KLD between two Poincaré distributions is asymmetric in general. The situation is completely different from the Cauchy distribution whose f-divergences are always symmetric [19,24].

Recently, Tojo and Yoshino [23] introduced a notion of deformed exponential family associated with their G/H method in representation theory. As an example of it, they considered a family of *deformed Poincaré distributions* with index $q > 1$. For $x \in I_q := \{x \in \mathbb{R} : (1-q)x + 1 > 0\}$, let

$$\exp_q(x) := ((1-q)x + 1)^{1/(1-q)}, \quad x \in I_q.$$

For $q \in [1, 2)$, let a q-deformed Poincaré distribution be the distribution

$$p_\theta(x, y) := c_q(\sqrt{|\theta|}) \exp_q\left(-\frac{a(x^2+y^2) + 2bx + c}{y}\right) \frac{1}{y^2}, \tag{11}$$

where $\theta \in \Theta$ and $c_q(x) := \dfrac{(2-q)x}{\pi(\exp_q(-2x))^{2-q}}$. In this case, Proposition 2 holds for q-deformed Poincaré distributions, so we also obtain that

Theorem 2. *Let $q \in [1, 2)$. Every f-divergence between two q-deformed Poincaré distributions p_θ and $p_{\theta'}$ is a function of $\left(|\theta|, |\theta'|, \mathrm{tr}\left(\theta'\theta^{-1}\right)\right)$.*

We can show this by Theorem 4 below and the correspondence principle in §4.

3 The Two-Dimensional Hyperboloid Distributions

We first give the definition of the Lobachevskii space (in reference to Minkowski hyperboloid model of hyperbolic geometry also called the Lorentz model) and the parameter space of the hyperboloid distribution. We focus on the bidimensional case $d = 2$. Let

$$\mathbb{L}^2 := \left\{(x_0, x_1, x_2) \in \mathbb{R}^3 : x_0 = \sqrt{1 + x_1^2 + x_2^2}\right\},$$

and

$$\Theta_{\mathbb{L}^2} := \left\{(\theta_0, \theta_1, \theta_2) \in \mathbb{R}^3 : \theta_0 > \sqrt{\theta_1^2 + \theta_2^2}\right\}.$$

Let the Minkowski inner product [12] be

$$[(x_0, x_1, x_2), (y_0, y_1, y_2)] := x_0 y_0 - x_1 y_1 - x_2 y_2.$$

We have $\mathbb{L}^2 = \{x \in \mathbb{R}^3 \ : \ [x, x] = 1\}$.

Now we define the *hyperboloid distribution* by following [5,7,9]. Hereafter, for ease of notation, we let $|\theta| := [\theta, \theta]^{1/2},\ \theta \in \Theta_{\mathbb{L}^2}$. For $\theta \in \Theta_{\mathbb{L}^2}$, we define a probability measure P_θ on $\mathbb{L}^d \simeq \mathbb{R}^d$ by

$$P_\theta(\mathrm{d}x_1\mathrm{d}x_2) := c_2(|\theta|)\exp(-[\theta, \widetilde{x}])\mu(\mathrm{d}x_1\mathrm{d}x_2), \tag{12}$$

where we let $c_2(t) := \frac{t\exp(t)}{2(2\pi)^{1/2}}$, $t > 0$, $\widetilde{x} := \left(\sqrt{1+x_1^2+x_2^2}, x_1, x_2\right)$, and $\mu(\mathrm{d}x_1\mathrm{d}x_2) := \frac{1}{\sqrt{1+x_1^2+x_2^2}}\mathrm{d}x_1\mathrm{d}x_2$.

Remark 2. The 1D hyperboloid distribution was first introduced in statistics in 1977 [6] to model the log-size distributions of particles from aeolian sand deposits, but the 3D hyperboloid distribution was later found already studied in statistical physics in 1911 [17]. The 2D hyperboloid distribution was investigated in 1981 [10].

Now we consider group actions on the space of parameters $\Theta_{\mathbb{L}^2}$. Let the indefinite special orthogonal group be

$$\mathrm{SO}(1,2) := \left\{A \in \mathrm{SL}(3,\mathbb{R}) : [Ax, Ay] = [x, y]\ \forall x, y \in \mathbb{R}^3\right\},$$

and $\mathrm{SO}_0(1,2) := \left\{A \in \mathrm{SO}(1,2) : A(\mathbb{L}^2) = \mathbb{L}^2\right\}$.

An action of $\mathrm{SO}_0(1,2)$ to $(\Theta_{\mathbb{L}^2})^2$ is defined by

$$\mathrm{SO}_0(1,2) \times (\Theta_{\mathbb{L}^2})^2 \ni (A, (\theta, \theta')) \mapsto (A\theta, A\theta') \in (\Theta_{\mathbb{L}^2})^2.$$

Proposition 4. *$(\theta, \theta') \mapsto ([\theta, \theta], [\theta', \theta'], [\theta, \theta'])$ is maximal invariant for the action of $\mathrm{SO}_0(1,2)$ to $(\Theta_{\mathbb{L}^2})^2$.*

In the following proof, all vectors are column vectors.

Proof. It is clear that the map is invariant with respect to the group action. Assume that

$$\left([\theta^{(1)}, \theta^{(1)}], [\theta^{(2)}, \theta^{(2)}], [\theta^{(1)}, \theta^{(2)}]\right) = \left(\left[\widetilde{\theta^{(1)}}, \widetilde{\theta^{(1)}}\right], \left[\widetilde{\theta^{(2)}}, \widetilde{\theta^{(2)}}\right], \left[\widetilde{\theta^{(1)}}, \widetilde{\theta^{(2)}}\right]\right).$$

Let $\psi_i := \frac{\theta^{(i)}}{|\theta^{(i)}|}$, $\widetilde{\psi_i} := \frac{\widetilde{\theta^{(i)}}}{\left|\widetilde{\theta^{(i)}}\right|}$, $i = 1, 2$. Then, $[\psi_1, \psi_2] = \left[\widetilde{\psi_1}, \widetilde{\psi_2}\right]$.

We first consider the case that $\psi_1 = \widetilde{\psi_1} = (1, 0, 0)^\top$. Let $\psi_i = (x_{i0}, x_{i1}, x_{i2})^\top, \widetilde{\psi_i} = (\widetilde{x_{i0}}, \widetilde{x_{i1}}, \widetilde{x_{i2}})^\top$, $i = 1, 2$. Then, $x_{20} = \widetilde{x_{20}} > 0$, $x_{21}^2 + x_{22}^2 = \widetilde{x_{21}}^2 + \widetilde{x_{22}}^2$ and hence there exists a special orthogonal matrix P such that $P(x_{21}, x_{22})^\top = (\widetilde{x_{21}}, \widetilde{x_{22}})^\top$. Let $A := \begin{pmatrix} 1 & 0 \\ 0 & P \end{pmatrix}$. Then, $A \in \mathrm{SO}_0(1,2)$, $A\psi_1 = (1, 0, 0)^\top = \widetilde{\psi_1}$ and $A\psi_2 = \widetilde{\psi_2}$.

We second consider the general case. Since the action of $\mathrm{SO}_0(1,2)$ to $\mathbb{L}^2$ defined by $(A, \psi) \mapsto A\psi$ is transitive, there exist $A, B \in \mathrm{SO}_0(1,2)$ such that $A\psi_1 = B\widetilde{\psi_1} = (1, 0, 0)^\top$. Thus this case is attributed to the first case.

We regard μ as a probability measure on $\mathbb{L}^2$. We recall that $[A\theta, A\widetilde{x}] = [\theta, \widetilde{x}]$ for $A \in \mathrm{SO}_0(1,2)$. We remark that μ is an $\mathrm{SO}(1,2)$-invariant Borel measure [16] on $\mathbb{L}^2$. Now we have that

Theorem 3. *Every f-divergence between p_θ and $p_{\theta'}$ is invariant with respect to the action of $\mathrm{SO}_0(1,2)$, and is a function of the triplet $([\theta,\theta],[\theta',\theta'],[\theta,\theta'])$, i.e., the pairwise Minkowski inner products of θ and θ'.*

There is a clear geometric interpretation of this fact: The side-angle-side theorem for triangles in Euclidean geometry states that if two sides and the included angle of one triangle are equal to two sides and the included angle of another triangle, then the triangles are congruent. This is also true for the hyperbolic geometry and it corresponds to Proposition 4 above. Every f-divergence is determined by the triangle formed by a pair of the parameters (θ, θ') when f is fixed.

Proposition 5. *We have the following results for two hyperboloid distributions p_θ and $p_{\theta'}$.*

(i) (Kullback-Leibler divergence) Let $f(u) = -\log u$. Then,

$$D_f[p_\theta : p_{\theta'}] = \log\left(\frac{|\theta|}{|\theta'|}\right) - |\theta'| + \frac{[\theta,\theta']}{[\theta,\theta]} + \frac{[\theta,\theta']}{|\theta|} - 1. \tag{13}$$

(ii) (squared Hellinger divergence) Let $f(u) = (\sqrt{u}-1)^2/2$. Then,

$$D_f[p_\theta : p_{\theta'}] = 1 - \frac{2|\theta|^{1/2}|\theta'|^{1/2}\exp\left(|\theta|/2 + |\theta'|/2\right)}{|\theta+\theta'|\exp\left(|\theta+\theta'|/2\right)}. \tag{14}$$

(iii) (Neyman chi-squared divergence) Let $f(u) := (u-1)^2$. Assume that $2\theta' - \theta \in \Theta_{\mathbb{L}^2}$. Then,

$$D_f[p_\theta : p_{\theta'}] = \frac{|\theta'|^2\exp(2|\theta'|)}{|\theta||2\theta'-\theta|\exp(|\theta| + |2\theta'-\theta|)} - 1. \tag{15}$$

Now we consider deformations of the hyperboloid distribution. For $q \in [1,2)$, we let a q-deformed hyperboloid distribution be the distribution

$$p_\theta(x_1,x_2) := c_q(|\theta|)\exp_q\left(-[\theta,\widetilde{x}]\right)\frac{1}{\sqrt{1+x_1^2+x_2^2}}, \tag{16}$$

where $c_q(z) := \dfrac{(2-q)z}{2\pi(\exp_q(-z))^{2-q}}$.

In the same manner as in the derivation of Theorem 3, we obtain that

Theorem 4 (Canonical terms of the f-divergences between deformed hyperboloid distributions). *Let $q \in [1,2)$. Then, every f-divergence between q-deformed hyperboloid distributions p_θ and $p_{\theta'}$ is invariant with respect to the action of $\mathrm{SO}_0(1,2)$, and is a function of the triplet $([\theta,\theta],[\theta',\theta'],[\theta,\theta'])$.*

4 Correspondence Principle

It is well-known that there exists a correspondence between the 2D Lobachevskii space $\mathbb{L} = \mathbb{L}^2$ (hyperboloid model) and the Poincaré upper-half plane $\mathbb{H}$.

Proposition 6 (Correspondence between the parameter spaces). *For* $\theta = (a, b, c) \in \Theta_{\mathbb{H}} := \left\{(a, b, c) : a > 0, c > 0, ac > b^2\right\}$, *let* $\theta_{\mathbb{L}} := (a+c, a-c, 2b) \in \Theta_{\mathbb{L}}$. *We denote the f-divergence on $\mathbb{L}$ and $\mathbb{H}$ by $D_f^{\mathbb{L}}[\cdot : \cdot]$ and $D_f^{\mathbb{H}}[\cdot : \cdot]$ respectively. Then,*

(i) For $\theta, \theta' \in \Theta_{\mathbb{H}}$,

$$|\theta_{\mathbb{L}}|^2 = [\theta_{\mathbb{L}}, \theta_{\mathbb{L}}] = 4|\theta|, \;\; |\theta'_{\mathbb{L}}|^2 = [\theta'_{\mathbb{L}}, \theta'_{\mathbb{L}}] = 4|\theta'|, \;\; [\theta_{\mathbb{L}}, \theta'_{\mathbb{L}}] = 2|\theta|\mathrm{tr}(\theta'\theta^{-1}). \quad (17)$$

(ii) For every f and $\theta, \theta' \in \mathbb{H}$,

$$D_f^{\mathbb{L}}\left[p_{\theta_{\mathbb{L}}} : p_{\theta'_{\mathbb{L}}}\right] = D_f^{\mathbb{H}}\left[p_\theta : p_{\theta'}\right]. \quad (18)$$

For (i), at its first glance, there seems to be an inconsistency in notation. However, $|\theta|$ is the Minkowski norm for $\theta \in \theta_{\mathbb{L}}$, and, $|\theta|$ is the determinant for $\theta \in \Theta_v$, so the notation is consistent in each setting. By this assertion, it suffices to compute the f-divergences between the hyperboloid distributions on $\mathbb{L}$.

Let $\mu_{\mathbb{H}}(\mathrm{d}x\mathrm{d}y) := \frac{\mathrm{d}x\mathrm{d}y}{y^2}$ and $\mu_{\mathbb{L}}(\mathrm{d}x\mathrm{d}y) := \frac{\mathrm{d}x\mathrm{d}y}{\sqrt{1+x^2+y^2}}$. By the change of variable

$$\mathbb{H} \ni (x, y) \mapsto (X, Y) = \left(\frac{1-x^2-y^2}{2y}, \frac{x}{y}\right) \in \mathbb{R}^2,$$

by recalling the correspondence between the parameters in Eq. (17), it holds that $y^2 p_\theta(x, y) = \sqrt{1+X^2+Y^2} p_{\theta^{\mathbb{L}}}(X, Y)$, and $\mu_{\mathbb{H}}(\mathrm{d}x\mathrm{d}y) = \mu_{\mathbb{L}}(\mathrm{d}X\mathrm{d}Y)$.

Acknowledgements. The authors are grateful to two anonymous referees for valuable comments. It is worth of special mention that Remark 1 is suggested by one referee and the proof of Proposition 2 is suggested by the other referee.

References

1. Ali, S.M., Silvey, S.D.: A general class of coefficients of divergence of one distribution from another. J. Roy. Stat. Soc.: Ser. B (Methodol.) **28**(1), 131–142 (1966)
2. Anderson, J.W.: Hyperbolic geometry. Springer Science & Business Media (2006)
3. Banerjee, A., Dhillon, I.S., Ghosh, J., Sra, S., Ridgeway, G.: Clustering on the Unit Hypersphere using von Mises-Fisher Distributions. J. Mach. Learn. Res. **6**(9) (2005)
4. Barbaresco, F.: Lie group machine learning and gibbs density on Poincaré unit disk from souriau lie groups thermodynamics and su(1,1) coadjoint orbits. In: Nielsen, F., Barbaresco, F. (eds.) GSI 2019. LNCS, vol. 11712, pp. 157–170. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-26980-7_17
5. Barndorff-Nielsen, O.: Hyperbolic distributions and distributions on hyperbolae. Scand. J. Stat. **5**, 151–157 (1978)

6. Barndorff-Nielsen, O.: Exponentially decreasing distributions for the logarithm of particle size. Proc. Royal Society of London. A. Math. Phys. Sci. **353**(1674), 401–419 (1977)
7. Barndorff-Nielsen, O.: The hyperbolic distribution in statistical physics. Scandinavian J. Stat. **9**(1), 43–46 (1982)
8. Barndorff-Nielsen, O.: Information and exponential families: in statistical theory. John Wiley & Sons (2014)
9. Barndorff-Nielsen, O., Blaesild, P.: Hyperbolic distributions and ramifications: Contributions to theory and application. In: Statistical distributions in scientific work, pp. 19–44. Springer (1981). https://doi.org/10.1007/978-94-009-8549-0_2
10. Blæsild, P.: The two-dimensional hyperbolic distribution and related distributions, with an application to Johannsen's bean data. Biometrika **68**(1), 251–263 (1981)
11. Cannon, J.W., Floyd, W.J., Kenyon, R., Parry, W.R., et al.: Hyperbolic geometry. Flavors of geometry **31**(59–115), 2 (1997)
12. Cho, H., DeMeo, B., Peng, J., Berger, B.: Large-margin classification in hyperbolic space. In: The 22nd International Conference on Artificial Intelligence and Statistics, pp. 1832–1840. PMLR (2019)
13. Cohen, T., Welling, M.: Harmonic exponential families on manifolds. In: International Conference on Machine Learning, pp. 1757–1765. PMLR (2015)
14. Csiszár, I.: Eine informationstheoretische ungleichung und ihre anwendung auf beweis der ergodizitaet von markoffschen ketten. Magyer Tud. Akad. Mat. Kutato Int. Koezl. 8, 85–108 (1964)
15. Eaton, M.L.: Group invariance applications in statistics. Hayward, CA: Institute of Mathematical Statistics; Alexandria, VA: American Statistical Association (1989)
16. Jensen, J.L.: On the hyperboloid distribution. Scandinavian J. Stat.**8**(4), 193–206 (1981)
17. Jüttner, F.: Das Maxwellsche Gesetz der Geschwindigkeitsverteilung in der Relativtheorie. Ann. Phys. **339**(5), 856–882 (1911)
18. Nielsen, F., Okamura, K.: Information measures and geometry of the hyperbolic exponential families of Poincaré and hyperboloid distributions. arXiv preprint arXiv:2205.13984 (2022)
19. Nielsen, F., Okamura, K.: On f-divergences between Cauchy distributions. IEEE Trans. Inf. Theory **69**(5), 3150–3171 (2023)
20. Sarkar, R.: Low distortion delaunay embedding of trees in hyperbolic plane. In: van Kreveld, M., Speckmann, B. (eds.) GD 2011. LNCS, vol. 7034, pp. 355–366. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-25878-7_34
21. Tojo, K., Yoshino, T.: An exponential family on the upper half plane and its conjugate prior. In: Barbaresco, F., Nielsen, F. (eds.) SPIGL 2020. SPMS, vol. 361, pp. 84–95. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77957-3_4
22. Tojo, K., Yoshino, T.: Harmonic exponential families on homogeneous spaces. Inform. Geometry **4**(1), 215–243 (2021)
23. Tojo, K., Yoshino, T.: Classification problem of invariant q-exponential families on homogeneous spaces. In: Mathematical optimization and statistical theories using geometric methods OCAMI Reports, Vol. 8, pp. 85–104. Osaka Central Advanced Mathematical Institute (OCAMI) (2022). https://doi.org/10.24544/ocu.20221208-007
24. Verdú, S.: The Cauchy Distribution in Information Theory. entropy **25**(2), 346 (2023)