entropy

MDPI

Article

# On Voronoi Diagrams on the Information-Geometric Cauchy Manifolds

**Frank Nielsen**

Sony Computer Science Laboratories, Tokyo 141-0022, Japan; nielsen@lix.polytechnique.fr or Frank.Nielsen@acm.org

Received: 11 June 2020; Accepted: 24 June 2020; Published: 28 June 2020


**Abstract:** We study the Voronoi diagrams of a finite set of Cauchy distributions and their dual complexes from the viewpoint of information geometry by considering the Fisher-Rao distance, the Kullback-Leibler divergence, the chi square divergence, and a flat divergence derived from Tsallis entropy related to the conformal flattening of the Fisher-Rao geometry. We prove that the Voronoi diagrams of the Fisher-Rao distance, the chi square divergence, and the Kullback-Leibler divergences all coincide with a hyperbolic Voronoi diagram on the corresponding Cauchy location-scale parameters, and that the dual Cauchy hyperbolic Delaunay complexes are Fisher orthogonal to the Cauchy hyperbolic Voronoi diagrams. The dual Voronoi diagrams with respect to the dual flat divergences amount to dual Bregman Voronoi diagrams, and their dual complexes are regular triangulations. The primal Bregman Voronoi diagram is the Euclidean Voronoi diagram and the dual Bregman Voronoi diagram coincides with the Cauchy hyperbolic Voronoi diagram. In addition, we prove that the square root of the Kullback-Leibler divergence between Cauchy distributions yields a metric distance which is Hilbertian for the Cauchy scale families.

**Keywords:** Cauchy distribution; Fisher-Rao distance; Kullback-Leibler divergence; chi square divergence; Bregman divergence; Jensen-Bregman divergence; Legendre-Fenchel divergence; metrization; Voronoi diagram; hyperbolic geometry; $q$-Gaussian; conformal flattening

## 1. Introduction

Let $\mathcal{P} = \{P_1, \ldots, P_n\}$ be a finite set of points in a space $\mathbb{X}$ equipped with a measure of *dissimilarity* $D(\cdot,\cdot) : \mathbb{X} \times \mathbb{X} \to \mathbb{R}_+$. The *Voronoi diagram* [1] of $\mathcal{P}$ partitions $\mathbb{X}$ into elementary *Voronoi cells* $\mathrm{Vor}(P_1), \ldots, \mathrm{Vor}(P_n)$ (also called Dirichlet cells [2]) such that

$$\mathrm{Vor}_D(P_i) := \left\{ X \in \mathbb{X}, \quad D(P_i, X) \leq D(P_j, X), \quad \forall j \in \{1, \ldots, n\} \right\} \tag{1}$$

denotes the *proximity cell* of point *generator* $P_i$ (also called *Voronoi site*), i.e., the locii of points $X \in \mathbb{X}$ closer with respect to $D$ to $P_i$ than to any other generator $P_j$.

When the dissimilarity $D$ is chosen as the Euclidean distance $\rho_E$, we recover the ordinary Voronoi diagram [1]. The Euclidean distance $\rho_E(P, Q)$ between two points $P$ and $Q$ is defined as

$$\rho_E(P, Q) = \|p - q\|_2, \tag{2}$$

where $p$ and $q$ denote the Cartesian coordinates of point $P$ and $Q$, respectively, and $\|\cdot\|_2$ the $\ell_2$-norm. Figure 1 (left) displays the Voronoi cells of an ordinary Voronoi diagram for a given set of generators.

The Voronoi diagram and its dual *Delaunay complex* [3] are fundamental data structures of computational geometry [4]. These core geometric data-structures find many applications in robotics, 3D reconstruction, geographic information systems (GISs), etc. See the textbook [1] for some of their

applications. The *Delaunay simplicial complex* is obtained by drawing a straight edge between two generators iff their Voronoi cells share an edge (Figure 1, right). In Euclidean geometry, the *Delaunay simplicial complex* triangulates the convex hull of the generators, and is therefore called the *Delaunay triangulation*. Figure 1 depicts the dual Delaunay triangulations corresponding to ordinary Voronoi diagrams. In general, when considering arbitrary dissimilarity $D$, the Delaunay simplicial complex may *not* triangulate the convex hull of the generators (see [5] and Section 4).

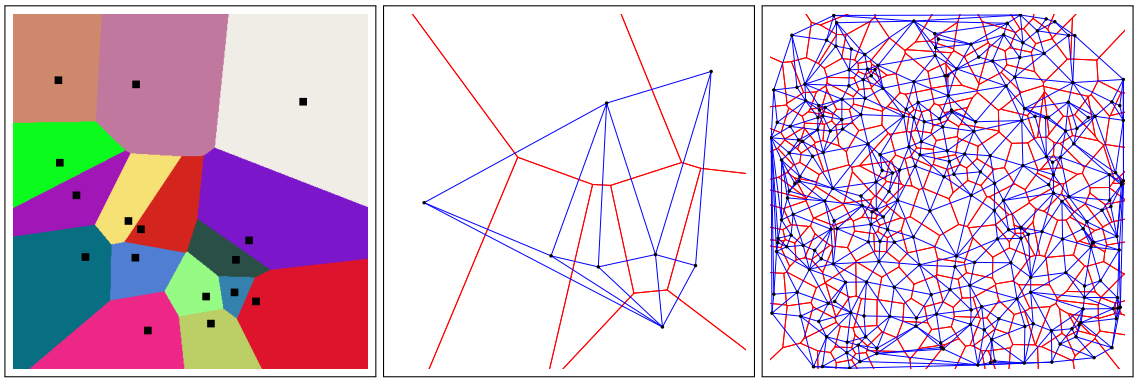


**Figure 1.** Euclidean Voronoi diagram of a set of generators (black square) in the plane with colored Voronoi cells (**left**). Euclidean Voronoi diagrams (red) and their dual Delaunay triangulations (blue) for $n = 8$ points (**middle**) and $n = 256$ points (**right**).

When the dissimilarity is *oriented* or *asymmetric*, i.e., $D(P, Q) \neq D(Q, P)$, one can define the *reverse* or *dual* dissimilarity $D^*(P, Q) := D(Q, P)$. This duality is termed *reference duality* in [6], and is an involution:

$$(D^*)^*(P, Q) = D(P, Q). \tag{3}$$

The dissimilarity $D(P : Q)$ is called the *forward* dissimilarity.

In the remainder, we shall use the ':' notational convention [7] between the arguments of the dissimilarity to emphasize that a dissimilarity $D$ is asymmetric: $D(P : Q) \neq D(Q : P)$. For an oriented dissimilarity $D(\cdot : \cdot)$, we can define two types of *dual Voronoi cells* as follows:

$$\mathrm{Vor}_D(P_i) := \left\{ X \in \mathbb{X}, \quad D(P_i : X) \leq D(P_j : X), \quad \forall j \in \{1, \ldots, n\} \right\}, \tag{4}$$

and

$$\begin{aligned} \mathrm{Vor}^*_D(P_i) &:= \left\{ X \in \mathbb{X} \quad D(X : P_i) \leq D(X : P_j), \quad \forall j \in \{1, \ldots, n\} \right\}, & (5)\\ &= \left\{ X \in \mathbb{X} \quad D^*(P_i : X) \leq D^*(P_j : X), \quad \forall j \in \{1, \ldots, n\} \right\}, & (6)\\ &= \mathrm{Vor}^*_D(P_i) = \mathrm{Vor}_{D^*}(P_i). & (7) \end{aligned}$$

That is, the dual Voronoi cell $\mathrm{Vor}^*_D(P_i)$ with respect to a dissimilarity $D$ is the primal Voronoi cell $\mathrm{Vor}_{D^*}(P_i)$ for the dual (reverse) dissimilarity $D^*$.

In general, we can build a Voronoi diagram as a *minimization diagram* [8] by defining the $n$ functions $f_i(X) := D(P_i : X)$. Then $X \in \mathrm{Vor}_D(P_i)$ iff $f_i(X) \leq f_j(X)$ for all $j \in \{1, \ldots, n\}$. Thus, by building the *lower envelope* [8] of the $n$ functions $f_1(X), \ldots, f_n(X)$, we can retrieve the Voronoi diagram.

An important class of *smooth* asymmetric dissimilarities are the Bregman divergences [9]. A *Bregman divergence* $B_F$ is defined for a smooth and strictly convex functional generator $F(\theta)$ by

$$B_F(\theta_1 : \theta_2) := F(\theta_1) - F(\theta_2) - (\theta_1 - \theta_2)^\top \nabla F(\theta_2), \tag{8}$$

where $\nabla F$ denotes the gradient of $F$. In information geometry [7,10,11], Bregman divergences are the canonical divergences of dually flat spaces [7]. Dually flat spaces generalize the (self-dual) Euclidean

geometry obtained for the generator $F_{\mathrm{Eucl}}(\theta) = \frac{1}{2}\theta^\top\theta$. In information sciences, dually flat spaces can be obtained, for example, as the induced information geometry of the Kullback-Leibler divergence [12] of an exponential family manifold [7,13] or a mixture manifold [14]. The dual Bregman Voronoi diagrams and their dual regular complexes have been studied in [15,16].

In this paper, we study the Voronoi diagrams induced by the Fisher-Rao distance [17–19], the Kullback-Leibler (KL) divergence [12] and the chi square distance [20] for the family $\mathcal{C}$ of Cauchy distributions. Cauchy distributions also called Lorentzian distributions in the literature [21,22].

The paper is organized with our main contributions as follows:

In Section 2, we concisely review the information geometry of the Cauchy family: We first describe the hyperbolic Fisher-Rao geometry in Section 2.1 and make a connection between the Fisher-Rao distance and the chi square divergence, then we point out the remarkable fact that any $\alpha$-geometry coincides with the Fisher-Rao geometry (Section 2.2), and we finally present dually flat geometric structures on the Cauchy manifold related to Tsallis' quadratic entropy [23,24] which amount to a conformal flattening of the Fisher-Rao geometry (Section 2.4). Section 3.3 proves that the square root of the KL divergence between any two Cauchy distributions yields a metric distance (Theorem 3), and that this metric distance can be isometrically embedded in a Hilbert space for the case of Cauchy scale families (Theorem 4). Section 4 shows that the Cauchy Voronoi diagrams induced either by the Fisher-Rao distance, the chi-square divergence, or the Kullback-Leibler divergence (and its square root metrization) all coincide with a hyperbolic Voronoi diagram [25] calculated on the Cauchy 2D location-scale parameters. This result yields a practical and efficient construction algorithm of hyperbolic Cauchy Voronoi diagrams [25,26] (Theorem 5) and their dual hyperbolic Cauchy Delaunay complexes (explained in detail in Section 6). We prove that the hyperbolic Cauchy Voronoi diagrams are Fisher orthogonal to the dual Cauchy Delaunay complexes (Theorem 6). In Section 4.2, we show that the primal Voronoi diagram with respect to the flat divergence coincides with the hyperbolic Voronoi diagram, and that the Voronoi diagram with respect to the reverse flat divergence matches the ordinary Euclidean Voronoi diagram. Finally, we conclude this work in Section 5.

## 2. Information Geometry of the Cauchy Family

We start by reporting the Fisher-Rao geometry of the Cauchy manifold (Section 2.1), then show that all $\alpha$-geometries coincide with the Fisher-Rao geometry (Section 2.2). Then we recall that we can associate an information-geometric structure to any parametric divergence (Section 2.3), and finally dually flatten this Fisher-Rao curved geometry using Tsallis's quadratic entropy [23,24] (Section 2.4) and a conformal Fisher metric.

### 2.1. Fisher-Rao Geometry of the Cauchy Manifold

*Information geometry* [7,10,11] investigates the geometry of families of probability measures. The 2D family $\mathcal{C}$ of Cauchy distributions

$$\mathcal{C} := \left\{ p_\lambda(x) := \frac{s}{\pi(s^2 + (x-l)^2)}, \quad \lambda := (l, s) \in \mathbb{H} := \mathbb{R} \times \mathbb{R}_+ \right\}, \tag{9}$$

is a *location-scale family* [27] (and also a univariate *elliptical distribution family* [28]) where $l \in \mathbb{R}$ and $s > 0$ denote the *location* parameter and the *scale* parameter, respectively:

$$p_{l,s}(x) := \frac{1}{s} p\left(\frac{x-l}{s}\right), \tag{10}$$

where

$$p(x) := \frac{1}{\pi(1+x^2)} =: p_{0,1}(x) \tag{11}$$

is the *Cauchy standard distribution*.

Let $l_\lambda(x) := \log p_\lambda(x)$ denote the *log density*. The parameter space $\mathbb{H} := \mathbb{R} \times \mathbb{R}_+$ of the Cauchy family is called the *upper plane*. The *Fisher-Rao geometry* [17,19,29] of $\mathcal{C}$ consists in modeling $\mathcal{C}$ as a Riemannian manifold $(\mathcal{C}, g_{\mathrm{FR}})$ by choosing the *Fisher Information metric* [7] (FIm)

$$g_{\mathrm{FR}}(\lambda) = [g_{ij}^{\mathrm{FR}}(\lambda)], \quad g_{ij}^{\mathrm{FR}}(\lambda) := E_{p_\lambda}\left[\partial_i l_\lambda(x) \partial_j l_\lambda(x)\right], \tag{12}$$

as the Riemannian metric tensor, where $\partial_m := \frac{\partial}{\partial \lambda_m}$ for $m \in \{1,2\}$ (i.e., $\partial_1 = \frac{\partial}{\partial l}$ and $\partial_2 = \frac{\partial}{\partial s}$). The matrix $[g_{ij}^{\mathrm{FR}}]$ is called the *Fisher Information Matrix* (FIM), and is the expression of the FIm tensor in a local coordinate system $\{e_1, e_2\}$: $g_{ij}^{\mathrm{FR}}(\lambda) = g(e_i, e_j)$ with $i, j \in \{1,2\}$.

The *Fisher-Rao distance* $\rho_{\mathrm{FR}}[p_{\lambda_1}, p_{\lambda_2}] = \rho_{\mathrm{FR}}[p_{l_1,s_1}, p_{l_2,s_2}]$ is then defined as the *Riemannian geodesic length distance* on the Cauchy manifold $(\mathcal{C}, g_{\mathrm{FR}})$:

$$\rho_{\mathrm{FR}}\left(p_{\lambda_1}(x), p_{\lambda_2}(x)\right) = \min_{\substack{\lambda(s) \\ \text{such that} \\ \lambda(0)=\lambda_1, \lambda(1)=\lambda_2}} \int_0^1 \sqrt{\left(\frac{\mathrm{d}\lambda(t)}{\mathrm{d}t}\right)^T g_{\mathrm{FR}}(\lambda(s)) \frac{\mathrm{d}\lambda(t)}{\mathrm{d}t}} \mathrm{d}t. \tag{13}$$

The Fisher information metric tensor for the Cauchy family [28] is

$$g_{\mathrm{FR}}(\lambda) = g_{\mathrm{FR}}(l, s) = \frac{1}{2s^2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{14}$$

where $\lambda = (l, s) \in \mathbb{H}$.

A *generic formula* for the Fisher-Rao distance between two univariate elliptical distributions is reported in [28]. This formula when instantiated for the Cauchy distributions yields the following closed-form formula for the Fisher-Rao distance:

$$\rho_{\mathrm{FR}}[p_{l_1,s_1}, p_{l_2,s_2}] = \frac{1}{\sqrt{2}} \left| \log \frac{\tan\left(\frac{\psi_1}{2}\right)}{\tan\left(\frac{\psi_2}{2}\right)} \right|, \tag{15}$$

where

$$\psi_i = \arcsin\left(\frac{s_i}{A}\right), \quad i \in \{1,2\}, \tag{16}$$

$$A^2 = s_1^2 + \frac{\left((l_2 - l_1)^2 - (s_1^2 - s_2^2)\right)^2}{4(l_2 - l_1)^2}. \tag{17}$$

However, by noticing that the metric tensor for the Cauchy family (Equation (14)) is equal to the *scaled* metric tensor $g_P$ of the Poincaré (P) hyperbolic upper plane [30]:

$$g_P(x, y) = \frac{1}{y^2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{18}$$

we get a relationship between the square infinitesimal lengths (line elements) $\mathrm{d}s_{\mathrm{FR}}^2 = \frac{\mathrm{d}l^2 + \mathrm{d}s^2}{2s^2}$ and $\mathrm{d}s_P^2 = \frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2}$ as follows:

$$\mathrm{d}s_{\mathrm{FR}} = \frac{1}{\sqrt{2}} \mathrm{d}s_P. \tag{19}$$

It follows that the Fisher-Rao distance between two Cauchy distributions is simply obtained by *rescaling* the 2D hyperbolic distance expressed in the Poincaré upper plane [30]:

$$\rho_{\mathrm{FR}}[p_{l_1,s_1}, p_{l_2,s_2}] = \frac{1}{\sqrt{2}} \rho_P(l_1, s_1; l_2, s_2) \tag{20}$$

where

$$\rho_P(l_1, s_1; l_2, s_2) := \operatorname{arccosh}\left(1 + \delta(l_1, s_1, l_2, s_2)\right), \tag{21}$$

with

$$\operatorname{arccosh}(x) := \log\left(x + \sqrt{x^2 - 1}\right), \quad x > 1, \tag{22}$$

and

$$\delta(l_1, s_1; l_2, s_2) := \frac{(l_2 - l_1)^2 + (s_2 - s_1)^2}{2 s_1 s_2}. \tag{23}$$

This latter term $\delta$ shall naturally appear in Section 2.4 when studying the dually flat space obtained by conformal flattening the Fisher-Rao geometry. The expression $\delta(l_1, s_1, l_2, s_2)$ of Equation (23) can be interpreted as a *conformal divergence* for the squared Euclidean distance [31–33].

We may also write the delta term using the 2D Cartesian coordinates $\lambda = (\lambda^{(1)}, \lambda^{(2)})$ as:

$$\delta(\lambda_1, \lambda_2) := \frac{(\lambda_2^{(1)} - \lambda_1^{(1)})^2 + (\lambda_2^{(2)} - \lambda_1^{(1)})^2}{2\lambda_1^{(2)}\lambda_2^{(2)}} = \frac{\|\lambda_1 - \lambda_2\|_2^2}{2\lambda_1^{(2)}\lambda_2^{(2)}}, \tag{24}$$

where $\lambda \in \mathbb{H}$.

In particular, when $l_1 = l_2$, we get the simplified Fisher-Rao distance for Cauchy scale families:

$$\rho_{\mathrm{FR}}[p_{l,s_1}, p_{l,s_2}] = \frac{1}{\sqrt{2}} \left| \log\left(\frac{s_1}{s_2}\right) \right|. \tag{25}$$

**Proposition 1.** *The Fisher-Rao distance between two Cauchy distributions is*

$$\rho_{\mathrm{FR}}[p_{l_1,s_1}, p_{l_2,s_2}] = \begin{cases} \frac{1}{\sqrt{2}} \left| \log \frac{s_1}{s_2} \right| & \text{when } l_1 = l_2, \\ \frac{1}{\sqrt{2}} \operatorname{arccosh}\left(1 + \frac{(l_2 - l_1)^2 + (s_2 - s_1)^2}{2 s_1 s_2}\right) & \text{when } l_1 \neq l_2. \end{cases}$$

The Fisher-Rao manifold of Cauchy distributions has *constant negative scalar curvature* $\kappa = -2$, see [28] for detailed calculations.

**Remark 1.** *It is well-known that the Fisher-Rao geometry of location-scale families amount to a hyperbolic geometry [27]. For d-variate scale-isotropic Cauchy distributions $p_\lambda(x)$ with $\lambda = (l, s) \in \mathbb{R}^d \times \mathbb{R}$, the Fisher information metric is $g_{\mathrm{FR}}(\lambda) = \frac{1}{2s^2} I$, where $I$ denotes the $(d+1) \times (d+1)$ identity matrix. It follows that*

$$\rho_{\mathrm{FR}}[p_{l_1,s_1}, p_{l_2,s_2}] = \frac{1}{\sqrt{2}} \operatorname{arccosh}\left(1 + \Delta(l_1, s_1, l_2, s_2)\right), \tag{26}$$

*where*

$$\Delta(l_1, s_1, l_2, s_2) := \frac{\|l_2 - l_1\|_2^2 + (s_2 - s_1)^2}{2 s_1 s_2}, \tag{27}$$

*where $\|\cdot\|_2$ is the d-dimensional Euclidean $\ell_2$-norm: $\|x\| = \sqrt{x^\top x}$. That is, $\rho_{\mathrm{FR}}[p_{l_1,s_1}, p_{l_2,s_2}]$ is the scaled d-dimensional real hyperbolic distance [30] expressed in the Poincaré upper space model.*

Let us mention that recently the Riemannian geometry of location-scale models was also studied from the complementary viewpoint of warped metrics [34,35].

**Remark 2.** *Li and Zhao [36] proposed to use the* Wasserstein Information metric *(WIm) expressed using the distribution parameter coordinates by the* Wasserstein Information Matrix *(WIM). They reported the explicit formula of the WIM for generic location-scale families:*

$$I^W(l,s) = \begin{bmatrix} \frac{E_{p_\lambda}[x^2]-2lE_{p_\lambda}[x]+l^2}{s^2} & 0 \\ 0 & 1 \end{bmatrix}. \tag{28}$$

*In particular, the WIM of the Gaussian family (a location-scale family) is the identity matrix and yields the Euclidean geometry (see the Wasserstein geometry of Gaussians [37]). Although the WIM can be calculated for the Gaussian location-scale family, let us notice that the moments greater or equal to one (i.e., $E[X]$ and $E[X^2]$) are* not *finite for the Cauchy distributions. Thus, the WIM is not well-defined for the Cauchy family since Equation* (28) *makes sense only for finite moments.*

### 2.2. The Dualistic $\alpha$-Geometry of the Statistical Cauchy Manifold

A *statistical manifold* [38] is a triplet $(M, g, T)$ where $g$ is a Riemannian metric tensor and $T$ is a cubic totally symmetric tensor (i.e., $T_{\sigma(i)\sigma(j)\sigma(k)} = T_{ijk}$ for any permutation $\sigma$). For a *parametric* family of probability densities $M = \{p_\lambda(x)\}$, the cubic tensor is called the *skewness tensor* [7], and defined by:

$$T_{ijk}(\theta) := E_{p_\lambda}\left[\partial_i l_\lambda(x)\partial_j l\lambda(x)\partial_k l\lambda(x)\right]. \tag{29}$$

A statistical manifold structure $(M, g, T)$ allows one to construct Amari's dualistic *$\alpha$-geometry* [7] for any $\alpha \in \mathbb{R}$: Namely a quadruplet $(M, g_{\mathrm{FR}}, \nabla^{-\alpha}, \nabla^{\alpha})$ where $\nabla^{-\alpha}$ and $\nabla^{\alpha}$ are *dual torsion-free affine connections* coupled to the Fisher metric $g_{\mathrm{FR}}$ (i.e., $\nabla^{-\alpha} = (\nabla^{\alpha})^*$). We refer the reader to the textbook [7] and the overview [11] for further details.

The Fisher-Rao geometry $(M, g_{\mathrm{FR}})$ corresponds to the 0-geometry, i.e., the self-dual geometry where $\nabla^0 := {}^g\nabla$ is the *Levi-Civita metric connection* [7] induced by the metric tensor (with $({}^g\nabla)^* = {}^g\nabla$). That is, we have

$$(\mathcal{C}, g_{\mathrm{FR}}) = (\mathcal{C}, g_{\mathrm{FR}}, \nabla^0, \nabla^0). \tag{30}$$

In information geometry, the *invariance principle* states that the geometry should be invariant under the transformation of a random variable $X$ to $Y$ provided that $Y = t(X)$ is a sufficient statistics [7] of $X$. The $\alpha$-geometry $(M, g_{\mathrm{FR}}, \nabla^{-\alpha}, \nabla^{\alpha})$ and its special case of Fisher-Rao geometry are invariant geometry [7,11] for any $\alpha \in \mathbb{R}$.

A remarkable fact is that all the $\alpha$-geometries of the Cauchy family coincide with the Fisher-Rao geometry since the cubic skewness tensor $T$ vanishes everywhere [28], i.e., $T_{ijk} = 0$. The non-zero coefficients of the Christoffel symbols of the $\alpha$-connections (including the Levi-Civita metric connection derived from the Fisher metric tensor) are:

$$ {}^{\alpha}\Gamma^1_{12} = {}^{\alpha}\Gamma^1_{21} = {}^{\alpha}\Gamma^2_{22} = -\frac{1}{s}, \tag{31}$$

$$ {}^{\alpha}\Gamma^2_{11} = \frac{1}{s}. \tag{32}$$

Thus, all $\alpha$-geometries coincide and have constant negative scalar curvature $\kappa = -2$. In other words, we cannot choose a value for $\alpha$ to make the Cauchy manifold dually flat [7]. To contrast with this result, Mitchell [28] reported values of $\alpha$ for which the $\alpha$-geometry is dually flat for some parametric location-scale families of distributions: For example, it is well known that the manifold $\mathcal{N}$ of univariate Gaussian distributions is $\pm 1$-flat [7]. The manifold $\mathcal{S}_k$ of $t$-Student's distributions with $k$ degrees of freedom is proven dually flat when $\alpha = \pm\frac{k+5}{k-1}$ [28]. Dually flat manifolds are Hessian manifolds [39] with dual geodesics being straight lines in one of the two dual global affine coordinate systems. On a global Hessian manifold, the canonical divergences are Bregman divergences. Thus, these dually flat Bregman manifolds are computationally friendly [15] as many techniques

of computational geometry [4] can be naturally extended to these Hessian spaces (e.g., the smallest enclosing balls [40]).

### 2.3. Dualistic Structures Induced by a Divergence

A *divergence* or *contrast function* [13] is a smooth parametric dissimilarity. Let $\mathcal{M}$ denote the manifold of its parameter space. Eguchi [13] showed how to associate to any divergence $D$ a canonical information-geometric structure $(\mathcal{M}, {}^{D}g, {}^{D}\nabla, {}^{D}\nabla^*)$. Moreover, the construction allows proving that ${}^{D}\nabla^* = {}^{D^*}\nabla$. That is the dual connection ${}^{D}\nabla^*$ for the divergence $D$ corresponds to the primal connection for the reverse divergence $D^*$ (see [7,11] for details).

Conversely, Matsumoto [41] proved that given an information-geometric structure $(\mathcal{M}, g, \nabla, \nabla^*)$, one can build a divergence $D$ such that $(\mathcal{M}, g, T) = (\mathcal{M}, {}^{D}g, {}^{D}T)$ from which we can derive the structure $(\mathcal{M}, {}^{D}g, {}^{D}\nabla, {}^{D}\nabla^*)$. Thus, when calculating the Voronoi diagram $\mathrm{Vor}_D$ for an arbitrary divergence $D$, we may use the induced information-geometric structure $(\mathcal{M}, {}^{D}g, {}^{D}\nabla, {}^{D}\nabla^*)$ to investigate some of the properties of the Voronoi diagram: For example, is the bisector $\mathrm{Bi}_D$ ${}^{D}\nabla$-autoparallel?, or is the bisector $\mathrm{Bi}_D$ of two generators orthogonal with respect to the metric ${}^{D}g$ to their ${}^{D}\nabla$-geodesic? Section 4 will study these questions in particular cases.

### 2.4. Dually Flat Geometry of the Cauchy Manifold by Conformal Flattening

The Cauchy distributions are usually handled in information geometry using the wider scope of *q-Gaussians* [7,22,42] (deformed exponential families [43]). The $q$-Gaussians also include the Student's $t$-distributions. Cauchy distributions are $q$-Gaussians for $q = 2$. These $q$-Gaussians are also called $q$-normal distributions [44], and they can be obtained as maximum entropy distributions with respect to *Tsallis' entropy* $T_q(\cdot)$ [23,24] (see Theorem 4.12 of [7]):

$$T_q(p) := \frac{1}{q-1}\left(1 - \int_{-\infty}^{\infty} p^q(x)\mathrm{d}x\right), \quad q \neq 1. \tag{33}$$

When $q = 2$, we have the following *Tsallis' quadratic entropy*:

$$T_2(p) := 1 - \int_{-\infty}^{\infty} p^2(x)\mathrm{d}x. \tag{34}$$

We have $\lim_{q\to 1} T_q(p) = S(p) := -\int p(x)\log p(x)\mathrm{d}x$, Shannon entropy.

Thus, $q$-Gaussians are $q$-exponential families [21], generalizing the MaxEnt exponential families derived from Shannon entropy [45]. The integral $E(p) := \int_{-\infty}^{\infty} p^2(x)\mathrm{d}x$ corresponds to *Onicescu's informational energy* [46,47]. Tsallis' entropy is considered in non-extensive statistical physics [24].

A dually flat structure construction for $q$-Gaussians is reported in [7] (Sec. 4.3, pp. 84–89). We instantiate this construction for the Cauchy distributions (2-Gaussians):

Let

$$\exp_{\mathcal{C}}(u) := \frac{1}{1-u}, \quad u \neq 1, \tag{35}$$

denote the *deformed q-exponential* and

$$\log_{\mathcal{C}}(u) := 1 - \frac{1}{u}, \quad u \neq 0, \tag{36}$$

its compositional inverse, the *deformed q-logarithm*.

The probability density of a 2-Gaussian can be factorized as

$$p_\theta(x) = \exp_{\mathcal{C}}(\theta^\top x - F(\theta)), \tag{37}$$

where $\theta$ denotes the 2D natural parameters. We have:

$$\log_{\mathcal{C}}(p_\theta(x)) = 1 - \frac{1}{s}\pi(s^2 + (x-l)^2) = 1 - \pi\left(s + \frac{(x-l)^2}{s}\right), \quad (38)$$

$$=: \theta^\top t(x) - F(\theta), \quad (39)$$

$$= \underbrace{\left(2\pi\frac{l}{s}\right)x + \left(-\frac{\pi}{s}\right)x^2}_{\theta^\top t(x)} - \underbrace{\left(\pi s + \pi\frac{l^2}{s} - 1\right)}_{F(\theta)}. \quad (40)$$

Therefore the natural parameter is $\theta(l,s) = (\theta_1, \theta_2) = \left(2\pi\frac{l}{s}, -\frac{\pi}{s}\right) \in \Theta = \mathbb{R} \times \mathbb{R}_-$ (for $t(x) = (x, x^2)$) and the *deformed log-normalizer* is

$$F(\theta(\lambda)) = \pi s + \pi\frac{l^2}{s} - 1 =: F_\lambda(\lambda), \quad (41)$$

$$F(\theta) = -\frac{\pi^2}{\theta_2} - \frac{\theta_1^2}{4\theta_2} - 1. \quad (42)$$

In general, we obtain a strictly convex and $C^3$-function $F_q(\theta)$, called the *q-free energy* for a $q$-Gaussian family. Here, we let $F(\theta) := F_2(q)$ for the Cauchy family: $F(\theta)$ is the *Cauchy free energy*.

We convert back the natural parameter $\theta \in \Theta$ to the ordinary parameter $\lambda \in \mathbb{H}$ as follows:

$$\lambda(\theta) = (l, s) = \left(-\frac{\theta_1}{2\theta_2}, -\frac{\pi}{\theta_2}\right). \quad (43)$$

The gradient of the deformed log-normalizer is:

$$\nabla F(\theta) = \begin{bmatrix} -\frac{\theta_1}{2\theta_2} \\ \frac{\pi^2}{\theta_2^2} + \frac{\theta_1^2}{4\theta_2^2} \end{bmatrix}. \quad (44)$$

The gradient $\nabla F(\theta)$ defines the dual global affine coordinate system $\eta := \nabla F(\theta)$ where $\eta \in H = \mathbb{R} \times \mathbb{R}_+$ is the dual parameter space.

It follows the following divergence $D_{\text{flat}}[p_{\lambda_1} : p_{\lambda_2}]$ [7] between Cauchy densities which is by construction equivalent to a Bregman divergence $B_F(\theta_1 : \theta_2)$ (canonical divergence in dually flat space) between their corresponding natural parameters (Eq. (4.95) of [7] instantiated for $q = 2$):

$$D_{\text{flat}}[p_{\lambda_1} : p_{\lambda_2}] := \frac{1}{\int p_{\lambda_2}^2(x)\mathrm{d}x}\left(\int \frac{p_{\lambda_2}^2(x)}{p_{\lambda_1}(x)}\mathrm{d}x - 1\right), \quad (45)$$

$$= 2\pi s_2\left(\frac{s_1^2 + s_2^2 + (l_1 - l_2)^2}{2s_1 s_2} - 1\right), \quad (46)$$

$$= 2\pi s_2\frac{(s_1 - s_2)^2 + (l_1 - l_2)^2}{2s_1 s_2}, \quad (47)$$

$$= 2\pi s_2\delta(l_1, s_1, l_2, s_2), \quad (48)$$

$$= B_F(\theta_1 : \theta_2), \quad (49)$$

where $\theta_1 := \theta(\lambda_1)$ and $\theta_2 := \theta(\lambda_2)$. We term $B_F(\theta_1 : \theta_2)$ the *Bregman-Tsallis (quadratic) divergence* ($B_{F_q}$ for general $q$-Gaussians).

We used a computer algebra system (CAS, see Section 7) to calculate the closed-form formulas of the following definite integrals:

$$\int p_{\lambda_2}^2(x)\mathrm{d}x = \frac{1}{2\pi s_2}, \tag{50}$$

$$\int \frac{p_{\lambda_2}^2(x)}{p_{\lambda_1}}\mathrm{d}x = \frac{s_1^2+s_2^2+(l_1-l_2)^2}{2s_1s_2}. \tag{51}$$

Here, observe that the equivalent Bregman divergence is *not* on swapped parameter order as it is the case for ordinary exponential families: $D_{\mathrm{KL}}[p_{\theta_1}:p_{\theta_2}]=B_F(\theta_2:\theta_1)$ where $F$ denotes the cumulant function of the exponential family, see [7,11].

We term the divergence $D_{\mathrm{flat}}$ the *flat divergence* because its induced affine connection [13] ${}^{D_{\mathrm{flat}}}\nabla$ has *zero curvature* (i.e., the 4D Riemann-Christofel curvature tensor induced by the connection vanishes, see [7] p. 134).

Since $D_{\mathrm{flat}}[p_{\lambda_1}:p_{\lambda_2}]=2\pi s_2\delta(l_1,s_1,l_2,s_2)=\frac{\pi}{s_1}\left((s_1-s_2)^2+(l_1-l_2)^2\right)$, the flat divergence is interpreted as a conformal squared Euclidean distance [33], with conformal factor $\frac{\pi}{s_1}$. In general, the Fisher-Rao geometry of $q$-Gaussians has scalar curvature [44] $\kappa=-\frac{q}{3-q}$. Thus, we recover the scalar curvature $\kappa=-2$ for the Fisher-Rao Cauchy manifold since $q=2$.

**Theorem 1.** *The flat divergence $D_{\mathrm{flat}}[p_{\lambda_1}:p_{\lambda_2}]$ between two Cauchy distributions is equivalent to a Bregman divergence $B_F(\theta_1:\theta_2)$ on the corresponding natural parameters, and yields the following closed-form formula using the ordinary location-scale parameterization:*

$$D_{\mathrm{flat}}[p_{\lambda_1}:p_{\lambda_2}]=2\pi s_2\delta(l_1,s_1,l_2,s_2)=\frac{\pi}{s_1}\left((s_1-s_2)^2+(l_1-l_2)^2\right)=\frac{\pi}{s_1}\|\lambda_1-\lambda_2\|_2^2. \tag{52}$$

The conversion of $\eta$-coordinates to $\theta$-coordinates are calculated as follows:

$$\theta(\eta)=\begin{bmatrix}\frac{2\pi\eta_1}{\sqrt{\eta_2-\eta_1^2}}\\ \frac{-\pi-}{\sqrt{\eta_2-\eta_1^2}}\end{bmatrix}:=\nabla F^*(\eta), \tag{53}$$

where

$$F^*(\eta):=\theta(\eta)^\top\eta-F(\theta(\eta)), \tag{54}$$

is the *Legendre-Fenchel convex conjugate* [7]:

$$F^*(\eta)=1-2\pi\sqrt{\eta_2-\eta_1^2}. \tag{55}$$

Since

$$\eta(\lambda)=\eta(\theta(\lambda))=(\lambda_1,\lambda_1^2+\lambda_2^2)=(l,l^2+s^2), \tag{56}$$

we have

$$F_\lambda^*(\lambda):=F^*(\eta(\lambda))=1-2\pi\sqrt{l^2+s^2-l^2}=1-2\pi s \tag{57}$$

that is *independent* of the location parameter $l$. Moreover, we have [7]

$$F_\lambda^*(\lambda):=1-\frac{1}{\int p^2(x)dx}=1-\frac{1}{\frac{1}{2\pi s}}=1-2\pi s. \tag{58}$$

We can convert the dual parameter $\eta$ to the ordinary parameter $\lambda\in\mathbb{H}$ as follows:

$$\lambda(\eta)=(l,s)=(\eta_1,\sqrt{\eta_2-\eta_1^2}). \tag{59}$$

It follows that we have the following equivalent expressions for the flat divergence:

$$D_{\mathrm{flat}}[p_{\lambda_1} : p_{\lambda_2}] = B_F(\theta_1 : \theta_2) = B_{F^*}(\eta_2 : \eta_1) = A_F(\theta_1 : \eta_2) = A_{F^*}(\eta_2 : \theta_1), \tag{60}$$

where

$$A_F(\theta_1 : \eta_2) := F(\theta_1) + F^*(\eta_2) - \theta_1^\top \eta_2, \tag{61}$$

is the *Legendre-Fenchel divergence* measuring the *inequality gap* of the Fenchel-Young inequality:

$$F(\theta_1) + F^*(\eta_2) \geq \theta_1^\top \eta_2. \tag{62}$$

That is, $A_F(\theta_1 : \eta_2) = \mathrm{rhs}(\theta_1 : \eta_2) - \mathrm{lhs}(\theta_1 : \eta_2) \geq 0$, where $\mathrm{rhs}(\theta_1 : \eta_2) := F(\theta_1) + F^*(\eta_2)$ and $\mathrm{lhs}(\theta_1 : \eta_2) = \theta_1^\top \eta_2$.

The Hessian metrics of the dual convex potential functions $F(\theta)$ and $F^*(\eta)$ are:

$$\nabla^2 F(\theta) = \begin{bmatrix} -\frac{1}{2\theta_2} & \frac{\theta_1}{2\theta_2^2} \\ \frac{\theta_1}{2\theta_2^2} & -\frac{\theta_1^2}{2\theta_2^2} - \frac{2\pi^2}{\theta_2^2} \end{bmatrix} =: g_F(\theta), \tag{63}$$

$$\nabla^2 F^*(\eta) = \begin{bmatrix} \frac{2}{\sqrt{\eta_2 - \eta_1^2}} + \frac{2\eta_1^2}{(\eta_2 - \eta_1^2)^{\frac{3}{2}}} & -\frac{\eta_1}{(\eta_2 - \eta_1^2)^{\frac{3}{2}}} \\ -\frac{\eta_1}{(\eta_2 - \eta_1^2)^{\frac{3}{2}}} & \frac{1}{2}(\eta_2 - \eta_1^2)^{\frac{3}{2}} \end{bmatrix} =: g_F^*(\eta). \tag{64}$$

We check the Crouzeix identity [11,48]:

$$\nabla^2 F(\theta) \nabla^2 F^*(\eta(\theta)) = \nabla^2 F(\theta(\eta)) \nabla^2 F^*(\eta) = I, \tag{65}$$

where $I$ denotes the $2 \times 2$ identity matrix.

The Hessian metric $\nabla^2 F(\theta)$ is also called the *q-Fisher metric* [44] (for $q = 2$). Let $g_{\mathrm{FR}}^\lambda(\lambda)$ and $g_{\mathrm{FR}}^\theta(\theta)$ denote the Fisher information metric expressed using the $\lambda$-coordinates and the $\theta$-coordinates, respectively. Then, we have

$$g_{\mathrm{FR}}^\theta(\theta) = \mathrm{Jac}_\lambda^\top(\theta) \times g_{\mathrm{FR}}^\lambda(\lambda(\theta)) \times \mathrm{Jac}_\lambda(\theta), \tag{66}$$

where $\mathrm{Jac}_\lambda(\theta)$ denotes the Jacobian matrix:

$$\mathrm{Jac}_\lambda(\theta) := \left[ \frac{\partial \lambda_i}{\partial \theta_j} \right]. \tag{67}$$

Similarly, we can express the Hessian metric $g_F := \nabla^2 F(\theta)$ using the $\lambda$-coordinate system:

$$g_F^\lambda(\lambda) = \mathrm{Jac}_\theta^\top(\lambda) \times g_F^\theta(\theta(\lambda)) \times \mathrm{Jac}_\theta(\lambda). \tag{68}$$

We calculate explicitly the following Jacobian matrices:

$$\mathrm{Jac}_\theta(\lambda) = \pi \begin{bmatrix} \frac{2}{\lambda_2} & -2\frac{\lambda_1}{\lambda_2^2} \\ 0 & \frac{1}{\lambda_2^2} \end{bmatrix}. \tag{69}$$

and

$$\mathrm{Jac}_\lambda(\theta) = \begin{bmatrix} -\frac{1}{2\theta_2} & \frac{\theta_1}{2\theta_2^2} \\ 0 & \frac{\pi}{\theta_2^2} \end{bmatrix}. \tag{70}$$

We check that we have

$$g_F^{\theta}(\theta) = -\frac{2\theta_2}{\pi^2} g_{\mathrm{FR}}^{\theta}(\theta), \quad (71)$$

$$g_F^{\lambda}(\lambda) = \frac{2}{\pi\sigma} g_{\mathrm{FR}}^{\lambda}(\lambda). \quad (72)$$

That is, the Riemannian metric tensors $g_{\mathrm{FR}}^{\lambda}(\lambda)$ and $g_F^{\lambda}(\lambda)$ (or $g_F^{\theta}(\theta)$ and $g_{\mathrm{FR}}^{\theta}(\theta)$) are conformally equivalent. This is, there exists a smooth function $u(\lambda) = \log \frac{2}{\pi\sigma}$ such that $g_F^{\lambda}(\lambda) = e^{u(\lambda)} g_{\mathrm{FR}}^{\lambda}(\lambda)$.

This dually flat space construction of the Cauchy manifold

$$\left(\mathcal{C}, g(\theta) = \nabla^2 F(\theta), {}^{D_{\mathrm{flat}}}\nabla, {}^{D_{\mathrm{flat}}}\nabla^* = {}^{D^*_{\mathrm{flat}}}\nabla\right)$$

can be interpreted as a conformal flattening of the curved $\alpha$-geometry [7,44,49]. The relationships between the curvature tensors of dual $\pm\alpha$-connections are studied in [50].

Notice that this dually flat geometry can be recovered from the divergence-based structure of Section 2.3 by considering the Bregman-Tsallis divergence. Figure 2 illustrates the relationships between the invariant $\alpha$-geometry and the dually flat geometry of the Cauchy manifold. The $q$-Gaussians can further be generalized by *$\chi$-family* with corresponding deformed logarithm and exponential functions [7,45]. The *$\chi$-family* unifies both the dually flat exponential family with the dually flat mixture family [45].

A statistical dissimilarity $D[p_{\lambda_1} : p_{\lambda_2}]$ between two parametric distributions $p_{\lambda_1}$ and $p_{\lambda_2}$ amounts to an equivalent dissimilarity $D(\theta_1 : \theta_2)$ between their parameters: $D(\theta_1 : \theta_2) := D[p_{\lambda_1} : p_{\lambda_2}]$. When the parametric dissimilarity is smooth, one can construct the divergence-based $\alpha$-geometry [11,51]. Thus, the dually flat space structure of the Cauchy manifold can also be obtained from the *divergence-based $\pm\alpha$-geometry* obtained from the flat divergence $D_{\mathrm{flat}}$ (see Figure 2). It can be shown that the dually flat space $q$-geometry is the unique geometry in the intersection of the conformal Fisher-Rao geometry with the deformed $\chi$-geometry (Theorem 13 of [45]) when the manifold is the positive orthant $\mathbb{R}^{d+1}$. Please note that a dually flat space in information geometry is usually not Riemannian flat (with respect to the Levi-Civita connection, e.g., the Gaussian manifold). In particular, Matsuzoe proved in [52] that the Riemannian manifold $(\mathcal{C}, \nabla^2 F(\theta))$ induced by the $q$-Fisher metric is of constant curvature $-1$ when $q = 2$.

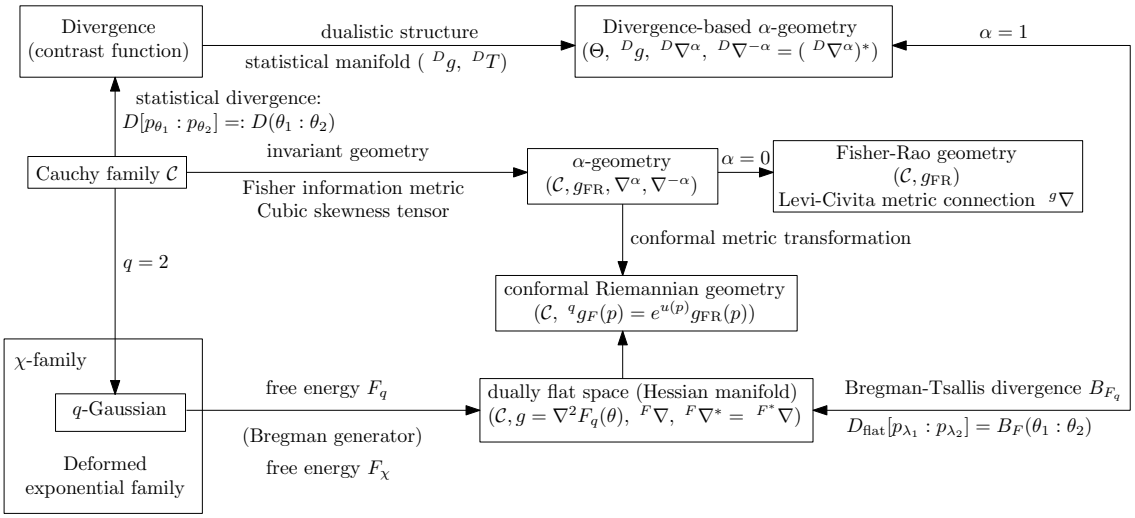



**Figure 2.** Information-geometric structures on the Cauchy manifold and their relationships.

There are many alternative possible ways to build a dually flat space from a $q$-Gaussian family once a convex Bregman generator $F(\theta)$ has been built from the density $p_q(\theta)$ of a $q$-Gaussian. The method presented above is a natural generalization of the dually flat space construction for exponential families.

To give another approach, let us mention that Matsuzoe [52] also introduced another Hessian metric $g^M(\theta) = [g^M_{ij}(\theta)]$ defined by:

$$g^M_{ij}(\theta) := \int \partial_i p_\theta(x) \partial_j \log_q p_\theta(x) \mathrm{d}x. \tag{73}$$

This metric is conformal to both the Fisher metric and the $q$-Fisher metric, and is obtained by generalizing equivalent representations of the Fisher information matrix (see $\alpha$-representations in [7]).

## 3. Invariant Divergences: $f$-Divergences and $\alpha$-Divergences

### 3.1. Invariant Divergences in Information Geometry

The *$f$-divergences* [20,53] between two densities $p(x)$ and $q(x)$ is defined for a positive convex function $f$, strictly convex at 1, with $f(1) = 0$ as:

$$I_f[p:q] := \int_{\mathcal{X}} p(x) f\left(\frac{q(x)}{p(x)}\right) \mathrm{d}x, \tag{74}$$

The KL divergence is a $f$-divergence obtained for the generator $f(u) = -\log(u)$.

An *invariant divergence* is a divergence $D$ which satisfies the *information monotonicity* [7]: $D[p_X : p_Y] \geq D[p_{t(X)} : p_{t(Y)}]$ with equality iff $t(X)$ is a sufficient statistic. The invariant divergences are the $f$-divergences for the simplex sample space [7]. Moreover, the *standard $f$-divergences* (calibrated with $f(1) = 0$ and $f'(1) = f''(1) = 1$) induce the Fisher information metric (FIm) for its metric tensor ${}^{I_f}g$ when the sample space is the probability simplex: ${}^{I_f}g = g_{\mathrm{FR}}$, see [7].

### 3.2. $\alpha$-Divergences between Location-Scale Densities

Let $I_\alpha[p:q]$ denote the *$\alpha$-divergence* [7,54,55] between $p$ and $q$:

$$I_\alpha[p:q] := \frac{1}{\alpha(1-\alpha)}(1 - C_\alpha[p:q]), \quad \alpha \notin \{0,1\} \tag{75}$$

where $C_\alpha[p:q]$ is *Chernoff $\alpha$-coefficient* [56,57]:

$$\begin{aligned} C_\alpha[p:q] &:= \int p^\alpha(x) q^{1-\alpha}(x) \mathrm{d}x, & (76)\\ &= \int q(x) \left(\frac{p(x)}{q(x)}\right)^\alpha, & (77)\\ &= C_{1-\alpha}[q:p]. & (78) \end{aligned}$$

We have $I_\alpha[p:q] = I_{1-\alpha}[q:p] = {I_\alpha}^*[p:q]$.

The $\alpha$-divergences include the *chi square divergence* ($\alpha = 2$), the *squared Hellinger divergence* ($\alpha = \frac{1}{2}$, symmetric) and in the limit cases the Kullback-Leibler (KL) divergence ($\alpha \to 1$) and the reverse KL divergence ($\alpha \to 0$). The $\alpha$-divergences are $f$-divergences for the generator:

$$f_\alpha(u) = \begin{cases} \frac{u^{1-\alpha}-u}{\alpha(\alpha-1)}, & \text{if } \alpha \neq 0, \alpha \neq 1 \\ u\log(u), & \text{if } \alpha = 0 \quad \text{(reverse Kullback-Leibler divergence)}, \\ -\log(u), & \text{if } \alpha = 1 \quad \text{(Kullback-Leibler divergence)}. \end{cases} \tag{79}$$

For location scale families, let

$$C_\alpha(l_1, s_1; l_2, s_2) := C_\alpha\left[p_{l_1,s_1} : p_{l_2,s_2}\right]. \tag{80}$$

Using change of variables in the integrals, one can show the following identities:

$$\begin{aligned} C_\alpha(l_1, s_1; l_2, s_2) &= C_\alpha\left(0, 1; \frac{l_2 - l_1}{s_1}, \frac{s_2}{s_1}\right), && (81)\\ &= C_\alpha\left(\frac{l_1 - l_2}{s_2}, \frac{s_1}{s_2}; 0, 1\right), && (82)\\ &= C_{1-\alpha}\left(0, 1; \frac{l_1 - l_2}{s_2}, \frac{s_1}{s_2}\right), && (83)\\ &= C_{1-\alpha}(l_2, s_2; l_1, s_1). && (84) \end{aligned}$$

For the location-scale families which include the normal family $\mathcal{N}$, the Cauchy family $\mathcal{C}$ and the $t$-Student families $\mathcal{S}_k$ with fixed degree of freedom $k$, the $\alpha$-divergences are *not* symmetric in general (e.g., $\alpha$-divergences between two normal distributions). However, we have shown that the chi square divergences and the KL divergence are symmetric when densities belong to the Cauchy family. Thus, it is of interest to prove whether the $\alpha$-divergences between Cauchy densities are symmetric or not, and report their closed-form formula for all $\alpha \in \mathbb{R}$.

Using symbolic integration described in Section 7, we found that

$$C_3(p_{\lambda_1}; p_{\lambda_2}) = \frac{3s_2^4 + (2s_1^2 + 6l_2^2 - 12l_1l_2 + 6l_1^2)s_2^2 + 3s_1^4 + (6l_2^2 - 12l_1l_2 + 6l_1^2)s_1^2 + 3l_2^4 - 12l_1l_2^3 + 18l_1^2l_2^2 - 12l_1^3l_2 + 3l_1^4)}{8s_1^2s_2^2}, \quad (85)$$

and checked that this Chernoff similarity coefficient is symmetric:

$$C_3(p_{\lambda_1}; p_{\lambda_2}) = C_3(p_{\lambda_2}; p_{\lambda_1}). \quad (86)$$

Therefore the 3-divergence $I_3$ between two Cauchy distributions is symmetric. In particular, when $l_1 = l_2 = l$, we find that

$$\begin{aligned} C_3(p_{l,s_1}; p_{l,s_2}) &= \frac{3(s_1^4 + s_2^4) + 2s_1^2s_2^2}{8s_1^2s_2^2}, && (87)\\ &= 1 + \frac{3}{4}\frac{(s_1^2 - s_2^2)^2}{2s_1^2s_2^2}, && (88)\\ &= 1 + \frac{3}{4}\delta(l^2, s_1^2, l_2^2, s_2^2). && (89) \end{aligned}$$

In the Section7, we proved by symbolic calculations that the $\alpha$-divergences are symmetric for $\alpha \in \{0, 1, 2, 3, 4\}$.

**Remark 3.** *The Cauchy family can also be interpreted as a family of univariate elliptical distributions [28]. A* univariate elliptical distribution *has canonical parametric density:*

$$q_{\mu,\sigma}(x) := \frac{1}{\sigma} h\left(\left(\frac{x - \mu}{\sigma}\right)^2\right), \quad (90)$$

*for some function $h(u)$. For example, the Gaussian distributions are elliptical distributions obtained for $h(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u\right)$. Location-scale densities $p_{l,s}$ with standard density $p_{0,1}$ can be interpreted as univariate elliptical distributions $q_{\mu,\sigma}$ with $h(u) = p_{0,1}(u^2)$ and $(\mu, \sigma) = (l, s)$: $p_{l,s} = q_{\mu,\sigma}$. It follows that the Cauchy densities are elliptical distributions for $h(u) = \frac{1}{\pi(1+u)}$. By doing a change of variable in the KL divergence integral, we find again the following identity:*

$$D_{\mathrm{KL}}\left[q_{\mu_1,\sigma_1} : q_{\mu_2,\sigma_2}\right] = D_{\mathrm{KL}}\left[q_{0,1} : q_{\frac{\mu_2 - \mu_1}{\sigma_2}, \frac{\sigma_1}{\sigma_2}}\right]. \quad (91)$$

### 3.3. Metrization of the Kullback-Leibler Divergence

The *Kullback-Leibler divergence* [12] $D_{\mathrm{KL}}[p:q]$ between two continuous probability densities $p$ and $q$ defined over the real line support is an oriented dissimilarity measure defined by:

$$D_{\mathrm{KL}}[p:q] := \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x. \tag{92}$$

The closed-form formula for the KL divergence between two Cauchy distributions requires to perform a (non-trivial) integration task. The following closed-form expression has been reported in [58] using advanced symbolic integration:

$$D_{\mathrm{KL}}[p_{l_1,s_1} : p_{l_2,s_2}] = \log\left(1 + \frac{(s_1+s_2)^2 + (l_1-l_2)^2}{4s_1s_2}\right). \tag{93}$$

Although the KL divergence is usually asymmetric, it is a remarkable fact that it is symmetric between any two Cauchy densities. However, the KL divergence of Equations (92) and (93) does *not* satisfy the triangle inequality, and therefore although symmetric, it is not a metric distance.

The KL divergence between two Cauchy distributions is related to the Pearson $D_{\chi^2_P}[p:q]$ and Neyman $D_{\chi^2_N}[p:q]$ chi square divergences [20]:

$$D_{\chi^2_P}[p:q] := \int \frac{(q(x)-p(x))^2}{p(x)} \mathrm{d}x, \tag{94}$$

$$D_{\chi^2_N}[p:q] := \int \frac{(q(x)-p(x))^2}{q(x)} \mathrm{d}x = D^*_{\chi^2_P}[p:q] = D_{\chi^2_P}[q:p]. \tag{95}$$

Indeed, the formula for the Pearson and Neyman chi square divergences between two Cauchy distributions coincide, and (surprisingly) amount to the $\delta$ distance:

$$D_{\chi^2_P}[p_{l_1,s_1} : p_{l_2,s_2}] = D_{\chi^2_N}[p_{l_1,s_1} : p_{l_2,s_2}], \tag{96}$$

$$= \frac{(s_1+s_2)^2 + (l_2-l_1)^2}{2s_1s_2}, \tag{97}$$

$$=: \delta(l_1,s_1;l_2,s_2). \tag{98}$$

Since the Pearson and Neyman chi square divergences are symmetric, let us write $D_{\chi^2}[p:q] = D_{\chi^2_P}[p:q]$ in the remainder. We can rewrite the Fisher-Rao distance between two Cauchy distributions using the $D_{\chi^2}$ divergence as follows:

$$\rho_{\mathrm{FR}}[p_{l_1,s_1}, p_{l_2,s_2}] = \frac{1}{\sqrt{2}} \operatorname{arccosh}\left(1 + D_{\chi^2}[p_{l_1,s_1} : p_{l_2,s_2}]\right). \tag{99}$$

Figure 3 plots the strictly increasing chi-to-Fisher-Rao conversion function:

$$t_{\chi\to\mathrm{FR}}(u) := \frac{1}{\sqrt{2}} \operatorname{arccosh}(1+u). \tag{100}$$

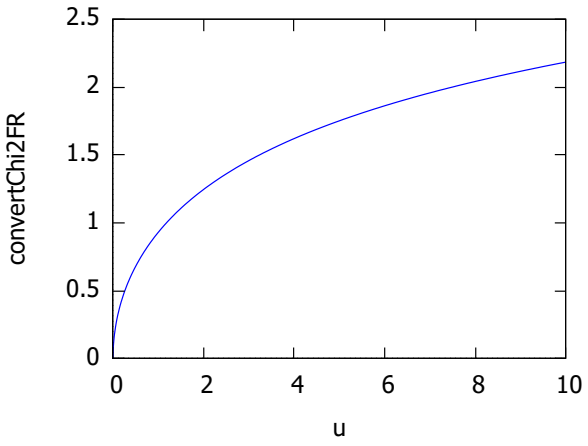



**Figure 3.** Plot of the chi-to-Fisher-Rao conversion function: A strictly increasing function.

Since the Cauchy family is a location-scale family, we have the following general invariance property of $f$-divergences:

**Theorem 2.** *The $f$-divergence [53] between two location-scale densities $p_{l_1,s_1}$ and $p_{l_2,s_2}$ can be reduced to the calculation of the $f$-divergence between one standard density with another location-scale density:*

$$I_f[p_{l_1,s_1} : p_{l_2,s_2}] = I_f\left[p : p_{\frac{l_2-l_1}{s_1},\frac{s_2}{s_1}}\right] = I_f\left[p_{\frac{l_1-l_2}{s_2},\frac{s_1}{s_2}} : p\right]. \tag{101}$$

**Proof.** The proof follows from changes of the variable $x$ in the definite integral of Equation (74): Consider $y = \frac{x-l_1}{s_1}$ with $\mathrm{d}x = s_1\mathrm{d}y$, $x = s_1y + l_1$ and $\frac{x-l_2}{s_2} = \frac{s_1y+l_1-l_2}{s_2} = \frac{y-\frac{l_2-l_1}{s_1}}{\frac{s_2}{s_1}}$. We have

$$\begin{aligned}
I_f[p_{l_1,s_1} : p_{l_2,s_2}] &:= \int_{\mathcal{X}} p_{l_1,s_1}(x) f\left(\frac{p_{l_2,s_2}(x)}{p_{l_1,s_1}(x)}\right)\mathrm{d}x, && (102)\\
&= \int_{\mathcal{Y}} \frac{1}{s_1}p(y) f\left(\frac{\frac{1}{s_2}p\left(\frac{y-\frac{l_2-l_1}{s_1}}{\frac{s_2}{s_1}}\right)}{\frac{1}{s_1}p(y)}\right) s_1\mathrm{d}y, && (103)\\
&= \int p(y) f\left(\frac{p_{\frac{l_2-l_1}{s_1},\frac{s_2}{s_1}}(y)}{p(y)}\right)\mathrm{d}y, && (104)\\
&= I_f\left[p : p_{\frac{l_2-l_1}{s_1},\frac{s_2}{s_1}}\right]. && (105)
\end{aligned}$$

The proof for $I_f[p_{l_1,s_1} : p_{l_2,s_2}] = I_f(p_{\frac{l_1-l_2}{s_2},\frac{s_1}{s_2}} : p)$ is similar. One can also use the *conjugate generator* $f^*(u) := uf(\frac{1}{u})$ which yields the *reverse $f$-divergence*: $I_{f^*}[p : q] = I_f[q : p] = I_{f^*}[p : q]$. □

Since the KL divergence is expressed by $D_{\mathrm{KL}}[p_{l_1,s_1} : p_{l_2,s_2}] = \log\left(1+\frac{1}{2}\delta(l_1,s_1;l_2,s_2)\right)$, we also check that

$$\begin{aligned}\delta(l_1,s_1;l_2,s_2) &= \delta\left(0,1;\frac{l_1-l_2}{s_2},\frac{s_1}{s_2}\right), &(106)\\ &= \delta\left(\frac{l_2-l_1}{s_1},\frac{s_2}{s_1};0,1\right), &(107)\\ &=: \delta(a,b), &(108)\end{aligned}$$

where

$$\delta(a,b) := \frac{a^2+(b-1)^2}{4b}. \tag{109}$$

It follows the following corollary for scale families:

**Corollary 1.** *The $f$-divergences between scale densities is scale-invariant and amount to a scalar scale-invariant divergence $D_f(s_1 : s_2) := I_f[p_{s_1} : p_{s_2}]$.*

**Proof.**

$$\begin{aligned}D_f(s_1 : s_2) := I_f[p_{s_1} : p_{s_2}] &= I_f\left(p : p_{\frac{s_2}{s_1}}\right) =: D_f\left(1 : \frac{s_2}{s_1}\right), &(110)\\ &= I_f\left[p_{\frac{s_1}{s_2}} : q\right] =: D_f\left(\frac{s_1}{s_2} : 1\right). &(111)\end{aligned}$$

□

Many algorithms and data-structures can be designed efficiently when dealing with metric distances: For example, the metric ball tree [59] or the vantage point tree [60,61] are two such data structures for querying efficiently nearest neighbors in metric spaces. Thus, it is of interest to consider *statistical dissimilarities* which are metric distances. The total variation distance [12] and the square-root of the Jensen-Shannon divergence [62] are two common examples of *statistical metric distances* often met in the literature. In general, the metrization of $f$-divergences was investigated in [63,64].

We shall prove the following theorem:

**Theorem 3.** *The square root of the Kullback-Leibler divergence between two Cauchy density $p_{l_1,s_1}$ and $p_{l_2,s_2}$ is a metric distance:*

$$\rho_{\mathrm{KL}}[p_{l_1,s_1}, p_{l_2,s_2}] := \sqrt{D_{\mathrm{KL}}[p_{l_1,s_1} : p_{l_2,s_2}]} = \sqrt{\log\left(1+\frac{(s_1-s_2)^2+(l_1-l_2)^2}{4s_1s_2}\right)}. \tag{112}$$

**Proof.** The proof consists in showing that the square root of the conversion function of the Fisher-Rao distance to the KL divergence is a metric transform [65]. A *metric transform* $t(u) : \mathbb{R}_+ \to \mathbb{R}_+$ is a transform which preserves the metric distance $\rho$, i.e., $(t \circ \rho)(p,q) = t(\rho(p,q))$ is a metric distance. The following are sufficient conditions for function $t(u)$ to be a metric transform:

1. $t$ is a strictly increasing function,
2. $t(0) = 0$,
3. $t$ satisfies that *subadditive property*: $t(a+b) \leq t(a)+t(b)$ for all $a, b \geq 0$.

For example, strictly concave functions $t(u)$ with $t(0) = 0$ are metric transforms. In general, one can check that $t(u)$ is subadditive by verifying that the ratio of functions $\frac{t(u)}{u}$ is non-decreasing.

The following transform $\sqrt{t_{\mathrm{FR}\to\mathrm{KL}}(u)}$ converts the Fisher-Rao distance $\rho_{\mathrm{FR}}$ to the Kullback-Leibler divergence $D_{\mathrm{KL}}$:

$$t_{\mathrm{FR}\to\mathrm{KL}}(u) := \log\left(\frac{1}{2}+\frac{1}{2}\cosh(\sqrt{2}u)\right), \tag{113}$$

where

$$\cosh(x) := \frac{e^x + e^{-x}}{2}. \tag{114}$$

The square root of that conversion function is a subadditive function since $\frac{\sqrt{t_{\mathrm{FR}\to\mathrm{KL}}(u)}}{u}$ is non-decreasing (see Figure 4) and $\sqrt{t_{\mathrm{FR}\to\mathrm{KL}}(0)} = 0$.

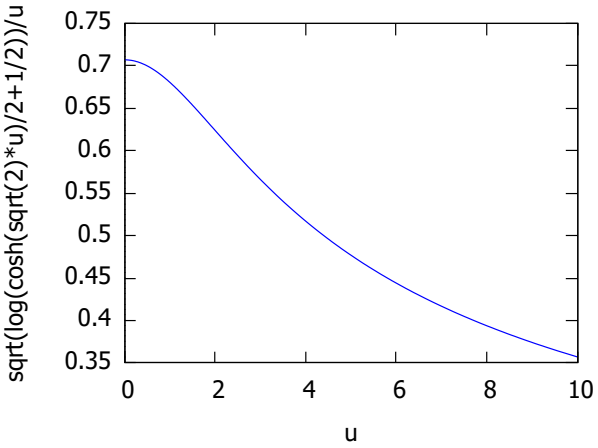



**Figure 4.** Plot of the function $\frac{\sqrt{t_{\mathrm{FR}\to\mathrm{KL}}(u)}}{u}$.

Since the Fisher-Rao distance is a metric distance and since $\sqrt{t_{\mathrm{FR}\to\mathrm{KL}}(u)}$ is a metric transform, we conclude that

$$\rho_{\mathrm{KL}}[p_{l_1,s_1} : p_{l_2,s_2}] := \sqrt{D_{\mathrm{KL}}[p_{l_1,s_1} : p_{l_2,s_2}]} = \sqrt{t_{\mathrm{FR}\to\mathrm{KL}}(\rho_{\mathrm{FR}}[p_{l_1,s_1} : p_{l_2,s_2}])} \tag{115}$$

is a metric distance. □

A metric distance $\rho(p,q)$ is said to be *Hilbertian* if there exists an *embedding* $\phi(\cdot)$ into a Hilbert space such that $\rho(p,q) = \|\phi(p) - \phi(q)\|_H$, where $\|\cdot\|_H$ is a norm. A metric is said to be Euclidean if there exists an *embedding* with associated norm $\ell_2$, the Euclidean norm. For example, the square root of the celebrated Jensen-Shannon divergence is a Hilbertian distance [62].

Let us prove the following:

**Theorem 4.** *The square root of the KL divergence between to Cauchy densities of the same scale family is a Hilbertian distance.*

**Proof.** For Cauchy distributions with fixed location parameter $l$, the KL divergence of Equation (93) simplifies to:

$$D_{\mathrm{KL}}[p_{l,s_1} : p_{l,s_2}] = \log\left(\frac{(s_1+s_2)^2}{4s_1s_2}\right). \tag{116}$$

We can rewrite this KL divergence as

$$D_{\mathrm{KL}}[p_{l,s_1} : p_{l,s_2}] = 2\log\left(\frac{A(s_1,s_2)}{G(s_1,s_2)}\right), \tag{117}$$

where $A(s_1,s_2) = \frac{s_1+s_2}{2}$ and $G(s_1,s_2) = \sqrt{s_1s_2}$ are the *arithmetic mean* and the *geometric mean* of $s_1$ and $s_2$, respectively. Then we use Lemma 3 of [66] to conclude that $\sqrt{D_{\mathrm{KL}}[p_{l,s_1} : p_{l,s_2}]}$ is a Hilbertian metric distance.

Another proof consists in rewriting the KL divergence as a scaled *Jensen-Bregman divergence* [66,67]:

$$D_{\mathrm{KL}}[p_{l,s_1} : p_{l,s_2}] = 2\,\mathrm{JB}_F(s_1, s_2), \tag{118}$$

where

$$\mathrm{JB}_F(\theta_1, \theta_2) := \frac{F(\theta_1) + F(\theta_2)}{2} - F\left(\frac{\theta_1 + \theta_2}{2}\right), \tag{119}$$

for a strictly convex generator $F$. We use $F(\theta) = -\log(u)$, i.e., the *Burg information* yielding the *Jensen-Burg divergence* $\mathrm{JB}_F$. Then we use Corollary 1 of [66] (i.e., $F$ is the cumulant of an infinitely divisible distribution) to conclude that $\sqrt{\mathrm{JB}_F(\theta_1, \theta_2)}$ is a metric distance (and hence, $\rho_{\mathrm{KL}}(l, s_1, l, s_2) = \sqrt{D_{\mathrm{KL}}[p_{l,s_1} : p_{l,s_2}]} = \sqrt{2}\sqrt{\mathrm{JB}_F(s_1, s_2)}$ is a Hilbertian metric distance). □

The *α-skewed Jensen-Bregman divergence* is defined by

$$\mathrm{JB}_F^{\alpha}(\theta_1 : \theta_2) := \alpha F(\theta_1) + (1-\alpha)F(\theta_2) - F\left(\alpha\theta_1 + (1-\alpha)\theta_2\right), \tag{120}$$

and the maximal $\alpha$-skewed Jensen-Bregman divergence is called the *Jensen-Chernoff divergence*:

$$\mathrm{JB}_F^{\alpha^*}(\theta_1 : \theta_2) := \max_{\alpha \in (0,1)} \mathrm{JB}_F^{\alpha}(\theta_1 : \theta_2). \tag{121}$$

The maximal exponent $\alpha^*$ corresponds to the *error exponent* in Bayesian hypothesis testing on exponential family manifolds [57]. In general, the metrization of Jensen-Bregman divergence (and Jensen-Chernoff) was studied in [68].

Furthermore, by combining Corollary 1 of [66] with Theorem 3 of [67], we get the following proposition:

**Proposition 2.** *The square root of the Bhattacharyya divergence between two densities of an exponential family is a metric distance when the exponential family is infinitely divisible.*

This proposition holds because the Bhattacharyya divergence

$$D_{\mathrm{Bhat}}[p, q] = -\log\left(\int \sqrt{p(x)q(x)}\mathrm{d}x\right), \tag{122}$$

between two parametric densities $p(x) = p_{\theta_1}(x)$ and $q(x) = p_{\theta_2}(x)$ of an exponential family with cumulant function $F$ amounts to a Jensen-Bregman divergence [67] (Theorem 3 of [67]):

$$D_{\mathrm{Bhat}}[p_{\theta_1}(x), p_{\theta_2}(x)] = \mathrm{JB}_F(\theta_1, \theta_2). \tag{123}$$

Notice that Proposition 2 recovers the fact that the square root of the Bhattacharyya divergence between two zero-centered normal distributions is a metric (proved differently in [69]) since the set of normal distributions form an infinitely divisible exponential family.

## 4. Cauchy Voronoi Diagrams and Dual Cauchy Delaunay Complexes

Let us consider the Voronoi diagram [1] of a finite set $\mathcal{P} = \{p_{\lambda_1}, \ldots p_{\lambda_n}\}$ of $n$ Cauchy distributions with the location-scale parameters $\lambda_i = (l_i, s_i) \in \mathbb{H}$ for $i \in \{1, \ldots, n\}$. We shall consider the Fisher-Rao distance $\rho_{\mathrm{FR}}$, the KL divergence $D_{\mathrm{KL}}$ and its square root metrization $\rho_{\mathrm{KL}}$, the chi square divergence $D_{\chi^2}$, and the flat divergence $D_{\mathrm{flat}}$.

### *4.1. The Hyperbolic Cauchy Voronoi Diagrams*

Observe that the Voronoi diagram does not change under any strictly increasing function $t$ of the dissimilarity measure (e.g., square root function): $\mathrm{Vor}_{D\circ t}(\mathcal{P}) = \mathrm{Vor}_D(\mathcal{P})$. Thus, we get the following theorem:

**Theorem 5.** *The Cauchy Voronoi diagrams under the Fisher-Rao distance, the the chi-square divergence and the Kullback-Leibler divergence all coincide, and amount to a hyperbolic Voronoi diagram on the corresponding location-scale parameters.*

**Proof.** The KL divergence can be expressed as

$$D_{\mathrm{KL}}[p_{l_1,s_1} : p_{l_2,s_2}] = \log\left(1 + \frac{1}{2}\delta(l_1, s_1, l_2, s_2)\right). \tag{124}$$

Thus, both the $D_{\mathrm{KL}}$ and $\rho_{\mathrm{FR}}$ dissimilarities are expressed as strictly increasing functions of $\delta$ (a synonym for the $D_{\chi^2}$ divergence). Therefore the *Voronoi bisectors* between two Cauchy distributions $p_{l_1,s_1}$ and $p_{l_2,s_2}$ for $D \in \{\rho_{\mathrm{FR}}, D_{\mathrm{KL}}, \sqrt{D_{\mathrm{KL}}}, D_{\chi^2}\}$ amounts to the same expression:

$$\mathrm{Bi}_D(p_{\lambda_1} : p_{\lambda_2}) = \{\lambda \in \mathbb{H} : \delta(\lambda, \lambda_1) = \delta(\lambda, \lambda_2)\}, \tag{125}$$

$$\mathrm{Bi}_D(p_{l_1,s_1} : p_{l_2,s_2}) = \{(l, s) \in \mathbb{H} : \delta(l, s, l_1, s_1) = \delta(l, s, l_2, s_2)\}. \tag{126}$$

□

It follows that we can calculate the Cauchy Voronoi diagram of $n$ Cauchy distributions in optimal $\Theta(n \log n)$ time by calculating the 2D hyperbolic Voronoi diagram [25,26] on the location-scale parameters (see Section 6 for details). Figure 5 displays the Voronoi diagram of a set of Cauchy distributions by its equivalent parameter hyperbolic Voronoi diagram in the Poincaré upper plane model, the Poincaré disk model, and the Klein disk model. Figure 6 shows the hyperbolic Voronoi diagram in the upper plane with colored Voronoi cells. A model of hyperbolic geometry is said to be *conformal* if it preserves angles, i.e., its underlying Riemannian metric tensor is a scalar positive function of the Euclidean metric tensor. The Poincaré disk model and the Poincaré upper plane model are both conformal models [30]. The Klein model is *not* conformal, except at the disk origin. Let $\mathbb{D} = \{p : \|p\| < 1\}$ denote the open unit disk domain for the Poincaré and Klein disk models. Indeed, the Riemannian metric corresponding to the Klein disk model is

$$\mathrm{d}s^2_{\mathrm{Klein}}(p) = \frac{\mathrm{d}s^2_{\mathrm{Eucl}}}{1 - \|p\|^2} + \frac{\langle p, \mathrm{d}p\rangle}{(1 - \|p\|^2)^2}, \tag{127}$$

where $\mathrm{d}p = \mathrm{d}x + \mathrm{d}y$ and $\mathrm{d}s_{\mathrm{Eucl}} = \sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}$ denotes the Euclidean line element. Since $\mathrm{d}s^2_{\mathrm{Klein}}(0) = \mathrm{d}s^2_{\mathrm{Eucl}}$, we deduce that Klein model is conformal at the origin (when measuring the angles between two vectors $v_1$ and $v_2$ of the tangent plane $T_0$).

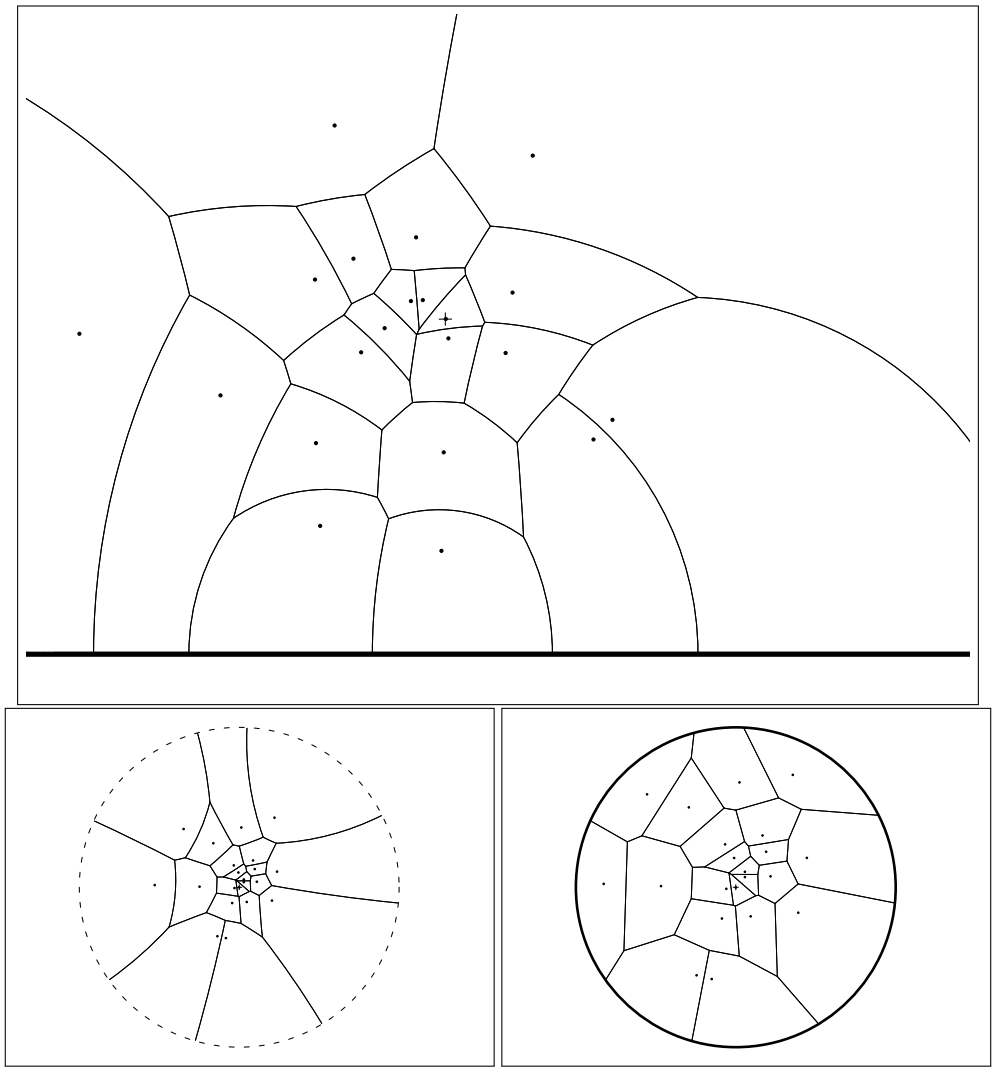


**Figure 5.** Hyperbolic Voronoi diagram of a set of Cauchy distributions in the Poincaré upper plane (**top**), the Poincaré disk model (**bottom left**), and the Klein disk model (**bottom right**).

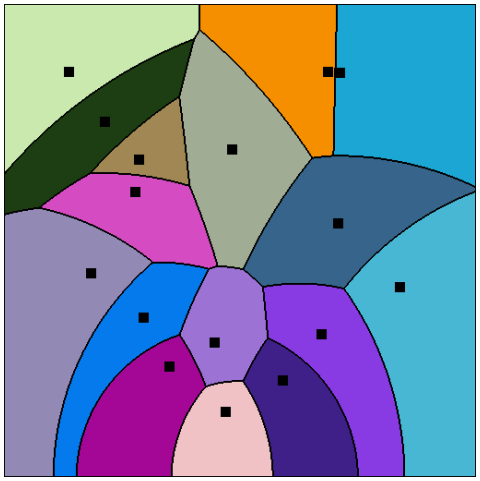


**Figure 6.** A hyperbolic Cauchy Voronoi diagram of a finite set of Cauchy distributions (black square generators, colored Voronoi cells, and black cell borders).

The dual of the Voronoi diagram is called the *Delaunay (simplicial) complex* [4,5]: We build the Delaunay complex by drawing an edge between generators whose Voronoi cells are adjacent. For the ordinary Euclidean Delaunay complex with points in general position (i.e., no $d+2$ cospherical points in dimension $d$), the Delaunay complex triangulates the convex hull of the points [8,70]. Therefore it is called the *Delaunay triangulation* [1,3,8]. Figure 7 displays an Euclidean Voronoi diagram with its dual Delaunay triangulation.

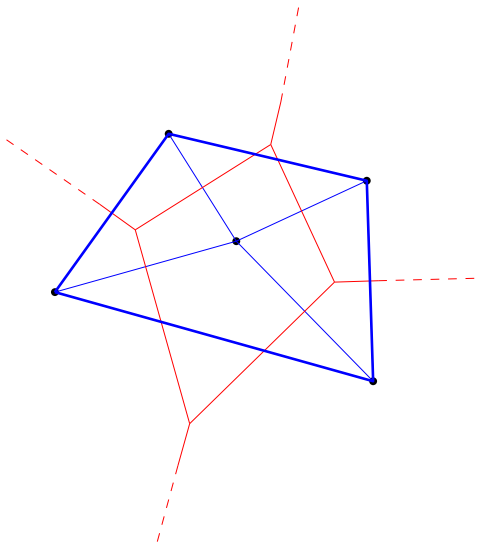


**Figure 7.** Duality between the ordinary Euclidean Voronoi diagram and the Delaunay structures: The Voronoi diagram partitions the space into Voronoi proximity cells. The Delaunay complex triangulates the convex hull of the generators. A Delaunay edge is drawn between the generators of adjacent Voronoi cells. Observe that the Delaunay edges cuts orthogonally the corresponding Voronoi bisectors in Euclidean geometry.

Similarly, for the hyperbolic Voronoi diagram, we construct the *hyperbolic Delaunay complex* by drawing a hyperbolic geodesic edge between any two generators whose Voronoi cells are adjacent. However, we do not necessarily obtain anymore a geodesic triangulation of the hyperbolic geodesic convex hull but rather a simplicial complex, hence the name *hyperbolic Delaunay complex* [5,71,72]. In extreme cases, the hyperbolic Delaunay complex has a tree structure. See Figure 8 for examples of a hyperbolic Delaunay triangulation and a hyperbolic Delaunay complex which is not a triangulation In fact, hyperbolic geometry is very well-suited for embedding isometrically with low distortion weighted tree graphs [73]. Hyperbolic embeddings of hierarchical structures [74] has become a hot topic in machine learning.

Let us now prove that these Cauchy hyperbolic Voronoi/Delaunay structures are *Fisher orthogonal*:

**Theorem 6.** *The Cauchy Voronoi diagram is Fisher orthogonal to the Cauchy Delaunay complex.*

**Proof.** It is enough to prove that the corresponding hyperbolic geodesic $\gamma(p_{\lambda_1}, p_{\lambda_2})$ is orthogonal to the bisector $\mathrm{Bi}(p_{\lambda_1} : p_{\lambda_2})$. The distance in the Klein disk model is

$$\rho_{\mathrm{Klein}}(p,q) = \rho_K(p,q) := \operatorname{arccosh}\left(\frac{1-\langle p,q\rangle}{\sqrt{(1-\|p\|^2)(1-\|q\|^2)}}\right). \tag{128}$$

The equation of the hyperbolic bisector in the Klein disk model [25] is

$$\mathrm{Bi}_{\rho_{\mathrm{Klein}}}(\lambda_1:\lambda_2) = \left\{\lambda\in\mathbb{D} \;:\; \lambda^\top\left(\sqrt{1-\|\lambda_1\|^2}\lambda_2 - \sqrt{1-\|\lambda_2\|^2}\lambda_1\right) + \sqrt{1-\|\lambda_2\|^2} - \sqrt{1-\|\lambda_1\|^2} = 0\right\}. \tag{129}$$

Using a Möbius transformation [25] (i.e., a hyperbolic "rigid motion"), we may consider without loss of generality that $p_{\lambda_1} = -p_{\lambda_2}$. It follows that the bisector equation writes simply as

$$\mathrm{Bi}_{\rho_{\mathrm{Klein}}} = \left\{ \lambda \ : \ 2\sqrt{1 - \|p_{\lambda_1}\|}\lambda^\top \lambda_1 = 0 \right\}. \tag{130}$$

Since the Klein disk model is conformal at the origin, we deduce from Equation (130) that we have $\gamma(p_{\lambda_1}, p_{\lambda_2}) \perp \mathrm{Bi}(p_{\lambda_1} : p_{\lambda_2})$. □

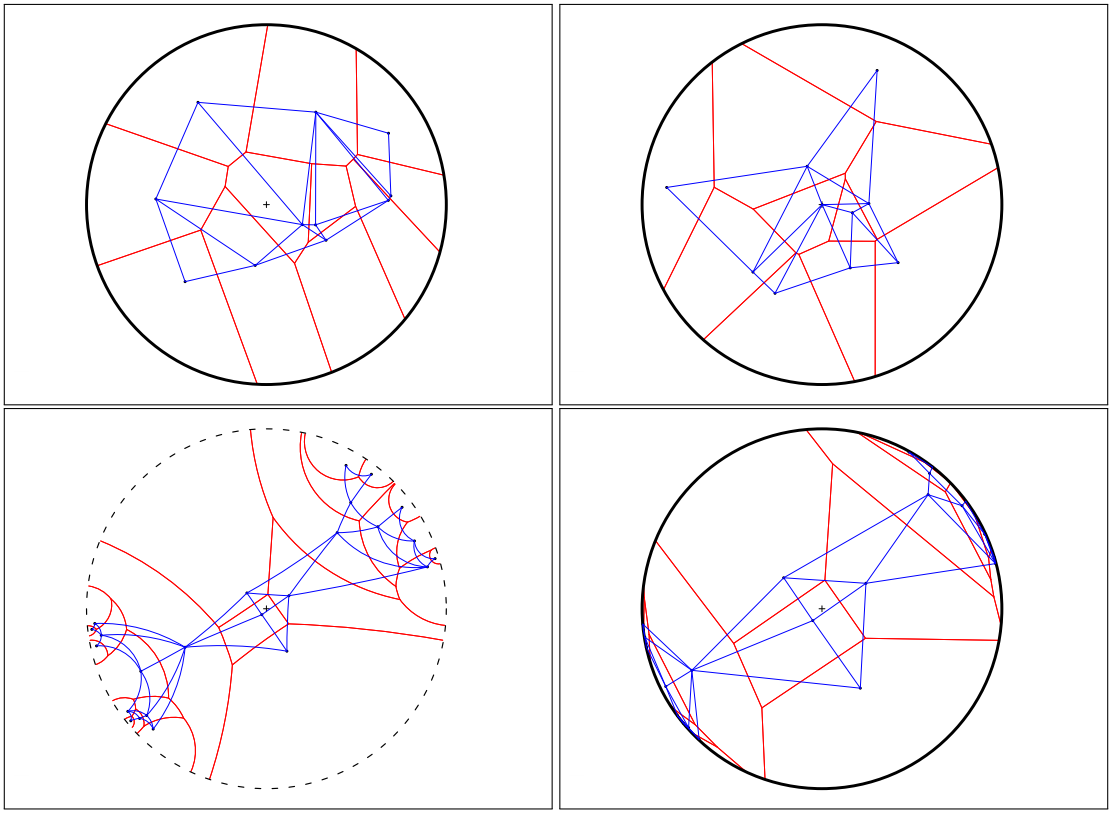


**Figure 8.** Examples of hyperbolic Voronoi Delaunay complexes drawn in the Klein model: Delaunay complex triangulates the convex hull yielding the Delaunay triangulation (**top left**), and Delaunay complex which does not triangulate the convex hull, (**top right**). **Bottom**: A hyperbolic Voronoi diagram and its dual Delaunay complex displayed in the Poincaré disk model (**left**) and in the Klein disk model (**right**).

Figure 9 displays two bisectors with their corresponding geodesics in the Klein model. We check that the Euclidean angles are deformed when the intersection point is not at the disk origin. Section 6 provides further details for the efficient construction of the hyperbolic Voronoi diagram in the Klein model.

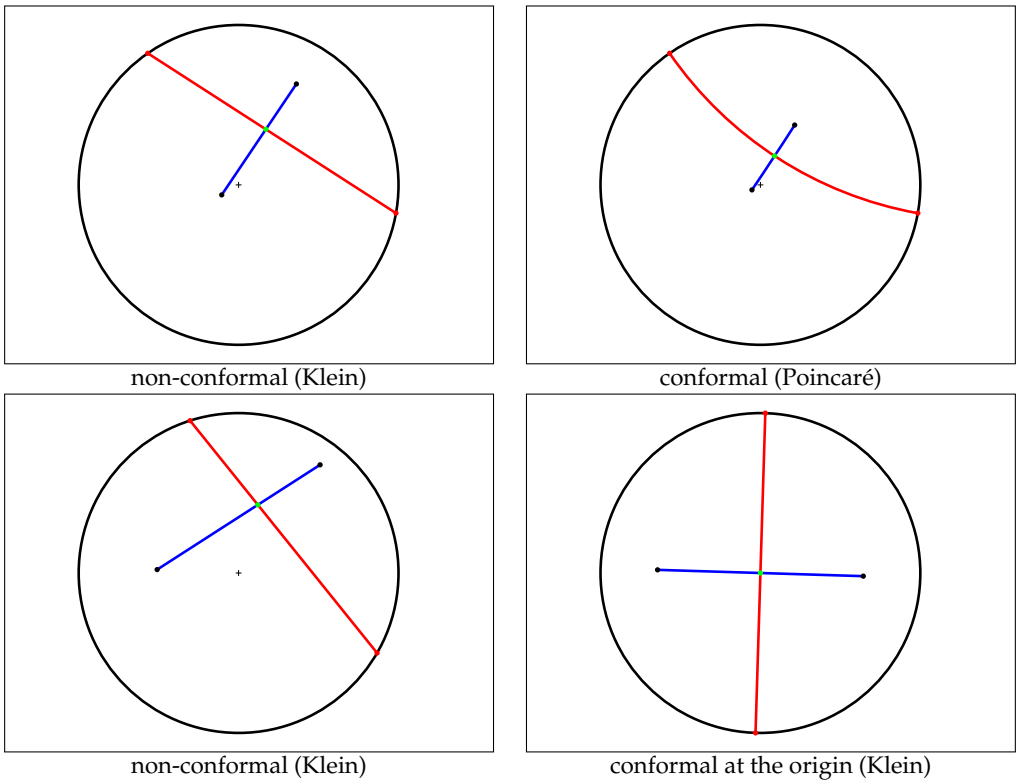



**Figure 9.** In hyperbolic geometry, the Voronoi bisector between two generators is orthogonal to the geodesic linking them. The top figures display a pair of (bisector,geodesic) in the Klein model (left), and the same pair in the Poincaré model (right). When viewed in Klein non-conformal model, the bisector does not intersect orthogonally (with respect to the Euclidean geometry) the geodesic (left) except when the intersection point is at the disk origin (bottom right).

**Remark 4.** *The hyperbolic Cauchy Voronoi diagram can be used for* classification tasks *in statistics as originally motivated by C.R. Rao in his celebrated paper [17]: Let $p_{\lambda_1}, \ldots, p_{\lambda_n}$ be $n$ Cauchy distributions, and $x_1, \ldots, x_s$ be $s$ identically and independently samples drawn from a Cauchy distribution $p_\lambda$. We can estimate $\hat{\lambda}$ the location-scale parameters from the $s$ samples [75], and then decide the multiple test hypothesis $H_i : p_\lambda = p_{\lambda_i}$ by choosing the hypothesis $H_i$ such that $\rho_{\mathrm{FR}}(p_{\lambda_i}, p_\lambda) \leq \rho_{\mathrm{FR}}(p_{\lambda_j}, p_\lambda)$ for all $j \in \{1, \ldots, n\}$. This classification task amounts to perform a nearest neighbor query in the Fisher-Rao hyperbolic Cauchy Voronoi diagram. Hypothesis testing for comparing location parameters based on Rao's distance is investigated in [76].*

Figure 10 displays the hyperbolic Voronoi Cauchy diagram induced by 300 Cauchy distribution generators.

Notice that it is possible to construct a set of points such that all hyperbolic Voronoi cells for that point set are unbounded. See Figure 11 for such an example.

The ordinary Euclidean Delaunay triangulation satisfies the *empty sphere* property [4,77]: That is the *circumscribing spheres* passing through the vertices of the Delaunay triangles of the Delaunay complex are empty of any other Voronoi site. This property still holds for the hyperbolic Delaunay complex which is obtained by a filtration of the ordinary Euclidean Delaunay triangulation in [5]. A *hyperbolic ball* in the Poincaré conformal disk model or the upper plane model has the shape of a Euclidean ball with *displaced center* [71]. Figure 12 displays the Delaunay complex with the empty sphere property in the Poincaré and Klein disk models. The centers of these circumscribing spheres are located at the $T$-junctions of the Voronoi diagrams.

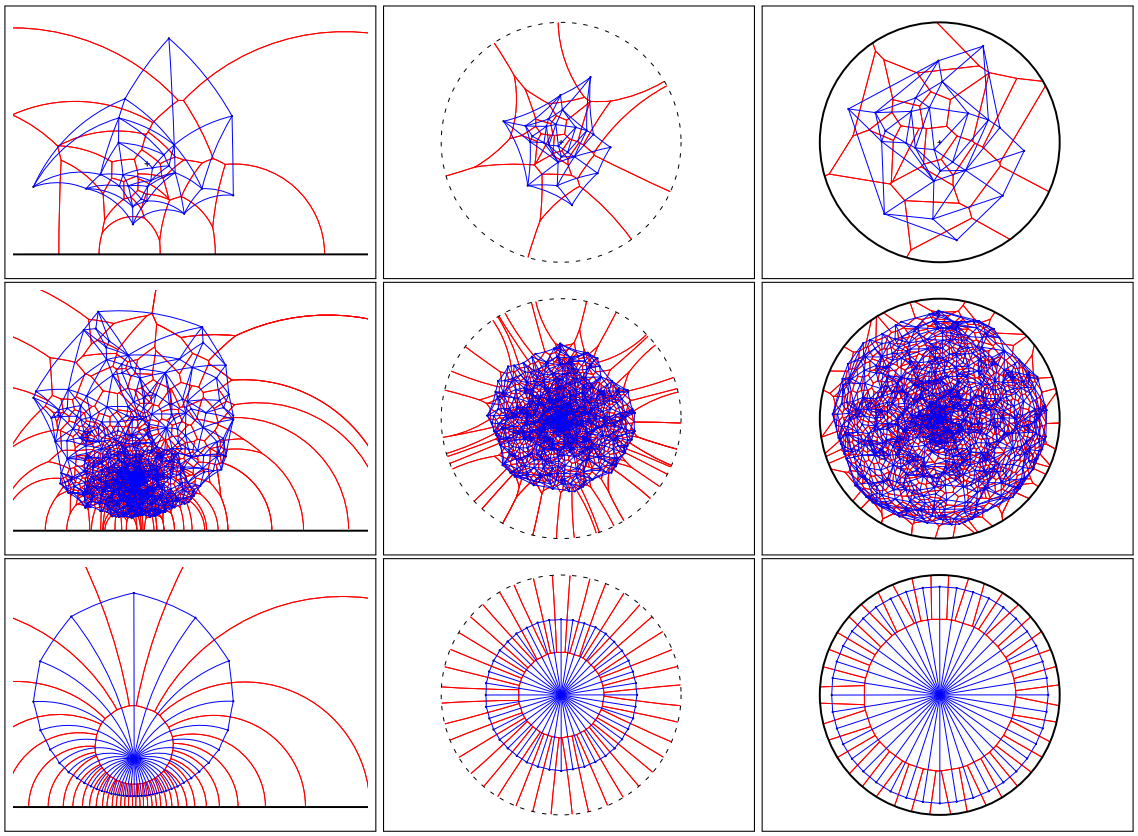


**Figure 10.** Equivalent hyperbolic Voronoi diagram and dual Delaunay complex of a set of Cauchy distributions in the Poincaré upper plane (**left**), the Poincaré disk model (**middle**), and the Klein disk model (**right**). Top row figures for $n = 24$ Cauchy distributions, middle row figures for $n = 1024$ distributions and bottom row figures for a quasi-regular set of $n = 25$ Cauchy distributions.

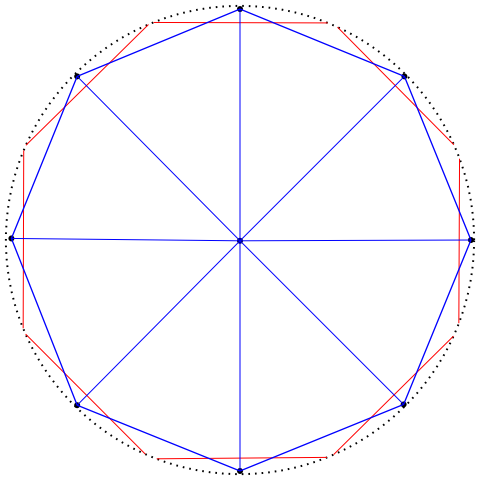


**Figure 11.** A hyperbolic Voronoi diagram with all unbounded Voronoi cells.

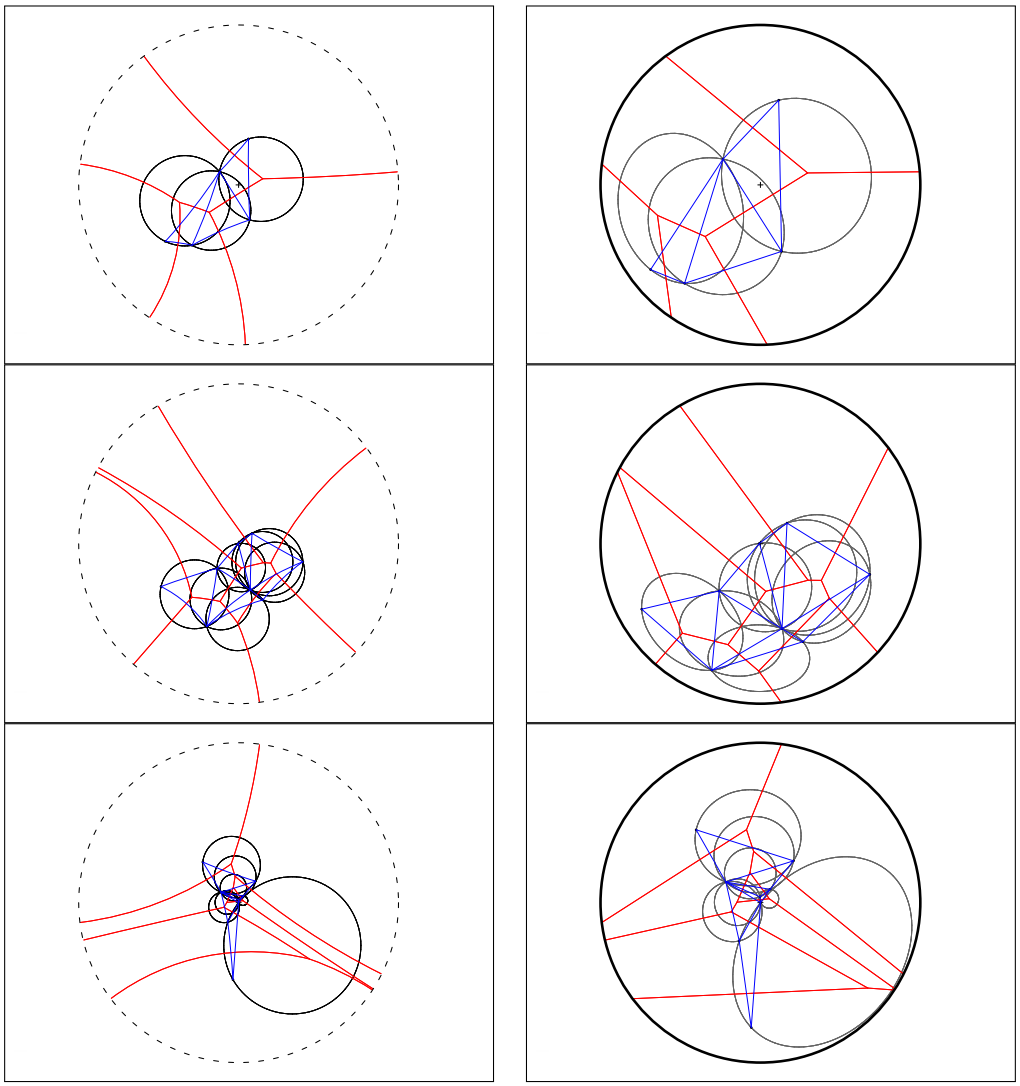


**Figure 12.** Delaunay triangles of the hyperbolic Delaunay complex satisfy the empty circumscribing sphere property. The empty sphere centers are located on the Voronoi $T$-junction vertices. The hyperbolic spheres are displayed as ordinary Euclidean sphere (with displaced center) in the Poincaré model (**left column**) and as ellipsoids (with displaced center) in the Klein model (**right column**). The centers of the empty hyperbolic spheres are located at the Voronoi $T$-junctions.

### *4.2. The Dual Voronoi Diagrams on the Cauchy Dually Flat Manifold*

The dual Cauchy Voronoi diagrams with respect to the flat divergence $D_{\text{flat}}$ (and dual reverse flat divergence $D^*_{\text{flat}}$ which corresponds to a dual Bregman-Tsallis divergence) of Section 2.4 amount to calculate 2D dual Bregman Voronoi diagrams [15,16]. We get the following dual bisectors: The primal bisector with respect to the dual flat divergence is:

$$\begin{aligned} \mathrm{Bi}_{D_{\text{flat}}}(p_{\lambda_1} : p_{\lambda_2}) &= \left\{ p_\lambda \ : \ D_{\text{flat}}[p_{\lambda_1} : p_\lambda] = D_{\text{flat}}[p_{\lambda_2} : p_\lambda] \right\}, & (131) \\ &= \left\{ \lambda \ : \ \delta(l_1, s_1; l, s) = \delta(l_2, s_2; l, s) \right\}. & (132) \end{aligned}$$

Thus, this primal bisector with respect to the flat divergence corresponds to the hyperbolic bisector of the Fisher-Rao distance/chi square/ KL divergences:

$$\mathrm{Bi}_{D_{\text{flat}}}(p_{\lambda_1} : p_{\lambda_2}) = \mathrm{Bi}_{\rho_{\text{FR}}}(p_{\lambda_1} : p_{\lambda_2}) = \mathrm{Bi}_{D_{\text{KL}}}(p_{\lambda_1} : p_{\lambda_2}) = \mathrm{Bi}_{D_{\chi^2}}(p_{\lambda_1} : p_{\lambda_2}). \quad (133)$$

The dual bisector with respect to the dual flat divergence (reverse Bregman-Tsallis divergence) is:

$$\begin{aligned} \mathrm{Bi}^*_{D_{\mathrm{flat}}}(p_{\lambda_1} : p_{\lambda_2}) &= \left\{ p_\lambda \ : \ D_{\mathrm{flat}}[p_\lambda : p_{\lambda_1}] = D_{\mathrm{flat}}[p_\lambda : p_{\lambda_2}] \right\}, & (134) \\ &= \left\{ \lambda \ : \ \|\lambda - \lambda_1\| = \|\lambda - \lambda_2\| \right\}. & (135) \end{aligned}$$

That is, the dual bisector corresponds to an ordinary Euclidean bisector:

$$\mathrm{Bi}^*_{D_{\mathrm{flat}}}(p_{\lambda_1} : p_{\lambda_2}) = \mathrm{Bi}_{\rho_E}(p_{\lambda_1}, p_{\lambda_2}). \quad (136)$$

Notice that $\mathrm{Bi}^*_{D_{\mathrm{flat}}}(p_{\lambda_1} : p_{\lambda_2}) = \mathrm{Bi}_{D^*_{\mathrm{flat}}}(p_{\lambda_1} : p_{\lambda_2})$.

To summarize, one primal bisector coincides with the Fisher-Rao bisector while the dual bisector amounts to the ordinary Euclidean bisector.

**Theorem 7.** *The dual Cauchy Voronoi diagrams with respect to the flat divergence can be calculated efficiently in $\Theta(n \log n)$-time.*

The construction of 2D Bregman Voronoi diagrams is described in [15].

### *4.3. The Cauchy Voronoi Diagrams with Respect to α-Divergences*

The dual bisectors with respect to the $\alpha$-divergences between any two parametric probability densities $p_{\lambda_1}(x)$ and $p_{\lambda_2}(x)$ are

$$\begin{aligned} \mathrm{Bi}_{I_\alpha}(p_{\lambda_1} : p_{\lambda_2}) &= \left\{ p_\lambda \ : \ I_\alpha[p_{\lambda_1} : p_\lambda] = I_\alpha[p_{\lambda_2} : p_\lambda] \right\}, & (137) \\ &= \left\{ \lambda \ : \ C_\alpha(p_{\lambda_1}; p_\lambda) = C_\alpha(p_{\lambda_2}; p_\lambda) \right\}, & (138) \end{aligned}$$

and

$$\begin{aligned} \mathrm{Bi}^*_{I_\alpha}(p_{\lambda_1} : p_{\lambda_2}) &= \left\{ p_\lambda \ : \ I_\alpha[p_\lambda : p_{\lambda_1}] = I_\alpha[p_\lambda : p_{\lambda_2}] \right\}, & (139) \\ &= \mathrm{Bi}_{I_{1-\alpha}}(p_{\lambda_1} : p_{\lambda_2}). & (140) \end{aligned}$$

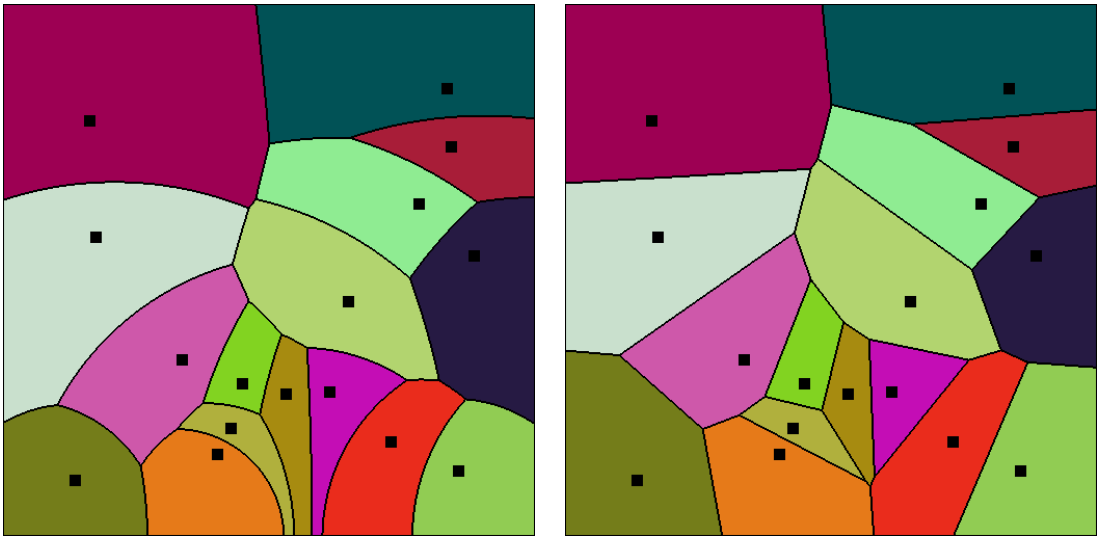
It is an open problem to prove *when* the dual $\alpha$-bisectors coincide for the Cauchy family. We have shown it is the case for the $\chi^2$-divergence and the KL divergence. In theory, the Risch semi-algorithm [78] allows one to answer whether a definite integral has a closed-form formula or not. However, the Risch semi-algorithm is only a semi-algorithm as it requires to implement an oracle to check whether some mathematical expressions are equivalent to zero or not.

## 5. Conclusions

In this paper, we have considered the construction of Voronoi diagrams of finite sets of Cauchy distributions with respect to some common statistical distances. Since statistical distances can potentially be asymmetric, we defined the dual Voronoi diagrams with respect to the forward and reverse/dual statistical distances. From the viewpoint of information geometry [7], we have reported the construction of two types of geometry on the Cauchy manifold: (1) The *invariant α-geometry* equipped with the Fisher metric tensor $g_{\mathrm{FR}}$ and the skewness tensor $T$ from which we can build a family of pairs of torsion-free affine connections coupled with the metric, and (2) a *dually flat geometry* induced by a Bregman generator defined by the free energy $F_q$ of the $q$-Gaussians (here, instantiated to $q = 2$ when dealing with the Cauchy family). The metric tensor of the latter geometry is called the $q$-Fisher information metric, and is a Riemannian conformal metric of the Fisher information metric. We have shown that the Fisher-Rao distance amount to a scaled hyperbolic distance in the Poincaré upper plane model (Proposition 1), and that all Amari's $\alpha$-geometries [7] coincide with the Fisher-Rao geometry since the cubic tensor vanishes, thus yielding a hyperbolic manifold of negative constant

scalar curvature $\kappa = -2$ for the Cauchy $\alpha$-geometric manifolds. We noticed that the Fisher-Rao distance and the KL divergence can be expressed as a strictly increasing function of the chi square divergence. Then we explained how to conformally flatten the curved Fisher-Rao geometry to obtain a dually flat space where the flat divergence amounts to a canonical Bregman divergence built from Tsallis' quadratic entropy (Theorem 1). We reported the Hessian metrics of the dual potential functions of the dually flat space, and showed that there are other alternative choices for building Hessian structures [52].

Table 1 summarizes the various closed-form formula of statistical dissimilarities obtained for the Cauchy family. We proved that the square root of the KL divergence between any two Cauchy distributions is a metric distance (Theorem 3) in general, and more precisely a Hilbertian metric for the scale Cauchy families (Theorem 4). It follows that the Cauchy Voronoi diagram for the Fisher-Rao distance coincides with the Voronoi diagram with respect to the KL divergence or the chi square divergence (Figure 13). We showed how to build this hyperbolic Cauchy diagram from an equivalent hyperbolic Voronoi diagram on the corresponding location-scale parameters (see also Section 6). Then we proved that the dual hyperbolic Cauchy Delaunay complex is Fisher orthogonal to the Fisher-Rao hyperbolic Cauchy Voronoi diagram (Theorem 6). The dual Voronoi diagrams with respect to the dual flat divergences can be built from the corresponding dual Bregman-Tsallis divergences with the primal Voronoi diagram coinciding with the hyperbolic Voronoi diagram and the dual diagram coinciding with the ordinary Euclidean Voronoi diagram (Figure 13). These results are particular to the special case of the Cauchy location-scale family, and do not hold in general for arbitrary location-scale families since the cubic tensor may not vanish [28] and the KL divergence is usually asymmetric (e.g., the Gaussian location-scale family). However, the Fisher-Rao geometry of any location-scale family amounts after a potential rescaling to hyperbolic geometry [27,79].



$\mathrm{Vor}_{\rho_{\mathrm{FR}}} = \mathrm{Vor}_{\rho_{\mathrm{KL}}} = \mathrm{Vor}_{\rho^2_{\chi}} = \mathrm{Vor}_{D_{\mathrm{flat}}}$ $\mathrm{Vor}_{D^*_{\mathrm{flat}}} = \mathrm{Vor}_{\rho_E}$.

**Figure 13.** Voronoi diagrams of a set of Cauchy distributions with respect to the Fisher-Rao (FR) distance $\rho_{\mathrm{FR}}$, the Kullback-Leibler (KL) divergence $D_{\mathrm{KL}}$, the $\chi^2$-divergence $D_{\chi^2}$, and the asymmetric Bregman-Tsallis flat divergence $D_{\mathrm{flat}}$.

**Table 1.** Summary of the main closed-form formula for the statistical distances between Cauchy densities and their induced Voronoi diagrams.

| Formula | Voronoi |
|---|---|
| $D_{\chi^2}[p_{l_1,s_1}, p_{l_2,s_2}] = \frac{(l_2-l_1)^2+(s_2-s_1)^2}{2s_1s_2}$ | $\mathrm{Vor}_{D_{\chi^2}}$ hyperbolic Voronoi |
| $\rho_{\mathrm{FR}}[p_{l_1,s_1}, p_{l_2,s_2}] = \frac{1}{\sqrt{2}}\mathrm{arccosh}(1 + D_{\chi^2}[p_{l_1,s_1}, p_{l_2,s_2}])$ | $\mathrm{Vor}_{\rho_{\mathrm{FR}}}$ hyperbolic Voronoi |
| $D_{\mathrm{KL}}[p_{l_1,s_1}, p_{l_2,s_2}] = \log\left(1 + \frac{1}{2}D_{\chi^2}[p_{l_1,s_1}, p_{l_2,s_2}]\right)$ | $\mathrm{Vor}_{D_{\mathrm{KL}}}$ hyperbolic Voronoi |
| $\rho_{\mathrm{KL}}[p_{l_1,s_1}, p_{l_2,s_2}] = \sqrt{D_{\mathrm{KL}}[p_{l_1,s_1}, p_{l_2,s_2}]}$ (metric) | $\mathrm{Vor}_{\rho_{\mathrm{KL}}}$ hyperbolic Voronoi |
| $D_{\mathrm{flat}}[p_{l_1,s_1}, p_{l_2,s_2}] = 2\pi s_2 D_{\chi^2}[p_{l_1,s_1}, p_{l_2,s_2}]$ | Bregman Voronoi: $\mathrm{Vor}_{D_{\mathrm{flat}}}$ hyperbolic Voronoi, $\mathrm{Vor}^*_{D_{\mathrm{flat}}}$ Euclidean Voronoi. |

## 6. Klein Hyperbolic Voronoi Diagram from a Clipped Power Diagram

We concisely recall the efficient construction of the hyperbolic Voronoi diagram in the Klein disk model [25]. Let $\mathcal{P} = \{p_1, \ldots, p_n\}$ be a set of $n$ points in the $d$-dimensional open unit ball domain $\mathbb{D} = \left\{x \in \mathbb{R}^d \ : \ \|x\|_2 < 1\right\}$, where $\|\cdot\|_2$ denotes the Euclidean $\ell_2$-norm. The hyperbolic distance between two points $p$ and $q$ is expressed in the Klein model as follows:

$$\rho_K(p,q) := \mathrm{arccosh}\left(\frac{1 - \langle p, q\rangle}{\sqrt{\left(1 - \|p\|_2^2\right)\left(1 - \|q\|_2^2\right)}}\right). \tag{141}$$

It follows that the *Klein bisector* between any two points in the Klein disk is an hyperplane (affine equation) clipped to $\mathbb{D}$:

$$\mathrm{Bi}_{\rho_K}(\lambda_1 : \lambda_2) = \left\{\lambda \in \mathbb{D} \ : \ \lambda^\top\left(\sqrt{1 - \|\lambda_1\|_2^2}\lambda_2 - \sqrt{1 - \|\lambda_2\|_2^2}\lambda_1\right) + \sqrt{1 - \|\lambda_2\|_2^2} - \sqrt{1 - \|\lambda_1\|_2^2} = 0\right\}. \tag{142}$$

The Klein bisector is a hyperplane (i.e., line in 2D) restricted to the disk domain $\mathbb{D}$. A Voronoi diagram is said to be *affine* [8] when all bisectors are hyperplanes. It is known that *affine Voronoi diagrams* can be constructed from equivalent *power diagrams* [8]. Thus, the Klein hyperbolic Voronoi diagram is equivalent to a *clipped power diagram*:

$$\mathrm{Vor}_{\rho_K}(\mathcal{P}) = \mathrm{Vor}_{D_{\mathrm{PD}}}(\mathcal{S}) \cap \mathbb{D}, \tag{143}$$

where

$$D_{\mathrm{PD}}(\sigma, x) := \|x - c\|^2 - w, \tag{144}$$

denotes the *power "distance"* between a point $x$ (and more generally a *weighted point* [80] when the weight can be negative) to a sphere $\sigma = (c, w)$, and $\mathcal{S} = \{\sigma_1 = (c_1, w_1), \ldots, \sigma_n = (c_n, w_n)\}$ is the equivalent set of weighted points. The power distance is a *signed distance* since we have the following property: $D_{\mathrm{PD}}(\sigma, x) < 0$ iff $x \in \mathrm{int}(\sigma)$, i.e., the point $x$ falls inside the sphere $\sigma = \{x \ : \ \|x - c\|_2^2 = w\}$. The *power bisector* is a hyperplane of equation

$$\mathrm{Bi}_{\mathrm{PD}}(\sigma_i, \sigma_j) = \left\{x \in \mathbb{R}^d \ : \ 2x^\top(c_j - c_i) + w_i - w_j = 0\right\} \tag{145}$$

Notice that by shifting all weights by a predefined constant $a$, we obtain the same power bisector since $(w_i + a) - (w_j + a) = w_i - w_j$ is kept invariant. Thus, we may consider without loss of generality that all weights are non-negative, and that the weighted points correspond to spheres with non-negative radius $r_i^2 = w_i$.

By identifying Equation (142) with Equation (145), we get the following equivalent spheres $\sigma_i = (c_i, w_i)$ [25] for the points in the Klein disk:

$$c_i = \frac{p_i}{2\sqrt{1 - \|p_i\|^2}}, \tag{146}$$

$$w_i = \frac{\|p_i\|_2^2}{4\left(1 - \|p_i\|_2^2\right)} - \frac{1}{\sqrt{1 - \|p_i\|_2^2}}. \tag{147}$$

We can then shift all weights by the constant $a = \min_{i \in \{1,\ldots,n\}} w_i$ so that $w_i' = w_i + a \geq 0$.

Thus, the Klein hyperbolic Voronoi diagram is a *power diagram* clipped to the unit ball $\mathbb{D}$ [80–82]. In computational geometry [4], the power diagram can be calculated from the intersection of $n$ halfspaces by lifting the spheres $\sigma_i$ to corresponding halfspaces $H_i^+$ of $\mathbb{R}^{d+1}$ as follows: Let $\mathcal{F} = \{(x, z) \in \mathbb{R}^{d+1} \ : \ z \geq \sum_{i=1}^d x_i^2\}$ be the epigraph of the paraboloid function, and $\partial\mathcal{F}$ denotes its boundary. We lift a point $x \in \mathbb{R}^d$ to $\partial\mathcal{F}$ using the upper arrow operator $x^\uparrow = (x, z = \sum_{i=1}^d x_i^2)$, and we project orthogonally a point $(x, z)$ of the potential function $\mathcal{F}$ by dropping its last $z$-coordinate so that we have $\downarrow (x^\uparrow) = x$. Now, when we lift a sphere $\sigma = (c, w)$ to $\mathcal{F}$, the set of lifted points $\sigma^\uparrow$ all belong to a hyperplane $H_\sigma$, called the *polar hyperplane* of equation:

$$H_\sigma : z = 2c^\top x - c^\top c + w. \tag{148}$$

Let $H_\sigma^+$ denote the upper halfspace with bounding hyperplane $H_\sigma$: $H_\sigma^+ : z \geq 2c^\top x - c^\top c + w$. Then one can show [4] that $\mathrm{Vor}_{D_{\mathrm{PD}}}(\mathcal{S})$ is obtained as the vertical projection $\downarrow$ of the intersection of all these polar halfspaces $\mathcal{H}_i$ with $\partial\mathcal{F}$:

$$\mathrm{Vor}_{D_{\mathrm{PD}}}(\mathcal{S}) = \downarrow \left( \left( \cap_{i=1}^n \mathcal{H}_i^+ \right) \cap \partial\mathcal{F} \right). \tag{149}$$

Transforming back and forth non-vertical $(d+1)$-dimensional hyperplanes to corresponding $d$-dimensional spheres allows one to design various efficient algorithms, e.g., computing the intersection or the union of spheres [4], useful primitives for molecular chemistry [1].

Let $H_{\mathbb{D}}^-$ denote the lower halfspace (containing the origin $(x = 0, z = 0)$) supported by the polar hyperplane associated with the boundary sphere of the disk domain $\mathbb{D}$. Computing the clipped power diagram $\mathrm{Vor}_{D_{\mathrm{PD}}}(\mathcal{S}) \cap \mathbb{D}$ can be done equivalently as follows:

$$\mathrm{Vor}_{D_{\mathrm{PD}}}(\mathcal{S}) \cap \mathbb{D} = \downarrow \left( \left( \left( \cap_{i=1}^n \mathcal{H}_i^+ \right) \cap \partial\mathcal{F} \right) \cap H_{\mathbb{D}}^- \right), \tag{150}$$

$$= \downarrow \left( \left( \left( \cap_{i=1}^n \mathcal{H}_i^+ \right) \cap H_{\mathbb{D}}^- \right) \cap \partial\mathcal{F} \right), \tag{151}$$

using the commutative property of the set intersection.

The advantage of the method of Equation (151) is that we begin to clip the power diagram using $H_{\mathbb{D}}^-$ before explicitly calculating it. Indeed, we first compute the *intersection polytope* of $n+1$ hyperplanes $\mathcal{P}_K := \left( \cap_{i=1}^n \mathcal{H}_i^+ \right) \cap H_{\mathbb{D}}^-$. Then we project down orthogonally the intersection of $\mathcal{P}_K$ with $\partial\mathcal{F}$ to get the clipped power diagram equivalent to the hyperbolic Klein Voronoi diagram:

$$\mathrm{Vor}_{\rho_K}(\mathcal{P}) = \downarrow \left( \mathcal{P}_K \cap \partial\mathcal{F} \right). \tag{152}$$

By doing so, we potentially reduce the algorithmic complexity by avoiding to compute some of the vertices of $\mathcal{P}_{\mathrm{PD}} := \left( \cap_{i=1}^n \mathcal{H}_i^+ \right)$ whose orthogonal projection fall outside the domain $\mathbb{D}$.

More generally, a Bregman Voronoi diagram [15] can be calculated equivalently as a power diagram (and intersection of $d+1$-dimensional halfspaces) using an arbitrary smooth and strictly convex potential function $F$ instead of the the paraboloid potential function of Euclidean geometry [25]. The non-empty intersection of halfspaces can in turn be calculated as an equivalent *convex hull* [4]. Thus,

we can compute in practice the hyperbolic Voronoi diagram in the Klein model using the Quickhull algorithm [83].

## 7. Symbolic Calculations with a Computer Algebra System

We use the open source computer algebra system MAXIMA(can be freely downloaded at http://maxima.sourceforge.net/) to calculate the gradient (partial derivatives) and Hessian of the deformed log-normalizer, and some definite integrals based on the Cauchy location-scale densities.

```
/* Written in Maxima */
assume(s>0);
CauchyStd(x) := (1/(%pi*(x**2+1)));
Cauchy(x,l,s) := (s/(%pi*((x-l)**2+s**2)));
/* check that we get a probability density (=1) */
integrate(Cauchy(x,l,s),x,-inf,inf);
/* calculate the the deformed log-normalizer */
logC(u):=1-(1/u);
logC(Cauchy(x,l,s));
ratsimp(%);
/* calculate partial derivatives of the deformed log-normalizer */
theta(l,s):=[2*%pi*l/s,-%pi/s];
F(theta):=(-%pi**2/theta[2])-(theta[1]**2/(4*theta[2]))-1;
derivative(F(theta),theta[1],1);
derivative(F(theta),theta[2],1);
/* calculated definite integrals */
assume(s1>0);
assume(s2>0);
integrate(Cauchy(x,l2,s2)**2,x,-inf,inf);
integrate(Cauchy(x,l2,s2)**2/Cauchy(x,l1,s1),x,-inf,inf);
```

We calculate the function $\theta(\eta)$ by solving the following system of equations:

```
solve([-t1/(2*t2)=e1, (%pi/t2)**2+ (t1/t2)**2/4=e2],[t1, t2]);
```

The Hessian metrics of the dual potential functions $F$ and $F^*$ (denoted by $G$ in the code) can be calculated as follows:

```
F(theta):=(-%pi**2/theta[2])-(theta[1]**2/(4*theta[2]))-1;
hessian(F(theta),[theta[1], theta[2]]);
G(eta):=1-2*%pi*sqrt(eta[2]-eta[1]**2);
hessian(G(eta),[eta[1], eta[2]]);
```

The plot of the Fisher-Rao to the square root KL divergence can be plotted using the following commands:

```
t(u):=sqrt(log((1/2)+(1/2)*cosh(sqrt(2)*u)));
plot2d(t(u)/u,[u,0,10]);
```

Symbolic calculations for the $\alpha$-Chernoff coefficient between two Cauchy distributions prove that the $\alpha$-Chernoff coefficient is symmetric for $\alpha = 3$ and $\alpha = 4$ as exemplified by the MAXIMA code below:

```
assume(s1>0);
assume(s2>0);
assume(s>0);
CauchyStd(x) := (1/(%pi*(x**2+1)));
```

```
Cauchy(x,l,s) := (s/(%pi*((x-l)**2+s**2)));
/* closed-form */
a: 3;
integrate((Cauchy(x,l2,s2)**a) * (Cauchy(x,l1,s1)**(1-a)),x,-inf,inf);
term1(l1,s1,l2,s2):=ratsimp(%);
integrate((Cauchy(x,l2,s2)**(1-a)) * (Cauchy(x,l1,s1)**(a)),x,-inf,inf);
term2(l1,s1,l2,s2):=ratsimp(%);
/* Is the a-divergence symmetric? */
term1(l1,s1,l2,s2)-term2(l1,s1,l2,s2);
ratsimp(%);
```

**Funding:** This research received no external funding.

**Conflicts of Interest:** The author declares no conflict of interest.

## References


1. Okabe, A.; Boots, B.; Sugihara, K.; Chiu, S.N. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*; John Wiley & Sons: Hoboken, NJ, USA, 2009; Volume 501.
2. Aurenhammer, F. Voronoi diagrams: A survey of a fundamental geometric data structure. *ACM Comput. Surv. (CSUR)* **1991**, *23*, 345–405. [CrossRef]
3. Cheng, S.W.; Dey, T.K.; Shewchuk, J. *Delaunay Mesh Generation*; CRC Press: Boca Raton, FL, USA, 2012.
4. Boissonnat, J.D.; Yvinec, M. *Algorithmic Geometry*; Cambridge University Press: Cambridge, UK, 1998.
5. Bogdanov, M.; Devillers, O.; Teillaud, M. Hyperbolic Delaunay complexes and Voronoi diagrams made practical. In Proceedings of the twenty-ninth Annual Symposium on Computational Geometry, Rio de Janeiro, Brazil, 17–20 June 2013; pp. 67–76.
6. Zhang, J. Divergence function, duality, and convex analysis. *Neural Comput.* **2004**, *16*, 159–195. [CrossRef] [PubMed]
7. Amari, S.i. *Information Geometry and Its Applications*; Springer: Berlin, Germany, 2016; Volume 194.
8. Boissonnat, J.D.; Wormser, C.; Yvinec, M. Curved Voronoi diagrams. In *Effective Computational Geometry for Curves and Surfaces*; Springer: Berlin, Germany, 2006; pp. 67–116.
9. Bregman, L.M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* **1967**, *7*, 200–217. [CrossRef]
10. Calin, O.; Udrişte, C. *Geometric Modeling in Probability and Statistics*; Springer: Berlin, Germany, 2014.
11. Nielsen, F. An elementary introduction to information geometry. *arXiv* **2018**, arXiv:1808.08271.
12. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
13. Eguchi, S. Geometry of minimum contrast. *Hiroshima Math. J.* **1992**, *22*, 631–647. [CrossRef]
14. Nielsen, F.; Hadjeres, G. Monte Carlo information geometry: The dually flat case. *arXiv* **2018**, arXiv:1803.07225.
15. Boissonnat, J.D.; Nielsen, F.; Nock, R. Bregman Voronoi diagrams. *Discrete Comput. Geom.* **2010**, *44*, 281–307. [CrossRef]
16. Nielsen, F.; Boissonnat, J.D.; Nock, R. Visualizing Bregman voronoi diagrams. In Proceedings of the twenty-third annual symposium on Computational geometry (SoCG), Gyeongju, Korea, 6–8 June 2007; pp. 121–122.
17. Rao, C.R. Information and the Accuracy Attainable in the Estimation of Statistical Parameters. *Bull. Cal. Math. Soc.* **1945**, *37*, 81–91.
18. Atkinson, C.; Mitchell, A.F. Rao's distance measure. *Sankhyā The Indian J. Stat. Series A* **1981**, *43*, 345–365.
19. Pinele, J.; Strapasson, J.E.; Costa, S.I. The Fisher-Rao Distance between Multivariate Normal Distributions: Special Cases, Bounds and Applications. *Entropy* **2020**, *22*, 404. [CrossRef]
20. Nielsen, F.; Nock, R. On the chi square and higher-order chi distances for approximating $f$-divergences. *IEEE Signal Process. Lett.* **2013**, *21*, 10–13. [CrossRef]
21. Naudts, J. The $q$-exponential family in statistical physics. *Cent. Eur. J. Phys.* **2009**, *7*, 405–413. [CrossRef]

22. Matsuzoe, H.; Henmi, M. Hessian structures and divergence functions on deformed exponential families. In *Geometric Theory of Information*; Springer: Berlin, Germany, 2014; pp. 57–80.
23. Tsallis, C. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [CrossRef]
24. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*; Springer Science & Business Media: Berlin, Germany, 2009.
25. Nielsen, F.; Nock, R. Hyperbolic Voronoi diagrams made easy. In Proceedings of the 2010 International Conference on Computational Science and Its Applications, Fukuoka, Japan, 23–26 March 2010; pp. 74–80.
26. Nielsen, F.; Nock, R. Visualizing hyperbolic Voronoi diagrams. In Proceedings of the thirtieth annual symposium on Computational geometry, Kyoto, Japan, 8–11 June 2014; pp. 90–91.
27. Murray, M.K.; Rice, J.W. *Differential Geometry and Statistics*; CRC Press: Boca Raton, FL, USA, 1993; Volume 48.
28. Mitchell, A.F. Statistical manifolds of univariate elliptic distributions. *Int. Stat. Rev.* **1988**, *56*, 1–16. [CrossRef]
29. Hotelling, H. Spaces of statistical parameters. *Bull. Am. Math. Soc* **1930**, *36*, 191.
30. Anderson, J.W. *Hyperbolic Geometry*; Springer Science & Business Media: Berlin, Germany, 2006.
31. Nielsen, F.; Nock, R. Total Jensen divergences: Definition, properties and $k$-means++ clustering. *arXiv* **2013**, arXiv:1309.7109.
32. Nielsen, F.; Nock, R. Total Jensen divergences: Definition, properties and clustering. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Queensland, Australia, 19–24 April 2015; pp. 2016–2020.
33. Nock, R.; Nielsen, F.; Amari, S.i. On conformal divergences and their population minimizers. *IEEE Trans. Inf. Theory* **2015**, *62*, 527–538. [CrossRef]
34. Chen, B.Y. *Differential Geometry of Warped Product Manifolds And Submanifolds*; World Scientific Singapore: Singapore, 2017.
35. Said, S.; Bombrun, L.; Berthoumieu, Y. Warped Riemannian metrics for location-scale models. In *Geometric Structures of Information*; Springer: Berlin, Germany, 2019; pp. 251–296.
36. Li, W.; Zhao, J. Wasserstein information matrix. *arXiv* **2019**, arXiv:1910.11248.
37. Takatsu, A. Wasserstein geometry of Gaussian measures. *Osaka J. Math.* **2011**, *48*, 1005–1026.
38. Lauritzen, S.L. Statistical manifolds. *Differ. Geom. Stat. Inference* **1987**, *10*, 163–216.
39. Shima, H. *The Geometry of Hessian Structures*; World Scientific: Singapore, 2007.
40. Nielsen, F.; Nock, R. On the smallest enclosing information disk. *Inf. Process. Lett.* **2008**, *105*, 93–97. [CrossRef]
41. Matumoto, T. Any statistical manifold has a contrast function—on the $C^3$-functions taking the minimum at the diagonal of the product manifold. *Hiroshima Math. J.* **1993**, *23*, 327–332. [CrossRef]
42. Naudts, J. *Generalised Thermostatistics*; Springer Science & Business Media: Berlin, Germany, 2011.
43. Vigelis, R.F.; Cavalcante, C.C. On $\varphi$-families of probability distributions. *J. Theor. Probab.* **2013**, *26*, 870–884. [CrossRef]
44. Tanaya, D.; Tanaka, M.; Matsuzoe, H. Notes on geometry of $q$-normal distributions. In *Recent Progress in Differential Geometry and Its Related Fields*; World Scientific: Singapore, 2012; pp. 137–149.
45. Amari, S.i.; Ohara, A.; Matsuzoe, H. Geometry of deformed exponential families: Invariant, dually-flat and conformal geometries. *Physi. A Stat. Mech. Its Appl.* **2012**, *391*, 4308–4319. [CrossRef]
46. Onicescu, O. Théorie de l'information énergie informationelle. *Comptes rendus de l'Academie des Sci. Series AB* **1966**, *263*, 841–842.
47. Nielsen, F. A note on Onicescu's informational energy and correlation coefficient in exponential families. *arXiv* **2020**, arXiv:2003.13199.
48. Crouzeix, J.P. A relationship between the second derivatives of a convex function and of its conjugate. *Math. Program.* **1977**, *13*, 364–365. [CrossRef]
49. Ohara, A. Conformal Flattening on the Probability Simplex and Its Applications to Voronoi Partitions and Centroids. In *Geometric Structures of Information*; Springer: Berlin, Germany, 2019; pp. 51–68.
50. Zhang, J. A note on curvature of $\alpha$-connections of a statistical manifold. *Ann. Inst. Stat. Math.* **2007**, *59*, 161–170. [CrossRef]
51. Amari, S.i.; Cichocki, A. Information geometry of divergence functions. *Bull. Pol. Acad. Sci. Tech. Sci.* **2010**, *58*, 183–195. [CrossRef]
52. Matsuzoe, H. Hessian structures on deformed exponential families and their conformal structures. *Differ. Geom. Its Appl.* **2014**, *35*, 323–333. [CrossRef]

53. Csiszár, I. Information-type measures of difference of probability distributions and indirect observation. *Studia Sci. Math. Hungar* **1967**, *2*, 229–318.
54. Schwander, O.; Nielsen, F. Non-flat clustering with alpha-divergences. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; pp. 2100–2103.
55. Nielsen, F.; Sun, K. Combinatorial bounds on the $\alpha$-divergence of univariate mixture models. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 4476–4480.
56. Chernoff, H. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Stat.* **1952**, *23*, 493–507. [CrossRef]
57. Nielsen, F. An information-geometric characterization of Chernoff information. *IEEE Signal Process. Lett.* **2013**, *20*, 269–272. [CrossRef]
58. Chyzak, F.; Nielsen, F. A closed-form formula for the Kullback–Leibler divergence between Cauchy distributions. *arXiv* **2019**, arXiv:1905.10965.
59. Uhlmann, J.K. Metric trees. *Appl. Math. Lett.* **1991**, *4*, 61–62. [CrossRef]
60. Yianilos, P.N. Data structures and algorithms for nearest neighbor seach in general metric spaces. In Proceedings of the Symposium on Discrete Algorithms (SODA), Austin, TX, USA, 25–27 January 1993; pp. 311–321.
61. Nielsen, F.; Piro, P.; Barlaud, M. Bregman vantage point trees for efficient nearest neighbor queries. In Proceedings of the IEEE International Conference on Multimedia and Expo, Cancun, Mexico, 28 June–3 July 2009; pp. 878–881.
62. Fuglede, B.; Topsoe, F. Jensen-Shannon divergence and Hilbert space embedding. In Proceedings of the International Symposium on Information Theory (ISIT), Chicago, IL, USA, 27 June–2 July 2004; p. 31.
63. Kafka, P.; Österreicher, F.; Vincze, I. On powers of $f$-divergences defining a distance. *Studia Sci. Math. Hungar* **1991**, *26*, 415–422.
64. Vajda, I. On metric divergences of probability measures. *Kybernetika* **2009**, *45*, 885–900.
65. Duin, R.P.W.; Elzbieta, P. *Dissimilarity Representation for Pattern Recognition: The Foundations and Applications*; World Scientific: Singapore, 2005; Volume 64.
66. Acharyya, S.; Banerjee, A.; Boley, D. Bregman divergences and triangle inequality. In Proceedings of the SIAM International Conference on Data Mining, San Diego, CA, USA, 8–12 July 2013; pp. 476–484.
67. Nielsen, F.; Boltz, S. The Burbea-Rao and Bhattacharyya centroids. *IEEE Trans. Inf. Theory* **2011**, *57*, 5455–5466. [CrossRef]
68. Chen, P.; Chen, Y.; Rao, M. Metrics defined by Bregman divergences: Part 2. *Commun. Math. Sci.* **2008**, *6*, 927–948. [CrossRef]
69. Sra, S. Positive definite matrices and the *S*-divergence. *Proc. Am. Math. Soc.* **2016**, *144*, 2787–2797. [CrossRef]
70. Nielsen, F.; Yvinec, M. An output-sensitive convex hull algorithm for planar objects. *Int. J. Comput. Geom. Appl.* **1998**, *8*, 39–65. [CrossRef]
71. Tanuma, T.; Imai, H.; Moriyama, S. Revisiting hyperbolic Voronoi diagrams in two and higher dimensions from theoretical, applied and generalized viewpoints. In *Transactions on Computational Science XIV*; Springer: Berlin, Germany, 2011; pp. 1–30.
72. DeBlois, J. The Delaunay tessellation in hyperbolic space. *Math. Proc. Camb. Philos. Soc.* **2018**, *164*, 15–46. [CrossRef]
73. Sarkar, R. Low distortion Delaunay embedding of trees in hyperbolic plane. In *International Symposium on Graph Drawing*; Springer: Berlin, Germany, 2011; pp. 355–366.
74. Nickel, M.; Kiela, D. Poincaré embeddings for learning hierarchical representations. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6338–6347.
75. Haas, G.; Bain, L.; Antle, C. Inferences for the Cauchy distribution based on maximum likelihood estimators. *Biometrika* **1970**, *57*, 403–408.
76. Giménez, P.; López, J.N.; Guarracino, L. Geodesic Hypothesis Testing for Comparing Location Parameters in Elliptical Populations. *Sankhya A* **2016**, *78*, 19–42. [CrossRef]
77. Delaunay, B. Sur la sphère vide. *Izv. Akad. Nauk SSSR, Otdelenie Matematicheskii i Estestvennyka Nauk* **1934**, *7*, 1–2.

78. Risch, R.H. The solution of the problem of integration in finite terms. *Bull. Am. Math. Soc.* **1970**, *76*, 605–608. [CrossRef]
79. Komaki, F. Bayesian prediction based on a class of shrinkage priors for location-scale models. *Ann. Inst. Stat. Math.* **2007**, *59*, 135–146. [CrossRef]
80. Boissonnat, J.D.; Delage, C. Convex hull and Voronoi diagram of additively weighted points. In *European Symposium on Algorithms*; Springer: Berlin, Germany, 2005; pp. 367–378.
81. Nielsen, F. Grouping and querying: A paradigm to get output-sensitive algorithms. In *Japanese Conference on Discrete and Computational Geometry*; Springer: Berlin, Germany, 1998; pp. 250–257.
82. Yan, D.M.; Wang, W.; LéVy, B.; Liu, Y. Efficient computation of clipped Voronoi diagram for mesh generation. *Comput.-Aided Des.* **2013**, *45*, 843–852. [CrossRef]
83. Barber, C.B.; Dobkin, D.P.; Huhdanpaa, H. The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw. (TOMS)* **1996**, *22*, 469–483. [CrossRef]