MEDIANS AND MEANS IN FINSLER GEOMETRY

MARC ARNAUDON AND FRANK NIELSEN

Abstract. We investigate existence and uniqueness of p-means e_p and the median e_1 of a probability measure μ on a Finsler manifold, in relation with the convexity of the support of μ. We prove that e_p is the limit point of a continuous time gradient flow. Under some additional condition which is always satisfied for $p \geq 2$, a discretization of this path converges to e_p. This provides an algorithm for determining those Finsler center points.

Contents

1. Introduction

The geometric barycenter of a set of points is the point which minimizes the sum of the *squared* distances to these points. It is the most traditional estimator is statistics that is however sensitive to outliers [17]. Thus it is natural to replace the average distance squaring (power 2) by taking the power of p for some $p \in [1, 2)$. This leads to the definition of p-means. When $p = 1$, the minimizer is the median of the set of points, very often used in robust statistics [17]. In many applications, p-means with some $p \in (1, 2)$ give the best compromise. For existence and uniqueness in Riemannian manifolds under convexity conditions on the support of the measure, see Afsari [1].

The Fermat-Weber problem concerns finding the median e_1 of a set of points in an Euclidean space. Numerous authors worked out algorithms for computing e_1. The first algorithm was proposed by Weiszfeld in [34] (see also [33]). It has been extended to sufficiently small domains in Riemannian manifolds with nonnegative curvature by Fletcher and al. in [13]. A complete generalization to manifolds with positive or negative curvature (under some convexity conditions in positive curvature), has been recently given by Yang in [36].

The Riemannian barycenter or Karcher mean of a set of points in a manifold or more generally of a probability measure has been extensively studied, see e.g. [18], [19], [20], [11], [30], [4], [9], where questions of existence, uniqueness, stability, relation with martingales in manifolds, behavior when measures are pushed by stochastic flows have been considered. The Riemannian barycenter corresponds to $p = 2$ in the above description. Computation of Riemannian barycenters by gradient descent has been performed by Le in [22].

The aim of this paper is to extend to the context of Finsler manifolds the results on existence and uniqueness of p-means of probability measures, as well as algorithms for computing them. Some convexity is needed, and as we shall see the fact that comparison results for triangles as Alexandroff and Toponogov theorems do not exist impose more restrictions on the support of the probability measure. As a consequence, the sharp results on existence and uniqueness established by Afsari [1] and the algorithm for computing means of Yang in [36] do not extend to Finsler manifolds.

The motivation for this work primarily comes from signal filtering and denoising in the context of Diffusion Tensor Imaging (DTI), High Angular Resolution Imaging (HARDI, see [32], [28], [5]), Orientation Distribution Function (ODF), active contours [24]. Applications with experimental results of an implementation will be reported in forthcoming papers.

Information geometry at its heart considers the differential geometry nature of probability distributions induced by a divergence function. In probability theory, invariance by monotonic re-parameterization and sufficient statistics yields the class of f-divergences [2] $I_f(p,q) = \int p(x) f(\frac{q(x)}{p(x)})\mathrm{d}x$ that includes the Kullback-Leibler (KL) information-theoretic divergence $\mathrm{KL}(p,q) = \int p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x$ as its prominent member (for $f(t) = -\log t$). It is well-known that the KL divergence (better known as the relative entropy) yields a dually flat structure [2] generalizing the (self-dual) Euclidean space.

Because divergences are usually asymmetric and violate the triangle inequality they have not been extensively considered from an *algorithmic* point of view. Indeed, the triangle inequality property is often used in computational geometry to design *efficient* algorithms by allowing various "pruning" techniques [10, 23]. Computational geometry has thus mostly considered metric spaces for keeping the triangle inequality properties.

One can metrize divergences. The KL divergence can be symmetrized either into the Jeffreys divergence $J(p,q) = \mathrm{KL}(p,q) + \mathrm{KL}(q,p)$ or the Jensen-Shannon (JS) divergence:

$$\begin{aligned}\mathrm{JS}(p,q) &= \mathrm{KL}(p,\frac{p+q}{2}) + \mathrm{KL}(q,\frac{p+q}{2})\\ &= \int (p(x)\log\frac{2p(x)}{p(x)+q(x)} + q(x)\log\frac{2q(x)}{p(x)+q(x)})\mathrm{d}x.\end{aligned}$$

The latter is preferred in practice because it is bounded, and its square root yields a metric that can be embedded into a Hilbert space [14].

Finsler distances, arising from the underlying the Finsler metrics, are attractive as they preserve the triangle inequality [32] for efficient algorithmics but potentially model asymmetric distances.

In information geometry, the regular divergence D associated to a Finslerian metric distance d can be defined as $D(p,q) = d^2(p,q)$. Observe that the Finslerian-based divergence looses then the triangle inequality property [32]. (eg., the squared Euclidean distance does not satisfy the triangle inequality).

2. Preliminaries

Let M be a smooth manifold. On M we consider a Finsler structure $F : TM \to \mathbb{R}_+$. For any $x \in M$, $V, X, Y, Z \in T_xM$ such that $V \neq 0$, let

$$g_V(X,Y) := \frac{1}{2}\frac{\partial^2}{\partial s \partial t}\Big|_{(s,t)=(0,0)} F^2(V + sX + tY). \tag{2.1}$$

(we shall also use the notation $<X,Y>_V = g_V(X,Y)$) and

$$<X,Y,Z>_V := \frac{1}{4}\frac{\partial^3}{\partial r \partial s \partial t}\Big|_{(r,s,t)=(0,0,0)} F^2(V + rX + sY + tZ). \tag{2.2}$$

We have

$$<X,Y,Z>_V = \frac{1}{2}\frac{\partial}{\partial r}\Big|_{r=0} g_{V+rX}(Y,Z) \tag{2.3}$$

and in particular since F^2 is 2-homogeneous and $V \mapsto g_V(X,Y)$ is 0-homogeneous,

$$<V,Y,Z>_V = 0. \tag{2.4}$$

Let V be a non-vanishing vector field on M. The Chern connection ∇^V is torsionfree and almost metric, and can be characterized by

$$X<Y,Z>_V = <\nabla^V_X Y, Z>_V + <Y, \nabla^V_X Z>_V + 2<\nabla^V_X V, Y, Z>_V. \tag{2.5}$$

More precisely, parameterizing locally TM by coordinates

$$(x^1, \dots, x^m, y^1 = dx^1, \dots, y^m = dx^m),$$

defining the geodesic coefficients as

$$G^i(y) = \frac{1}{4} g^{ik}(y)\left(2\frac{\partial g_{jk}}{\partial x^l} - \frac{\partial g_{jl}}{\partial x^k}\right) y^j y^l, \quad y \in TM \backslash \{0\}, \tag{2.6}$$

letting

$$N^i_j = \frac{\partial G^i}{\partial y^j}, \quad \frac{\delta}{\delta x^i} = \frac{\partial}{\partial x^i} - N^k_i(y)\frac{\partial}{\partial y^k} \in T_y\left(TM\backslash\{0\}\right), \tag{2.7}$$

then the Christoffel symbols of the Chern connection are given by

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl}\left(\frac{\delta g_{lj}}{\delta x^i} + \frac{\delta g_{il}}{\delta x^j} - \frac{\delta g_{ij}}{\delta x^l}\right) \tag{2.8}$$

(see [8]). Note that defining

$$\delta y^i = dy^i + N^i_j(y) dx^j \tag{2.9}$$

we have for a smooth function $f : TM\backslash\{0\} \to \mathbb{R}$

$$df = \frac{\delta f}{\delta x^i} dx^i + \frac{\partial f}{\partial y^i}\delta y^i. \tag{2.10}$$

The Chern curvature tensor is defined by the equation

$$R^V(X,Y)Z := \nabla^V_X \nabla^V_Y Z - \nabla^V_Y \nabla^V_X Z - \nabla^V_{[X,Y]} Z, \tag{2.11}$$

and the flag curvature is

$$\mathscr{K}(V,W) := \frac{<R^V(V,W)W, V>}{<V,V>_V <W,W>_V - <V,W>_V^2}, \tag{2.12}$$

for two non collinear $V, W \in T_xM$.

We say that M has nonpositive flag curvature if for all V, W, $\mathscr{K}(V,W) \leq 0$.

The tangent curvature of two vectors $V, W \in T_xM$ is defined as

$$\mathscr{T}_V(W) = <\nabla_W^W \tilde{W} - \nabla_W^V \tilde{W}, V>_V \tag{2.13}$$

where $\tilde{W}$ is a vector field satisfying $\tilde{W}_x = W$. For a nonnegative constant $\delta \geq 0$ we say that $\mathscr{T} \geq -\delta$ or $\mathscr{T} \leq \delta$ if respectively

$$\mathscr{T}_V(W) \geq -\delta F(V)F(W)^2 \quad \text{or} \quad \mathscr{T}_V(W) \leq \delta F(V)F(W)^2. \tag{2.14}$$

For $x \in M$ we define

$$\mathscr{C}(x) = \sup_{v,w \in T_xM\backslash\{0\}} \sqrt{\frac{<v,v>_v}{<v,v>_w}}, \quad \mathscr{D}(x) = \sup_{v,w \in T_xM\backslash\{0\}} \sqrt{\frac{<v,v>_w}{<v,v>_v}}. \tag{2.15}$$

Remark 2.1. For applications in active contours, a "Wulff shape" is given which does not depend on x and defines the Finsler structure. From this shape $\mathscr{C}$ and $\mathscr{D}$ can easily be calculated. See e.g [24] and [37].

A geodesic in M is a curve $t \mapsto c(t)$ satisfying for all t, $\nabla_{\dot{c}(t)}^{\dot{c}(t)} \dot{c} = 0$. It is well known that a geodesic has constant speed, and that it locally minimizes the distance ([8]). If so, letting $\rho(x,y)$ the forward distance from x to y, then

$$\rho^2(x,y) = <\dot{c}(0), \dot{c}(0)>_{\dot{c}(0)} \tag{2.16}$$

where $t \mapsto c(t)$ is the minimal geodesic satisfying $c(0) = x$ and $c(1) = y$. By definition, the backward distance from x to y is $\rho(y,x)$.

For $x \in M$ and $v \in T_xM$, we let whenever it exists $\exp_x(v) := c(1)$ where $t \mapsto c(t)$ is the geodesic satisfying $\dot{c}(0) = v$.

If M is complete, analytic, simply connected and has nonpositive flag curvature (we say that M is an analytic Cartan-Hadamard manifold), then $\exp_x : T_xM \to M$ is an homeomorphism ([6] theorem 4.7). Under these assumption, letting for $x, y \in M$, $\overrightarrow{xy} = \exp_x^{-1}(y)$, we have

$$\rho^2(x,y) = <\overrightarrow{xy}, \overrightarrow{xy}>_{\overrightarrow{xy}}. \tag{2.17}$$

For $x_0 \in M$ and $R > 0$,, let us denote by $B(x_0, R)$ (resp. $\bar{B}(x_0, R)$) the (forward) open (resp. closed) ball with center x_0 and radius R:

$$B(x_0, R) = \{y \in M,\ \rho(x_0, y) < R\} \quad (\text{resp.} \quad \bar{B}(x_0, R) = \{y \in M,\ \rho(x_0, y) \leq R\}) \tag{2.18}$$

Now let $(t,s) \mapsto c(t,s)$ a family of minimizing geodesics $t \mapsto c(t,s)$, $t \in [0,1]$, parametrized by $s \in I$, I an interval in $\mathbb{R}$. Define

$$E(s) = \frac{1}{2}\rho^2(c(0,s), c(1,s)). \tag{2.19}$$

The computation of $E'(s)$ and $E''(s)$ is well-known, see e.g. [7]. We have

$$E'(s) = <\partial_s c(1,s), \partial_t c(1,s)>_{\partial_t c(1,s)} - <\partial_s c(0,s), \partial_t c(0,s)>_{\partial_t c(0,s)}. \tag{2.20}$$

As for the second derivative, letting $c = c(\cdot, 0)$, and for X, Y vector fields along c

$$I(X,Y) = \int_0^1 \left(<\nabla_T^T X, \nabla_T^T Y>_T - <R^T(X,T)T, Y>_T\right) dt \tag{2.21}$$

the index of X and Y, we have

$$\begin{aligned} E''(0) = {} & <\nabla_{\partial_s c(1,0)}^{\partial_t c(1,0)} \partial_s c(1,\cdot), \partial_t c(1,0)>_{\partial_t c(1,0)} - <\nabla_{\partial_s c(0,0)}^{\partial_t c(0,0)} \partial_s c(0,\cdot), \partial_t c(0,0)>_{\partial_t c(0,0)} \\ & + I(\partial_s c(\cdot,0), \partial_s c(\cdot,0)). \end{aligned} \tag{2.22}$$

Assuming $s \mapsto c(0,s)$ and $s \mapsto c(1,s)$ are geodesics, we obtain

$$E''(0) = \mathscr{T}_{\partial_t c(0,0)}(\partial_s c(0,0)) - \mathscr{T}_{\partial_t c(1,0)}(\partial_s c(1,0)) + I(\partial_s c(\cdot,0), \partial_s c(\cdot,0)). \tag{2.23}$$

We are interested in the situation where $c(1,s) \equiv z$ a constant. In this case we have

$$E''(0) = \mathscr{T}_{\partial_t c(0,0)}(\partial_s c(0,0)) + I(\partial_s c(\cdot,0), \partial_s c(\cdot,0)). \tag{2.24}$$

For $p \geq 1$, define

$$D_p(s) = \rho^p(c(0,s), z) \tag{2.25}$$

Proposition 2.2. *Assume $\mathscr{K} \leq k$, $\mathscr{T} \geq -\delta$, $\mathscr{C} \leq C$ for some $k, \delta \geq 0$, $C \geq 1$. Let $p > 1$. Then writing $r = \rho(x,z)$,*

$$D_p''(0) \geq p r^{p-2} \left(\min\left(p-1, \frac{\sqrt{k} r \cos(\sqrt{k} r)}{\sin(\sqrt{k} r)} \right) C^{-2} - \delta r \right). \tag{2.26}$$

If z and $x = c(0,0)$ satisfy $\rho(x,z) < R(p,k,\delta,C)$ with

$$R(p,k,\delta,C) = \min\left(\frac{p-1}{C^2 \delta}, \frac{1}{\sqrt{k}} \arctan\left(\frac{\sqrt{k}}{C^2 \delta} \right) \right) \tag{2.27}$$

and the injectivity radius at x is strictly larger than $R(p,k,\delta,C)$, then $D_p''(0) > 0$.

Remark 2.3. Note if $p \geq 2$ then

$$R(p,k,\delta,C) = R(2,k,\delta,C) = \frac{1}{\sqrt{k}} \arctan\left(\frac{\sqrt{k}}{C^2 \delta} \right).$$

Proof. Define $T(t) = \partial_t c(t,0)$, $J(t) = \partial_s c(t,0)$,

$$J^T(t) = \frac{1}{F(T(t))^2} {<} J(t), T(t) {>}_{T(t)} T(t), \quad J^N(t) = J(t) - J^T(t).$$

Using successively [7] Lemma 9.5.1 which compares the index $I(J,J)$ with the one of its "transplant" into a manifold with constant curvature k^2 and the index lemma [7] Lemma 7.3.2 which compares the index of the transplant to the one of the Jacobi field with same boundary values, we get, letting $r = \rho(x,z) = D_1(0)$,

$$I(J,J) \geq \frac{\sqrt{k} r \cos(\sqrt{k} r)}{\sin(\sqrt{k} r)} {<} J^N(0), J^N(0) {>}_{T(0)} + {<} J^T(0), J^T(0) {>}_{T(0)}. \tag{2.28}$$

Using the expression (2.24) for $E''(0)$ we obtain

$$E''(0) \geq -\delta r + \frac{\sqrt{k} r \cos(\sqrt{k} r)}{\sin(\sqrt{k} r)} {<} J^N(0), J^N(0) {>}_{T(0)} + {<} J^T(0), J^T(0) {>}_{T(0)}. \tag{2.29}$$

We have from (2.20)

$$E'(0)^2 = r^2 {<} J^T(0), J^T(0) {>}_{T(0)}. \tag{2.30}$$

Now from $D_1(s) = \sqrt{2E(s)}$ we get

$$D_1'(s) = \frac{E'(s)}{D_1(s)}, \quad D_1''(s) = \frac{E''(s)}{D_1(s)} - \frac{E'(s)^2}{D_1^3(s)}, \tag{2.31}$$

and this yields

$$
\begin{aligned}
&D_p''(0)\\
&= pD_1(0)^{p-2}\left((p-1)D_1'(0)^2 + D_1(0)D_1''(0)\right)\\
&= pr^{p-2}\left((p-2)<J^T(0), J^T(0)>_{T(0)} + E''(0)\right)\\
&\ge pr^{p-2}\left((p-1)<J^T(0), J^T(0)>_{T(0)} - \delta r + \frac{\sqrt{k}r\cos(\sqrt{k}r)}{\sin(\sqrt{k}r)}<J^N(0), J^N(0)>_{T(0)}\right)\\
&\ge pr^{p-2}\left(\min\left(p-1, \frac{\sqrt{k}r\cos(\sqrt{k}r)}{\sin(\sqrt{k}r)}\right)<J(0), J(0)>_{T(0)} - \delta r\right)\\
&\ge pr^{p-2}\left(\min\left(p-1, \frac{\sqrt{k}r\cos(\sqrt{k}r)}{\sin(\sqrt{k}r)}\right)C^{-2} - \delta r\right).
\end{aligned}
$$

¿From this bound the rest of the proof follows easily. □

Similarly, we can obtain an upper bound for $D_p''(0)$:

Proposition 2.4. *Assume the sectional curvatures $\mathscr{K}$ have a lower bound $-\beta^2$ for some $\beta > 0$, and $\mathscr{T} \le \delta'$ for some $\delta' > 0$, $\mathscr{D} \le D$ for some $D \ge 1$. Again let $r = \rho(x, z)$, assume that the injectivity radius at x is larger than r. Then*

$$D_p''(0) \le pr^{p-2}\left(D^2 \max\left(p-1, \beta r\coth(\beta r)\right) + \delta' r\right). \tag{2.32}$$

Proof. We have by (2.24) and (2.31) together with the fact that

$$I(J, J) = <J^T(0), J^T(0)> + I(J^N, J^N), \tag{2.33}$$

$$D_p''(0) = pr^{p-2}\left(p-1)<J^T(0), J^T(0)>_{T(0)} + I(J^N, J^N) + \mathscr{T}_T(J)\right).$$

Let $t \mapsto X(t)$ the parallel vector field along $t \mapsto c(t, 0)$ with initial condition $J^N(0)$, and for $t \in [0, 1]$, let

$$G(t) = \cosh(r\beta t) - \coth(r\beta)\sinh(r\beta t).$$

This is the solution of $G'' = r\beta G$ with conditions $G(0) = 1$ and $G(1) = 0$. The vector field $t \mapsto Y(t)$ along $t \mapsto c(t, 0)$ defined by

$$Y(t) = G(t)X(t) \tag{2.34}$$

has same boundary values as $t \mapsto J^N(t)$, so by the index lemma [7] Lemma 7.3.2 we have

$$I(J^N, J^N) \le I(Y, Y). \tag{2.35}$$

On the other hand

$$
\begin{aligned}
&I(Y, Y)\\
&= \int_0^1 \left(G'(t)^2<J^N(0), J^N(0)>_{T(0)} - G(t)^2<R^T(X(t), T(t))T(t), X(t)>_{T(t)}\right)\, dt\\
&\le <J^N(0), J^N(0)>_{T(0)} \int_0^1 \left(G'(t)^2 + r^2\beta^2 G(t)^2\right)\, dt\\
&= <J^N(0), J^N(0)>_{T(0)} \left([G'(t)G(t)]_0^1 + \int_0^1 G(t)\left(-G''(t)^2 + r^2\beta^2 G(t)\right)\, dt\right)\\
&= <J^N(0), J^N(0)>_{T(0)} r\beta\coth(r\beta).
\end{aligned}
$$

So
(2.36)
$$\begin{aligned}
&D_p''(0)\\
&\le pr^{p-2}\left((p-1)<J^T(0),J^T(0)>_{T(0)}+r\beta\coth(r\beta)<J^N(0),J^N(0)>_{T(0)}\right)+\delta' r\\
&\le pr^{p-2}\left(\max\left((p-1),r\beta\coth(r\beta)\right)<J(0),J(0)>_{T(0)}+\delta' r\right)\\
&\le pr^{p-2}\left(D^2\max\left((p-1),r\beta\coth(r\beta)\right)+\delta' r\right)
\end{aligned}$$
since $F(J(0))=1$.

□

For $x\in M$, let $\ell_x : T_xM\to T_x^*M$ be the Legendre transformation, defined as

$$\ell_x(V)=g_V(V,\cdot)\quad\text{if}\quad V\neq 0,\quad \ell_x(0)=0. \tag{2.37}$$

It is well-known that ℓ_x is a bijection. The global Legendre transformation on TM is defined as

$$\mathscr{L}(V)=\ell_{\pi(V)}(V) \tag{2.38}$$

where $\pi : TM\to M$ is the canonical projection. If we define the dual Minkowski norm F^* on T_x^*M as

$$F^*(\xi)=\max\{\xi(y),\ y\in T_xM,\ F(y)=1\}, \tag{2.39}$$

then

$$F=F^*\circ\mathscr{L} \tag{2.40}$$

and for non zero $V\in TM$ and $\alpha\in T^*M$,

$$\langle\mathscr{L}(V),V\rangle=F(V)^2,\qquad \langle\alpha,\mathscr{L}^{-1}(\alpha)\rangle=F^*(\alpha)^2 \tag{2.41}$$

(see e.g. [3]).

For f a C^1 function on M we may define the gradient of f

$$\operatorname{grad} f=\mathscr{L}^{-1}(df). \tag{2.42}$$

3. Forward p-means

Let μ be a compactly supported probability measure in M. For $p>1$ and $x\in M$ we define

$$\mathscr{E}_{\mu,p}(x)=\int_M\rho^p(x,z)\,\mu(dz). \tag{3.1}$$

The (forward) p-mean of μ is the point e_p of M where $\mathscr{E}_{\mu,p}$ reaches its minimum whenever it exists and is unique.

In this paper we will consider forward p-means and we will call them p-means. Similarly we could define the backward p-mean $\overleftarrow{e}_p$ as the point which minimizes

$$x\mapsto\overleftarrow{\mathscr{E}}_{\mu,p}(x):=\int_M\rho^p(z,x)\,\mu(dz).$$

Depending on the context, forward or backward mean is more appropriate. One should note that defining the reverse (or adjoint) Finsler structure $\overleftarrow{F}(v)=F(-v)$, $v\in TM$, it is easy to check that the associated distance $\overleftarrow{\rho}$ satisfies $\overleftarrow{\rho}(z,x)=\rho(x,z)$, and forward p-mean for $\overleftarrow{F}$ is backward p-mean for F. So without loss of generality we can consider only the forward p-means.

One should also note that in High Angular Resolution Imaging the Finsler structure is symmetric, so both notions coincide. It is not the case for the application concerning active contours where it is natural to consider non symmetric F.

Even if it is in a non-Finslerian context, one can give the example of right-sided and left-sided Kullback-Leibler divergences for families of Gaussian probability densities (see [26]). The left-sided centroid focuses on the highest mode (it is zero-forcing), and the right-sided centroid tries to cover the support of both normals (it is zero-avoiding as depicted in Fig.2 of [26]). Furthermore, the left-sided Kullback-Leibler centroid yields the best single Gaussian distribution approximating the mixture model induced by the (weighted) family of Gaussians (Theorem 4.1 of [29]). The left-sided Kullback-Leibler centroid has been successfully used for simplifying statistical mixture models [16].

Proposition 3.1. *Assume there exists $C > 0$ such that $\mathscr{C}(x) \le C$ for all $x \in M$, where $\mathscr{C}(x)$ is defined in* (2.15). *Assume furthermore that* $\mathrm{supp}(\mu) \subset B(x_0, R)$ *for some $x_0 \in M$ and $R > 0$. Then $x \mapsto \mathscr{E}_{\mu,p}(x)$ has at least one global minimum in $\bar{B}(x_0, C(1+C)R)$.*

Proof. We begin with establishing that for all $y_1, y_2 \in M$,

$$\frac{1}{C}\rho(y_2, y_1) \le \rho(y_1, y_2) \le C\rho(y_2, y_1). \tag{3.2}$$

It is sufficient to establish the second inequality and then to exchange y_1 and y_2. If $t \mapsto \varphi(t)$ is a path from $y_1 = \varphi(0)$ and $y_2 = \varphi(1)$ then its length $L(\varphi)$ satisfies

$$\begin{aligned}
L(\varphi) &= \int_0^1 \sqrt{<\dot\varphi(t), \dot\varphi(t)>_{\dot\varphi(t)}}\, dt \\
&= \int_0^1 \sqrt{\frac{<-\dot\varphi(t), -\dot\varphi(t)>_{\dot\varphi(t)}}{<-\dot\varphi(t), -\dot\varphi(t)>_{-\dot\varphi(t)}}}\sqrt{<-\dot\varphi(t), -\dot\varphi(t)>_{-\dot\varphi(t)}}\, dt \\
&\le \int_0^1 \mathscr{C}(\varphi(t))\sqrt{<-\dot\varphi(t), -\dot\varphi(t)>_{-\dot\varphi(t)}}\, dt \\
&\le C\int_0^1 \sqrt{<-\dot\varphi(t), -\dot\varphi(t)>_{-\dot\varphi(t)}}\, dt \\
&= CL(\hat\varphi)
\end{aligned}$$

where $\hat\varphi$ is the path from y_2 to y_1 defined by $\hat\varphi(t) = \varphi(1-t)$. Minimizing over all paths $\hat\varphi$ from y_2 to y_1 we get

$$\rho(y_1, y_2) \le C\rho(y_2, y_1). \tag{3.3}$$

Now if $\mathrm{supp}(\mu) \subset B(x_0, R)$ then $\mathscr{E}_{\mu,p}(x_0) \le R^p$. On the other hand, if $x \notin \bar{B}(x_0, C(1+C)R)$ then for all $y \in B(x_0, R)$

$$\begin{aligned}
\rho(x, y) &\ge \rho(x, x_0) - \rho(y, x_0) \\
&\ge \frac{1}{C}\rho(x_0, x) - C\rho(x_0, y) \\
&\ge (1+C)R - CR = R
\end{aligned}$$

and this clearly implies that $\mathscr{E}_{\mu,p}(x) \ge R^p$. From this we get the conclusion. □

Concerning the uniqueness of the global minimum of $\mathscr{E}_{\mu,p}$, we also have the following easy result.

Proposition 3.2. *Assume that μ is supported by a compact ball $\bar{B}(x_0,R)$, and that for all $z\in\bar{B}(x_0,R)$, the function $x\mapsto\rho^p(x,z)$ is strictly convex in $\bar{B}(x_0,C(1+C)R)$. Then μ has a unique p-mean in $\bar{B}(x_0,C(1+C)R)$.*

Proof. If $x\mapsto\rho^p(x,z)$ is strictly convex for all z in the support of μ then $\mathscr{E}_{\mu,p}$ is strictly convex, and this implies that it has a unique minimum, which is attained at a unique point e_p. □

Corollary 3.3. *Assume $\mathscr{K}\le k$, $\mathscr{T}\ge-\delta$, $\mathscr{C}\le C$ for some $k,\delta\ge 0$, $C\ge 1$. Let $p>1$. Again let*

$$R(p,k,\delta,C)=\min\left(\frac{p-1}{C^2\delta},\frac{1}{\sqrt{k}}\arctan\left(\frac{\sqrt{k}}{C^2\delta}\right)\right)$$

If μ is supported by a geodesic ball $B(x_0,R)$ with

$$R\le\frac{1}{C(C+1)^2}R(p,k,\delta,C) \tag{3.4}$$

and the injectivity radius at any $x\in B(x_0,C(1+C)R)$ is strictly larger than $R(p,k,\delta,C)$ then μ has a unique p-mean e_p satisfying

$$e_p\in\bar{B}\left(x_0,\frac{1}{C+1}R(p,k,\delta,C)\right). \tag{3.5}$$

Proof. If $x,z\in B(x_0,C(1+C)R)$ then

$$\rho(x,z)\le\rho(x,x_0)+\rho(x_0,z)\le(1+C)^2CR\le R(p,k,\delta,C).$$

Using proposition 2.2, we obtain that $\mathscr{E}_{\mu,p}$ is strictly convex on $B(x_0,C(1+C)R)$. So by proposition 3.2 μ has a unique p-mean in $\bar{B}(x_0,C(1+C)R)$. □

Remark 3.4. Letting $x_0\in M$, D be a relatively compact neighborhood of x_0, then $\mathscr{K}$ and $\mathscr{C}$ are bounded above on D by, say k_D and C_D, and $\mathscr{T}$ is bounded below on D by $-\delta_D$. Using these bounds instead of k, C and δ, we can find R sufficiently small so that the conditions of corollary 3.3 are fulfilled. So we can say any measure μ with sufficiently small support has a unique p-mean.

Remark 3.5. If M is a Cartan-Hadamard manifold, we recover the fact that we can take $R(p,k,\delta,C)$ as large as we want.

More generally, in the Riemannian case, Afsari [1] proved existence and uniqueness of p-means, $p\ge 1$ on geodesic balls with radius $r<\frac{1}{2}\min\left\{\mathrm{inj}(M),\frac{\pi}{2\sqrt{k}}\right\}$ if $p\in[1,2)$, and $r<\frac{1}{2}\min\left\{\mathrm{inj}(M),\frac{\pi}{\sqrt{k}}\right\}$ if $p\ge 2$. Even taking $\delta=0$ and $C=1$ in Corollary 3.3 the support of μ has half the size of the one in [1] for $p\in(1,2)$ due to the fact that we have an additional condition (3.4) coming from the non optimality of Proposition 3.2 in the Riemannian context. As for $p\ge 2$ another factor two is gained in [1] with repeated use of Toponogov and Alexandroff theorems which are not available in our context.

Remark 3.6. The condition on injectivity radius is the same as in the Riemannian case. The cut locus of any point of $x \in B(x_0, C(1+C)R)$ has to be at distance larger than $R(p,k,\delta,C)$. As for Riemannian manifold there is no general condition which insures this property for cut points, but for conjugate points the same condition holds, due to Rauch comparison theorem, see Theorem 9.6.1 in [7]. In the particular case when M is a Cartan-Hadamard Finsler manifold, i.e. it has nonpositive flag curvature and it is simply connected, then the injectivity radius in everywhere infinite (see Theorem 9.4.1 in [7]).

Proposition 3.7. *Let $a \mapsto x(a)$ solve the equation*

$$x(0) = x_0 \quad \text{and for } a \geq 0 \quad x'(a) = \mathrm{grad}_{x(a)}(-\mathscr{E}_{\mu,p}). \tag{3.6}$$

Under the conditions of Corollary 3.3, the path $a \mapsto x(a)$ converges as $a \to \infty$ to the p-mean of μ.

Proof. If $f(a) = (-\mathscr{E}_\mu)(x(a))$ we have as soon as $\mathrm{grad}_{x(a)}(-\mathscr{E}_{\mu,p}) \neq 0$,

$$\begin{aligned} f'(a) &= \langle d_{x(a)}(-\mathscr{E}_{\mu,p}), x'(a)\rangle \\ &= \left\langle d_{x(a)}(-\mathscr{E}_{\mu,p}), \mathrm{grad}_{x(a)}(-\mathscr{E}_{\mu,p})\right\rangle \\ &= \langle d_{x(a)}(-\mathscr{E}_{\mu,p}), \mathscr{L}^{-1}(d_{x(a)}(-\mathscr{E}_{\mu,p}))\rangle \\ &= F^*\left(d_{x(a)}(-\mathscr{E}_{\mu,p})\right)^2 \end{aligned}$$

by (2.40) and (2.41).

On the other hand, we have $f(0) \geq -R^p$ and f is nondecreasing. This implies that for all $a \geq 0$, $x(a) \in \bar{B}(x_0, C(1+C)R)$, since for all $x \notin \bar{B}(x_0, C(1+C)R)$, $\mathscr{E}_{\mu,p}(x) \geq R^p$. As a consequence $x(a)$ has limit points as a goes to infinity, and since $f(a)$ converges, any limit point is a critical point of $x \mapsto \mathscr{E}_{\mu,p}(x)$. But by Proposition 3.2 $\mathscr{E}_{\mu,p}$ has a unique critical point in $\bar{B}(x_0, C(1+C)R)$ which is the mean e_p of μ. So we can conclude that $x(a)$ converges to e_p. □

4. Forward median

Let μ be a compactly supported probability measure in M. For $x \in M$ we define

$$\mathscr{F}_\mu(x) = \int_M \rho(x,z)\,\mu(dz). \tag{4.1}$$

The median of μ is the point in M where $\mathscr{F}_\mu$ reaches its minimum whenever it exists and is unique.

In the terminology of Section 3 we have $\mathscr{F}_\mu = \mathscr{E}_{\mu,1}$. The following result extends Proposition 3.1 to the case $p = 1$ and has exactly the same proof.

Proposition 4.1. *Assume there exists $C > 0$ such that $\mathscr{C}(x) \leq C$ for all $x \in M$. Assume furthermore that $\mathrm{supp}(\mu) \subset B(x_0, R)$ for some $x_0 \in M$ and $R > 0$. Then $x \mapsto \mathscr{F}_\mu(x)$ has at least one global minimum in $\bar{B}(x_0, C(1+C)R)$.*

Proposition 4.2. *Assume that μ is supported by a compact ball $\bar{B}(x_0, R)$, that the support of μ is not contained in a single geodesic and that for all $z \in \bar{B}(x_0, R)$, the forward distance to z is convex, and strictly convex in any geodesic of $\bar{B}(x_0, C(1+C)R)$ which does not contain z. Then μ has a unique median $m \in \bar{B}(x_0, C(1+C)R)$.*

Proof. Clearly under these assumptions $\mathscr{F}_\mu$ is strictly convex, so it has a unique local minimum, this minimum is global and is attained at a unique point $m \in \bar{B}(x_0, C(1+C)R)$. $\square$

Remark 4.3. Contrarily to the case of p-means for $p > 1$, we cannot say at this stage that any probability measure μ with sufficiently small support has a unique median, since we don't know whether $\mathscr{F}_\mu$ is strictly convex or not. In the next proposition we give a sufficient condition for strict convexity of $\mathscr{F}_\mu$.

Proposition 4.4. *Assume $\mathscr{K} \le k$ and $\mathscr{T} \ge -\delta$ for some $k, \delta > 0$. Assume that the injectivity radius at any point of $\bar{B}(x_0, C(1+C)R)$ is larger than $(C^2+C+1)R$. Define*

$$\begin{aligned}\eta = \min\Big\{ &\int_M \sqrt{k}\cot\left(\sqrt{k}\rho(\pi(v), z)\right) <v^N, v^N>_{\overrightarrow{\pi(v)z}} \mu(dz), \\ & v \in TM \text{ satisfying } \pi(v) \in \bar{B}(x_0, C(1+C)R),\ F(v) = 1\Big\}\end{aligned} \tag{4.2}$$

where v^N is the normal part of v with respect to the vector $\overrightarrow{\pi(v)z}$ and the scalar product $<\cdot,\cdot>_{\overrightarrow{\pi(v)z}}$. If $\eta - \delta > 0$ then $\mathscr{F}_\mu$ is strictly convex on $\bar{B}(x_0, C(1+C)R)$. More precisely, for all $x \in B(x_0, C(1+C)R)$ and for all unit speed geodesic γ starting at x,

$$(\mathscr{F}_\mu \circ \gamma)''(0) \ge \eta - \delta. \tag{4.3}$$

Proof. With the notations of section 2, from (2.31) we have

$$D_1''(0) = r^{-1}\left(E''(0) - <J^T(0), J^T(0)>_{T(0)}\right). \tag{4.4}$$

Let $\gamma(s) = c(0,s)$, the unit speed geodesic with initial condition $v = J(0)$, and $f(s) = \mathscr{F}_\mu(\gamma(s))$. Equation (4.4) together with (2.29) gives

$$f''(0) \ge -\delta + \int_M \sqrt{k}\cot\left(\sqrt{k}\rho(\pi(v), z)\right) <v^N, v^N>_{\overrightarrow{\pi(v)z}} \mu(dz) \ge \eta - \delta. \tag{4.5}$$

From this we get the condition for the strict convexity of $\mathscr{F}_\mu$.

$\square$

For $x \in M$ define the measure $\mu_x = \mu - \mu(\{x\})\delta_x$. Then the map $y \mapsto \mathscr{F}_{\mu_x}(y)$ is differentiable at $y = x$.

Since

$$\mathscr{F}_\mu(y) = \mathscr{F}_{\mu_x}(y) + \mu(\{x\})\rho(y, x) \tag{4.6}$$

and for $v \in T_xM$, $\mathscr{F}_\mu$ is differentiable in the direction v with derivative

$$\langle d\mathscr{F}_\mu, v\rangle = \langle d\mathscr{F}_{\mu_x}, v\rangle + \mu(\{x\})F(-v), \tag{4.7}$$

we see that x is a local minimum of $\mathscr{F}_\mu$ if and only if for all nonzero $v \in T_xM$

$$\mu(\{x\})F(-v) \ge \langle d\mathscr{F}_{\mu_x}, -v\rangle \tag{4.8}$$

which is equivalent to

$$\mu(\{x\}) \ge \frac{\left(F^*\left(d\mathscr{F}_{\mu_x}\right)\right)^2}{F\left(\mathscr{L}^{-1}(d\mathscr{F}_{\mu_x})\right)} \tag{4.9}$$

(take $-v = \dfrac{\mathscr{L}^{-1}(d\mathscr{F}_{\mu_x})}{F\left(\mathscr{L}^{-1}(d\mathscr{F}_{\mu_x})\right)}$). But since $F^* = F \circ \mathscr{L}^{-1}$, we get

Proposition 4.5. *A point x in M is a local minimum of $\mathscr{F}_\mu$ if and only if*

$$\mu(\{x\}) \ge F^*\left(d\mathscr{F}_{\mu_x}\right). \tag{4.10}$$

Note that for the Riemannian case this result is due to Le Yang [36].

Define the vector

$$H(x) = \mathrm{grad}_y\left(\mathscr{F}_{\mu_x}(y)\right)|_{y=x}. \tag{4.11}$$

Alternatively,

$$H(x) = \mathscr{L}^{-1}\left(\int_{M\backslash\{x\}} \mathscr{L}\left(-\frac{1}{\rho(x,z)}\overrightarrow{xz}\right)\mu(dz)\right). \tag{4.12}$$

Let $a \mapsto x(a)$ be the path in M defined by $x(0) = x_0$ and

$$\begin{aligned} \dot{x}(a) = &\quad -H(x(a)) &&\text{if for all } a' \le a,\ \mu(\{x(a')\}) < F^*\left(d\mathscr{F}_{\mu_{x(a')}}\right);\\ \dot{x}(a) = &\quad 0 &&\text{if for some } a' \le a,\ \mu(\{x(a')\}) \ge F^*\left(d\mathscr{F}_{\mu_{x(a')}}\right). \end{aligned} \tag{4.13}$$

Define

$$f(a) = \mathscr{F}_\mu(x(a)). \tag{4.14}$$

We have for the right derivative of f when $x(a)$ is not a minimal point of $\mathscr{F}_\mu$:

$$\begin{aligned} f'_+(a) &= \left\langle d_{x(a)}(\mathscr{F}_{\mu_{x(a)}}), \dot{x}(a)\right\rangle + \mu(\{x(a)\})F\left(-\dot{x}(a)\right)\\ &= -\left\langle d_{x(a)}(\mathscr{F}_{\mu_{x(a)}}), \mathscr{L}^{-1}(d_{x(a)}(\mathscr{F}_{\mu_{x(a)}}))\right\rangle\\ &\quad + \mu(\{x(a)\})F\left(\mathscr{L}^{-1}(d_{x(a)}(\mathscr{F}_{\mu_{x(a)}}))\right)\\ &= -F^*\left(d_{x(a)}(\mathscr{F}_{\mu_{x(a)}})\right)^2 + \mu(\{x(a)\})F\left(\mathscr{L}^{-1}(d_{x(a)}(\mathscr{F}_{\mu_{x(a)}}))\right). \end{aligned}$$

We get

$$f'_+(a) = -F^*\left(d_{x(a)}(\mathscr{F}_{\mu_{x(a)}})\right)\left(F^*\left(d_{x(a)}(\mathscr{F}_{\mu_{x(a)}})\right) - \mu(\{x(a)\})\right) \tag{4.15}$$

which is negative as soon as $x(a)$ is not a minimal point of $\mathscr{F}_\mu$. From this we get the following

Proposition 4.6. *Assume that μ is supported by a compact ball $\bar{B}(x_0, R)$, that the support of μ is not contained in a single geodesic and that for all $z \in \bar{B}(x_0, R)$, the forward distance to z is convex, and strictly convex in any geodesic of $\bar{B}(x_0, C(1+C)R)$ which does not contain z. Then the path $a \mapsto x(a)$ converges to the median m of μ.*

Proof. Similar to the proof of proposition 3.7 □

5. An algorithm for computing p-means

Lemma 5.1. *Assume $\mathscr{K} \ge -\beta^2$, $\mathscr{T} \le \delta'$, $\mathscr{D} \le D$ with $\beta > 0$, $\delta' \ge 0$ $D \ge 1$. For $p > 1$, $r > 0$, define*

$$H(r) = H_{p,\beta,D,\delta'}(r) := pr^{p-2}\left(D^2\max\left((p-1), r\beta\coth(r\beta)\right) + \delta' r\right). \tag{5.1}$$

If μ is a probability measure on M with bounded support and $x \in M$, define

$$H_\mu(x) = H_{\mu,p,\beta,D,\delta'}(x) := \int_M H_{p,\beta,D,\delta'}(\rho(x,y))\,d\mu. \tag{5.2}$$

If $t \mapsto \gamma(t)$ is a unit speed geodesic then for all t

$$(\mathscr{E}_{\mu,p} \circ \gamma)''(t) \le H_\mu(\gamma(t)). \tag{5.3}$$

Proof. For $x, y \in M$, $r = \rho(x,y)$, $s \mapsto \gamma(s) = c(0,s)$ a unit speed geodesic started at $x = c(0,0)$, $t \mapsto c(t,s)$ the geodesic satisfying $c(1,s) = y$, we have

$$D_p''(0) \le pr^{p-2}\left(D^2 \max\left((p-1), r\beta\coth(r\beta)\right) + \delta' r\right).$$

Integrating with respect to y this equation gives the result. □

Remark 5.2. If $p \ge 2$ or μ has a smooth density then the function H_μ is bounded on all compact sets.

The main result is the following (see [22] for a similar result in a Riemannian manifold).

Proposition 5.3. *Assume $-\beta^2 \le \mathscr{K} \le k$, $-\delta \le \mathscr{T} \le \delta'$, $\mathscr{C} \le C$ and $\mathscr{D} \le D$ for some $\beta, k, \delta, \delta' > 0$ and $C, D \ge 1$. Let $p > 1$. Assume the support of μ is contained in $B(x_0, R)$ and $\mathscr{E}_{\mu,p}$ is strictly convex on $\bar{B}(x_0, C(C+1)R)$. Assume furthermore that the function $H_\mu = H_{\mu,p,\beta,D,\delta'}$ is bounded on $\bar{B}(x_0, C(C+1)R)$ by a constant $C_H > 0$, and that the injectivity radius at any point of $\bar{B}(x_0, C(C+1)R)$ is larger than $C^2 + C + 1$. Define the gradient algorithm as follows:*

Step 1 *Start from a point $x_1 \in B(x_0, C(C+1)R)$ such that $\mathscr{E}_{\mu,p}(x_1) \le R^p$ (take for instance $x_1 = x_0$) and let $k = 1$.*

Step 2 *Let*

$$v_k = \frac{\mathrm{grad}(-\mathscr{E}_{\mu,p}(x_k))}{F\left(\mathrm{grad}(-\mathscr{E}_{\mu,p}(x_k))\right)}, \qquad t_k = \frac{F\left(\mathrm{grad}(-\mathscr{E}_{\mu,p}(x_k))\right)}{C_H}. \tag{5.4}$$

and let γ_k be the geodesic satisfying $\gamma_k(0) = x_k$, $\dot\gamma_k(0) = v_k$. Define

$$x_{k+1} = \gamma_k(t_k) \tag{5.5}$$

then do again step 2 with $k = k+1$.

Then the sequence $(x_k)_{k\ge 1}$ converges to e_p.

Proof. We first prove that the sequence $(\mathscr{E}_{\mu,p}(x_k))_{k\in\mathbb{N}}$ is nonincreasing. For this we write

$$\begin{aligned}
\mathscr{E}_{\mu,p}(\gamma_k(t_k)) &\le \mathscr{E}_{\mu,p}(\gamma_k(0)) + \langle d\mathscr{E}_{\mu,p}, v_k\rangle\, t_k + C_H \frac{t_k^2}{2} \\
&\le \mathscr{E}_{\mu,p}(\gamma_k(0)) - F\left(\mathrm{grad}(-\mathscr{E}_{\mu,p}(x_k))\right) \frac{1}{C_H} F\left(\mathrm{grad}(-\mathscr{E}_{\mu,p}(x_k))\right) \\
&\quad + \frac{C_H}{2}\left(\frac{F\left(\mathrm{grad}(-\mathscr{E}_{\mu,p}(x_k))\right)}{C_H}\right)^2 \\
&= \mathscr{E}_{\mu,p}(\gamma_k(0)) - \frac{C_H}{2}\left(\frac{F\left(\mathrm{grad}(-\mathscr{E}_{\mu,p}(x_k))\right)}{C_H}\right)^2.
\end{aligned} \tag{5.6}$$

This proves that the sequence is nonincreasing. As a consequence, for all $k \ge 1$, $x_k \in \bar{B}(x_0, C(C+1)R)$, since $\mathscr{E}_{\mu,p}(x_k) \le R^p$ and for all $x \notin \bar{B}(x_0, C(C+1)R)$, $\mathscr{E}_{\mu,p}(x) > R^p$.

Next we prove that $\mathscr{E}_{\mu,p}(x_k)$ converges to $\mathscr{E}_{\mu,p}(e_p)$. We know that $\mathscr{E}_{\mu,p}(x_k)$ converges to $a \ge \mathscr{E}_{\mu,p}(e_p)$. Extracting a subsequence x_{k_ℓ} converging to some $x_\infty \in \bar{B}(x_0, C(C+1)R)$, this implies that t_{k_ℓ} converges to 0. But this is possible only if $x_\infty = e_p$, which implies that $a = \mathscr{E}_{\mu,p}(e_p)$. As a consequence, any converging subsequence has e_p as a limit, and this implies that x_k converges to e_p. □

Remark 5.4. For this result we need the Hessian of $\mathscr{E}_{\mu,p}$ to be bounded, and the subgradient algorithm in Riemannian manifolds as developed in [36] does not work. The reason is that for this algorithm, we would need to take

$$v_k = \frac{\mathrm{grad}_{\overrightarrow{x_k e_p}}(-\mathscr{E}_{\mu,p}(x_k)))}{F\left(\mathrm{grad}_{\overrightarrow{x_k e_p}}(-\mathscr{E}_{\mu,p}(x_k))\right)}$$

where $\mathrm{grad}_{\overrightarrow{x_k e_p}}$ denotes the gradient with respect to the metric $<\cdot,\cdot>_{\overrightarrow{x_k e_p}}$. So we would need to know e_p!

Corollary 5.5. *Let $p = 2$. If $R \le \dfrac{1}{C(C+1)^2\sqrt{k}} \arctan\left(\dfrac{\sqrt{k}}{C\delta^2}\right)$ or M has non-positive flag curvature, then the algorithm of Proposition 5.3 can be applied with the appropriate constants*

Proof. With this assumption, by Proposition 2.2 the function $\mathscr{E}_{\mu,2}$ is strictly convex on $\bar{B}(x_0, C(C+1)R)$. □

References

[1] B. Afsari, *Riemannian L^p center of mass : existence, uniqueness, and convexity*, Proceedings of the American Mathematical Society, S 0002-9939(2010)10541-5, Article electronically published on August 27, 2010.

[2] S.I. Amary and H. Nagaoka, *Methods of Information Geometry* Translations of Mathematical Monographs, Vol. 191, AMS, Oxford University Press, 2000.

[3] J.C. Alvarez Paiva, *Some problems on Finsler geometry*, Handbook of differential geometry, Vol. II, pp. 1–33, Elsevier/North-Holland, 2006.

[4] M. Arnaudon and X.M. Li, *Barycenters of measures transported by stochastic flows*, The Annals of Probability, 33 (2005), no. 4, 1509–1543

[5] L. Astola and L. Florack, *Finsler geometry on higher oredre tensor fields and applications to high angular resolution diffusion imaging*, Scale Space and Variational Methods in Computer Vision, Second International Conference, SSVM 2009, Voss, Norway, June 1-5, 2009. Proceedings (SSVM), pp. 224–234, 2009

[6] L. Auslander, *On curvature in Finsler geometry*, Transactions of the American Mathematical Society; Vol. 79, No.2 (Jul. 1955), pp. 378–388

[7] D. Bao, S.S. Chern and Z. Shen, *An introduction to Riemann-Finsler geometry*, Graduate Texts in Mathematics, Springer, 2000

[8] S.S. Chern and Z. Shen, *Riemann-Finsler geometry*, Nankai Tracts in Mathematics, Vol. 6, World Scientific, 2005

[9] J.M. Corcuera and W.S. Kendall, *Riemannian barycentres and geodesic convexity* Math. Proc. Cambridge Philos. Soc. 127 (1999), no. 2, 253269.

[10] Charles Elkan, *Using the Triangle Inequality to Accelerate k-Means*, Proceedings of the Twentieth International Conference on Machine Learning (ICML), (2003), pp. 147-153

[11] M. Emery and G. Mokobodzki, *Sur le barycentre d'une probabilité dans une variété*, Séminaire de Probabilités XXV, Lecture Notes in Mathematics 1485 (Springer, Berlin, 1991), pp. 220–233

[12] Ehtibar Dzhafarov and Hans Colonius, *Fechnerian metrics in unidimensional and multidimensional stimulus spaces*, Psychonomic Bulletin & Review, Volume 6, Issue 2 Springer New York, 1999

[13] P.T. Fletcher, S. Venkatasubramanian, S. Joshi, *The geometric median on Riemannian manifolds with application to robust atlas estimation*, NeuroImage, 45 (2009), pp. S143–S152

[14] Bent Fuglede and Flemming Topsoe, *Jensen-Shannon divergence and Hilbert space embedding*, IEEE International Symposium on Information Theory (2004) pp. 31–31

[15] R. Gallego Torrome, *On the generalization of theorems from Riemannian to Finsler geometry I: Metric Theorems*, arXiv:math/0503704v3, [math.DG] 3 Apr 2008

[16] V. Garcia and F. Nielsen, *Simplification and hierarchical representations of mixtures of exponential families.* Signal Processing, vol 90, no.12 (2010), pp. 3197–3212

[17] F.R. Hampel, P.J. Rousseeuw, E.M. Ronchetti, W.A. Stahel, *Robust Statistics The Approach Based on Influence Function*, Wiley, New York, 1986
[18] H. Karcher, *Riemannian center of mass and mollifier smoothing*, Communications on Pure and Applied Mathematics, vol XXX (1977), 509–541
[19] W.S. Kendall, *Probability, convexity and harmonic maps with small image I: uniqueness and fine existence*, Proc. London Math. Soc. (3) 61 no. 2 (1990) pp. 371–406
[20] W.S. Kendall, *Convexity and the hemisphere*, J. London Math. Soc., (2) 43 no.3 (1991), pp. 567–576
[21] H.W. Kuhn, *A note on Fermat's problem*, Mathematical Programmming, 4 (1973), pp. 98–107
[22] H. Le, *Estimation of Riemannian barycentres*, LMS J. Comput. Math. 7 (2004), pp. 193–200
[23] , Fuhry, David and Jin, Ruoming and Zhang, Donghui, *Efficient skyline computation in metric space*, Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology, EDBT '09, ACM, New York, NY, USA (2009), pp. 1042–1051,
[24] J. Melonakos, E. Pichon, S. Angenent and A. Tannenbaum, Finsler active contours, IEEE Trans. Pattern Anal. Marc. Intell. 30(3), pp. 412–423, 2008.
[25] S-i Ohta, *Uniform convexity and smoothness, and their applications in Finsler geometry*, Math. Ann. 343 (2009), pp. 669–699
[26] F. Nielsen and R. Nock, *Sided and symmetrized Bregman centroids*, IEEE transactions on Information Theory, Vol. 55 Issue 66 (June 2009), pp. 2882–2904
[27] L.M.JR. Ostresh, *On the convergence of a class of iterative methods for solving Weber location problem*, Operation. Research, 26 (1978), no. 4
[28] M. Pchaud, R. Keriven and M. Descoteaux, *Brain connectivity using geodesics in hardi*, IEEE International Conference on Medical Image Computing and Computed Assisted Intervention (MICCAI), London, Sept. 2009
[29] B. Pelletier, *Informative barycentres in statistics*, Annals of the Institute of Statistical Mathematics, vol. 57, no. 4, (2005), pp. 767-780.
[30] J. Picard, *Barycentres et martingales dans les variétés*, Ann. Inst. H. Poincaré Probab. Statist. 30 (1994), pp. 647–702
[31] A. Sahib, *Espérance d'une variable aléatoire à valeurs dans un espace métrique*, Thèse de l'université de Rouen (1998)
[32] Z. Shen, *Riemann-Finsler geometry, with applications to information geometry*, Chinese Annals of Mathematics, Series B, 27, pp. 73–94, 2006
[33] Vardi, Y. and Zhang, C.H. *The multivariate L_1-median and associated data dept*, Proc. Natl. Acad. Sci, USA, 97, (2000), pp.1423-1426.
[34] E. Weiszfeld, *Sur le point pour lequel la somme des distances de n points donnés est minimum*, Tôhoku Math. J. 43 (1937), pp. 355–386
[35] B.Y. Wu and Y.L. Xin, *Comparison theorems in Finsler geometry and their applications*, Math. Ann. 337 (2007), pp. 177–196
[36] L.Yang, *Riemannian median and its estimation*, to appear in LMS Journal of Computation and Mathematics
[37] C.Zach, L. Shan and M. Niethammer, *Globally Optimal Finsler Active Contours*, Lecture Notes in Computer Science, Vol. 5748 (2009), pp. 552–561

Laboratoire de Mathématiques et Applications
CNRS: UMR 6086
Université de Poitiers, Téléport 2 - BP 30179
F–86962 Futuroscope Chasseneuil Cedex, France
E-mail address: `marc.arnaudon@math.univ-poitiers.fr`

Laboratoire d'informatique (LIX)
École Polytechnique
91128 Palaiseau Cedex, France
and
Sony Computer Science Laboratories, Inc
Tokyo, Japan
E-mail address: `Frank.Nielsen@acm.org`