On Approximating the Riemannian 1-Center

Marc Arnaudon[a], Frank Nielsen[b,c,*]

[a]*Laboratoire de Mathématiques et Applications*
CNRS: UMR 6086, Université de Poitiers
Téléport 2 - BP 30179
F–86962 Futuroscope Chasseneuil Cedex, France
[b]*Ecole Polytechnique*
Computer Science Department (LIX)
Palaiseau, France.
[c]*Sony Computer Science Laboratories, Inc. (FRL),*
3-14-13 Higashi Gotanda 3F, Shinagawa-Ku,
Tokyo 141-0022, Japan.

Abstract

We generalize the Euclidean 1-center approximation algorithm of Bădoiu and Clarkson (2003) to arbitrary Riemannian geometries, and study the corresponding convergence rate. We then show how to instantiate this generic algorithm to two particular settings: (1) the hyperbolic geometry, and (2) the Riemannian manifold of symmetric positive definite matrices.

Keywords: 1-center; minimax center; Riemannian geometry; core-set; approximation

1. Introduction and prior work

Finding the unique smallest enclosing ball (SEB) of a finite Euclidean point set $P = \{p_1, ..., p_n\}$ is a fundamental problem that was first posed by Sylvester (1857). This problem has been thoroughly investigated in the computational geometry community by Welzl (1991) and Nielsen and Nock (2009), where it is also known as the minimum enclosing ball (MEB), the

*Corresponding author

Email addresses: Marc.Arnaudon@math.univ-poitiers.fr (Marc Arnaudon), Frank.Nielsen@acm.org (Frank Nielsen)

Second revision. Source codes for reproducible research available at http://www.informationgeometry.org/RiemannMinimax/

1-center problem, or the minimax optimization problem in operations research. In practice, since computing the SEB exactly is intractable in high dimensions, efficient approximation algorithms have been proposed. An algorithmic breakthrough was achieved by Bădoiu and Clarkson (2008) that proved the existence of a *core-set* $C \subseteq P$ of *optimal size* $|C| = \lceil \frac{1}{\epsilon} \rceil$ so that $r(C) \leq (1+\epsilon)r(P)$ (for any arbitrary $\epsilon > 0$), where $r(S)$ denotes the radius of the SEB of S. Let $c(S)$ denote the ball center, i.e. the minimax center. Since the size of the core-set depends *only* on the approximation precision ϵ and is *independent* of the dimension, core-sets have become widely popular in high-dimensional applications such as supervised classification in machine learning (see for example, the core vector machines of Tsang et al. (2007)). In the work of Bădoiu and Clarkson (2003), a fast and simple approximation algorithm is designed as follows:

BC-ALG:

- Initialize the center $c_1 \in P$, and
- Iteratively update the current center using the rule

$$c_{i+1} \leftarrow c_i + \frac{f_i - c_i}{i+1},$$

where f_i denotes the farthest point of P to c_i.

It can be proved that a $(1+\epsilon)$-approximation of the SEB is obtained after $\lceil \frac{1}{\epsilon^2} \rceil$ iterations, thereby showing the existence of a core-set $C = \{f_1, f_2, ...\}$ of a size at most $\lceil \frac{1}{\epsilon^2} \rceil$: $r(C) \leq (1+\epsilon)r(P)$. This simple algorithm runs in time $O(\frac{dn}{\epsilon^2})$, and has been generalized to Bregman divergences by Nock and Nielsen (2005) which include the (squared) Euclidean distance, and are the canonical distances of dually flat spaces, including the particular case of self-dual Euclidean geometry. (Note that if we start from the optimal center $c_1 = c(S)$, the first iteration yields a center c_2 away from $c(S)$ but it will converge in the long run to $c(S)$.) Bădoiu and Clarkson (2008) proved the existence of optimal ϵ-core-set of size $\lceil \frac{1}{\epsilon} \rceil$. Since finding tight core-sets requires as a black box primitive the computation of the exact smallest enclosing balls of small-size point sets, we rather consider the Riemmanian generalization of the BC-ALG, although that even in the Euclidean case it does not deliver optimal size core-sets.

Many data-sets arising in medical imaging (see Pennec (2008)) or in computer vision (refer to Turaga and Chellappa (2010)) cannot be considered as emanating from vectorial spaces but rather as lying on curved manifolds. For example, the space of rotations or the space of invertible matrices are not flat, as the arithmetic average of two elements does not necessarily lie inside the space.

In this work, we extend the Euclidean BC-ALG algorithm to Riemannian geometry. In the remainder, we assume the reader familiar with basic notions of Riemannian geometry (see Berger (2003) for an introductory textbook) in order not to burden the paper with technical Riemannian definitions. However in the appendix, we recall some specific notions which play a key role in the paper, such as geodesics, sectional curvature, injectivity radius, Alexandrov and Toponogov theorems, and cosine laws for triangles. Furthermore, we consider probability measures instead of finite point sets[2] so as to study the most general setting.

Let M be a complete Riemannian manifold and ν a probability measure on M. Denote by $\rho(x, y)$ the Riemannian distance from x to y on M that satisfies the metric axioms. Assume the measure support $\mathrm{supp}(\nu)$ is included in a geodesic ball $B(o, R)$.

Recall that if $p \in [1, \infty)$ and $f : M \to \mathbb{R}$ is a measurable function then

$$\|f\|_{L^p(\nu)} = \left(\int_M |f(y)|^p \, \nu(dy) \right)^{1/p}$$

and

$$\|f\|_{L^\infty(\nu)} = \inf \left\{ a > 0, \quad \nu \left(\{ y \in M, \ |f(y)| > a \} \right) = 0 \right\}.$$

Let

$$R_{\alpha,p} = \begin{cases} \frac{1}{2} \min \left\{ \mathrm{inj}(M), \frac{\pi}{2\alpha} \right\} & \text{if} \quad 1 \leq p < 2, \\ \frac{1}{2} \min \left\{ \mathrm{inj}(M), \frac{\pi}{\alpha} \right\} & \text{if} \quad 2 \leq p \leq \infty \end{cases} \tag{1}$$

where $\mathrm{inj}(M)$ is the injectivity radius (see the appendix) and $\alpha > 0$ is such that α^2 is an upper bound for the sectional curvatures in M (in fact replacing M by $B(o, 2R)$ is sufficient, so that we can always assume that $\alpha > 0$). For $p \in [1, \infty]$, under the assumption that

$$R < R_{\alpha,p} \tag{2}$$

[2] We view finite point sets as discrete uniform probability measures.

it has been proved by Afsari (2011) that there exists a *unique* point c_p which minimizes the following cost function

$$\begin{aligned} H_p \quad : \quad & M \to [0, \infty] \\ & x \mapsto \|\rho(x, \cdot)\|_{L^p(\nu)} \end{aligned} \tag{3}$$

with $c_p \in B(o, R)$ (in fact, lying inside the closure of the convex hull of the support of ν).

For a discrete uniform measure viewed as a "point cloud" in an Euclidean space and $p \in [1, \infty)$, we have $H_p(x) = \left(\frac{1}{n}\sum_{i=1}^n \|p_i - x\|_p^p\right)^{1/p}$, with $\|\cdot\|_p$ denoting the L_p norm, and $H_\infty(x)$ is the distance from x to its farthest point in the cloud.

In the general situation the point c_p that realizes the minimum represents a notion of centrality of the measure (eg., median for $p = 1$, mean for $p = 2$, and minimax center for $p = \infty$). This center is a *global* minimizer (not only in $B(o, R)$), and this explains why a bound for the sectional curvature is required on the whole manifold M (in fact $B(o, 2R)$ is sufficient, see Afsari (2011)).

Deterministic subgradient algorithms for finding c_p have been considered by Yang (2010) for the median case ($p = 1$). Stochastic algorithms have been investigated by Arnaudon et al. (2010) for the case $p \in [1, \infty)$, and a central limit theorem (CLT) for the suitably renormalized process is derived (in fact a convergence in law to a diffusion process). See also for similar algorithms minimizing other cost functions, the work of Bonnabel (2011).

In this work, we consider the case $p = \infty$, with c_∞ denoting the minimax center. Hereafter we use c for c_∞, H for H_∞ and R_α for $R_{\alpha,\infty}$. In this case there is no canonical deterministic algorithm which generalizes the gradient descent algorithms considered for $p \in [1, \infty)$. Following Eq. 3, $H(x)$ denotes the farthest distance from x to a point of the support of the measure (L^∞-norm).

To give an example of a Riemannian manifold, consider the space of symmetric positive definite matrices with associated Riemannian distance (see Section 4)

$$\rho(P, Q) = \|\log(P^{-1}Q)\|_F = \sqrt{\sum_i \log^2 \lambda_i} \tag{4}$$

where λ_i are the eigenvalues of matrix $P^{-1}Q$. This is a non-compact Riemannian symmetric space of nonpositive curvature (Cartan-Hadamard manifold, see Lang (1999), chapter 12). In this context *any* measure ν with *bounded support* satisfies. Eq. 2 (since we can take $\alpha > 0$ as small as we like), and consequently the minimizer c of H exists and is unique. We call it the 1-center or minimax center of ν.

We generalized the BC-ALG by noticing that the iterative update is a barycenter of the current minimax center with the current farthest point. Thus the new position of the minimax center falls along the straight line joining these two points in Euclidean geometry. In Riemannian geometry, the shortest path linking two points is called a geodesic (for example, arc of a great circle for spherical geometry). Instead of walking on a straight line, we instead walk on the geodesic to the farthest point as follows:

GEO-ALG:

- Initialize the center with $c_1 \in P$, and
- Iteratively update the current minimax center as

$$c_{i+1} = \text{Geodesic}\left(c_i, f_i, \frac{1}{i+1}\right),$$

where f_i denotes the farthest point of P to c_i, and $\text{Geodesic}(p, q, t)$ denotes the intermediate point m on the geodesic passing through p and q such that $\rho(p, m) = t \times \rho(p, q)$.

Note that GEO-ALG generalized BC-ALG by taking the Euclidean distance $\rho(p, q) = \|p - q\|$.

The paper is organized as follows: Section 2 gives and proves a crucial lemma. It is followed by the description and convergence rate analysis of our generic Riemannian algorithm in Section 3. Section 4 instantiates the algorithm for the particular cases of the hyperbolic manifold and the manifold of symmetric positive definite matrices. Section 5 concludes the paper and hints at further perspectives. To make the paper self-contained, the appendix recalls the fundamental notions of Riemannian geometry used throughout the paper.

2. A key lemma

In this section, we assume[3] that $\operatorname{supp}(\nu) \subset B(o, R)$ and

$$R < R_\alpha = \frac{1}{2} \min \left\{ \operatorname{inj}(M), \frac{\pi}{\alpha} \right\}$$

with $\alpha > 0$ such that α^2 is an upper bound for the sectional curvatures in M. The following lemma is essential for proving the convergence of the algorithm determining the minimax of ν.

Lemma 1. *There exists $\tau > 0$ such that for all $x \in B(o, R)$,*

$$H(x) - H(c) \geq \tau \rho^2(x, c). \tag{5}$$

Proof:
The point c is the center of the smallest ball which contains $\operatorname{supp}(\nu)$ and the radius of this ball is exactly $r^* := H(c)$ (see Afsari (2009)). An immediate consequence is that $r^* \leq R$. Denoting by $S(c, r^*)$ the boundary of this ball and by $S_c M$ the set of unitary vectors in $T_c M$, for all $v \in S_c M$ there exists $y \in S(c, r^*) \cap \operatorname{supp}(\nu)$ such that

$$\langle \overrightarrow{cy}, v \rangle \leq 0 \tag{6}$$

where $t \mapsto \gamma_t(c, y)$ is the geodesic from c to y in time one, $\dot{\gamma}_t(c, y)$ denotes derivative with respect to t and $\overrightarrow{cy} = \dot{\gamma}_0(c, y)$. Indeed, if this was not true it would contradict the minimality of $S(c, r^*)$ (refer to Afsari (2009)).

Now letting $t \mapsto \gamma_t(v) = \exp_x(tv)$ the geodesic satisfying $\dot{\gamma}_0(v) = v$, we prove Eq. 5 for $x = \gamma_t(v)$. We have

$$H(\gamma_t(v)) - H(c) \geq \rho(\gamma_t(v), y) - \rho(c, y) = \rho(\gamma_t(v), y) - r^* \tag{7}$$

by definition of H.

Then we consider a 2-dimensional sphere $S^2_{\alpha^2}$ with *constant* curvature α^2, distance function $\tilde{\rho}$, and in $S^2_{\alpha^2}$ a *comparison triangle* $\tilde{\gamma}_t(\tilde{v})\tilde{y}\tilde{c}$ such that $\tilde{\rho}(\tilde{y}, \tilde{c}) = r^*$, $\tilde{v}$ is a unitary vector in $T_{\tilde{c}} S^2_{\alpha^2}$ satisfying

$$\left\langle \overrightarrow{\tilde{c}\tilde{y}}, \tilde{v} \right\rangle = \langle \overrightarrow{cy}, v \rangle \tag{8}$$

[3] Any bounded measure on a Cartan-Hadamard manifold satisfies this assumption.

Let us prove that

$$\tilde{\rho}(\tilde{\gamma}_t(\tilde{v}),\tilde{y}) - r^* = \tilde{\rho}(\tilde{\gamma}_t(\tilde{v}),\tilde{y}) - \rho(\tilde{c},\tilde{y}) \geq \tau_\alpha \tilde{\rho}^2(\tilde{\gamma}_t(\tilde{v}),\tilde{c}) \tag{9}$$

for some $\tau_\alpha > 0$ provided condition Eq. 6 is realized: for simplicity we will write $\tilde{d} = \tilde{\rho}(\tilde{\gamma}_t(\tilde{v}),\tilde{y})$. Using Eq. 6 and the first law of cosines (Theorem 4 in the appendix), we get

$$0 \geq \cos\left(\overrightarrow{\tilde{c}\tilde{y}},\tilde{v}\right) = \frac{\cos\left(\alpha\tilde{d}\right) - \cos\left(\alpha r^*\right)\cos(\alpha t)}{\sin\left(\alpha r^*\right)\sin(\alpha t)} \tag{10}$$

which yields

$$\cos\left(\alpha\tilde{d}\right) - \cos\left(\alpha r^*\right)\cos(\alpha t) \leq 0.$$

On the other hand since $0 < \alpha r^* < \frac{\pi}{2}$ and $0 \leq \alpha\tilde{d} < \pi$ we have

$$0 \leq 2\sin\left(\alpha\tilde{d}\right)\cos(\alpha r^*)\sin(\alpha r^*).$$

So we get

$$\cos\left(\alpha\tilde{d}\right) - \cos\left(\alpha r^*\right)\cos(\alpha t) \leq 2\sin\left(\alpha\tilde{d}\right)\cos(\alpha r^*)\sin(\alpha r^*)$$

which is equivalent to

$$\begin{gathered}\cos^2(\alpha r^*)\cos(\alpha\tilde{d}) + \cos(\alpha r^*)\sin(\alpha r^*)\sin(\alpha\tilde{d}) - \cos(\alpha r^*)\cos(\alpha t) \\ \leq \sin(\alpha\tilde{d})\cos(\alpha r^*)\sin(\alpha r^*) - \sin^2(\alpha r^*)\cos(\alpha\tilde{d})\end{gathered}$$

and this in turn implies

$$\sin\left(\alpha\left(\tilde{d}-r^*\right)\right) \geq \operatorname{cotan}(\alpha r^*)\left(\cos\left(\alpha\left(\tilde{d}-r^*\right)\right) - \cos(\alpha t)\right)$$

so

$$\liminf_{t\searrow 0}\frac{\tilde{\rho}(\tilde{\gamma}_t(\tilde{v}),\tilde{y}) - r^*}{t^2} \geq \frac{\alpha}{2}\operatorname{cotan}(\alpha r^*) \geq \frac{\alpha}{2}\operatorname{cotan}(\alpha R_\alpha)$$

uniformly in $\tilde{v}$. Consequently Eq. 9 is true for $\tilde{\gamma}_t(\tilde{v})$ in a neighborhood of $\tilde{c}$, and since $\tilde{\rho}(\tilde{\gamma}_t(\tilde{v}),\tilde{y}) - r^*$ does not vanish outside this neighbourhood, by a compactness argument we prove that Eq. 9 is true in any compact included in $\tilde{B}(\tilde{c},R_\alpha)$, if τ_α is sufficiently small.

To finish the proof we are left to use the Alexandrov comparison theorem (Theorem 2 in the appendix) with triangles $\gamma_t(v)yc$ and $\tilde{\gamma}_t(\tilde{v})\tilde{y}\tilde{c}$ to check that the right hand side of Eq. 7 in M is larger than the left hand side of Eq. 9. This proves Eq. 5 in $B(c,R)\cap B(o,R)$, and for proving it in $B(o,R)$ we just have to notice that H is continuous and positive on the compact set $\bar{B}(o,R)\backslash B(c,R)$, hence it has a positive lower bound. □

3. Riemannian approximation algorithm

For $x \in B(o, R)$, denote by $t \mapsto \gamma_t(v(x,\nu))$ a unit speed geodesic from $\gamma_0(v(x,\nu)) = x$ to one point $y = \gamma_{H(x)}(v(x,\nu))$ in $\mathrm{supp}(\nu)$ which realizes the maximum of the distance from x to $\mathrm{supp}(\nu)$. So $v = \dfrac{1}{H(x)} \exp_x^{-1}(y)$. A measurable choice is always possible. Note that if ν has finite support, when there is a finite number of possibilities for y it is natural to make a random uniform choice. However in a generic situation this should never happen, there should be only one choice.

We consider the following stochastic algorithm.

RIE-ALG:
Fix some $\delta > 0$.
Step 1 Choose a starting point $x_0 \in \mathrm{supp}(\nu)$ and let $k = 0$
Step 2 Choose a step size $t_{k+1} \in (0, \delta]$ and let $x_{k+1} = \gamma_{t_{k+1}}(v(x_k, \nu))$, then do again step 2 with $k \leftarrow k + 1$.

This algorithm generalizes the Euclidean scheme of Bădoiu and Clarkson (2003) and algorithm GEO-ALG for probability measures. Indeed, if GEO-ALG is initialized with $c_{k_0} \in P$ with k_0 the first integer larger than $1/\delta$, then it suffices to take $t_k = 1/k$ for $k \geq k_0$ in RIE-ALG.

Let $a \wedge b$ denote the minimum operator $a \wedge b = \min(a, b)$.

Let

$$R_0 = \frac{R_\alpha - R}{2} \wedge \frac{R}{2}. \tag{11}$$

Theorem 1. *Assume $\alpha, \beta > 0$ are such that $-\beta^2$ is a lower bound and α^2 an upper bound of the sectional curvatures in M.*

If the step sizes $(t_k)_{k\geq 1}$ satisfy

$$\delta \leq \frac{R_0}{2} \wedge \frac{2}{\beta} \operatorname{arctanh}\left(\tanh(\beta R_0/2)\cos(\alpha R)\tan(\alpha R_0/4)\right), \tag{12}$$

$$\lim_{k\to\infty} t_k = 0, \qquad \sum_{k=1}^{\infty} t_k = +\infty \quad \text{and} \quad \sum_{k=1}^{\infty} t_k^2 < \infty. \tag{13}$$

then the sequence $(x_k)_{k\geq 1}$ generated by the algorithm satisfies

$$\lim_{k\to\infty} \rho(x_k, c) = 0. \tag{14}$$

Remark 1. *In practice ν is given and one takes any ball $B(o,R)$ which contains its support. We need the condition $R < R_\alpha$. One should take R as small as possible for R_0 and then δ being not too small. The best choice is $o = c$ and $R = H(c)$ but they are not known a priori. If ν has a finite support one can take for o a point of the support of ν and for R the maximal distance from this point to another point of the support. It always works in a simply connected manifold of negative curvature since in this case α can be taken as small as we want. This is the case in our two main examples considered in Section 4, namely the hyperbolic space and the set of positive definite symmetric matrices with our specific choice of metric. Note that in this situation R_0 and δ can also be taken as large as we want.*

Proof:

First we prove that for all $r \in [R_0, R]$, if $x_k \in B(c,r)$ then $x_{k+1} \in B(c,r)$: if $\rho(x_k, c) \leq R_0/2$ it is clear since $\delta \leq R_0/2$. If $\rho(x_k, c) \geq R_0/2$ we prove that $\rho(x_{k+1}, c) \leq \rho(x_k, c)$. Let $y_{k+1} = \gamma_{H(x_k)}(v(x_k, \nu))$: $y_{k+1} \in \text{supp}(\nu)$ is such that $H(x_k) = \rho(x_k, y_{k+1})$; consider the triangle cx_ky_{k+1}. Let $a = \rho(x_k, y_{k+1})$, $b = \rho(y_{k+1}, c)$ and $r = \rho(c, x_k)$, $\hat{x}_k$ the angle corresponding to the point x_k. By Alexandrov comparison theorem (in fact Corollary 1 in the appendix) $\hat{x}_k$ is smaller than the same in constant curvature α^2. This together with the law of cosines in spherical geometry (Theorem 4 in the appendix) yields

$$\cos \hat{x}_k \geq \frac{\cos \alpha b - \cos \alpha r \cos \alpha a}{\sin \alpha r \sin \alpha a}.$$

Now $r \geq R_0/2$, $b \leq r^*$ and $a \geq r^*$ so

$$\cos \hat{x}_k \geq \frac{\cos \alpha r^*(1 - \cos(\alpha R_0/2))}{\sin(\alpha R_0/2)} = \cos \alpha r^* \tan(\alpha R_0/4) \geq \cos \alpha R \tan(\alpha R_0/4). \tag{15}$$

Consider now the triangle cx_kx_{k+1} and let $f = \rho(c, x_{k+1})$. Recall $\rho(x_k, x_{k+1}) = t_{k+1}$. Now by Toponogov theorem (Theorem 3 in the appendix) f is smaller than the same in constant curvature $-\beta^2$. This together with first law of cosines in hyperbolic geometry (Theorem 4 in the appendix) yields

$$\cosh \beta f \leq \cosh \beta r \cosh \beta t_{k+1} - \cos \hat{x}_k \sinh \beta r \sinh \beta t_{k+1} \tag{16}$$

which implies by Eq. 15

$$\cosh \beta f \leq \cosh(\beta r) \cosh \beta t_{k+1} - \cos \alpha R \tan(\alpha R_0/4) \sinh(\beta r) \sinh \beta t_{k+1}. \tag{17}$$

Let us check that the condition on δ implies that the right hand side is smaller than $\cosh \beta r$: we want to prove

$$\cosh(\beta r)(\cosh \beta t_{k+1} - 1) \leq \cos \alpha R \tan(\alpha R_0/4) \sinh(\beta r) \sinh \beta t_{k+1}$$

or equivalently

$$\frac{\cosh \beta t_{k+1} - 1}{\sinh \beta t_{k+1}} \leq \cos \alpha R \tan(\alpha R_0/4) \tanh(\beta r). \tag{18}$$

But

$$\frac{\cosh \beta t_{k+1} - 1}{\sinh \beta t_{k+1}} = \tanh\left(\frac{\beta t_{k+1}}{2}\right)$$

and $t_{k+1} \leq \delta$, $r \geq R_0/2$, so that Eq. 18 is implied by

$$\tanh\left(\frac{\beta\delta}{2}\right) \leq \cos \alpha R \tan(\alpha R_0/4) \tanh\left(\frac{\beta R_0}{2}\right). \tag{19}$$

Now clearly the condition on δ implies Eq. 19.

So we have proved that $\rho(c, x_{k+1}) \leq \rho(c, x_k)$.

Then we prove that there exists $\eta > 0$ such that if $x_k \in B(c, R)\backslash B(c, R_0)$ then

$$\frac{\cosh(\beta\rho(c, x_{k+1}))}{\cosh(\beta\rho(c, x_k))} \leq 1 - \eta t_{k+1}. \tag{20}$$

From Eq. 17, we obtain

$$\begin{aligned}
\frac{\cosh \beta f}{\cosh \beta r} &\leq \cosh \beta t_{k+1} - \cos \alpha R \tan(\alpha R_0/4) \tanh(\beta r) \sinh \beta t_{k+1} \\
&\leq \cosh \beta t_{k+1} - \cos \alpha R \tan(\alpha R_0/4) \tanh(\beta R_0) \sinh \beta t_{k+1} \\
&\leq 1 - 2\left(\cos \alpha R \tan(\alpha R_0/4) \tanh(\beta R_0) \cosh(\beta t_{k+1}/2)\right. \\
&\quad \left. - \sinh(\beta t_{k+1}/2)\right) \sinh(\beta t_{k+1}/2) \\
&\leq 1 - \left(\cos \alpha R \tan(\alpha R_0/4) \tanh(\beta R_0) \cosh(\beta t_{k+1}/2) - \sinh(\beta t_{k+1}/2)\right) \beta t_{k+1} \\
&\leq 1 - \left(\cos \alpha R \tan(\alpha R_0/4) \tanh(\beta R_0)\right. \\
&\quad \left. - \cos \alpha R \tan(\alpha R_0/4) \tanh(\beta R_0/2)\right) \cosh(\beta t_{k+1}/2) \beta t_{k+1}
\end{aligned}$$

where we used Eq. 12 in the last inequality. So

$$\begin{aligned}\frac{\cosh \beta\rho(c, x_{k+1})}{\cosh \beta\rho(c, x_k)} \leq\ & 1 - (\cos \alpha R \tan(\alpha R_0/4) \tanh(\beta R_0) \\ & - \cos \alpha R \tan(\alpha R_0/4) \tanh(\beta R_0/2))\beta t_{k+1}\end{aligned} \tag{21}$$

and this gives Eq. 20.

At this stage, since $\sum_{k=1}^{\infty} t_k = \infty$, we can conclude that there exists k_0 such that $\cosh(\beta\rho(c, x_{k_0})) \leq \cosh(\beta R_0)$ so $x_{k_0} \in B(c, R_0)$. Moreover from the first part of the proof we have that for all $k \geq k_0$, $x_k \in B(c, R_0)$.

Now we use the fact that on $B(c, R_0)$, H is convex and satisfies Eq. 5. By boundedness of the Hessian of square distance to c (see Yang (2010) Lemma 1.1 for details), we have for $k \geq k_0$

$$\begin{aligned}&\rho^2(c, x_{k+1}) \leq \\ &\quad \rho^2(c, x_k) - 2t_{k+1} \left\langle \exp_{x_k}^{-1} c, \dot{\gamma}_0(v(x_k, \nu)) \right\rangle + C\left(\frac{R_\alpha + R}{2}, \beta\right) t_{k+1}^2\end{aligned} \tag{22}$$

with

$$C(r, \beta) = 2r\beta \mathrm{cotanh}(2\beta r). \tag{23}$$

Now letting $y_{k+1} = \gamma_{H(x_k)}(v(x_k, \nu))$ we have $H \geq \rho(\cdot, y_{k+1})$ since $y_{k+1} \in \mathrm{supp}(\nu)$. We remark that $\rho^2(\cdot, y_{k+1})$ is convex on $B(c, R_0)$ by the fact that for all $z \in B(c, R_0)$ and $y \in \mathrm{supp}(\nu)$, $\rho(z, y) < R_\alpha$. Moreover we have $H(x_k) = \rho(x_k, y_{k+1})$. As a consequence, we get

$$\begin{aligned}H(c) - H(x_k) \quad & \geq \rho^2(c, y_{k+1}) - \rho^2(x_k, y_{k+1}) \\ & \geq -2\left\langle \exp_{x_k}^{-1} c, \dot{\gamma}_0(v(x_k, \nu)) \right\rangle\end{aligned}$$

and this implies by Lemma 1

$$-2\left\langle \exp_{x_k}^{-1} c, \dot{\gamma}_0(v(x_k, \nu)) \right\rangle \leq -\tau\rho^2(c, x_k). \tag{24}$$

Plugging into Eq. 22 yields

$$\rho^2(c, x_{k+1}) \leq (1 - \tau t_{k+1})\rho^2(c, x_k) + C\left(\frac{R_\alpha + R}{2}, \beta\right) t_{k+1}^2. \tag{25}$$

We recall from here the standard argument to prove that $\rho^2(c, x_k)$ converges to 0. Let

$$a = \limsup_{k\to\infty} \rho^2(c, x_k).$$

Iterating Eq. 25 yields for $\ell \geq 1$

$$\rho^2(c, x_{k+\ell}) \leq \prod_{j=1}^{\ell}(1 - \tau t_{k+j})\rho^2(c, x_k) + C\sum_{j=1}^{\ell} t_{k+j}^2$$

with $C = C\left(\frac{R_\alpha+R}{2}, \beta\right)$. Letting $\ell \to \infty$ and using the fact that $\sum_{j=1}^{\infty} t_{k+j} = \infty$, which implies

$$\prod_{j=1}^{\infty}(1 - \tau t_{k+j}) = 0,$$

we get

$$a \leq C\sum_{j=1}^{\infty} t_{k+j}^2.$$

Finally using $\sum_{j=1}^{\infty} t_j^2 < \infty$ we obtain that $\lim_{k\to\infty}\sum_{j=1}^{\infty} t_{k+j}^2 = 0$, so $a = 0$.

□

Remark 2. *In Theorem 1, it looks difficult to find a larger δ. The choice is almost optimal to have $\rho(c, x_{k+1}) \leq \rho(c, x_k)$ outside $B(c, R_0)$. On the other hand Eq. 21 yields an explicit value for η in Eq. 20 and this in turn can be used to find an explicit $\eta' > 0$ such that*

$$\rho^2(c, x_{k+1}) \leq (1 - \eta' t_{k+1})\rho^2(c, x_k), \quad t_{k+1} \leq \delta \wedge 1/\eta'. \tag{26}$$

For the speed of convergence, taking $t_k = \dfrac{r}{k+1}$, we proceed as in Proposition 4.10 of Yang (2010). We use the following lemma, borrowed from the paper of Nedic and Bertsekas (2000):

Lemma 2. *Let $(u_k)_{k\geq 1}$ be a sequence of nonnegative real numbers such that*

$$u_{k+1} \leq \left(1 - \frac{\lambda}{k+1}\right)u_k + \frac{\xi}{(k+1)^2}$$

where λ and ξ are positive constants. Then

$$u_{k+1} \leq \begin{cases} \frac{1}{(k+1)^{\lambda}}\left(u_0 + \frac{2^{\lambda}\xi(2-\lambda)}{1-\lambda}\right) & \text{if} \quad 0 < \lambda < 1; \\ \frac{\xi(1+\ln(k+1))}{k+1} & \text{if} \quad \lambda = 1; \\ \frac{1}{(\lambda-1)(k+2)}\left(\xi + \frac{(\lambda-1)u_0-\xi}{(k+2)^{\lambda-1}}\right) & \text{if} \quad \lambda > 1. \end{cases}$$

Proposition 1. *Choosing $t_k = \dfrac{r}{k+1}$, letting k_0 such that for all $k \geq k_0$, $x_k \in B(c, R_0)$,*

$$\rho^2(x_{k_0+k}, c) \leq \begin{cases} \frac{1}{(k+1)^{\lambda}}\left(R_0^2 + \frac{2^{\lambda}\xi(2-\lambda)}{1-\lambda}\right) & \text{if} \quad 0 < \lambda < 1; \\ \frac{\xi(1+\ln(k+1))}{k+1} & \text{if} \quad \lambda = 1; \\ \frac{1}{(\lambda-1)(k+2)}\left(\xi + \frac{(\lambda-1)R_0^2-\xi}{(k+2)^{\lambda-1}}\right) & \text{if} \quad \lambda > 1. \end{cases}$$

where $\lambda = \tau r$ (with τ given in Lemma 1) and $\xi = r^2 C\left(\frac{R_\alpha+R}{2}, \beta\right)$.

Proof:
This is a direct consequence of lemma 2 and inequality Eq. 25, valid for $k \geq k_0$. □

Remark 3. *From the estimate of η given by Eq. 21 one can get an estimate of k_0. Another possibility is to replace τ by $\tau \wedge \eta'$ in Eq. 25 with η' defined in Eq. 26. Then Proposition 1 is valid for all $k \geq 1$ without the condition $x_k \in B(c, R_0)$.*

Remark 4. *The proof of Theorem 1 works for R_0 defined in Eq. 11. It also works for any smaller positive value. It is better to have R_0 large so that x_k rapidly enters the ball $B(c, R_0)$. On the other hand when R_0 is small and x_k is already in this ball then one can take τ close to $\frac{\alpha}{2}\text{cotan}(\alpha R_\alpha)$. Again explicit estimates are possible.*

4. Two case studies

In order to implement algorithm GEO-ALG (a specialization of RIE-ALG for point clouds with step sizes $t_i = \frac{1}{i+1}$), we need to describe the geodesics of the underlying manifold, and find an intermediate point $m = \text{Geodesic}(p, q, t)$ on the geodesic passing through p and q such that $\rho(p, m)$=t $\rho(p, q)$.

4.1. Hyperbolic manifold

A hyperbolic manifold is a complete Riemannian d-dimensional manifold of constant sectional curvature -1 that is isometric to the real hyperbolic space. There exists several models of hyperbolic geometry. Here, we consider the planar non-conformal Klein model where geodesics are straight lines. See Nielsen and Nock (2010). Although there exists no known closed-form formula for the hyperbolic centroid ($p = 2$), Welzl's minimax algorithm generalizes to the Klein disk as described in Nielsen and Nock (2010) to compute exactly the hyperbolic 1-center. The Klein Riemannian distance on the unit disk is defined by

$$\rho(p, q) = \operatorname{arccosh} \frac{1 - p^\top q}{\sqrt{(1 - p^\top p)(1 - q^\top q)}} \tag{27}$$

where $\operatorname{arccosh}(x) = \log(x + \sqrt{x^2 - 1})$, and the geodesic passing through p and q is the straight line segment

$$\gamma_t(p, q) = (1 - t)p + tq, \; t \in [0, 1]. \tag{28}$$

Finding m such that $\rho(p, m)$=t $\rho(p, q)$ cannot be solved in closed-form solution (except for $t = \frac{1}{2}$, see Nielsen and Nock (2010)), so that we rather proceed by a bisection search algorithm on parameter t up to machine precision. Figure 1 shows the snapshots of our implementation in Java Processing.[4]

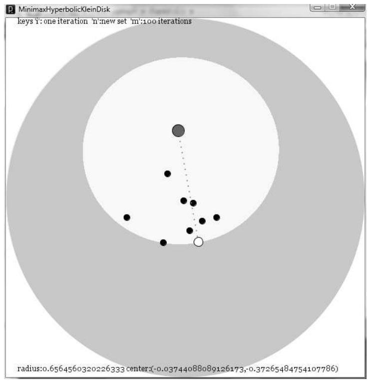
Figure 2 plots the convergence rate of the GEO-ALG algorithm. The code is publicly available on-line for reproducible research.

4.2. Manifold of symmetric positive definite matrices

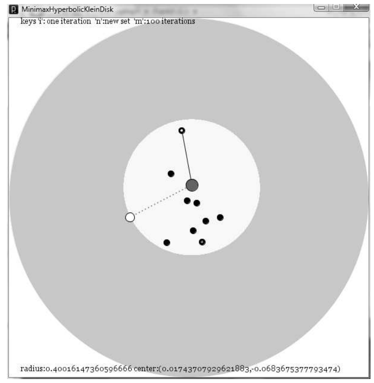
A $d \times d$ matrix M with real entries is said symmetric positive definite (SPD) iff. it is symmetric ($M = M^\top$), and that for all $x \neq 0$, $x^\top M x > 0$. The set of $d \times d$ SPD matrices forms a smooth manifold of dimension $\frac{d(d+1)}{2}$. We refer to Lang (1999) (Chapter 12) for a description of the geometry of SPD matrices. See also the work of Ji (2007) for optimization on matrix manifolds. The geodesic linking (matrix) point P to point Q is given by

$$\gamma_t(P, Q) = P^{\frac{1}{2}} \left(P^{-\frac{1}{2}} Q P^{-\frac{1}{2}} \right)^t P^{\frac{1}{2}}, \tag{29}$$

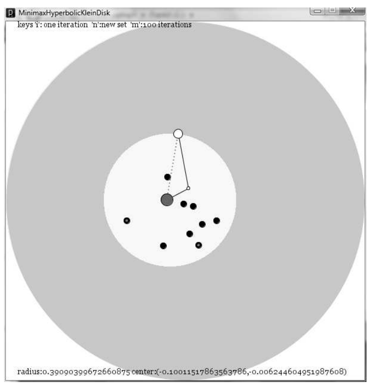
[4] `processing.org`



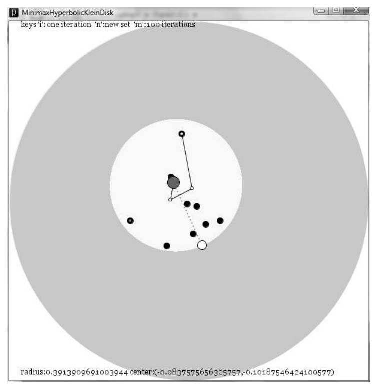
Initialization



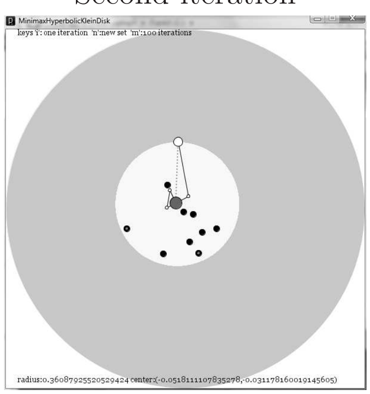
First iteration



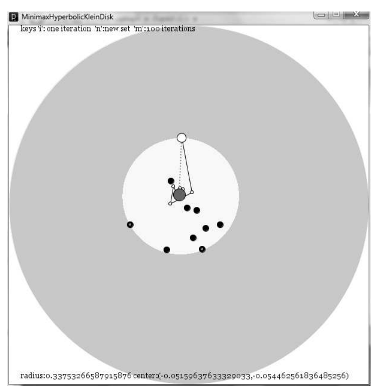
Second iteration



Third iteration

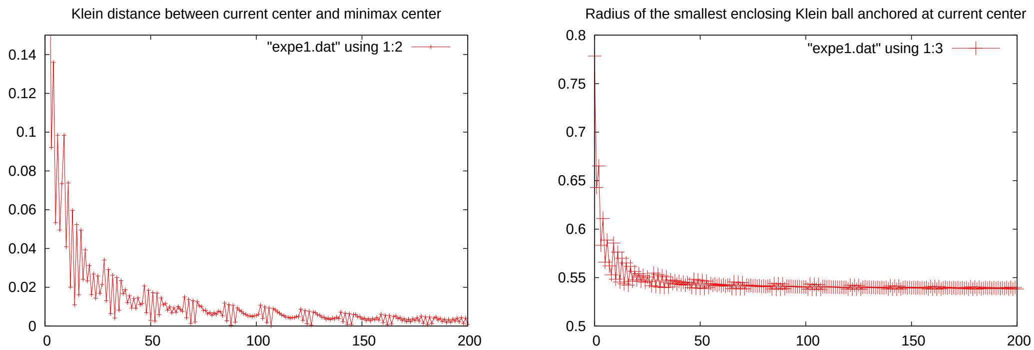


Fourth iteration



after 104 iterations

Figure 1: Snapshots of the GEO-ALG algorithm implemented for the hyperbolic Klein disk: The large black disk and the white disk denote the current center and farthest point, respectively. The linked path shows the trajectory of the centers as the number of iterations increase. On-line demo available at `http://www.informationgeometry.org/RiemannMinimax/`



(a) (b)

Figure 2: Convergence rate of the GEO-ALG algorithm for the hyperbolic disk for the first 200 iterations. The horizontal axis denotes the number of iterations and the vertical axis (a) the relative Klein distance between the current center and the optimal 1-center (approximated for a large number of iterations), (b) the radius of the smallest enclosing ball anchored at the current center.

where the matrix function $h(M)$ is computed from the singular value decomposition $M = UDV^\top$ (with U and V unitary matrices and $D = \mathrm{diag}(\lambda_1, ..., \lambda_d)$ a diagonal matrix of eigenvalues) as $h(M) = U\mathrm{diag}(h(\lambda_1), ..., h(\lambda_d))V^\top$. For example, the square root function of a matrix is computed as $M^{\frac{1}{2}} = U\mathrm{diag}(\sqrt{\lambda_1}, ..., \sqrt{\lambda_d})V^\top$.

In this case, finding t such that

$$\|\log(P^{-1}Q)^t\|_F^2 = r\|\log P^{-1}Q\|_F^2, \tag{30}$$

where $\|\cdot\|_F$ denotes the Fröbenius norm yields to $t = r$. Indeed, consider $\lambda_1, ..., \lambda_d$ the eigenvalues of $P^{-1}Q$, then Eq. 30 amounts to find

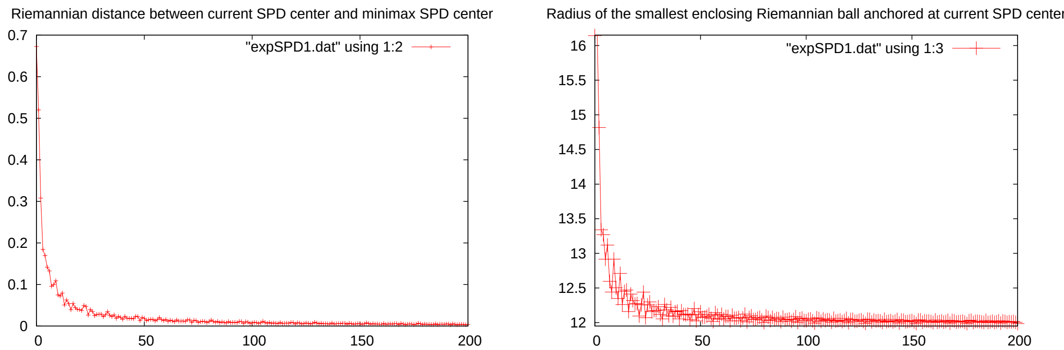
$$\sum_{i=1}^{d} \log^2 \lambda_i^t = t^2 \sum_{i=1}^{d} \log^2 \lambda_i = r^2 \sum_{i=1}^{d} \log^2 \lambda_i. \tag{31}$$

That is $t = r$.

Figure 3 displays the plots of the convergence rate of the algorithm for the SPD manifold.

5. Concluding remarks and discussion

We described a generalization of the 1-center algorithm of Bădoiu and Clarkson (2003) to arbitrary Riemannian geometry, and proved the conver-



(a) (b)

Figure 3: Convergence rate of the GEO-ALG algorithm for the SPD Riemannian manifold (dimension 5) for the first 200 iterations. The horizontal axis denotes the number of iterations i and the vertical axis (a) the relative Riemannian distance between the current center c_i and the optimal 1-center c^* ($\frac{\rho(c*, c_i)}{r^*}$, where $\rho*$ and r^* are approximated for a large number of iterations), (b) the radius r_i of the smallest enclosing SPD ball anchored at the current center.

gence under mild assumptions. This proves the existence of Riemannian coresets for optimization. This 1-center building block can be used for k-center clustering. Furthermore, the algorithm can be straightforwardly extended to sets of geodesic balls.

An open-source source code implementation in Java™ for reproducible research is available on-line at

`http://www.informationgeometry.org/RiemannMinimax/`

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. FN (5793b870) thanks Mr. Prasenjit Saha for discussions related to this topic, and gratefully acknowledge financial support from French funding agency ANR (GAIA 07-BLAN-0328-01) and Sony Computer Science Laboratories, Inc.

6. Appendix: Some notions of Riemannian geometry

In this section, we recall some basic notions of Riemannian geometry used throughout the paper. For a complete presentation, we refer to Cheeger and

Ebin (1975).

We let M be a Riemannian manifold and $\langle\cdot,\cdot\rangle$ the Riemannian metric, which is a definite positive bilinear form on each tangent space T_xM, and depends smoothly on x. The associated norm in T_xM will be denoted by $\|\cdot\|$: $\|u\| = \langle u, u\rangle^{1/2}$. We denote by $\rho(x, y)$ the distance between two points on the manifold M:

$$\rho(x,y) = \inf\left\{\int_0^1 \|\dot\varphi(t)\|\,\mathrm{d}t, \quad \varphi \in C^1([0,1], M), \quad \varphi(0) = x, \quad \varphi(1) = y\right\}.$$

A *geodesic* in M is a smooth path which locally minimizes the distance between two points. In general such a curve does not minimize it globally. However it is true in all the sets we are considering in this paper. Given a vector $v \in TM$ with base point x, there is a unique geodesic started at x with speed v at time 0. It is denoted by $t \mapsto \exp_x(tv)$ or compactly by $t \mapsto \gamma_t(v)$. It depends smoothly on v but it has in general finite lifetime. A geodesic defined on a time interval $[a, b]$ is said to be *minimal* if it minimizes the distance from the image of a to the image of b. If the manifold is complete, taking $x, y \in M$, there exists a minimal geodesic from x to y in time 1. In all the scenarii we are considering in this paper, the minimal geodesic is unique and depends smoothly on x and y, and we denote it by $\gamma_\cdot(x, y) : [0, 1] \to M$, $t \mapsto \gamma_t(x, y)$ with the conditions $\gamma_0(x, y) = x$ and $\gamma_1(x, y) = y$. A subset U of M is said to be *convex* if for any $x, y \in U$, there exists a unique minimal geodesic $\gamma_\cdot(x, y)$ in M from x to y, this geodesic fully lies in U and depends smoothly on x, y, t.

The *injectivity radius* of M, denoted by $\mathrm{inj}(M)$, is the largest $r > 0$ such that for all $x \in M$, the map $\exp_x$ restricted to the open ball in T_xM centered at 0 with radius r is an embedding.

Given $x \in M$, u, v two non collinear vectors in T_xM, the *sectional curvature* $\mathrm{Sect}(u, v) = K$ is a number which gives information on how the geodesics issued from x behave near x. More precisely the image by $\exp_x$ of the circle centered at 0 of radius $r > 0$ in $\mathrm{Span}(u, v)$ has length

$$2\pi S_K(r) + o(r^3) \quad \text{as} \quad r \to 0$$

with

$$S_K(r) = \begin{cases} \frac{\sin(\sqrt{K}r)}{\sqrt{K}} & \text{if} \quad K > 0, \\ r & \text{if} \quad K = 0, \\ \frac{\sinh(\sqrt{-K}r)}{\sqrt{-K}} & \text{if} \quad K < 0. \end{cases}$$

For instance, if $K > 0$, $\exp_x(\mathrm{Span}(u, v))$ is near x approximatively a 2-dimensional sphere with radius $\frac{1}{\sqrt{K}}$. In fact, if M is simply connected and all the sectional curvatures are equal to the same $K > 0$, then M is a d-dimensional sphere with radius $\frac{1}{\sqrt{K}}$, where d is the dimension of M. If M is simply connected and all the sectional curvatures are equal to the same $K < 0$, we say that M is a d-dimensional hyperbolic space with curvature K.

An upper bound (resp. lower bound) of sectional curvatures is a number a such that for all non collinear u, v in the same tangent space, $\mathrm{Sect}(u, v) \leq a$ (resp. $\mathrm{Sect}(u, v) \geq a$). In the paper, we used a positive upper bound α^2 and a negative lower bound $-\beta^2$, $\alpha, \beta > 0$.

The existence of the upper bound α^2 for sectional curvatures makes possible to compare geodesic triangles, by *Alexandrov* theorem (see Chavel (2003)).

Theorem 2. *Let $x_1, x_2, x_3 \in M$ satisfy $x_1 \neq x_2$, $x_1 \neq x_3$ and*

$$\rho(x_1, x_2) + \rho(x_2, x_3) + \rho(x_3, x_1) < 2 \min \left\{ \mathrm{inj} M, \frac{\pi}{\alpha} \right\}$$

where $\alpha > 0$ is such that α^2 is an upper bound of sectional curvatures. Let the minimizing geodesic from x_1 to x_2 and the minimizing geodesic from x_1 to x_3 make an angle θ at x_1. Denoting by $S^2_{\alpha^2}$ the 2-dimensional sphere of constant curvature α^2 (hence of radius $1/\alpha$) and $\tilde{\rho}$ the distance in $S^2_{\alpha^2}$, we consider points $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3 \in S^2_{\alpha^2}$ such that $\rho(x_1, x_2) = \tilde{\rho}(\tilde{x}_1, \tilde{x}_2)$, $\rho(x_1, x_3) = \tilde{\rho}(\tilde{x}_1, \tilde{x}_3)$. Assume that the minimizing geodesic from $\tilde{x}_1$ to $\tilde{x}_2$ and the minimizing geodesic from $\tilde{x}_1$ to $\tilde{x}_3$ also make an angle θ at $\tilde{x}_1$.

Then we have $\rho(x_2, x_3) \geq \tilde{\rho}(\tilde{x}_2, \tilde{x}_3)$.

Instead of prescribing the angle in the comparison triangle in the sphere, it is possible to prescribe the third distance:

Corollary 1. *The assumption are the same as in Theorem 2 except that we assume that $\rho(x_2, x_3) = \tilde{\rho}(\tilde{x}_2, \tilde{x}_3)$ (all the distances are equal), but the minimizing geodesic from $\tilde{x}_1$ to $\tilde{x}_2$ and the minimizing geodesic from $\tilde{x}_1$ to $\tilde{x}_3$ now make an angle $\tilde{\theta}$ at $\tilde{x}_1$.*

Then we have $\tilde{\theta} \geq \theta$.

There also exists a comparison result in the other direction, called Topogonov's theorem.

Theorem 3. *Assume $\beta > 0$ is such that $-\beta^2$ is a lower bound for sectional curvatures in M. Let $x_1, x_2, x_3 \in M$ satisfy $x_1 \neq x_2$, $x_1 \neq x_3$. Let the minimizing geodesic from x_1 to x_2 and the minimizing geodesic from x_1 to x_3 make an angle θ at x_1. Denoting by $H^2_{-\beta^2}$ the hyperbolic 2-dimensional space of constant curvature $-\beta^2$ and $\tilde{\rho}$ the distance in $H^2_{-\beta^2}$, we consider points $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3 \in H^2_{-\beta^2}$ such that $\rho(x_1, x_2) = \tilde{\rho}(\tilde{x}_1, \tilde{x}_2)$, $\rho(x_1, x_3) = \tilde{\rho}(\tilde{x}_1, \tilde{x}_3)$. Assume that the minimizing geodesic from $\tilde{x}_1$ to $\tilde{x}_2$ and the minimizing geodesic from $\tilde{x}_1$ to $\tilde{x}_3$ also make an angle θ at $\tilde{x}_1$.*

Then we have $\rho(x_2, x_3) \leq \tilde{\rho}(\tilde{x}_2, \tilde{x}_3)$.

Triangles in the sphere $S^2_{\alpha^2}$ and in the hyperbolic space $H^2_{-\beta^2}$ have explicit relations between distance and angles as we will see below. This combined with Theorems 2 and 3 and Corollary 1 allow to find related bounds in M, which are intensively used in our proofs.

In this paper, we only use the *first law of cosines* in $S^2_{\alpha^2}$ and in $H^2_{-\beta^2}$ (see e.g., the paper of Ratcliffe (1994) Theorem 2.5.3 and Theorem 3.5.3).

Theorem 4. *If $\theta_1, \theta_2, \theta_3$ are the angles of a triangle in $S^2_{\alpha^2}$ and x_1, x_2, x_3 are the lengths of the opposite sides, then*

$$\cos\theta_3 = \frac{\cos(\alpha x_3) - \cos(\alpha x_1)\cos(\alpha x_2)}{\sin(\alpha x_1)\sin(\alpha x_2)}.$$

If $\theta_1, \theta_2, \theta_3$ are the angles of a triangle in $H^2_{-\beta^2}$ and x_1, x_2, x_3 are the lengths of the opposite sides, then

$$\cos\theta_3 = \frac{\cosh(\beta x_1)\cosh(\beta x_2) - \cosh(\beta x_3)}{\sinh(\beta x_1)\sinh(\beta x_2)}.$$

References

Afsari, B., 2009. Means and averaging on Riemannian manifolds. Ph.D. thesis, University of Maryland.

Afsari, B., February 2011. Riemannian L^p center of mass: existence, uniqueness, and convexity. Proceedings of the American Mathematical Society 139, 655–674.

Arnaudon, M., Dombry, C., Phan, A., Yang, L., 2010. Stochastic algorithms for computing means of probability measures.
Stochastic Processes and their Applications 122 (2012), pp. 1437-1455

Berger, M., 2003. A panoramic view of Riemannian geometry. Springer Verlag, Berlin.

Bonnabel, S., 2011. Stochastic gradient descent on manifolds. arXiv:1111.5280v2

Bădoiu, M., Clarkson, K. L., 2003. Smaller core-sets for balls. In Proceedings of the fourteenth annual ACM-SIAM symposium on Discrete algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, pp. 801–802.

Bădoiu, M., Clarkson, K. L., May 2008. Optimal core-sets for balls. Computational Geometry: Theory and Applications 40, 14–22.

Chavel, I., 2006. Riemannian geometry: A modern introduction. Cambridge University Press, 2nd edition, 2006.

Cheeger, J., Ebin, D.G., 1975. Comparison Theorems in Riemannian Geometry. North-Holland mathematical library, Vol. 9.

Ji, H., 2007. Optimization approaches on smooth manifolds. PhD thesis, Australian National University.

Lang, S., 1999. Fundamentals of differential geometry. Vol. 191 of Graduate Texts in Mathematics. Springer-Verlag, New York.

Nedic, A., Bertsekas, D., 2000. Convergence rate of incremental subgradient algorithms. In Stochastic Optimization: Algorithms and Applications. Kluwer, pp. 263–304.

Nielsen, F., Nock, R., 2009. Approximating smallest enclosing balls with applications to machine learning. Int. J. Comput. Geometry Appl. 19 (5), 389–414.

Nielsen, F., Nock, R., 2010. Hyperbolic Voronoi diagrams made easy. In: International Conference on Computational Science and its Applications (ICCSA). IEEE Computer Society, Los Alamitos, CA, USA, pp. 74–80.

Nock, R., Nielsen, F., 2005. Fitting the smallest enclosing Bregman ball. In: European Conference on Machine Learning (ECML). pp. 649–656.

Turaga, P., Veeraraghavan, A., Srivastava, A., Chellappa, R., 2010. Statistical computations on Grassmann and Stiefel manifolds for image and video based recognition. IEEE Trans. Pattern Anal. Mach. Intell. (PAMI).

Pennec, X., 2008. Statistical computing on manifolds: From Riemannian geometry to computational anatomy. In Emerging Trends in Visual Computing (ETVC), F. Nielsen (Ed). pp. 347–386.

Ratcliffe, J., 1994. Foundations of hyperbolic manifolds. Graduate texts in Mathematics, Springer-Verlag, 1994.

J.J. Sylvester, 1857. A Question in the Geometry of Situation, Quarterly Journal of Mathematics, Vol. 1, p. 79

Tsang, I. W., Kocsor, A., Kwok, J. T., 2007. Simpler core vector machines with enclosing balls. In Proceedings of the 24th international conference on Machine learning. ACM, New York, NY, USA, pp. 911–918.

Welzl, E., 1991. Smallest enclosing disks (balls and ellipsoids). In: Maurer, H. (Ed.), New Results and New Trends in Computer Science. LNCS. Springer.

Yang, L., 2010. Riemannian Median and Its Estimation. In: LMS Journal of Computation and Mathematics, Vol 13 (2010), pp 461–479.