

Riemannian median and its estimation

Le Yang

ABSTRACT

In this paper, we define the geometric median for a probability measure on a Riemannian manifold, give its characterization and a natural condition to ensure its uniqueness. In order to compute the geometric median in practical cases, we also propose a subgradient algorithm and prove its convergence as well as estimating the error of approximation and the rate of convergence. The convergence property of this subgradient algorithm, which is a generalization of the classical Weiszfeld algorithm in Euclidean spaces to the context of Riemannian manifolds, also improves a recent result of P. T. Fletcher *et al.* [*NeuroImage* 45 (2009) S143–S152].

1. Introduction

The classical Fermat point of a triangle in the plane is the point minimizing the sum of distances to the three vertices. This is a prototype of the more general Fermat–Weber problem concerning the same question but for more than three points and in higher dimensions. The solution to this problem is called the geometric median of these points and provides them with a notion of centrality. For this reason, the geometric median is a natural estimator in statistics which possesses another important property called robustness, that is, not sensitive to outliers. As a consequence, the geometric median is a widely used robust estimator in both theoretical and practical theory of robust statistics.

Naturally, one can also ask the question of how to find a point that minimizes the sum of distances to a set of given points in a much more general space as long as it carries a distance. This has been done in [30] where the existence of the geometric median of a probability measure on a complete, separable and finitely compact metric space is proved. Recently, there has been a growing interest in the methods that characterize statistical data lying on a Riemannian manifold and their applications; see for example [6–9, 15, 16, 28], where the centrality of empirical data is modeled by the Riemannian barycenter which was first introduced by Karcher in his paper [19] and then has been extensively studied and generalized by many other authors during the past thirty years [3–5, 13, 21, 22, 29]. As is well known, the barycenter is not a robust estimator and is sensitive to outliers: in order to overcome this drawback Fletcher *et al.* [17] defined the weighted geometric median of a finite set of discrete sample points lying on a Riemannian manifold and proved its existence, uniqueness and robustness. For other statistically-based work related to medians, we refer to [32] and the references therein.

In many cases, especially in practice, one often needs to calculate or at least estimate the value of the geometric median. The first algorithm to do this computation was proposed by Weiszfeld in his paper [34], which is now known as the Weiszfeld algorithm. From then on, due to the fundamental importance of the Fermat–Weber problem in location theory, this algorithm has been studied, improved on and generalized by numerous authors; two examples of the huge literature on this subject are [23, 27].

In the context of Riemannian manifolds, Fletcher *et al.* [17] proposed a Riemannian analogue of the Weiszfeld algorithm to compute the geometric median. They proved a

Received 17 November 2009; revised 26 May 2010.

2000 *Mathematics Subject Classification* 58C05, 60D05 (primary), 90C25, 90B85 (secondary).

The author is supported by a PhD fellowship Allocation de Recherche MRT and Thales Air Systems.

convergence result under the condition that the manifold is non-negatively curved and conjectured a similar convergence result in the negatively curved case.

The aim of this paper is to define the geometric median of a probability measure on a complete Riemannian manifold and investigate the question of uniqueness as well as the problem of approximation. As in [19, 24] we suppose that the support of the probability measure is contained in a convex ball and give a characterization of the geometric median which is proved in the case of Euclidean space in [23] for a finite discrete set of sample points. Then we prove the uniqueness for geometric medians under a natural condition imposed on the probability measure and show that this condition yields a strong convexity property, which is useful in error estimates. By regarding the Weiszfeld algorithm as a subgradient procedure, we introduce a subgradient algorithm to estimate geometric medians and prove that this algorithm always converges without condition of the sign of curvatures by generalizing the fundamental inequality in [14] where it was proved in positively curved manifolds. Finally, the results of approximating errors and rate of convergence are also obtained.

Throughout this paper, M is a complete Riemannian manifold with Riemannian metric $\langle \cdot, \cdot \rangle$ and Riemannian distance d . The gradient operator and the Hessian operator on M are denoted by grad and Hess , respectively. Moreover, for every point p in M , let d_p denote the distance function to p defined by $d_p(x) = d(x, p)$, $x \in M$.

We fix an open geodesic ball

$$B(a, \rho) = \{x \in M : d(x, a) < \rho\}$$

in M centered at a with a finite radius ρ . Let δ and Δ denote, respectively, a lower and an upper bound of sectional curvatures K in $\bar{B}(a, \rho)$. The injectivity radius of $\bar{B}(a, \rho)$ is denoted by $\text{inj}(\bar{B}(a, \rho))$. Furthermore, we assume that the radius of the ball verifies

$$\rho < \min \left\{ \frac{\pi}{4\sqrt{\Delta}}, \frac{\text{inj}(\bar{B}(a, \rho))}{2} \right\}, \tag{1.1}$$

where if $\Delta \leq 0$, then $\pi/(4\sqrt{\Delta})$ is interpreted as $+\infty$. Hence $\bar{B}(a, \rho)$ is convex (see [11, Theorem 5.14]); that is, for every two points x and y in $\bar{B}(a, \rho)$, there is a unique shortest geodesic from x to y in M , and this geodesic lies in $\bar{B}(a, \rho)$. Moreover, the geodesics in $\bar{B}(a, \rho)$ vary continuously with its endpoints. As a consequence, the angle comparison theorem of Alexandrov (see, for example, [1], [2, p. 3] and [10, Proposition II.4.9]) can be applied in $\bar{B}(a, \rho)$. Similarly, it is easy to check that the angle comparison theorem of Toponogov (see [11, Theorem 2.2]) can also be applied in $\bar{B}(a, \rho)$. Thus we have to introduce some notation of model spaces that provide us with geometric information.

NOTATION. Let κ be a real number. The model space M_κ^2 is defined as follows:

- (1) if $\kappa > 0$, then M_κ^2 is obtained from the sphere \mathbb{S}^2 by multiplying the distance function by $1/\sqrt{\kappa}$;
- (2) if $\kappa = 0$, then M_0^2 is the Euclidean space \mathbb{E}^2 ;
- (3) if $\kappa < 0$, then M_κ^2 is obtained from the hyperbolic space \mathbb{H}^2 by multiplying the distance function by $1/\sqrt{-\kappa}$.

The diameter of M_κ^2 will be denoted by D_κ . More precisely,

$$D_\kappa = \begin{cases} \pi/\sqrt{\kappa} & \text{if } \kappa > 0; \\ +\infty & \text{if } \kappa \leq 0. \end{cases}$$

The distance between two points \bar{A} and \bar{B} in M_κ^2 will be written as $\bar{d}(\bar{A}, \bar{B})$.

Moreover, we write for $t \in \mathbf{R}$,

$$S_\kappa(t) = \begin{cases} \sin(\sqrt{\kappa}t)/\sqrt{\kappa} & \text{if } \kappa > 0; \\ t & \text{if } \kappa = 0; \\ \sinh(\sqrt{-\kappa}t)/\sqrt{-\kappa} & \text{if } \kappa < 0. \end{cases}$$

For the necessity of later comparison arguments, we recall some terminology about triangles (see [10, p. 158 and Lemma I.2.14]). A *geodesic triangle* $\triangle ABC$ in M is a figure consisting of three distinct points A, B, C of M called the vertices and a choice of three shortest geodesics AB, BC, CA joining them called the sides. A *comparison triangle* in M_κ^2 of $\triangle ABC$ is a geodesic triangle $\triangle \bar{A}\bar{B}\bar{C}$ in M_κ^2 such that $\bar{d}(\bar{A}, \bar{B}) = d(A, B)$, $\bar{d}(\bar{B}, \bar{C}) = d(B, C)$, $\bar{d}(\bar{C}, \bar{A}) = d(C, A)$. Note that if the perimeter of $\triangle ABC$ is less than $2D_\kappa$, that is, if $d(A, B) + d(B, C) + d(C, A) < 2D_\kappa$, then its comparison triangle in M_κ^2 exists and it is unique up to an isometry. Hence every geodesic triangle in $\bar{B}(a, \rho)$ has its comparison triangles in M_δ^2 and M_Δ^2 , respectively. Note that, by convexity of $\bar{B}(a, \rho)$, every geodesic triangle in $\bar{B}(a, \rho)$ is uniquely determined by its three vertices.

The two estimations given in the lemma below are very useful to our further discussions; they are direct corollaries to the classical Hessian comparison theorems (see, for example, [31, Lemma IV.2.9] and [18, Theorem 4.6.1]).

LEMMA 1.1. *Let $p \in \bar{B}(a, \rho)$ and $\gamma : [0, b] \rightarrow \bar{B}(a, \rho)$ be a geodesic, then:*

- (i)
$$\text{Hess } d_p(\dot{\gamma}(t), \dot{\gamma}(t)) \geq D(\rho, \Delta) |\dot{\gamma}_p^{\text{nor}}(t)|^2$$

for every $t \in [0, b]$ such that $\gamma(t) \neq p$, where $D(\rho, \Delta) = S'_\Delta(2\rho)/S_\Delta(2\rho) > 0$ and $\dot{\gamma}_p^{\text{nor}}(t)$ is the normal component of $\dot{\gamma}(t)$ with respect to the geodesic from p to $\gamma(t)$ in $\bar{B}(a, \rho)$;

- (ii)
$$\text{Hess } \frac{1}{2}d_p^2(\dot{\gamma}(t), \dot{\gamma}(t)) \leq C(\rho, \delta) |\dot{\gamma}|^2$$

for every $t \in [0, b]$, where the constant $C(\rho, \delta) \geq 1$ is defined by

$$C(\rho, \delta) = \begin{cases} 1 & \text{if } \delta \geq 0; \\ 2\rho\sqrt{-\delta} \coth(2\rho\sqrt{-\delta}) & \text{if } \delta < 0. \end{cases}$$

2. Definition of a Riemannian median

As in [19], we consider a probability measure μ on M whose support is contained in the open ball $B(a, \rho)$ and define a function

$$f : \bar{B}(a, \rho) \longrightarrow \mathbf{R}_+, \quad x \longmapsto \int_M d(x, p)\mu(dp).$$

This function is 1-Lipschitz, hence continuous on the compact set $\bar{B}(a, \rho)$. Moreover, by the first estimation in Lemma 1.1, it is also convex. The convexity of f yields that its local minima coincide with its global ones, so that we do not need to distinguish the two. Now we give the following definition.

DEFINITION 1. A minimum point of f is called a median of μ . The set of all the medians of μ will be denoted by \mathfrak{M}_μ . The minimal value of f will be denoted by f_* .

It is easily seen that \mathfrak{M}_μ is compact and convex. In order to give a characterization of \mathfrak{M}_μ , we need the following proposition, which implies that f is not differentiable in general.

PROPOSITION 2.1. *Let $\gamma : [0, b] \rightarrow \bar{B}(a, \rho)$ be a geodesic, then*

$$\begin{aligned} \left. \frac{d}{dt} f(\gamma(t)) \right|_{t=t_0+} &= \langle \dot{\gamma}(t_0), H(\gamma(t_0)) \rangle + \mu\{\gamma(t_0)\}|\dot{\gamma}|, \quad t_0 \in [0, b), \\ \left. \frac{d}{dt} f(\gamma(t)) \right|_{t=t_0-} &= \langle \dot{\gamma}(t_0), H(\gamma(t_0)) \rangle - \mu\{\gamma(t_0)\}|\dot{\gamma}|, \quad t_0 \in (0, b], \end{aligned}$$

where for $x \in \bar{B}(a, \rho)$,

$$H(x) = \int_{M \setminus \{x\}} \frac{-\exp_x^{-1} p}{d(x, p)} \mu(dp),$$

is a tangent vector at x satisfying $|H(x)| \leq 1$. Particularly, if $\mu\{x\} = 0$, then $\text{grad } f(x) = H(x)$. Moreover, H is continuous on $\bar{B}(a, \rho) \setminus \text{supp}(\mu)$.

Proof. We only prove the first identity since the proof of the second one is similar. Let $t_0 \in [0, b)$ and $\varepsilon > 0$ be sufficiently small, then

$$\begin{aligned} \frac{f(\gamma(t_0 + \varepsilon)) - f(\gamma(t_0))}{\varepsilon} &= \int_M \frac{d(\gamma(t_0 + \varepsilon), p) - d(\gamma(t_0), p)}{\varepsilon} \mu(dp) \\ &= \int_{M \setminus \{\gamma(t_0)\}} \frac{d(\gamma(t_0 + \varepsilon), p) - d(\gamma(t_0), p)}{\varepsilon} \mu(dp) + \mu\{\gamma(t_0)\}|\dot{\gamma}|. \end{aligned}$$

Letting $\varepsilon \rightarrow 0+$ and using bounded convergence we obtain

$$\begin{aligned} \left. \frac{d}{dt} f(\gamma(t)) \right|_{t=t_0+} &= \int_{M \setminus \{\gamma(t_0)\}} \left. \frac{d}{dt} d(\gamma(t), p) \right|_{t=t_0} \mu(dp) + \mu\{\gamma(t_0)\}|\dot{\gamma}| \\ &= \int_{M \setminus \{\gamma(t_0)\}} \langle \dot{\gamma}(t_0), \text{grad } d_p(\gamma(t_0)) \rangle \mu(dp) + \mu\{\gamma(t_0)\}|\dot{\gamma}| \\ &= \langle \dot{\gamma}(t_0), H(\gamma(t_0)) \rangle + \mu\{\gamma(t_0)\}|\dot{\gamma}|. \quad \square \end{aligned}$$

Now we give the characterization of \mathfrak{M}_μ , which is proved in [23], for a finite number of points in an Euclidean space.

THEOREM 2.2. *The set of all the medians of μ is characterized by*

$$\mathfrak{M}_\mu = \{x \in \bar{B}(a, \rho) : |H(x)| \leq \mu\{x\}\}.$$

Proof. (C) Let $x \in \mathfrak{M}_\mu$. If $H(x) = 0$, then there is nothing to prove. So we assume that $H(x) \neq 0$. Consider the geodesic in $\bar{B}(a, \rho)$:

$$\gamma(t) = \exp_x \left(-t \frac{H(x)}{|H(x)|} \right), \quad t \in [0, b].$$

By Proposition 2.1 and the definition of \mathfrak{M}_μ we obtain

$$|H(x)| = \mu\{x\} - \left. \frac{d}{dt} f(\gamma(t)) \right|_{t=0+} \leq \mu\{x\}.$$

(D) Let $x \in \bar{B}(a, \rho)$ such that $|H(x)| \leq \mu\{x\}$. For every geodesic $\gamma : [0, 1] \rightarrow \bar{B}(a, \rho)$ with $\gamma(0) = x$ and $\gamma(1) = y$, by the convexity of f , Proposition 2.1 and the Cauchy–Schwartz inequality, we obtain

$$f(y) - f(x) \geq \left. \frac{d}{dt} f(\gamma(t)) \right|_{t=0+} \geq |\dot{\gamma}|(-|H(x)| + \mu\{x\}) \geq 0,$$

so that $x \in \mathfrak{M}_\mu$. □

In order to describe the location of \mathfrak{M}_μ , we need the following geometric lemma which is also useful in the next section.

LEMMA 2.3. *Let $\triangle ABC$ be a geodesic triangle in $\bar{B}(a, \rho)$ such that $\angle A \geq \pi/2$, then $\angle B < \pi/2$ and $\angle C < \pi/2$.*

Proof. We prove this for the case when $\Delta > 0$. The proof for the case when $\Delta \leq 0$ is similar. It suffices to show that $\angle B < \pi/2$. Let $d(B, C) = a_1$, $d(C, A) = b_1$ and $d(A, B) = c_1$. Consider a comparison triangle $\triangle \bar{A}\bar{B}\bar{C}$ of $\triangle ABC$ in M_Δ^2 . Since $K \leq \Delta$ in $\bar{B}(a, \rho)$, Alexandrov’s theorem yields that $\angle A \leq \angle \bar{A}$, $\angle B \leq \angle \bar{B}$. By the following identity in M_Δ^2 ,

$$\sin(\sqrt{\Delta}a_1) \cos \angle \bar{B} = \cos(\sqrt{\Delta}b_1) \sin(\sqrt{\Delta}c_1) - \sin(\sqrt{\Delta}b_1) \cos(\sqrt{\Delta}c_1) \cos \angle \bar{A},$$

we obtain $\angle \bar{B} < \pi/2$ and this completes the proof. □

PROPOSITION 2.4. \mathfrak{M}_μ is contained in the smallest closed convex subset of $B(a, \rho)$ containing the support of μ .

Proof. Let V be this set. By Theorem 2.2 it suffices to show that if $x \in \bar{B}(a, \rho) \setminus V$ then $H(x) \neq 0$. In fact, let y be a point in V such that $d(x, y) = \inf\{d(x, p) : p \in V\}$, then the convexity of V yields $\angle xyp \geq \pi/2$ for every $p \in V$. Hence by Lemma 2.3 we obtain $\angle pxy < \pi/2$ and this gives that

$$\begin{aligned} \langle H(x), \exp_x^{-1} y \rangle &= \int_V \frac{\langle -\exp_x^{-1} p, \exp_x^{-1} y \rangle}{d(x, p)} \mu(dp) \\ &= -d(x, y) \int_V \cos \angle pxy \mu(dp) < 0. \end{aligned}$$

The proof is completed by observing that $\exp_x^{-1} y \neq 0$. □

3. Uniqueness of the Riemannian median

In the Euclidean case, it is well known that if the sample points are not colinear, then the geometric median is unique. Hence we obtain a natural condition for μ to ensure the uniqueness for medians in the Riemannian case:

* *The support of μ is not totally contained in any geodesic. This means that for every geodesic $\gamma : [0, 1] \rightarrow \bar{B}(a, \rho)$, we have $\mu(\gamma[0, 1]) < 1$.*

Before giving the uniqueness theorem for medians, we introduce a procedure for extending geodesics in $\bar{B}(a, \rho)$. Let $\gamma : [0, 1] \rightarrow \bar{B}(a, \rho)$ be a geodesic such that $\gamma(0) = x$ and $\gamma(1) = y$. By the completeness of M and the fact that the diameter of $\bar{B}(a, \rho)$ equals $2\rho < \text{inj}(\bar{B}(a, \rho))$, we can extend γ from its endpoint y , along the direction of $\dot{\gamma}(1)$, to the point \hat{y} where the extended geodesic first hits the boundary of $\bar{B}(a, \rho)$. Similarly, we can also apply this procedure in the opposite direction: extend γ from its starting point x , along the direction of $-\dot{\gamma}(0)$, to the point \hat{x} where the extended geodesic firstly hits the boundary of $\bar{B}(a, \rho)$. Then we write $\hat{\gamma} : [0, 1] \rightarrow \bar{B}(a, \rho)$, the geodesic such that $\hat{\gamma}(0) = \hat{x}$ and $\hat{\gamma}(1) = \hat{y}$. Obviously, $\gamma[0, 1] \subset \hat{\gamma}[0, 1]$. Furthermore, the strong convexity of $B(a, \rho)$ (see [11, Theorem 5.14]) yields $\dot{\gamma}_p^{\text{nor}}(t) \neq 0$ for every $p \in \bar{B}(a, \rho) \setminus \hat{\gamma}[0, 1]$ and $t \in [0, 1]$.

THEOREM 3.1. *If condition * holds, then the median of μ is unique.*

Proof. We will prove this by showing that f is strictly convex, that is, for every geodesic $\gamma : [0, 1] \rightarrow \bar{B}(a, \rho)$, the function $f \circ \gamma$ is strictly convex. By the first estimation in Lemma 1.1, for every $p \in \bar{B}(a, \rho) \setminus \hat{\gamma}[0, 1]$ the function $t \mapsto d(\gamma(t), p)$ is strictly convex, and for $p \in \hat{\gamma}[0, 1]$ it is trivially convex. Since the condition * yields that $\mu(\bar{B}(a, \rho) \setminus \hat{\gamma}[0, 1]) > 0$, by integration we obtain the strict convexity of f and the proof is completed. □

REMARK 1. For the Riemannian barycenters of μ in $B(a, \rho)$ to be unique, Kendall showed that (see [21, Theorem 7.3] and [22]) the assumption

$$\rho < \min \left\{ \frac{\pi}{2\sqrt{\Delta}}, \text{inj}(a) \right\} \tag{3.1}$$

suffices (certainly, without condition $*$). Naturally, one may wonder that, under condition $*$, whether the median of μ remains unique if condition (1.1) is replaced by the weaker one (3.1). Unfortunately, this is not true. A counterexample may be found in [20], which shows that if three points on the upper hemisphere S^2_+ are symmetrically located in a circle which is parallel and close to the equator, then there are at least three medians.

In the proof of Theorem 3.1, we have seen that f is strictly convex if condition $*$ holds. However, under the same condition, we can show that f is in fact strongly convex (see [33, Definition 6.1.1]). That is, there exists a constant $\tau > 0$ such that for every geodesic $\gamma : [0, 1] \rightarrow \bar{B}(a, \rho)$, the following inequalities hold:

$$f(\gamma(t)) \leq (1-t)f(\gamma(0)) + tf(\gamma(1)) - \tau|\dot{\gamma}|^2(1-t)t, \quad t \in [0, 1]. \tag{3.2}$$

This is equivalent to saying that for every geodesic $\gamma : [0, 1] \rightarrow \bar{B}(a, \rho)$, the function $t \mapsto f(\gamma(t)) - \tau|\dot{\gamma}|^2t^2$ is convex on $[0, 1]$. To see this, we begin with an equivalent formulation of condition $*$.

LEMMA 3.2. *Condition $*$ holds if and only if there exist two constants $\varepsilon_\mu > 0$ and $\eta_\mu > 0$, such that for every geodesic $\gamma : [0, 1] \rightarrow \bar{B}(a, \rho)$, we have*

$$\mu(B(\gamma, \varepsilon_\mu)) \leq 1 - \eta_\mu,$$

where for $\varepsilon > 0$, $B(\gamma, \varepsilon) = \{x \in \bar{B}(a, \rho) : d(x, \gamma[0, 1]) < \varepsilon\}$.

Proof. We only have to show the necessity since the sufficiency is trivial. Assume that this is not true, then for every $\varepsilon > 0$ and $\eta > 0$, there exists a geodesic $\gamma : [0, 1] \rightarrow \bar{B}(a, \rho)$ such that $\mu(B(\gamma, \varepsilon)) > 1 - \eta$. Then we obtain a sequence of geodesics $(\gamma_n)_{n \geq 1} : [0, 1] \rightarrow \bar{B}(a, \rho)$ verifying $\mu(B(\gamma_n, 1/n)) > 1 - 1/n$. Since the sequence $(\gamma_n(0), \dot{\gamma}_n(0))_n$ is contained in the compact set $E = \{(x, v) \in TM : x \in \bar{B}(a, \rho), |v| \leq 2\rho\}$, there is a subsequence $(\gamma_{n_k}(0), \dot{\gamma}_{n_k}(0))_k$ and a point $(x, v) \in E$, such that $(\gamma_{n_k}(0), \dot{\gamma}_{n_k}(0)) \rightarrow (x, v)$. Let γ be the geodesic starting from x with velocity v , then by the classical theory of ordinary differential equations we know that $\gamma_{n_k} \rightarrow \gamma$ uniformly on $[0, 1]$. Particularly, $\gamma[0, 1] \subset \bar{B}(a, \rho)$. Then for every $j \geq 1$, $B(\gamma_{n_k}, 1/n_k) \subset B(\gamma, 1/j)$ for sufficiently large k , hence $\mu(B(\gamma, 1/j)) \geq 1 - 1/n_k$. By letting $k \rightarrow \infty$ and then letting $j \rightarrow \infty$, we obtain $\mu(\gamma[0, 1]) = 1$. This contradicts condition $*$. \square

The lemma below gives a basic angle estimation, which is also useful in the next section. In the following, we write

$$\sigma = \sup\{d(p, a) : p \in \text{supp } \mu\}.$$

Note that $\sigma < \rho$, since the support of μ is contained in the open ball $B(a, \rho)$.

LEMMA 3.3. *Let $\triangle ABC$ be a geodesic triangle in $\bar{B}(a, \rho)$ such that $A = a$, $B \in \bar{B}(a, \sigma)$ and $C \in \bar{B}(a, \rho) \setminus \bar{B}(a, \sigma)$, then*

$$\cos \angle C \geq \frac{S_\Delta(d(C, A) - \sigma)}{S_\Delta(d(C, A) + \sigma)}.$$

Proof. We prove this lemma for the case when $\Delta > 0$. The proof for the case when $\Delta \leq 0$ is similar. Let $d(B, C) = a_1$, $d(C, A) = b_1$ and $d(A, B) = c_1$. Consider a comparison triangle

$\triangle \bar{A}\bar{B}\bar{C}$ of $\triangle ABC$ in M_Δ^2 . Since $K \leq \Delta$ in $\bar{B}(a, \rho)$, Alexandrov's theorem yields that $\angle C \leq \angle \bar{C}$. Hence, observing $c_1 \leq \sigma$ and $b_1 - \sigma \leq a_1 \leq b_1 + \sigma$, we obtain

$$\begin{aligned} \cos \angle C \geq \cos \angle \bar{C} &= \frac{\cos(\sqrt{\Delta}c_1) - \cos(\sqrt{\Delta}a_1) \cos(\sqrt{\Delta}b_1)}{\sin(\sqrt{\Delta}a_1) \sin(\sqrt{\Delta}b_1)} \\ &\geq \frac{\cos(\sqrt{\Delta}\sigma) - \cos(\sqrt{\Delta}(b_1 - \sigma)) \cos(\sqrt{\Delta}b_1)}{\sin(\sqrt{\Delta}(b_1 + \sigma)) \sin(\sqrt{\Delta}b_1)} \\ &= \frac{\sin(\sqrt{\Delta}(b_1 - \sigma))}{\sin(\sqrt{\Delta}(b_1 + \sigma))} = \frac{S_\Delta(d(C, A) - \sigma)}{S_\Delta(d(C, A) + \sigma)}. \end{aligned} \quad \square$$

LEMMA 3.4. Let $\triangle ABC$ be a geodesic triangle in $\bar{B}(a, \rho)$ such that $\angle A \geq \pi/2$. Consider a geodesic triangle $\triangle A_1B_1C_1$ in M_δ^2 such that $\bar{d}(A_1, C_1) = d(A, C)$, $\bar{d}(A_1, B_1) = d(A, B)$ and $\angle A_1 = \angle A$. Then $\sin \angle C \geq \sin \angle C_1$.

Proof. We prove this for the case when $\delta > 0$. The proof for the case when $\delta \leq 0$ is similar. First, observe that by Lemma 2.3 we have $\angle B < \pi/2$ and $\angle C < \pi/2$. Let $\triangle A_2B_2C_2$ be a comparison triangle of $\triangle ABC$ in M_δ^2 . Since $K \geq \delta$ in $\bar{B}(a, \rho)$, Toponogov's theorem yields that $\angle A_2 \leq \angle A = \angle A_1$ and $\angle B_2 \leq \angle B$. Assume that $d(B, C) = a_2$, $d(A, C) = b_2$, $d(A, B) = c_2$ and $\bar{d}(B_1, C_1) = a_1$. Since $\angle A_2 \leq \angle A_1$, we have $a_2 \leq a_1$. By the law of cosines in M_δ^2 we obtain

$$\begin{aligned} \frac{d}{da_1} \cos \angle C_1 &= \frac{d}{da_1} \frac{\cos(\sqrt{\delta}c_2) - \cos(\sqrt{\delta}a_1) \cos(\sqrt{\delta}b_2)}{\sin(\sqrt{\delta}b_2) \sin(\sqrt{\delta}a_1)} \\ &= \frac{\sqrt{\delta}(\cos(\sqrt{\delta}b_2) - \cos(\sqrt{\delta}a_1) \cos(\sqrt{\delta}c_2))}{\sin(\sqrt{\delta}b_2) \sin^2(\sqrt{\delta}a_1)} \\ &\geq \frac{\sqrt{\delta}(\cos(\sqrt{\delta}b_2) - \cos(\sqrt{\delta}a_2) \cos(\sqrt{\delta}c_2))}{\sin(\sqrt{\delta}b_2) \sin^2(\sqrt{\delta}a_1)} \\ &= \frac{\sqrt{\delta} \sin(\sqrt{\delta}a_2) \sin(\sqrt{\delta}c_2)}{\sin(\sqrt{\delta}b_2) \sin^2(\sqrt{\delta}a_1)} \cos \angle B_2 > 0 \end{aligned}$$

so that $\cos \angle C_1$ is non-decreasing with respect to a_1 when $a_1 \geq a_2$. Hence, $\cos \angle C_1 \geq \cos \angle C_2$, that is $\angle C_1 \leq \angle C_2$. Then observing $\angle C_2 \leq \angle C < \pi/2$, we obtain $\sin \angle C \geq \sin \angle C_1$. The proof is complete. \square

The following lemma is important for giving a lower bound of the Hessian of the distance function. Though the two estimations below are very simple, they are sufficient for our purposes.

LEMMA 3.5. Let $\triangle ABC$ be a geodesic triangle in $\bar{B}(a, \rho)$.

(i) If $\angle A = \pi/2$, then

$$\sin \angle C \geq L_1(\rho, \delta)d(A, B),$$

where the constant

$$L_1(\rho, \delta) = \begin{cases} 2\sqrt{\delta}/\pi & \text{if } \delta > 0; \\ 1/4\rho & \text{if } \delta = 0; \\ \sqrt{-\delta}/\sinh(4\rho\sqrt{-\delta}) & \text{if } \delta < 0. \end{cases}$$

(ii) If $A \in \partial\bar{B}(a, \rho)$, $B \in \bar{B}(a, \sigma)$ and $\angle A \geq \pi/2$, then

$$\sin \angle C \geq L(\sigma, \rho, \delta, \Delta)d(A, B),$$

where the constant

$$L(\sigma, \rho, \delta, \Delta) = L_1(\rho, \delta) \frac{S_\Delta(\rho - \sigma)}{S_\Delta(\rho + \sigma)}.$$

Proof. We prove this lemma for the case when $\delta > 0$. The proof for the case when $\delta \leq 0$ is similar. Let $\triangle A_1 B_1 C_1$ be as in Lemma 3.4, then $\sin \angle C \geq \sin \angle C_1$. Let $\bar{d}(B_1, C_1) = a_1$ and $d(A, B) = c_1$, then the law of sines in M_δ^2 yields

$$\sin \angle C_1 = \frac{\sin(\sqrt{\delta}c_1)}{\sin(\sqrt{\delta}a_1)} \sin \angle A \geq \frac{2\sqrt{\delta}c_1}{\pi} \sin \angle A,$$

since $\sin \theta \geq 2\theta/\pi$ for $\theta \in [0, \pi/2]$. Hence the first estimation holds if $\angle A = \pi/2$. To show the second estimation, by the convexity of $\bar{B}(a, \rho)$ we obtain

$$\angle A \leq \angle C A a + \angle a A B < \frac{\pi}{2} + \angle a A B.$$

Since $\angle A \geq \pi/2$, by Lemma 3.3,

$$\sin \angle A \geq \cos \angle a A B \geq \frac{S_\Delta(\rho - \sigma)}{S_\Delta(\rho + \sigma)}.$$

The second estimation follows immediately and the proof is complete. □

We know that f is convex, thus along every geodesic it has a second derivative in the sense of distribution; the following proposition gives its specific form as well as the Taylor formulae.

PROPOSITION 3.6. *Let $\gamma : [0, b] \rightarrow \bar{B}(a, \rho)$ be a geodesic. Then for $t_0 \in [0, b]$ and $t \in [t_0, b]$,*

$$f(\gamma(t)) = f(\gamma(t_0)) + \left. \frac{d}{ds} f(\gamma(s)) \right|_{s=t_0+} (t - t_0) + \int_{(t_0, t)} (t - s) \nu(ds) \tag{3.3}$$

and for $t_0 \in (0, b]$ and $t \in [0, t_0]$,

$$f(\gamma(t)) = f(\gamma(t_0)) + \left. \frac{d}{ds} f(\gamma(s)) \right|_{s=t_0-} (t - t_0) + \int_{(t, t_0)} (s - t) \nu(ds),$$

with ν being the second derivative of $f \circ \gamma$ on $(0, b)$ in the sense of distribution, which is a bounded positive measure given by

$$\nu = \left(\int_{M \setminus \gamma[0, b]} \text{Hess } d_p(\dot{\gamma}, \dot{\gamma}) \mu(dp) \right) \cdot \lambda|_{(0, b)} + 2|\dot{\gamma}| \cdot (\mu \circ \gamma)|_{(0, b)},$$

where $\lambda|_{(0, b)}$ and $(\mu \circ \gamma)|_{(0, b)}$ denote the restrictions of Lebesgue measure and the measure $\mu \circ \gamma$ on $(0, b)$, respectively.

Proof. We shall only prove the first identity, since the proof of the second one is similar. Observe that since γ is a homeomorphism of $(0, b)$ onto its image, $\mu \circ \gamma$ is a well-defined measure on $(0, b)$. By Taylor’s formula,

$$\begin{aligned} & \int_{M \setminus \gamma[0, b]} d(\gamma(t), p) \mu(dp) \\ &= \int_{M \setminus \gamma[0, b]} \left(d(\gamma(t_0), p) + \left. \frac{d}{ds} d(\gamma(s), p) \right|_{s=t_0} (t - t_0) + \int_{t_0}^t (t - s) \frac{d^2}{ds^2} d(\gamma(s), p) ds \right) \mu(dp) \end{aligned}$$

$$\begin{aligned}
 &= \int_{M \setminus \gamma[0,b]} d(\gamma(t_0), p) \mu(dp) + \left\langle \dot{\gamma}(t_0), \int_{M \setminus \gamma[0,b]} \frac{-\exp_{\gamma(t_0)}^{-1} p}{d(\gamma(t_0), p)} \mu(dp) \right\rangle (t - t_0) \\
 &\quad + \int_{t_0}^t (t - s) ds \int_{M \setminus \gamma[0,b]} \text{Hess } d_p(\dot{\gamma}(s), \dot{\gamma}(s)) \mu(dp).
 \end{aligned}$$

It is easily seen that

$$\int_{M \setminus \gamma[0,b]} d(\gamma(t_0), p) = f(\gamma(t_0)) - \int_{\gamma[0,t_0]} d(\gamma(t_0), p) \mu(dp) - \int_{\gamma(t_0,b]} d(\gamma(t_0), p) \mu(dp),$$

and, by Proposition 2.1, that

$$\begin{aligned}
 &\left\langle \dot{\gamma}(t_0), \int_{M \setminus \gamma[0,b]} \frac{-\exp_{\gamma(t_0)}^{-1} p}{d(\gamma(t_0), p)} \mu(dp) \right\rangle \\
 &= \langle \dot{\gamma}(t_0), H(\gamma(t_0)) \rangle - \left\langle \dot{\gamma}(t_0), \int_{\gamma[0,t_0] \cup \gamma(t_0,b]} \frac{-\exp_{\gamma(t_0)}^{-1} p}{d(\gamma(t_0), p)} \mu(dp) \right\rangle \\
 &= \langle \dot{\gamma}(t_0), H(\gamma(t_0)) \rangle - |\dot{\gamma}| \mu(\gamma[0, t_0]) + |\dot{\gamma}| \mu(\gamma(t_0, b]) \\
 &= \left. \frac{d}{ds} f(\gamma(s)) \right|_{s=t_0+} - |\dot{\gamma}| \mu(\gamma[0, t_0]) + |\dot{\gamma}| \mu(\gamma(t_0, b]).
 \end{aligned}$$

Since

$$f(\gamma(t)) = \int_{M \setminus \gamma[0,b]} d(\gamma(t), p) \mu(dp) + \int_{\gamma[0,b]} d(\gamma(t), p) \mu(dp),$$

we obtain

$$\begin{aligned}
 &f(\gamma(t)) - f(\gamma(t_0)) - \left. \frac{d}{ds} f(\gamma(s)) \right|_{s=t_0+} (t - t_0) - \int_{t_0}^t (t - s) ds \int_{M \setminus \gamma[0,b]} \text{Hess } d_p(\dot{\gamma}(s), \dot{\gamma}(s)) \mu(dp) \\
 &= - \int_{\gamma[0,t_0]} d(\gamma(t_0), p) \mu(dp) - \int_{\gamma(t_0,b]} d(\gamma(t_0), p) \mu(dp) - |\dot{\gamma}| \mu(\gamma[0, t_0]) (t - t_0) \\
 &\quad + |\dot{\gamma}| \mu(\gamma(t_0, b]) (t - t_0) + \int_{\gamma[0,b]} d(\gamma(t), p) \mu(dp) \\
 &= \left(- \int_{\gamma[0,t_0]} d(\gamma(t_0), p) \mu(dp) - |\dot{\gamma}| \mu(\gamma[0, t_0]) (t - t_0) + \int_{\gamma[0,t_0]} d(\gamma(t), p) \mu(dp) \right) \\
 &\quad + \left(- \int_{\gamma(t_0,b]} d(\gamma(t_0), p) \mu(dp) + |\dot{\gamma}| \mu(\gamma(t_0, b]) (t - t_0) + \int_{\gamma(t_0,b]} d(\gamma(t), p) \mu(dp) \right) \\
 &= \left(- \int_{\gamma[0,t_0]} d(\gamma(t_0), p) \mu(dp) - \int_{\gamma[0,t_0]} d(\gamma(t_0), \gamma(t)) \mu(dp) + \int_{\gamma[0,t_0]} d(\gamma(t), p) \mu(dp) \right) \\
 &\quad + \left(- \int_{\gamma(t_0,t]} d(\gamma(t_0), p) \mu(dp) + \int_{\gamma(t_0,t]} d(\gamma(t_0), \gamma(t)) \mu(dp) + \int_{\gamma(t_0,t]} d(\gamma(t), p) \mu(dp) \right) \\
 &\quad + \left(- \int_{\gamma[t,b]} d(\gamma(t_0), p) \mu(dp) + \int_{\gamma[t,b]} d(\gamma(t_0), \gamma(t)) \mu(dp) + \int_{\gamma[t,b]} d(\gamma(t), p) \mu(dp) \right) \\
 &= 2 \int_{\gamma(t_0,t]} d(\gamma(t), p) \mu(dp) = 2|\dot{\gamma}| \int_{(t_0,t]} (t - s) (\mu \circ \gamma)(ds).
 \end{aligned}$$

Hence (3.3) holds. To show that ν is the second derivative of $f \circ \gamma$ on $(0, b)$ in the sense of distribution, let $\varphi \in C_c^\infty(0, b)$ and choose $t_0 = 0$ in (3.3), then Fubini's theorem and integration by parts yield

$$\begin{aligned} \int_{(0,b)} f(\gamma(t))\varphi''(t) dt &= f(\gamma(0)) \int_{(0,b)} \varphi''(t) dt + \left. \frac{d}{ds} f(\gamma(s)) \right|_{s=0+} \int_{(0,b)} t\varphi''(t) dt \\ &\quad + \int_{(0,b)} \varphi''(t) dt \int_{(0,t)} (t-s)\nu(ds) \\ &= \int_{(0,b)} \nu(ds) \int_{(s,b)} (t-s)\varphi''(t) dt = \int_{(0,b)} \varphi(s)\nu(ds). \end{aligned}$$

The proof is complete. □

Now we are ready to show that condition $*$ yields the strong convexity of f . Certainly, this also gives a proof of the uniqueness of the median.

THEOREM 3.7. *If condition $*$ holds, then f is strongly convex. More precisely, (3.2) holds for $\tau = (1/2)\varepsilon_\mu^2 \eta_\mu D(\rho, \Delta)L(\sigma, \rho, \delta, \Delta)^2 > 0$. Moreover, with this choice of τ , for every $x \in \bar{B}(a, \rho)$,*

$$f(x) \geq f_* + \tau d^2(x, m),$$

where m is the unique median of μ .

Proof. Let $\gamma : [0, 1] \rightarrow \bar{B}(a, \rho)$ be a geodesic, then by the first estimation in Lemma 1.1 we obtain that for every $s \in [0, 1]$,

$$\begin{aligned} \int_{M \setminus \gamma[0,1]} \text{Hess } d_p(\dot{\gamma}(s), \dot{\gamma}(s))\mu(dp) &\geq \int_{\bar{B}(a,\sigma) \setminus B(\hat{\gamma}, \varepsilon_\mu)} D(\rho, \Delta) |\dot{\gamma}_p^{\text{nor}}(s)|^2 \mu(dp) \\ &= D(\rho, \Delta) |\dot{\gamma}|^2 \int_{\bar{B}(a,\sigma) \setminus B(\hat{\gamma}, \varepsilon_\mu)} \sin^2 \angle(\dot{\gamma}(s), \exp_{\gamma(s)}^{-1} p) \mu(dp). \end{aligned}$$

Then for every $p \in \bar{B}(a, \sigma) \setminus B(\hat{\gamma}, \varepsilon_\mu)$, let $q = q(p)$ be the metric projection of p onto $\hat{\gamma}[0, 1]$. If $q \in \hat{\gamma}(0, 1)$ and $\gamma(s) \neq q$, then the geodesic triangle $\Delta pq\gamma(s)$ is a right triangle with $\angle pq\gamma(s) = \pi/2$. Hence the first estimation in Lemma 3.5 yields that

$$\sin \angle(\dot{\gamma}(s), \exp_{\gamma(s)}^{-1} p) \geq L_1(\rho, \delta) d(p, q) \geq L_1(\rho, \delta) \varepsilon_\mu.$$

If $q \in \{\hat{\gamma}(0), \hat{\gamma}(1)\}$ and $\gamma(s) \neq q$, then $\angle pq\gamma(s) \geq \pi/2$. Hence by the second estimation in Lemma 3.5,

$$\sin \angle(\dot{\gamma}(s), \exp_{\gamma(s)}^{-1} p) \geq L(\sigma, \rho, \delta, \Delta) d(p, q) \geq L(\sigma, \rho, \delta, \Delta) \varepsilon_\mu.$$

Since $L(\sigma, \rho, \delta, \Delta) < L_1(\rho, \delta)$, we always have

$$\sin \angle(\dot{\gamma}(s), \exp_{\gamma(s)}^{-1} p) \geq L(\sigma, \rho, \delta, \Delta) \varepsilon_\mu.$$

Thus by Lemma 3.2 we obtain

$$\int_{M \setminus \gamma[0,1]} \text{Hess } d_p(\dot{\gamma}(s), \dot{\gamma}(s))\mu(dp) \geq \varepsilon_\mu^2 \eta_\mu D(\rho, \Delta)L(\sigma, \rho, \delta, \Delta)^2 |\dot{\gamma}|^2 = 2\tau |\dot{\gamma}|^2.$$

Then for every $t \in [0, 1]$ by Proposition 3.6,

$$\begin{aligned} f(\gamma(1)) &= f(\gamma(t)) + \left. \frac{d}{ds} f(\gamma(s)) \right|_{s=t+} (1-t) + \int_{(t,1)} (1-s)\nu(ds) \\ &\geq f(\gamma(t)) + \left. \frac{d}{ds} f(\gamma(s)) \right|_{s=t+} (1-t) + 2\tau|\dot{\gamma}|^2 \int_{(t,1)} (1-s) ds \\ &= f(\gamma(t)) + \left. \frac{d}{ds} f(\gamma(s)) \right|_{s=t+} (1-t) + \tau|\dot{\gamma}|^2(1-t)^2. \end{aligned} \tag{3.4}$$

Similarly, for every $t \in (0, 1]$,

$$f(\gamma(0)) \geq f(\gamma(t)) + \left. \frac{d}{ds} f(\gamma(s)) \right|_{s=t-} (-t) + \tau|\dot{\gamma}|^2 t^2. \tag{3.5}$$

It follows by (3.4), (3.5) and Proposition 2.1 that for every $t \in (0, 1)$,

$$\begin{aligned} f(\gamma(t)) &\leq (1-t)f(\gamma(0)) + tf(\gamma(1)) - \tau|\dot{\gamma}|^2(1-t)t \\ &\quad - \left(\left. \frac{d}{ds} f(\gamma(s)) \right|_{s=t+} - \left. \frac{d}{ds} f(\gamma(s)) \right|_{s=t-} \right) (1-t)t \\ &\leq (1-t)f(\gamma(0)) + tf(\gamma(1)) - \tau|\dot{\gamma}|^2(1-t)t. \end{aligned}$$

The strong convexity of f is proved. To show the last inequality, let $\gamma : [0, 1] \rightarrow \bar{B}(a, \rho)$ be the geodesic such that $\gamma(0) = m$ and $\gamma(1) = x$. Then (3.4) yields

$$f(x) \geq f(m) + \left. \frac{d}{ds} f(\gamma(s)) \right|_{s=0+} + \tau|\dot{\gamma}|^2 \geq f(m) + \tau|\dot{\gamma}|^2 = f_* + \tau d^2(x, m).$$

The proof is complete. □

4. A subgradient algorithm

To begin with, we recall the definition of the subgradient of a convex function on a Riemannian manifold. For our purpose, it suffices to consider this notion in a convex subset of the manifold.

DEFINITION 2. Let U be a convex subset of M and let h be a convex function defined on U . For every $x \in U$, a vector $v \in T_x M$ is called a subgradient of h at x if for every geodesic $\gamma : [0, b] \rightarrow U$ with $\gamma(0) = x$, we have

$$h(\gamma(t)) \geq h(x) + \langle \dot{\gamma}(0), v \rangle t, \quad t \in [0, b].$$

Our idea to approximate the median of μ by a subgradient method, stems from the following simple observation.

LEMMA 4.1. For every $x \in \bar{B}(a, \rho)$, $H(x)$ is a subgradient of f at x .

Proof. Let $\gamma : [0, b] \rightarrow \bar{B}(a, \rho)$ be a geodesic such that $\gamma(0) = x$, then by Proposition 2.1, together with the convexity of f , we get for every $t \in [0, b]$ that

$$\begin{aligned} f(\gamma(t)) &\geq f(\gamma(0)) + \left. \frac{d}{ds} f(\gamma(s)) \right|_{s=0+} t \\ &= f(x) + \langle \dot{\gamma}(0), H(x) \rangle + \mu\{x\}|\dot{\gamma}|t \\ &\geq f(x) + \langle \dot{\gamma}(0), H(x) \rangle t. \end{aligned} \tag{□}$$

We proceed to give some notation which is necessary for introducing the subgradient algorithm.

NOTATION. If $x \in \bar{B}(a, \rho)$ and $H(x) \neq 0$, then we write

$$\begin{aligned}\gamma_x(t) &= \exp_x \left(-t \frac{H(x)}{|H(x)|} \right), \quad t \geq 0. \\ r_x &= \sup \{ t \in [0, 2\rho] : \gamma_x(t) \in \bar{B}(a, \rho) \}.\end{aligned}$$

Note that for $x \in \bar{B}(a, \rho)$ such that $H(x) \neq 0$, by the convexity of $\bar{B}(a, \rho)$ and the fact that $2\rho < \text{inj}(\bar{B}(a, \rho))$, we have $\gamma_x[0, r_x] \subset \bar{B}(a, \rho)$.

The following simple lemma shows that every r_x is strictly positive. More importantly, it ensures theoretically the possibility of the choice of stepsizes in the convergence theorem of the subgradient algorithm.

LEMMA 4.2. *The following estimation holds:*

$$\inf \{ r_x : x \in \bar{B}(a, \rho), H(x) \neq 0 \} > 0.$$

Proof. Since the support of μ is contained in $B(a, \rho)$, for every $x \in \partial \bar{B}(a, \rho)$, $H(x)$ is transverse to $\partial \bar{B}(a, \rho)$ and hence $r_x > 0$ for every $x \in \bar{B}(a, \rho)$ such that $H(x) \neq 0$. Moreover, there exists $\varepsilon > 0$ such that $\text{supp}(\mu) \subset B(a, \rho - \varepsilon)$. Then for $x \in B(a, \rho - \varepsilon)$ such that $H(x) \neq 0$ we have $r_x \geq \rho - d(x, a) > \varepsilon$. On the other hand, since H is continuous on $\bar{B}(a, \rho) \setminus B(a, \rho - \varepsilon)$, r_x vary continuously with x on this compact set. Thus there exists a point $x_0 \in \bar{B}(a, \rho) \setminus B(a, \rho - \varepsilon)$ such that $\inf \{ r_x : x \in \bar{B}(a, \rho) \setminus B(a, \rho - \varepsilon) \} = r_{x_0}$. Hence we obtain $\inf \{ r_x : x \in \bar{B}(a, \rho), H(x) \neq 0 \} \geq \min \{ \varepsilon, r_{x_0} \} > 0$. \square

Now we introduce the subgradient algorithm to approximate the medians of the probability measure μ .

ALGORITHM 1. Subgradient algorithm for Riemannian medians:

Step 1:

Choose a starting point $x_0 \in \bar{B}(a, \rho)$ and let $k = 0$.

Step 2:

If $H(x_k) = 0$, then $x_k \in \mathfrak{M}_\mu$ and stop. If not, then go to Step 3.

Step 3:

Choose a stepsize $t_k \in (0, r_{x_k}]$ and let $x_{k+1} = \gamma_{x_k}(t_k)$, then come back to Step 2 with $k = k + 1$.

REMARK 2. It should be noted that, in the above algorithm, we have already restricted every stepsize t_k to be in the interval $(0, r_{x_k}]$. From now on, we shall always make this restriction implicitly to ensure that the sequence $(x_k)_k$ will never get out of the ball $\bar{B}(a, \rho)$.

Now we turn to the convergence proof of the above algorithm under some further conditions on the stepsizes. In Euclidean spaces, it is well known that the following type of inequalities are of fundamental importance to conclude the convergence of subgradient algorithms (see, for example, [12, 26]):

$$\|x_{k+1} - y\|^2 \leq \|x_k - y\|^2 + \alpha t_k^2 + \beta \frac{2t_k}{\|v_k\|} (f(y) - f(x_k)).$$

For a non-negatively curved Riemannian manifold, Ferreira and Oliveira obtained a generalization of the above inequality in [14] by using Toponogov's comparison theorem. But,

their method is not applicable in our case since $\bar{B}(a, \rho)$ is not assumed to be non-negatively curved. However, we can still obtain a similar result using a different method.

LEMMA 4.3. *Let $(x_k)_k$ be the sequence generated by Algorithm 1. If $H(x_k) \neq 0$, then for every point $y \in \bar{B}(a, \rho)$,*

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + C(\rho, \delta)t_k^2 + \frac{2t_k}{|H(x_k)|}(f(y) - f(x_k)).$$

Particularly,

$$d^2(x_{k+1}, \mathfrak{M}_\mu) \leq d^2(x_k, \mathfrak{M}_\mu) + C(\rho, \delta)t_k^2 + 2t_k(f_* - f(x_k)). \tag{4.1}$$

Proof. By Taylor’s formula and the second estimation in Lemma 1.1, there exists $\xi \in (0, t_k)$ such that

$$\begin{aligned} \frac{1}{2}d^2(x_{k+1}, y) &= \frac{1}{2}d^2(\gamma_{x_k}(t_k), y) \\ &= \frac{1}{2}d^2(\gamma_{x_k}(0), y) + \frac{d}{dt} \left[\frac{1}{2}d^2(\gamma_{x_k}(t), y) \right]_{t=0} t_k + \frac{1}{2} \frac{d^2}{dt^2} \left[\frac{1}{2}d^2(\gamma_{x_k}(t), y) \right]_{t=\xi} t_k^2 \\ &= \frac{1}{2}d^2(x_k, y) + \left\langle \dot{\gamma}_{x_k}(0), \text{grad} \frac{1}{2}d_y^2(x_k) \right\rangle t_k + \frac{1}{2} \text{Hess} \frac{1}{2}d_y^2(\dot{\gamma}_{x_k}(\xi), \dot{\gamma}_{x_k}(\xi)) t_k^2 \\ &\leq \frac{1}{2}d^2(x_k, y) + \frac{\langle H(x_k), \exp_{x_k}^{-1} y \rangle}{|H(x_k)|} t_k + \frac{C(\rho, \delta)}{2} t_k^2. \end{aligned}$$

By Lemma 4.1, $H(x_k)$ is a subgradient of f at point x_k and hence

$$\langle H(x_k), \exp_{x_k}^{-1} y \rangle \leq f(y) - f(x_k).$$

Consequently,

$$\frac{1}{2}d^2(x_{k+1}, y) \leq \frac{1}{2}d^2(x_k, y) + \frac{t_k}{|H(x_k)|}(f(y) - f(x_k)) + \frac{C(\rho, \delta)}{2} t_k^2,$$

and the first inequality holds. The second one follows from $f_* \leq f(x_k)$ and $|H(x_k)| \leq 1$. □

As in the Euclidean case, once the fundamental inequality is established, the convergence of the subgradient algorithm is soon achieved and the proof is elementary. Since the fundamental inequality (4.1) in Lemma 4.3 is very similar to the Euclidean case, the proof of the following convergence theorem is also very similar to that in the Euclidean case, hence we omit it and refer to [12, 14, 26].

THEOREM 4.4. *If the stepsizes $(t_k)_k$ verify*

$$\lim_{k \rightarrow \infty} t_k = 0 \quad \text{and} \quad \sum_{k=0}^{\infty} t_k = +\infty,$$

then the sequence $(x_k)_k$ generated by Algorithm 1 satisfies

$$\lim_{k \rightarrow \infty} d(x_k, \mathfrak{M}_\mu) = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} f(x_k) = f_*.$$

Moreover, if the stepsizes $(t_k)_k$ also verify

$$\sum_{k=0}^{\infty} t_k^2 < +\infty,$$

then there exists some $m \in \mathfrak{M}_\mu$ such that $x_k \rightarrow m$.

Now we consider the problem of the choice of stepsizes. By Lemma 4.2 we can choose $(t_k)_k$ that verifies the conditions of the preceding theorem and hence yields the desired convergence of our algorithm. For example, we may take $t_k = r_{x_k}/(k + 1)$ for every $k \geq 0$. But the drawback is that we do not know much about r_{x_k} . However, with further analysis we can obtain an explicit lower bound for it.

LEMMA 4.5. For every $x \in \bar{B}(a, \rho) \setminus \bar{B}(a, \sigma)$,

$$r_x \geq \frac{2d(x, a)S_\Delta(d(x, a) - \sigma)}{C(\rho, \delta)S_\Delta(d(x, a) + \sigma)}.$$

Proof. Since $x \in \bar{B}(a, \rho) \setminus \bar{B}(a, \sigma)$, we have $H(x) \neq 0$ and hence r_x is well defined. Moreover, the diameter of $\bar{B}(a, \rho)$ is $2\rho < \text{inj}(B(a, \rho))$, thus the definition of r_x yields that $\gamma_x(r_x) \in \partial\bar{B}(a, \rho)$. By Taylor’s formula and the second estimation in Lemma 1.1, there exists $\xi \in (0, r_x)$ such that

$$\begin{aligned} \frac{1}{2}\rho^2 &= \frac{1}{2}d^2(\gamma_x(r_x), a) \\ &= \frac{1}{2}d^2(x, a) + \frac{d}{dt} \left[\frac{1}{2}d^2(\gamma_x(t), a) \right]_{t=0} r_x + \frac{1}{2} \frac{d^2}{dt^2} \left[\frac{1}{2}d^2(\gamma_x(t), a) \right]_{t=\xi} r_x^2 \\ &= \frac{1}{2}d^2(x, a) + \left\langle \dot{\gamma}_x(0), \text{grad} \frac{1}{2}d_a^2(x) \right\rangle r_x + \frac{1}{2} \text{Hess} \frac{1}{2}d_a^2(\dot{\gamma}_x(\xi), \dot{\gamma}_x(\xi)) r_x^2 \\ &\leq \frac{1}{2}d^2(x, a) + \frac{\langle H(x), \exp_x^{-1} a \rangle}{|H(x)|} r_x + \frac{C(\rho, \delta)}{2} r_x^2. \end{aligned}$$

The Gauss lemma yields that $\langle \exp_x^{-1} p, \exp_x^{-1} a \rangle > 0$ for $p \in \text{supp } \mu$, hence

$$\langle H(x), \exp_x^{-1} a \rangle = - \int_{\text{supp } \mu} \frac{\langle \exp_x^{-1} p, \exp_x^{-1} a \rangle}{d(x, p)} \mu(dp) < 0.$$

Combine this with $d(x, a) \leq \rho$, $C(\rho, \delta) > 0$ and $r_x > 0$, we obtain that

$$\begin{aligned} r_x &\geq \frac{1}{C(\rho, \delta)} \left\{ -\frac{\langle H(x), \exp_x^{-1} a \rangle}{|H(x)|} + \sqrt{\frac{\langle H(x), \exp_x^{-1} a \rangle^2}{|H(x)|^2} + C(\rho, \delta)(\rho^2 - d^2(x, a))} \right\} \\ &\geq \frac{1}{C(\rho, \delta)} \left\{ -\frac{\langle H(x), \exp_x^{-1} a \rangle}{|H(x)|} + \frac{|\langle H(x), \exp_x^{-1} a \rangle|}{|H(x)|} \right\} \\ &= \frac{-2}{C(\rho, \delta)} \frac{\langle H(x), \exp_x^{-1} a \rangle}{|H(x)|} \geq \frac{-2}{C(\rho, \delta)} \langle H(x), \exp_x^{-1} a \rangle \\ &= \frac{2}{C(\rho, \delta)} \int_{\text{supp } \mu} \frac{\langle \exp_x^{-1} p, \exp_x^{-1} a \rangle}{d(x, p)} \mu(dp) \\ &= \frac{2d(x, a)}{C(\rho, \delta)} \int_{\text{supp } \mu} \cos \angle pxa \mu(dp). \end{aligned}$$

Now it suffices to use Lemma 3.3 to obtain that for every $p \in \text{supp}(\mu)$,

$$\cos \angle pxa \geq \frac{S_\Delta(d(x, a) - \sigma)}{S_\Delta(d(x, a) + \sigma)}.$$

□

We are ready to give the desired lower bound.

LEMMA 4.6. For every $x \in \bar{B}(a, \rho)$ such that $H(x) \neq 0$ we have

$$r_x \geq \frac{\rho - \sigma}{C(\rho, \delta)F(\rho, \Delta) + 1},$$

where the constant $F(\rho, \Delta) \geq 1$ is given by

$$F(\rho, \Delta) = \begin{cases} 1 & \text{if } \Delta \geq 0; \\ \cosh(2\rho\sqrt{-\Delta}) & \text{if } \Delta < 0. \end{cases}$$

Proof. We prove this for the case when $\Delta > 0$. The proof for the cases when $\Delta \leq 0$ is similar. For every $x \in \bar{B}(a, \rho) \setminus \bar{B}(a, \sigma)$, by Lemma 4.5 we have

$$r_x \geq \frac{2d(x, a) \sin(\sqrt{\Delta}(d(x, a) - \sigma))}{C(\rho, \delta) \sin(\sqrt{\Delta}(d(x, a) + \sigma))}.$$

Note that $0 < \sqrt{\Delta}(d(x, a) \pm \sigma) < 2\rho\sqrt{\Delta} < \pi/2$ and that $(\sin u/\sin v) \geq (u/v)$ for $0 < u \leq v \leq \pi/2$, then we obtain

$$r_x \geq \frac{2d(x, a) \frac{d(x, a) - \sigma}{d(x, a) + \sigma}}{C(\rho, \delta) \frac{d(x, a) + \sigma}{d(x, a) + \sigma}} \geq \frac{2d(x, a) \frac{d(x, a) - \sigma}{d(x, a) + \sigma}}{C(\rho, \delta) 2d(x, a)} = \frac{d(x, a) - \sigma}{C(\rho, \delta)}.$$

On the other hand, we always have $r_x \geq \rho - d(x, a)$ and hence

$$r_x \geq \max\left\{\rho - d(x, a), \frac{d(x, a) - \sigma}{C(\rho, \delta)}\right\}.$$

Observe that

$$\min\left\{\max\left\{\rho - d(x, a), \frac{d(x, a) - \sigma}{C(\rho, \delta)}\right\} : \sigma < d(x, a) \leq \rho\right\} = \frac{\rho - \sigma}{C(\rho, \delta) + 1},$$

then we obtain

$$r_x \geq \frac{\rho - \sigma}{C(\rho, \delta) + 1}.$$

Moreover, for every $x \in \bar{B}(a, \sigma)$ such that $H(x) \neq 0$,

$$r_x \geq \rho - \sigma > \frac{\rho - \sigma}{C(\rho, \delta) + 1}.$$

The proof is complete. □

Thanks to the above estimation, we get a practically useful version of Theorem 4.4.

THEOREM 4.7. Let $(a_k)_k$ be a sequence in $(0, 1]$ such that

$$\lim_{k \rightarrow \infty} a_k = 0 \quad \text{and} \quad \sum_{k=0}^{\infty} a_k = +\infty.$$

Then we can choose

$$t_k = \frac{(\rho - \sigma)a_k}{C(\rho, \delta)F(\rho, \Delta) + 1}$$

in Algorithm 1 and, with this choice of stepsizes, the generated sequence $(x_k)_k$ satisfies

$$\lim_{k \rightarrow \infty} d(x_k, \mathfrak{M}_\mu) = 0 \quad \text{and} \quad \lim_{k \rightarrow \infty} f(x_k) = f_*.$$

Moreover, if $(a_k)_k$ also verifies that

$$\sum_{k=0}^{\infty} a_k^2 < +\infty,$$

then there exists some $m \in \mathfrak{M}_\mu$ such that $x_k \rightarrow m$.

Proof. This is a simple corollary to Lemma 4.6 and Theorem 4.4. □

Now we turn to the questions of error estimates and the rate of convergence of the subgradient algorithm under condition *.

PROPOSITION 4.8. *Let condition * hold and let the stepsizes $(t_k)_k$ satisfy*

$$\lim_{k \rightarrow \infty} t_k = 0 \quad \text{and} \quad \sum_{k=0}^{\infty} t_k = +\infty.$$

Then there exists $N \in \mathbf{N}$, such that for every $k \geq N$,

$$d^2(x_k, m) \leq b_k,$$

where m is the unique median of μ and the sequence $(b_k)_{k \geq N}$ is defined by

$$b_N = (\rho + \sigma)^2 \quad \text{and} \quad b_{k+1} = (1 - 2\tau t_k)b_k + C(\rho, \delta)t_k^2, \quad k \geq N,$$

which converges to 0 when $k \rightarrow \infty$. More explicitly, for every $k \geq N$,

$$b_{k+1} = (\rho + \sigma)^2 \prod_{i=N}^k (1 - 2\tau t_i) + C(\rho, \delta) \left(\sum_{j=N+1}^k t_{j-1}^2 \prod_{i=j}^k (1 - 2\tau t_i) + t_k^2 \right).$$

Proof. Since $t_k \rightarrow 0$, there exists $N \in \mathbf{N}$ such that for every $k \geq N$ we have $2\tau t_k < 1$. By Theorem 3.7,

$$f(x_k) - f_* \geq \tau d^2(x_k, m).$$

Combining this and Lemma 4.3 we obtain

$$d^2(x_{k+1}, m) \leq (1 - 2\tau t_k)d^2(x_k, m) + C(\rho, \delta)t_k^2.$$

By Proposition 2.4, $d^2(x_N, m) \leq (\rho + \sigma)^2 = b_N$. Then by induction it is easily seen that $d^2(x_k, m) \leq b_k$ for every $k \geq N$. To prove $b_k \rightarrow 0$, we first show

$$\liminf_{k \rightarrow \infty} b_k = 0.$$

If this is not true, then there exist $N_1 \geq N$ and $\eta > 0$ such that for every $k \geq N_1$ we have $b_k > \eta$ and $C(\rho, \delta)t_k < \tau\eta$. Thus

$$b_{k+1} = b_k + t_k(C(\rho, \delta)t_k - 2\tau b_k) \leq b_k - \tau\eta t_k.$$

By summing the above inequalities we obtain

$$\tau\eta \sum_{i=N_1}^k t_i \leq b_{N_1} - b_{k+1} \leq b_{N_1},$$

which contradicts $\sum_{k=0}^{\infty} t_k = +\infty$ and the assertion is proved.

For every $k \geq N$, we consider the following two cases: if $b_k > C(\rho, \delta)t_k/(2\tau)$, then

$$b_{k+1} < b_k - 2\tau t_k(C(\rho, \delta)t_k/(2\tau)) + C(\rho, \delta)t_k^2 = b_k;$$

and if $b_k \leq C(\rho, \delta)t_k/(2\tau)$, then

$$b_{k+1} \leq (1 - 2\tau t_k)C(\rho, \delta)t_k/(2\tau) + C(\rho, \delta)t_k^2 = C(\rho, \delta)t_k/(2\tau).$$

Hence we always have

$$b_{k+1} \leq \max\{b_k, C(\rho, \delta)t_k/(2\tau)\},$$

which yields by induction that for every $n \geq k$,

$$b_{n+1} \leq \max\{b_k, (C(\rho, \delta)/(2\tau)) \max\{t_k, t_{k+1}, \dots, t_n\}\}.$$

Then by taking the superior limit on the left-hand side and then the inferior limit on the right-hand side we conclude that $b_k \rightarrow 0$.

Finally, the explicit expressions of $(b_k)_k$ are obtained by induction. □

We proceed to show that if $(t_k)_k$ is chosen to be the harmonic series, then the rate of convergence of our algorithm is sublinear. To do this, we use the following lemma in [25].

LEMMA 4.9. *Let $(u_k)_{k \geq 0}$ be a sequence of non-negative real numbers such that*

$$u_{k+1} \leq \left(1 - \frac{\alpha}{k+1}\right)u_k + \frac{\zeta}{(k+1)^2},$$

where α and ζ are positive constants. Then

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

PROPOSITION 4.10. *Let condition * hold and we choose $t_k = r/(k+1)$ for every $k \geq 0$ with some constant $r > 0$, then*

$$d^2(x_{k+1}, m) \leq \begin{cases} \frac{1}{(k+2)^\alpha} \left((\rho + \sigma)^2 + \frac{2^\alpha r^2 C(\rho, \delta)(2-\alpha)}{1-\alpha}\right) & \text{if } 0 < \alpha < 1; \\ \frac{r^2 C(\rho, \delta)}{k+1} (1 + \ln(k+1)) & \text{if } \alpha = 1; \\ \frac{1}{(\alpha-1)(k+2)} \left(r^2 C(\rho, \delta) + \frac{(\alpha-1)(\rho + \sigma)^2 - r^2 C(\rho, \delta)}{(k+2)^{\alpha-1}}\right) & \text{if } \alpha > 1, \end{cases}$$

where m is the unique median of μ and $\alpha = 2\tau r$.

Proof. As in the proof of Proposition 4.8, we have for every $k \geq 0$,

$$d^2(x_{k+1}, m) \leq \left(1 - \frac{2\tau r}{k+1}\right)d^2(x_k, m) + \frac{r^2 C(\rho, \delta)}{(k+1)^2}.$$

Then it suffices to use Lemma 4.9 with $\alpha = 2\tau r$ and $\zeta = r^2 C(\rho, \delta)$ by observing that $d(x_0, m) \leq \rho + \sigma$. □

Acknowledgements. The author is very grateful to his PhD advisors Marc Arnaudon and Frédéric Barbaresco for many helpful discussions and their penetrating comments. The author would like also to thank the referee whose detailed comments greatly improved the quality of the paper.

References

1. A. D. ALEXANDROV, ‘Über eine Verallgemeinerung der Riemannschen Geometrie’, *Schr. Forschungsinst. Math.* 1 (1957) 33–84.
2. A. D. ALEKSANDROV *et al.*, ‘Generalized Riemannian spaces’, *Russian Math. Surveys* 41 (1986) no. 3, 1–54.
3. M. ARNAUDON, ‘Espérances conditionnelles et C-martingales dans les variétés’, *Séminaire de Probabilités-XXVIII*, Lecture Notes in Mathematics 1583 (Springer, Berlin, 1994) 300–311.
4. M. ARNAUDON, ‘Barycentres convexes et approximations des martingales dans les variétés’, *Séminaire de Probabilités-XXIX*, Lecture Notes in Mathematics 1613 (Springer, Berlin, 1995) 70–85.

5. M. ARNAUDON and X. M. LI, 'Barycenters of measures transported by stochastic flows', *Ann. Probab.* 33 (2005) no. 4, 1509–1543.
6. F. BARBARESCO, 'Innovative tools for radar signal processing based on Cartan's geometry of SPD matrices and information geometry', *IEEE International Radar Conference*, May 26–30, Rome, Italy, 2008.
7. F. BARBARESCO, 'Interactions between symmetric cone and information geometries', *ETVC'08*, Lecture Notes in Computer Science 5416 (Springer, Berlin, 2009) 124–163.
8. F. BARBARESCO, 'New foundation of radar doppler signal processing based on advanced differential geometry of symmetric spaces: doppler matrix CFAR and radar application', *Radar'09 Conference*, Bordeaux, October 2009.
9. F. BARBARESCO and G. BOUYT, 'Espace Riemannien symétrique et géométrie des espaces de matrices de covariance : équations de diffusion et calculs de médianes', *XXII^e Colloque GRETSI*, Dijon, September 2009.
10. M. BRIDSON and A. HAEFLIGER, *Metric spaces of non-positive curvature* (Springer, Berlin, 1999).
11. J. CHEEGER and D. G. EBIN, *Comparison theorems in Riemannian geometry* (North Holland, Amsterdam, 1975).
12. R. CORREA and C. LEMARÉCHAL, 'Convergence of some algorithms for convex minimization', *Math. Program.* 62 (1993) 261–275.
13. 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) 220–233.
14. O. P. FERREIRA and P. R. OLIVEIRA, 'Subgradient algorithm on Riemannian manifolds', *J. Optim. Theory Appl.* 97 (1998) no. 1, 93–104.
15. P. T. FLETCHER and S. JOSHI, 'Principle geodesic analysis on symmetric spaces: statistics of diffusion tensors', *Proceedings of ECCV Workshop on Computer Vision Approaches to Medical Image Analysis* (2004), 87–98.
16. P. T. FLETCHER *et al.*, 'Statistics of shape via principle geodesic analysis on Lie groups', *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2003) 95–101.
17. P. T. FLETCHER *et al.*, 'The geometric median on Riemannian manifolds with application to robust atlas estimation', *NeuroImage* 45 (2009) S143–S152.
18. J. JOST, *Riemannian geometry and geometric analysis* (Springer, Berlin, 2005).
19. H. KARCHER, 'Riemannian center of mass and mollifier smoothing', *Commun. Pure Appl. Math.* xxx (1977) 509–541.
20. I. N. KATZ and L. COOPER, 'Optimal location on a sphere', *Comput Math. Appl.* 6 (1980) 175–196.
21. W. S. KENDALL, 'Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence', *Proc. London Math. Soc.* (3) 61 (1990) no. 2, 371–406.
22. W. S. KENDALL, 'Convexity and the hemisphere', *J. London Math. Soc.* (2) 43 (1991) no. 3, 567–576.
23. H. W. KUHN, 'A note on Fermat's problem', *Math. Program.* 4 (1973) 98–107, North-Holland Publishing Company.
24. H. LE, 'Estimation of Riemannian barycentres', *LMS J. Comput. Math.* 7 (2004) 193–200.
25. A. NEDIC and D. P. BERTSEKAS, 'Convergence rate of incremental subgradient algorithms', *Stochastic optimization: algorithms and applications* (eds S. Uryasev and P. M. Pardalos; Kluwer Academic Publishers, Dordrecht, 2000) 263–304.
26. A. NEDIC and D. P. BERTSEKAS, 'Incremental subgradient methods for non-differentiable optimization', *SIAM J. Optim.* 12 (2001) no. 1, 109–138.
27. L. M. JR. OSTRESH, 'On the convergence of a class of iterative methods for solving Weber location problem', *Oper. Res.* 26 (1978) no. 4, 597–609.
28. X. PENNEC, 'Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements', *J. Math. Imaging Vision* 25 (2006) 127–154.
29. J. PICARD, 'Barycentres et martingales sur une variété', *Ann. Inst. H. Poincaré Probab. Statist.* 30 (1994) no. 4, 647–702.
30. A. SAHIB, 'Espérance d'une variable aléatoire à valeur dans un espace métrique', Thèse de l'Université de Rouen, 1998.
31. T. SAKAI, *Riemannian geometry*, Translations of Mathematical Monographs 149 (American Mathematical Society, Providence, RI, 1996).
32. C. G. SMALL, 'Multidimensional medians arising from geodesics on graphs', *Ann. Statist.* 25 (1997) no. 2, 478–494.
33. C. UDRISTE, *Convex functions and optimization methods on Riemannian manifolds* (Kluwer Academic Publishers, Dordrecht, 1994).
34. E. WEISZFELD, 'Sur le point pour lequel la somme des distances de n points donnés est minimum', *Tohoku Math. J.* 43 (1937) 355–386.

Le Yang
Laboratoire de Mathématiques et
Applications (CNRS: UMR 6086)
Université de Poitiers
Téléport 2 – BP30179, Boulevard
Marie et Pierre Curie, F – 86962
Futuroscope Chasseneuil Cedex
France

Le.Yang@math.univ-poitiers.fr