Computing Riemannian Center of Mass on Hadamard Manifolds

Glaydston de Carvalho Bento[1] [ORCID] · **Sandro Dimy Barbosa Bitar[2]** ·
João Xavier da Cruz Neto[3] · **Paulo Roberto Oliveira[4]** ·
João Carlos de Oliveira Souza[3]

Abstract

In this paper, we perform the steepest descent method for computing Riemannian center of mass on Hadamard manifolds. To this end, we extend convergence of the method to the Hadamard setting for continuously differentiable (possible nonconvex) functions which satisfy the Kurdyka–Łojasiewicz property. Some numerical experiments computing L^1 and L^2 center of mass in the context of positive definite symmetric matrices are presented using two different stepsize rules.

Communicated by Alexandru Kristály.

✉ Glaydston de Carvalho Bento
glaydston@ufg.com

Sandro Dimy Barbosa Bitar
sandrobitar@gmail.com

João Xavier da Cruz Neto
jxavier@ufpi.edu.br

Paulo Roberto Oliveira
poliveir@cos.ufrj.br

João Carlos de Oliveira Souza
joaocos.mat@ufpi.edu.br

[1] Universidade Federal de Goiás, Goiás, Brazil

[2] Universidade Federal do Amazonas, Manaus, Brazil

[3] Universidade Federal do Piauí, Teresina, Brazil

[4] Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

1 Introduction

We consider the problem of computing the (global) Riemannian center of mass of the data set on Hadamard manifolds. This notion and its variants have been appeared with many different names in the literature, e.g., Riemannian mean or average and Fréchet or Karcher mean, but we prefer (cf. [1]) the former denomination. This concept was proposed by Grove and Karcher [2], and since then, it has been extensively studied in pure mathematics as well as applied fields such as computer vision, statistical analysis of shapes, medical imaging, sensor networks and many other general data analysis applications. Several works have been devoted to study this problem on the context of positive definite symmetric matrices, e.g., [3–7]. This manifold plays a key role in diffusion tensor imaging, as explained in [8,9], and it always admits an invariant Riemannian metric, so that it becomes a Hadamard manifold (a complete, simply connected Riemannian manifold of nonpositive sectional curvature); see [10, Theorem 1.2]. For an approach to this problem in a more general setting, namely Finsler manifolds, see Kristály et al. [11]. Bačák [12] studied convergence of the proximal point method for computing center of mass in a geodesic metric space of nonpositive curvature, which includes Hadamard manifolds. Afsari et al. [13] investigated the applicability of a gradient descent algorithm, which is the most popular and easiest version of gradient descent method, for finding a center of mass in the particular case where the objective function is twice-continuously differentiable (in a small region at least). As observed by the authors, the main challenge is that the underlying cost function is usually neither globally differentiable nor globally convex on the general manifold.

Computing the Riemannian center of mass is just one of the many optimization problems arising in various applications, which are posed on manifolds and require a manifold structure (not necessarily with linear structure). Therefore, a large amount of classical notions and methods have been extended for problems formulated on Riemannian manifolds as well as on other spaces with nonlinear structures, e.g., [12,14–21]. As far as we know, one first propose dealing with the steepest descent method for continuously differentiable functions in the Riemannian setting was presented by Luemberger [22] and later by Gabay [23] both in the particular case where M is the inverse image of regular value. Udriste [24], Smith [25], Rapcsák [26] also studied that method in the case that the Riemannian manifold is a any complete Riemannian manifold and partial convergence results were obtained. A result of full convergence, using Armijo's rule, was obtained by Cruz Neto et al. [27], in the convex case by assuming that the Riemannian manifold has a nonnegative sectional curvature. The influence of the sign of the sectional curvature in the convergence analysis presented in [27] combined with the assumption of convexity of the objective function is directly related to the classic demonstration script adopted in the linear scenario, namely based on the quasi-Fejér convergence to the minimizers set, and these manifolds have geometric properties which are favorable to the characterization of the quasi-Fejér convergence of any sequence generated by the method.

Recent research papers have provided an unified framework for the convergence analysis of classical descent methods when the real-valued objective function is not necessarily a convex function, but that satisfies the well-known Kurdyka–Łojasiewicz

inequality; see, for instance, Absil et al. [28], Attouch et al. [29] and Lageman [30, Theorem 2.1.22]. This inequality was first introduced by Lojasiewicz [31], for real analytic functions, and extended for differentiable definable functions in an o-minimal structure by Kurdyka [32]. For extensions of Kurdyka–Łojasiewicz inequality to subanalytic nonsmooth functions (defined in Euclidean spaces), see, for example, Bolte et al. [33], Attouch and Bolte [34]. A more general extension, again in Euclidean context, was developed by Bolte et al. [35] for Clarke's subdifferentiable of a lower semicontinuous function definable in an o-minimal structure. Lageman [30] extended the Kurdyka–Łojasiewicz inequality for analytic manifolds and differentiable $\mathcal{C}$-functions in an analytic–geometric category, and, in particular, a full convergence result is shown for discrete gradient-like algorithm on a general Riemannian manifold when the differentiable $\mathcal{C}$-functions satisfy a certain descent condition, namely angle and Wolfe–Powell conditions; see [30, Example of Theorem 2.1.22, page 96]). It is important to note that Kurdyka et al. [36] had already established an extension of the Kurdyka–Łojasiewicz inequality for analytic manifolds and analytic functions in order to solve R. Thom's conjecture. Bento et al. in [37] presented an abstract convergence result for inexact descent methods in the Riemannian context and, in particular, extended the applicability of the gradient (with Armijo's rule) method to solving any problem which may be formulated as of minimizing a definable function (e.g., analytic) restricted, for example, to Hadamard manifolds; see also [38].

Following [12], who studied convergence of the proximal point method for computing center of mass on Hadamard manifolds, in this present paper it is studied the convergence of the steepest descent method with respect to the same problems on Hadamard manifolds. Taking into account the results presented in [13], it is worth mentioning that the main novelty of this work is to study convergence of the steepest descent method to a specific case of the problem (L^1 center of mass), which is also frequently used and not considered in [13].

The paper is organized as follows. In Sect. 2, we take a brief look at basic facts and notations regarding Riemannian geometry and Kurdyka–Łojasiewicz property and we present the steepest descent method. In Sect. 3, we use the steepest descent method for computing Riemannian center of mass of a set of data points on Hadamard manifolds. In Sect. 4, the robustness and efficiency of the method are illustrated with some numerical experiments in the context of positive definite symmetric matrices using the fixed stepsize and the Armijo's rule.

2 Preliminaries

In this paper, M is assumed to be finite-dimensional Hadamard manifold. The notations, results and concepts of Riemannian geometry used throughout the paper can be found in Udriste [24] and Ferreira and Oliveira [14].

Let $f : M \rightarrow \mathbb{R}$ be a differentiable function. The function f is said to have the Kurdyka–Łojasiewicz property at a point $\bar{x} \in M$ iff there exists $\eta \in]0, +\infty]$, a neighborhood U of $\bar{x}$ and a continuous concave function $\varphi : [0, \eta[\rightarrow \mathbb{R}_+$ (called desingularizing function) such that:

(i) $\varphi(0) = 0$, $\varphi \in C^1(]0, \eta[)$ and, for all $s \in]0, \eta[$, $\varphi'(s) > 0$;

(ii) For all $x \in U \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$, the Kurdyka–Łojasiewicz inequality holds

$$\varphi'(f(x) - f(\bar{x}))||\mathrm{grad}\, f(x)|| \geq 1, \tag{1}$$

where $[f(\bar{x}) < f < f(\bar{x}) + \eta] := \{x \in M : f(\bar{x}) < f(x) < f(\bar{x}) + \eta\}$.

If $f(\bar{x}) = 0$, then (1) can be rewritten as

$$||\nabla(\varphi \circ f)(x)|| \geq 1.$$

One can easily check that the Kurdyka–Łojasiewicz property is automatically satisfied at any noncritical point; see [39, Lemma 3.6]. The most fundamental works on this subject are of course due to Łojasiewicz [31] and Kurdyka [32].

Remark 2.1 A C^2 function is said to be a Morse function if at each critical point $\bar{x}$ of f the Hessian of f at $\bar{x}$, denoted by Hess $f(\bar{x})$, has all its eigenvalues different from zero. The class of Morse functions form a dense and open set in the space of C^2-functions; see Hirsh [40, Theorem 1.2]. It is known that a Morse function satisfies the Kurdyka–Łojasiewicz property at every point with desingularizing function $\varphi(t) = c\sqrt{t}$, for some $c > 0$; see [39, Theorem 3.8]. In particular, it follows from [24, Theorem 1.2] that every C^2 strongly convex function satisfies the Kurdyka–Łojasiewicz property at every point. A very rich class of functions satisfying the Kurdyka–Łojasiewicz property are the analytic functions defined on analytic manifolds; see [36].

Next, we recall the steepest descent method for solving a minimization problem

$$\min_{x \in M} f(x), \tag{2}$$

where $f : M \to \mathbb{R}$ is continuously differentiable with Lipschitz gradient. The steepest descent method generates a sequence as follows:

Steepest Descent Method (SDM)

Step 1. Choose $x^0 \in M$.

Step 2. Given x^k, if x^k is a critical point of f, then set $x^{k+p} = x^k$ for all $p \in \mathbb{N}$. Otherwise, compute $t_k > 0$ some real stepsize and take as the next iterate $x^{k+1} \in M$ such that

$$x^{k+1} = \exp_{x^k}(-t_k \mathrm{grad}\, f(x^k)). \tag{3}$$

Note that from (3), we have

$$t_k||\mathrm{grad}\, f(x^k)|| = d(x^{k+1}, x^k). \tag{4}$$

Thus, if $x^{k+1} = x^k$, then one can easily check that x^k is a critical point of f taking into account that $t_k > 0$.

The next convergence results of the steepest descent method, in the context of Kurdyka–Łojasiewicz inequality in Hadamard manifolds, are proved in [37, Theorem

7.1]. In particular setting of Łojasiewicz inequality, it can be found in [38, Theorem 2.3]. In [37], they consider two possibilities of the stepsize rule:

– (Armijo's rule) Let $\{t_k\}$ be a sequence obtained by

$$t_k := \max \left\{ 2^{-j} : j \in \mathbb{N}, \ f\left(\exp_{x^k}(-2^{-j} \operatorname{grad} f(x^k))\right) \leq f(x^k) \right. $$
$$\left. - \alpha 2^{-j} \, ||\operatorname{grad} f(x^k)||^2 \right\},$$

with $\alpha \in]0, 1[$.
– (Fixed stepsize rule) Given $\delta_1, \delta_2 > 0$ such that $L\delta_1 + \delta_2 < 1$, where L is the Lipschitz constant associated with the gradient map of f. Take $\{t_k\}$ such that

$$t_k \in \left]\delta_1, \frac{2}{L}(1 - \delta_2)\right[, \quad \forall k \geq 0.$$

Theorem 2.1 *Assume that the sequence $\{x^k\}$ generated by SDM has an accumulation point $x^* \in M$ and f satisfies the Kurdyka–Łojasiewicz inequality at x^*. Then, $\lim_{k \to +\infty} f(x^k) = f(x^*)$ and $\{x^k\}$ converges to x^* which is a critical point of f.*

Remark 2.2 In the previous result, M is supposed to be a Hadamard manifold. However, it also works on any set where uniqueness of minimizing geodesics holds, for example in convex balls (without any assumption on the signal of the curvature); see [37, Remark 7.2]. Therefore, our results presented on the sequel also hold in this context. We choose to handle with Hadamard manifolds just for the sake of simplicity.

3 Computing Riemannian L^p Center of Mass on Hadamard Manifolds

Consider the problem of computing (global) Riemannian L^p center of mass of the data set $\{a_i\}_{i=1}^n \subset M$ on a Hadamard manifold with respect to weights $0 \leq w_i \leq 1$, such that $\sum_{i=1}^n w_i = 1$. The Riemannian L^p center of mass is defined as the solution set of the following problem:

$$\min_{x \in M} f_p(x) := \frac{1}{p} \sum_{i=1}^n w_i d^p(x, a_i), \tag{5}$$

for $1 \leq p < \infty$. If $p = \infty$, the center of mass is defined as the minimizers of $\max_i d(x, a_i)$ in M. As a convention, when referring to the center of mass of some data points, we usually do not refer to explicit weights unless needed. In this paper, our focus is limited to the cases $p = 1$ and $p = 2$, which are the most commonly used cases.

As mentioned before, the problem of computing the Riemannian center of mass has been extensively studied in both theory and applications. In [13], the authors investigate the convergence of constant stepsize gradient descent algorithms for computing L^2 center of mass in Riemannian manifolds with curvature bounded from above and below. In [3,4] is studied incremental gradient methods for computing L^2 center of

mass in the context of positive definite symmetric matrices, which is a particular instance of Hadamard manifolds. Therefore, in some sense, our results generalize the previously mentioned works. The main novelty of this paper, in relation to [13], is the extension of the convergence analysis of the steepest descent method for computing L^1 center of mass. In this case, the main challenge is that the underlying cost function is not globally differentiable. Recently, Bačák [12] performed the proximal point method for computing L^1 and L^2 center of mass in a more general nonlinear setting in which a Hadamard manifold is a particular instance, namely CAT(0) spaces. We provide an alternative way to solve this problem by using the steepest descent method which is the most popular and easiest algorithm for solving a differentiable minimization problem.

One can check that

$$M \ni x \longmapsto \operatorname{grad} f_p(x) = - \sum_{i=1}^{n} w_i d^{p-2}(x, a_i) \exp_x^{-1} a_i, \qquad (6)$$

having in mind that, in the particular case where $p < 2$, x should be different from any of the data points.

Next, we recall a useful estimate about the Hessian of the Riemannian distance function. Assume that $x, y \in M$ (y distinct from x) and $t \mapsto \gamma(t)$ with $\gamma(0) = x$ is a unit speed geodesic making, at x, an angle β with the (minimal) geodesic from y to x. It can be proved that (see, e.g., [41, pp. 152–154])

$$\frac{\sin^2 \beta}{d(x, y)} \leq \frac{d^2}{dt^2} d(\gamma(t), y)|_{t=0}. \qquad (7)$$

Recall that the points $a_1, \ldots, a_n$ are said to be collinear if they reside on the same geodesic, i.e., there exist $y \in M$, $v \in T_y M$ and $t_i \in \mathbb{R}$, $i = 1, \ldots, n$, such that $a_i = \exp_y t_i v$, for each $i = 1, \ldots, n$.

Proposition 3.1 *Let $\{a_i\}_{i=1}^{n} \subset M$ be the data set and let γ be a unit speed geodesic such that $\gamma(0) = x$, where $x \neq a_i$, for $i = 1, \ldots, n$. Then, there exists a constant $\alpha \geq 0$ such that*

$$\alpha \leq \frac{d^2}{dt^2} (f_1 \circ \gamma)(t)|_{t=0}.$$

Furthermore, if the points $a_1, \ldots, a_n$ are not collinear, then $\alpha > 0$.

Proof Let β_i be the angle which the unit speed geodesic γ such that $\gamma(0) = x$ makes with the geodesic ξ_i connecting the points x and a_i. Applying (7) for $y = a_i$, for each $i = 1, \ldots, n$ and summing up these inequalities, we obtain

$$\sum_{i=1}^{n} \frac{\sin^2 \beta_i}{d(x, a_i)} \leq \sum_{i=1}^{n} \frac{d^2}{dt^2} d(\gamma(t), a_i)|_{t=0}. \qquad (8)$$

Denote by

$$\alpha = \frac{1}{\max_i d(x, a_i)} \sum_{i=1}^{n} \sin^2 \beta_i \geq 0.$$

Thus, from (8), we obtain

$$\alpha \leq \frac{d^2}{dt^2}(f_1 \circ \gamma)(t)|_{t=0}.$$

Now, if the points $a_1, \ldots, a_n$ are not collinear, then there exists at least one point, namely a_{i_0}, with $i_0 \in \{1, \ldots, n\}$ such that $\sin^2 \beta_{i_0} > 0$ which implies that $\alpha > 0$. □

Before analyzing the convergence of an algorithm for computing Riemannian L^1 center of mass, we state and prove some preliminary results.

Proposition 3.2 *The following statements hold:*

(a) *The function $f_1(x) = \sum_{i=1}^{n} w_i d(x, a_i)$ is convex;*
(b) *The problem (5), for $p = 1$, always has a solution. Furthermore, if the points $a_1, \ldots, a_n$ are not collinear, then the solution is unique;*
(c) *Let $i_0 \in \{1, \ldots, n\}$ be an index such that $f_1(a_{i_0}) = \min_{i=1,\ldots,n} f_1(a_i)$. Then, a_{i_0} is a minimizer of f_1 on M if and only if*

$$\left\| \sum_{i=1,i\neq i_0}^{n} w_i \frac{\exp_{a_{i_0}}^{-1} a_i}{d(a_i, a_{i_0})} \right\| \leq w_{i_0}.$$

Proof (a) The proof follows from the convexity of the distance function and properties of convex functions; see for instance [17, Lemma 5.2] and [24, Theorem 3.3].

(b) Since f_1 is continuous and coercive function which is bounded from below, f_1 has a minimizer. From the second part of Proposition 3.1, we have that the eigenvalues of Hess $f_1(x)$ are positive which implies that f_1 is strictly convex. Thus, f_1 has a unique minimizer.

(c) Since f_1 is convex, we have that a_{i_0} is a minimizer of f_1 if and only if $0 \in \partial f_1(a_{i_0})$. On the other hand, combining (6) with [17, Theorem 5.3], we obtain

$$\partial f_1(a_{i_0}) = \sum_{i=1,i\neq i_0}^{n} w_i \frac{\exp_{a_{i_0}}^{-1} a_i}{d(a_i, a_{i_0})} + w_{i_0} \mathbb{B}_{a_{i_0}},$$

where $\mathbb{B}_{a_{i_0}}$ denotes the unit ball with center in a_{i_0}. Thus, a_{i_0} is a minimizer of f_1 if and only if there exists $\theta_{i_0} \in \mathbb{B}_{a_{i_0}}$ such that

$$\sum_{i=1, i \neq i_0}^{n} w_i \frac{\exp_{a_{i_0}}^{-1} a_i}{d(a_i, a_{i_0})} = -w_{i_0} \theta_{i_0},$$

and the assertion follows taking the norm in both sides of the last equation having in mind that $\|\theta_{i_0}\| \leq 1$. $\square$

Remark 3.1 We directly conclude from Proposition 3.1 that if in the data set $\{a_i\}_{i=1}^{n}$ the points are not collinear, then the Hessian of the function $f_1(x) = \sum_{i=1}^{n} w_i d(x, a_i)$ has all its eigenvalues positive which means that f_1 is a Morse function. Thus, from [39, Theorem 3.8], we have that f_1 satisfies the Kurdyka–Łojasiewicz property at every point.

Proposition 3.3 *Let C be a compact set such that $a_i \notin C$, for each $i = 1, \ldots, n$. Then, the vector field* grad $f_1 : M \to TM$ *is Lipschitz continuous on C.*

Proof From the fact that C is compact, we have the manifold M which has the sectional curvature bounded from above and below on C denoted by Δ and δ, respectively. Take any $x \in C$, whence $x \neq a_i$ for all $i = 1, \ldots, n$, and let $\gamma(t)$ be a unit speed geodesic, with $\gamma(0) = x$, making at x an angle β_i with the minimal geodesic from a_i to x. From Proposition 3.1, there exists a constant $\alpha \geq 0$ such that

$$\alpha \leq \frac{d^2}{dt^2}(f_1 \circ \gamma)(t)|_{t=0}.$$

On the other hand, since M is a Hadamard manifold, we have that $\delta \leq 0$. It follows from [41, pp. 152–154] that

$$\frac{d^2}{dt^2}(f_1 \circ \gamma)(t)|_{t=0} \leq \sum_{i=1}^{n} ct_\delta(d(x, a_i)) \sin^2 \beta_i, \qquad (9)$$

where $ct_\delta(l) = \dfrac{1}{l}$, if $\delta = 0$, and $ct_\delta(l) = \sqrt{|\delta|} \coth(\sqrt{|\delta|}l)$, if $\delta < 0$. Since $a_i \notin C$, for each $i = 1, \ldots, n$, we have that $d(x, a_i)$ is positive and from the fact that C is compact we obtain that the right-hand side of (9) has an upper bound. Thus, the Hessian of f_1 at every $x \in C$ is bounded which means that grad f_1 is Lipschitz continuous on C. $\square$

Now, we present a version of the steepest descent method for computing Riemannian L^1 center of mass of the data set $\{a_i\}_{i=1}^{n} \subset M$. To this end, we denote the index $q \in \{1, \ldots, n\}$ such that $f_1(a_q) = \min_{i=1,\ldots,n} f_1(a_i)$. We also assume that a_q is not a solution (Riemannian L^1 center of mass). Next, we consider $\{x^k\}$ a sequence generated by Algorithm 1 for computing Riemannian L^1 center of mass, i.e., taking $f = f_1$. Note that for any $x^0 \in M$, we have that $x^k \in L_{f_1}(x^0)$, for all $k \in \mathbb{N}$, and $L_{f_1}(x^0)$ is a nonempty and compact set. Then, we consider a direction d_q and t_q small enough such

that $f_1(\exp_{a_q} t_q d_q) < f_1(a_q)$. Setting $x^0 := \exp_{a_q} t_q d_q$, we have that $a_i \notin L_{f_1}(x^0)$, for each $i = 1, \ldots, n$.

Theorem 3.1 *The sequence $\{x^k\}$ converges to the unique Riemannian L^1 center of mass of the data set $\{a_i\}_{i=1}^n$ as long as the points a_i, for $i = 1, \ldots, n$, are not collinear.*

Proof Since $x^k \in L_{f_1}(x^0)$, for all $k \in \mathbb{N}$ and $L_{f_1}(x^0)$ is a compact set, we have that $\{x^k\}$ is bounded. Hence, let $\hat{x}$ be a cluster point of $\{x^k\}$. Since the points a_i, for $i = 1, \ldots, n$, are not collinear, from Remark 3.1 we have that $f_1(x) = \sum_{i=1}^n w_i d(x, a_i)$ satisfies the Kurdyka–Łojasiewicz inequality at $\hat{x}$. Thus, from Theorem 2.1, we obtain that $\{x^k\}$ converges to $\hat{x}$ which is a critical point of f_1. From Proposition 3.2 (a), we have that f_1 is convex which means that $\hat{x}$ is a minimizer of f_1. From Proposition 3.2 (c), such a minimizer is unique. This completes the proof. □

Before analyzing the convergence of an algorithm for computing Riemannian L^2 center of mass, we state and prove some preliminary results.

Proposition 3.4 *The following statements hold:*

(a) *The function $f_2(x) = \frac{1}{2} \sum_{i=1}^n w_i d^2(x, a_i)$ is strictly convex and continuously differentiable with its gradient Lipschitz on compact sets;*
(b) *The problem of computing Riemannian L^2 center of mass always has a unique solution.*

Proof (a) The strictly convexity and continuously differentiability of f_2 can be found for instance in [24, page 111]. Let $C \subset M$ be a compact set. Then, M has the sectional curvature bounded from above and below on C denoted by Δ and δ, respectively. Take any $x \in C$ and let $\gamma(t)$ be a unit speed geodesic, with $\gamma(0) = x$, making at x an angle β_i with the minimal geodesic from a_i to x. One can verify from [41, pp. 152–154] that

$$\frac{d^2}{dt^2}(f_2 \circ \gamma)(t)|_{t=0} \leq \sum_{i=1}^n c_\delta(d(x, a_i)) \sin^2 \beta_i, \qquad (10)$$

where $c_\delta(l) = 1$, if $\delta = 0$, and $c_\delta(l) = \sqrt{|\delta|} \coth(\sqrt{|\delta|}l)$, if $\delta < 0$. Therefore, an upper bound to (10) follows from the fact that C is compact. From [42, Corollary 3.1], we have that $d^2(\cdot, z)$ is a strongly convex function, and hence

$$0 \leq \frac{d^2}{dt^2}(f_2 \circ \gamma)(t)|_{t=0}.$$

Thus, the Hessian of f_2 at every $x \in C$ is bounded which means that grad f_2 is Lipschitz continuous on C.
(b) Since f_2 is continuous and coercive, f_2 has a minimizer. The uniqueness follows from the strictly convexity of f_2. □

Remark 3.2 As already mentioned in Introduction, in [11] the authors dealt with Problem 5. It is worth mentioning that, among other things, the authors in particular studied

the existence of solutions to the L^1 case, investigated in Proposition 3.2 item b), and L^2, dealt with in Proposition 3.4 item b). We chose to keep a proof for each of these items just for the sake of completeness.

It follows from [42, Corollary 3.1] that $d^2(\cdot, z)$ is a strongly convex function. Since the finite sum of strongly convex functions is still strongly convex, from Remark 2.1 one has the following result:

Proposition 3.5 *The function $f_2(x) = \frac{1}{2} \sum_{i=1}^{n} w_i d^2(x, a_i)$ satisfies the Kurdyka–Łojasiewicz inequality at every point of M.*

Next, we consider $\{x^k\}$ the sequence generated by SDM for computing the Riemannian L^2 center of mass, i.e., taking $f = f_2$. Similar to the case $p = 1$, we have that for any $x^0 \in M$, $x^k \in L_{f_2}(x^0)$, for all $k \in \mathbb{N}$, and $L_{f_2}(x^0)$ is a nonempty and compact set. Then, we consider a direction d_q and t_q small enough such that $f_2(\exp_{a_q} t_q d_q) < f_2(a_q)$, where q denotes the index in $\{1, \ldots, n\}$ such that $f_2(a_q) = \min_{i=1,\ldots,n} f_2(a_i)$. Setting $x^0 := \exp_{a_q} t_q d_q$, we have that $a_i \notin L_{f_2}(x^0)$, for each $i = 1, \ldots, n$.

Theorem 3.2 *The sequence $\{x^k\}$ converges to the unique Riemannian L^2 center of mass of the data set $\{a_i\}_{i=1}^{n}$.*

Proof Since $x^k \in L_{f_2}(x^0)$, for all $k \in \mathbb{N}$ and $L_{f_2}(x^0)$ is a compact set, we have that $\{x^k\}$ is bounded. Hence, let $\hat{x}$ be a cluster point of $\{x^k\}$. From Proposition 3.5, we have that $f_2(x) = \frac{1}{2} \sum_{i=1}^{n} w_i d^2(x, a_i)$ satisfies the Kurdyka–Łojasiewicz inequality at $\hat{x}$. Thus, from Theorem 2.1, we obtain that $\{x^k\}$ converges to $\hat{x}$ which is a critical point of f_2. Proposition 3.4 (a) implies that f_2 is strictly convex which means that $\hat{x}$ is the unique minimizer of f and the proof is completed. $\qquad\square$

4 Numerical Experiments

This section is devoted to some numerical experiments. We coded our simulations in MATLAB on a machine AMD Athlon(tm) X2 with a Dual Core Processor and 1.20 GHz CPU.

Let $\mathbb{S}^m$ be the set of the symmetric matrices, $\mathbb{S}^m_+$ be the cone of the symmetric positive semi-definite matrices and $\mathbb{S}^m_{++}$ be the cone of the symmetric positive definite matrices both $m \times m$. For $X, Y \in \mathbb{S}^m$, $Y \succeq X$ (or $X \preceq Y$) means that $Y - X \in \mathbb{S}^m_+$ and $Y \succ X$ (or $X \prec Y$) means that $Y - X \in \mathbb{S}^m_{++}$. Let $M := (\mathbb{S}^m_{++}, \langle\,,\,\rangle)$ be the Riemannian manifold endowed with the Riemannian metric induced by the Euclidean Hessian of $\Psi(X) = -\ln \det X$,

$$\langle U, V \rangle = \mathrm{tr}\,(V \Psi''(X) U) = \mathrm{tr}\,(V X^{-1} U X^{-1}), \quad X \in M, \quad U, V \in T_X M, \quad (11)$$

where $\mathrm{tr}\,(A)$ denotes the trace of matrix $A \in \mathbb{S}^m$ and $T_X M \approx \mathbb{S}^m$, with the corresponding norm denoted by $\|\,.\,\|$; see Rothaus [43]. In this case, for any $X, Y \in M$ the unique geodesic joining those two points is given by:

$$\gamma(t) = X^{1/2} \left(X^{-1/2} Y X^{-1/2} \right)^t X^{1/2}, \qquad t \in [0, 1];$$

see, for instance, [44]. More precisely, M is a Hadamard manifold; see, for example, [10, Theorem 1.2., page 325]. One can compute the curvature of M and verify that M has nonconstant curvature; see [45, page 428]. From the above equality, it is immediate that

$$\gamma'(0) = X^{1/2} \ln \left(X^{-1/2} Y X^{-1/2} \right) X^{1/2}.$$

Thus, for each $X \in M$, $\exp_X : T_X M \to M$ and $\exp_X^{-1} : M \to T_X M$ are given, respectively, by

$$\exp_X V = X^{1/2} e^{\left(X^{-1/2} Y X^{-1/2} \right)} X^{1/2}, \quad \exp_X^{-1} Y = X^{1/2} \ln \left(X^{-1/2} Y X^{-1/2} \right) X^{1/2}. \tag{12}$$

Now, since the Riemannian distance d is given by $d(X, Y) = || \exp_X^{-1} Y ||$, from (11) along with second expression in (12), we conclude

$$d^2(X, Y) = \mathrm{tr} \left(\ln^2 X^{-1/2} Y X^{-1/2} \right) = \sum_{i=1}^{n} \ln^2 \lambda_i \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right), \tag{13}$$

where $\lambda_i (X^{-\frac{1}{2}} Y X^{-\frac{1}{2}})$ denotes the ith eigenvalue of the matrix $X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}$.

The numerical experiment aims to verify the efficiency of the gradient method in locating L^1 and L^2 center of mass on Hadamard manifolds. In our simulations, we consider different scenes taking into account three parameters: the number of matrices n in the data set $\{Q_i\}_{i=1}^{n}$, the size $m \times m$ of the matrices and the stopping rule $\varepsilon > 0$. The random matrices we use for our test are generated with an uniform (well-conditioned) and non-uniform (ill-conditioned) distribution of the eigenvalues of each matrix of the data set. Hence, the non-uniform distribution satisfies $\dfrac{\lambda_{\max}}{\lambda_{\min}} > 10^2$, where $\lambda_{\max}$ and $\lambda_{\min}$ denote the largest and the smallest eigenvalues of each matrix of the data set, respectively. Table 1 reports the number of iterations, CPU time in seconds and gradient norm of the sequence generated by Algorithm 1 for finding L^2 center of mass of the data set $\{Q_i\}_{i=1}^{n}$ with $m \times m$ matrices. The constant stepsize $\{t_k\}$ satisfying the fixed stepsize rule and weights $\{w_i\}_{i=1}^{n}$ were taken as $t_k = w_i = \frac{1}{n}$, for all $k \in \mathbb{N}$ and $i = 1, \ldots, n$ with $\delta_1 = \frac{1}{2n}$ and $\delta_2 = \frac{1}{4}$. In these simulations, we use the stopping rule $d(x^{k+1}, x^k) < 10^{-8}$ (Figs. 1, 2).

We report to the substantial reduction in the number of iterations obtained from non-uniform to uniform distribution of the eigenvalues.

Table 1 Computing L^2 center of mass

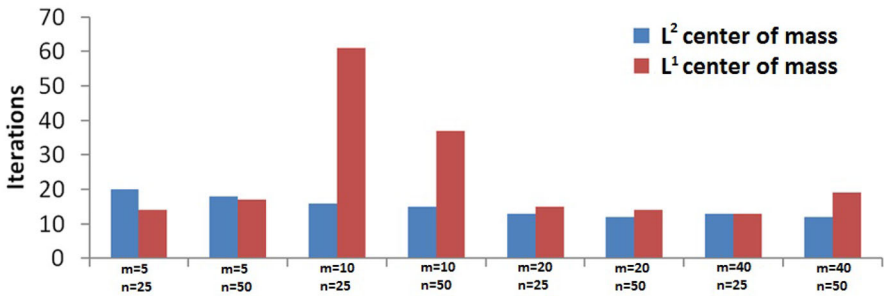| Scene		Distribution of eigenvalues					
		Non-uniform			Uniform		
m	n	Iterations	Time (s)	Gradient norm	Iterations	Time (s)	Gradient norm
5	25	25	3.3750	8.749664E−09	4	0.5010	1.595908E−09
5	50	24	5.6740	5.934015E−09	4	0.9740	1.060440E−10
5	75	21	7.4900	5.727201E−09	4	1.4600	1.631894E−10
5	100	22	10.4980	5.910573E−09	4	1.9100	1.906906E−10
10	25	20	8.1450	5.070213E−09	4	1.5840	1.440047E−09
10	50	19	14.7270	6.388001E−09	4	3.2520	8.882587E−10
10	75	17	19.6690	6.927831E−09	4	4.6280	6.153625E−10
10	100	18	27.2100	3.801648E−09	4	6.1710	5.657227E−10
20	25	17	26.3320	9.863386E−09	4	6.4640	2.814672E−09
20	50	16	48.2870	5.947555E−09	4	12.3150	1.669254E−09
20	75	15	67.5530	4.778214E−09	4	17.8580	1.793198E−09
20	100	15	90.3390	5.772144E−09	4	24.1080	2.184789E−09
40	25	16	105.4830	4.922728E−09	4	27.0140	4.602772E−09
40	50	14	180.8780	9.535475E−09	4	53.6340	3.741021E−09
40	75	14	269.8180	6.516032E−09	4	79.8530	3.070095E−09
40	100	14	360.3390	3.418268E−09	4	106.3220	2.906204E−09

Fig. 1 Non-uniform distribution
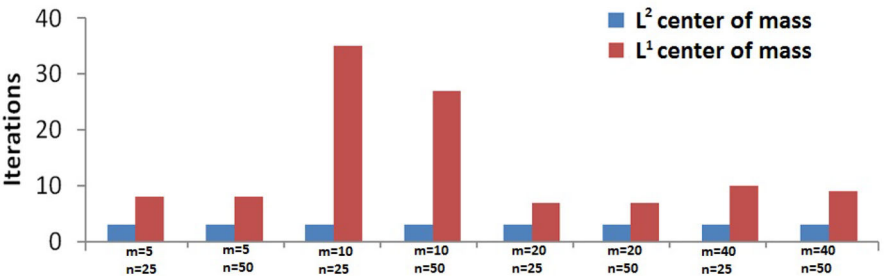
Fig. 2 Uniform distribution

Table 2 Computing L^1 center of mass

| Scene		Distribution of eigenvalues					
		Non-uniform			Uniform		
m	n	Iterations	Time (s)	Gradient norm	Iterations	Time (s)	Gradient norm
5	25	14	5.4600	1.615722E−07	8	2.5270	6.431948E−07
5	50	17	13.5570	4.462231E−08	8	4.9760	2.722158E−07
10	25	61	61.5860	8.960679E−01	35	23.0250	7.577621E−07
10	50	37	74.6150	3.736156E−07	27	35.4900	6.410049E−07
20	25	15	58.5940	3.808496E−07	7	17.8150	3.572900E−07
20	50	14	108.6230	3.813695E−01	7	35.0690	3.833297E−07
40	25	13	216.6850	1.659700E−07	10	107.8430	6.909771E−01
40	50	19	626.5430	9.074286E−09	9	195.2500	5.180544E−07
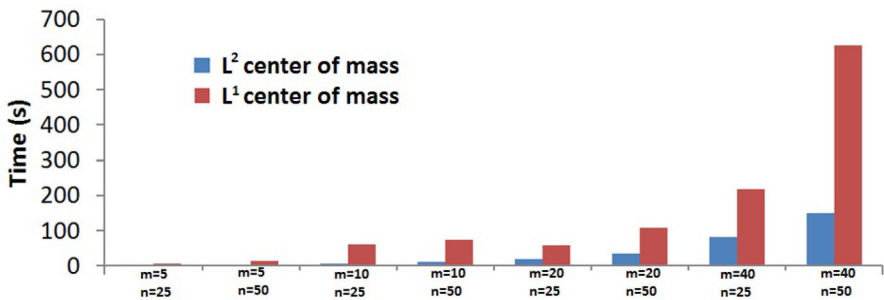
Fig. 3 Non-uniform distribution

Next, we perform Algorithm 1 for computing L^1 center of mass. The main challenge is that often the underlying function is not globally differentiable and despite this one would like to have guaranteed effectiveness of the method. Therefore, before starting the method, we determine a direction d_q and t_q small enough such that $f(\exp_{a_q} t_q d_q) < f(a_q)$, where $q \in \{1, \ldots, n\}$ satisfies $f(a_q) = \min_{i=1,\ldots,n} f(a_i)$, and hence set $x^0 = \exp_{a_q} t_q d_q$. Thus, we avoid the non-differentiability of the function under consideration. Table 2 reports the number of iterations, CPU time in seconds and gradient norm of the sequence generated by Algorithm 1 for finding L^1 center of mass of the data set $\{Q_i\}_{i=1}^n$ with $m \times m$ matrices and using the stepsize given by the Armijo's rule. We use the stopping rule $d(x^{k+1}, x^k) < 10^{-6}$.

We can note by comparing Tables 1 and 2 that, even with a uniform distribution of the eigenvalues, computing L^1 center of mass demands more efforts in both number of iterations and CPU time than to compute L^2 center of mass. Next figures plot side by side the number of iterations and CPU time necessary for computing L^1 and L^2 center of mass of a data set. In this simulation, we consider the eight scenarios presented in Table 2 for values of n and m (Figs. 3, 4).
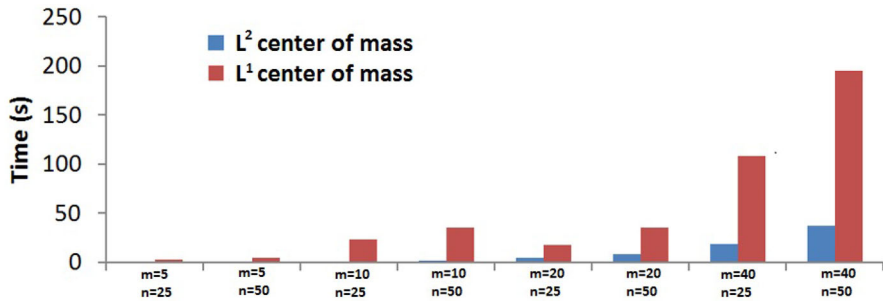
Fig. 4 Uniform distribution

5 Conclusions

In this paper, we have considered the steepest descent method for computing Riemannian center of mass on Hadamard manifolds. As mentioned in Remark 2.2, we choose to handle with Hadamard manifolds just for the sake of simplicity because our results also work on any set where the uniqueness of minimizing geodesics holds, for example in convex balls (without any assumption on the signal of the curvature). Some numerical experiments computing L^1 and L^2 center of mass in the context of positive definite symmetric matrices are presented using two different stepsize rules.

References

1. Karcher, H.: Riemannian Center of Mass and so called Karcher mean, pp. 1–3. arXiv:1407.2087 (2014)
2. Grove, K., Karcher, H.: How to conjugate C^1-close group actions. Math. Z. **132**(1), 11–20 (1973)
3. Kum, S., Sangwoon, Y.: Incremental gradient method for Karcher mean on symmetric cones. J. Optim. Theory Appl. **172**, 141–155 (2017)
4. Bini, D.A., Iannazzo, B.: Computing the Karcher mean of symmetric positive definite matrices. Linear Algebra Appl. **438**, 1700–1710 (2013)
5. Moakher, M.: A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM J. Matrix Anal. Appl. **26**, 735–747 (2005)
6. Moakher, M.: On the averaging of symmetric positive-definite tensors. J. Elast. **82**, 273–296 (2006)
7. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton (2008)
8. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Geometric means in a novel vector space structure on symmetric positive-definite matrices. SIAM J. Matrix Anal. Appl. **29**, 328–347 (2007)
9. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. Int. J. Comput. Vis. **66**, 41–66 (2006)
10. Lang, S.: Fundamentals of Differential Geometry. Springer, Berlin (1998)
11. Kristály, A., Morosanu, G., Róth, A.: Optimal placement of a deposit between markets: Riemannian–Finsler geometrical approach. J. Optim. Theory Appl. **139**(2), 263–276 (2008)
12. Bačák, M.: Computing medians and means in Hadamard spaces. SIAM J. Optim. **24**, 1542–1566 (2014)
13. Afsari, B., Tron, R., Vidal, R.: On the convergence of gradient descent for finding the riemannian center of mass. SIAM J. Control Optim. **51**, 2230–2260 (2013)

14. Ferreira, O.P., Oliveira, P.R.: Proximal point algorithm on Riemannian manifolds. Optimization **51**, 257–270 (2002)
15. Hosseini, S., Pouryayevali, M.R.: On the metric projection onto prox-regular subsets of Riemannian manifolds. Proc. Am. Math. Soc. **141**, 233–244 (2013)
16. Kristály, A.: Nash-type equilibria on Riemannian manifolds: a variational approach. J. Math. Pures Appl. **101**(5), 660–688 (2014)
17. Li, C., Mordukhovich, B.S., Wang, J., Yao, J.C.: Weak sharp minima on Riemannian manifolds. SIAM J. Optim. **21**, 1523–1560 (2011)
18. Li, C., López, G., Martín-Márquez, V.: Monotone vector fields and the proximal point algorithm on Hadamard manifolds. J. Lond. Math. Soc. **79**, 663–683 (2009)
19. Li, C., Yao, J.C.: Variational inequalities for set-valued vector fields on Riemannian manifolds: convexity of the solution set and the proximal point algorithm. SIAM J. Control Optim. **50**, 2486–2514 (2012)
20. Németh, S.Z.: Variational inequalities on Hadamard manifolds. Nonlinear Anal. Theory Methods Appl. **52**(5), 1491–1498 (2003)
21. Wang, X., Li, C., Wang, J., Yao, J.C.: Linear convergence of subgradient algorithm for convex feasibility on Riemannian manifolds. SIAM J. Optim. **25**, 2334–2358 (2015)
22. Luenberger, D.G.: The gradient projection method along geodesics. Manag. Sci. **18**, 620–631 (1972)
23. Gabay, D.: Minimizing a differentiable function over a differentiable manifold. J. Optim. Theory Appl. **37**, 177–219 (1982)
24. Udriste, C.: Convex functions and optimization algorithms on Riemannian manifolds. In: Mathematics and its Applications, vol. 297. Kluwer Academic Publishers, Dordrecht (1994)
25. Smith, S.T.: Optimization Techniques on Riemannian Manifolds, vol. 3, pp. 113–146. Fields Institute Communications, American Mathematical Society, Providence (1994)
26. Rapcsák, T.: Smooth Nonlinear Optimization in $\mathbb{R}^n$. Kluwer Academic Publishers, Dordrecht (1997)
27. Cruz Neto, J.X., Lima, L.L., Oliveira, P.R.: Geodesic algorithms in riemannian geometry. Balkan J. Geom. Appl. **3**, 89–100 (1998)
28. Absil, P.-A., Mahony, R., Andrews, B.: Convergence of the iterates of descent methods for analytic cost functions. SIAM J. Optim. **16**, 531–547 (2005)
29. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods. Math. Program. Ser. A **137**, 91–129 (2013)
30. Lageman, C.: Convergence of gradient-like dynamical systems and optimization algorithms. Ph.D. thesis, Universitat Wurzburg (2007)
31. Łojasiewicz, S.: Une propriété topologique des sous-ensembles analytiques réels. Les Équations aux Dérivées Partielles, Éditions du centre National de la Recherche Scientifique, 87–89 (1963)
32. Kurdyka, K.: On gradients of functions definable in o-minimal structures. Ann. Inst. Fourier **48**, 769–783 (1998)
33. Bolte, J., Danillidis, J.A., Lewis, A.: The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. SIAM J. Optim. **17**, 1205–1223 (2006)
34. Attouch, H., Bolte, J.: On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math. Program. Ser. B **116**(1–2), 5–16 (2009)
35. Bolte, J., Danillidis, J.A., Lewis, A., Shiota, M.: Clarke subgradients of stratifiable functions. SIAM J. Optim. **18**, 556–572 (2007)
36. Kurdyka, K., Mostowki, T., Parusinki, A.: Proof of the gradient conjecture of R. Thom. Ann. Math. **152**, 763–792 (2000)
37. Bento, G. C., Cruz Neto, J. X., Oliveira, P. R.: Convergence of inexact descent methods for nonconvex optimization on Riemannian manifolds, pp. 1–28. arXiv:1103.4828v1 (2011)
38. Schneider, R., Uschmajew, A.: Convergence Results for projected line-search methods on varieties of low-rank matrices via Łojasiewicz inequality. SIAM J. Optim. **25**(1), 622–646 (2015)
39. Cruz Neto, J.X., Oliveira, P.R., Soares Jr., P.A., Soubeyran, A.: Learning how to play Nash, potential games and alternating minimization method for structured nonconvex problems on riemannian manifolds. J. Convex Anal. **20**, 395–438 (2013)
40. Hirch, M.W.: Differential Topology. Spring, New York (1976)
41. Sakai, T.: Riemannian Geometry. Translations of Mathematical Monographs. American Mathematical Society, Providence (1996)

42. Cruz Neto, J.X., Ferreira, O.P., Lucambio Pérez, L.R.: Contributions to the study of monotone vector fields. Acta Math. Hung. **94**, 307–320 (2002)
43. Rothaus, O.S.: Domains of positivity. Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg **24**, 189–235 (1960)
44. Nesterov, Y.E., Todd, M.J.: On the Riemannian geometry defined by self-concordant barriers and interior-point methods. Found. Comput. Math. **2**, 333–361 (2002)
45. Lenglet, C., Rousson, M., Deriche, R., Faugeras, O.: Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor MRI processing. J. Math. Imaging Vis. **25**, 423–444 (2006)