Convergence Analysis of Gradient Algorithms on Riemannian Manifolds Without Curvature Constraints and Application to Riemannian Mass

Chong Li∗ Xiangmei Wang† Jinhua Wang‡ Jen-Chih Yao§

Abstract We study the convergence issue for the gradient algorithm (employing general step sizes) for optimization problems on general Riemannian manifolds (without curvature constraints). Under the assumption of the local convexity/quasi-convexity (resp. weak sharp minima), local/global convergence (resp. linear convergence) results are established. As an application, the linear convergence properties of the gradient algorithm employing the constant step sizes and the Armijo step sizes for finding the Riemannian $L^p$ ($p \in [1, +\infty)$) centers of mass are explored, respectively, which in particular extend and/or improve the corresponding results in [3].

Keywords Riemannian manifold; sectional curvature; gradient algorithm; local convergence; global convergence; linear convergence; Riemannian center of mass

AMS subject classifications. 53C20; 53C22

∗School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, P. R. China (cli@zju.edu.cn). Research of this author was supported in part by the National Natural Science Foundation of China (grant number 11571308, 11971429) and Zhejiang Provincial Natural Science Foundation of China (grant numbers LY18A010004).

†College of Mathematics and Statistics, Guizhou University, Guiyang 550025, P. R. China (xmwang2@gzu.edu.cn). Research of this author was supported in part by the National Natural Science Foundation of China (grant numbers 11661019) and Guizhou Provincial Natural Science Foundation of China (grant number 20161039).

‡Department of Mathematics, Zhejiang University of Technology, Hangzhou 310032, P R China (wjh@zjut.edu.cn). Research of this author was supported in part by the National Natural Science Foundation of China (grant numbers 11771397) and Zhejiang Provincial Natural Science Foundation of China (grant number LY17A010021).

§China Medical University, Taichung, Taiwan 40402, R.O.C. (yaojc@mail.cmu.edu.tw). Research of this author was supported in part by the Grant MOST 108-2115-M-039-005-MY3.
1 Introduction

Let $f : M \to \mathbb{R}$ be a locally Lipschitz continuous function defined on a Riemannian manifold $M$. The following optimization problem:

$$\min_{x \in M} f(x)$$  \hspace{1cm} (1.1)

has been extensively studied in the literature, which not only has applications in various areas, such as computer vision, machine learning system balancing, electronic structure computation, model reduction and robot manipulation, low-rank approximation (see, e.g., [2, 3, 26] and the references therein), but also is a useful tool to treat some nonsmooth/nonconvex and/or constrained optimization problems appeared on the Euclidean space. As explained in [14], the Riemannian geometry framework can be used to decrease/overcome the difficulties caused by nonsmoothness/constraints and to enhance the performances of numerical methods by exploiting the intrinsic reduction of the dimensionality of the problem and the method’s insight about the problem structure; see also [3, 5, 7, 10, 16, 17, 18, 30] and the references therein for more details. One of the most typical and important examples is the well-known problem of finding the Riemannian $L^p$ centers of mass of given points $\{y_i : 1 \leq i \leq N\} \subseteq M$, which can be formulated as a special case of problem (1.1) with the objective function $f$ defined by

$$f(x) := \left\{ \begin{array}{ll}
\frac{1}{p} \sum_{i=1}^{N} w_i d^p(x, y_i), & 1 \leq p < +\infty \\
\max_{1 \leq i \leq N} d(x, y_i), & p = +\infty
\end{array} \right. \quad \text{for any } x \in M,$$  \hspace{1cm} (1.2)

where $\{w_i : 1 \leq i \leq N\} \subseteq (0, +\infty)$ are the weights. This problem has various applications in the field of general data analysis, including computer graphics and animation, statistical analysis of shapes, medical imaging and sensor networks (see, e.g., [5, 15] and the references therein). As mentioned in [3], the first study of the problem could be traced back to 1920s (the work due to Cartan) regarding the existence and uniqueness issue of the Riemannian $L^2$ centers of mass on Hadamard manifolds. After that, this problem was extensively studied in the literature, including the existence and uniqueness results for the Riemannian $L^p$ centers of mass and some methods for locating the Riemannian centers of mass such as gradient algorithm, subgradient algorithm, stochastic gradient algorithm, Newton’s method; see, e.g., [4, 5, 6, 15, 35].

Related to the optimization problem (1.1), some important notions and techniques, such as weak sharp minima and variational analysis, have been developed in [17, 18]; while the classical numerical methods for solving optimization problems on the Euclidean space, such as Newton’s method, gradient algorithm, subgradient algorithm, trust region method, proximal point method, etc., have been extended to the Riemannian manifold setting; see, e.g., [1, 14, 26, 27, 29, 31]. In the present paper, we are particularly interested in the gradient algorithm, which is one of the most classical and important numerical algorithms for solving problem (1.1).
The original idea of the gradient algorithm dates back to at least the work in 1972 due to Luenberger [20], where the gradient projection method employing the exact line search carried out along a geodesic was proposed for solving the constrained optimization problem on the Euclidean space, that is, problem (1.1) with \( M := \{ x \in \mathbb{R}^n : h(x) = 0 \} \) and \( h : \mathbb{R}^n \to \mathbb{R} \) being also continuously differentiable; the global (linear) convergence results were established under the assumption that the Hessian of the corresponding Lagrangian function \( f(\cdot) + \lambda h(\cdot) \) (in the sense of the Euclidean setting) is uniformly bounded and uniformly positive definite on all tangent subspaces; see [20, Theorem 1] for more details. This work was developed by Gabay in [14] with the weaker assumption that the sub-level set of \( f \) associated to \( f(x_0) \) is bounded and the values of \( f \) at all critical points are distinct; moreover the linear convergence rate is estimated under the assumptions that \( f \) is third continuous differentiable and that the generated sequence converges to a nondegenerate point at which the Hessian form of \( f \) is positive definite (see [14, (57)] for the definition of the Hessian form).

One important development in this direction is the work of Smith in [26], where he developed the gradient algorithm (together with other algorithms such as Newton-type algorithm and the conjugate gradient algorithm) for solving problem (1.1), with \( f \) being continuously differentiable on a general Riemannian manifold. By using the pure differential geometry language (which is free from local coordinate systems), he obtained the linear convergence result for the gradient algorithm (employing the exact line search) in the case when the generated sequence converges to a nondegenerate point; see [26, Theorem 2.3]. Later, Yang studied the gradient algorithm employing the Armijo step sizes on a general Riemannian manifold, and established in [36, Theorem 3.4] the global convergence result under the assumption that the generated sequence \( \{ x_k \} \) satisfies \( \lim_{k \to +\infty} d(x_k, x_{k+1}) = 0 \) and has a cluster point \( \bar{x} \) such that \( \bar{x} \) is an isolated critical point, and in [36, Theorem 4.1], the linear convergence result under the assumption that the generated sequence converges to a nondegenerate point.

To relax the isolatedness assumption for the cluster points of the generated sequence, the following two crucial assumptions were introduced in [22] and [23] to establish the global convergence results for the gradient algorithm (employing the Armijo step sizes) for the convex case and quasi-convex case, respectively:

(A1) The curvatures of the Riemannian manifold \( M \) are nonnegative.

(A2) The function \( f \) is continuously differentiable and convex/quasi-convex on the whole manifold \( M \).

As explained in the following, either assumption (A1) or (A2) is clearly too stringent.

- Assumption (A1) prevents the application to a class of Hadamard manifolds including the Poincaré plane, hyperbolic spaces \( \mathbb{H}^n \), and the symmetric positive definite matrix manifolds \( \mathbb{R}^{n \times n}_{++} \).
• Assumption (A2) prevents the application to some special but important Riemannian manifolds, such as compact Stiefel manifolds $\text{St}(p, n)$ and Grassmann manifolds $\text{Grass}(p, n)$ ($p < n$) since there is no non-trivial (quasi-)convex function (with full domain) on a complete manifold with finite volume (see, e.g., [37]).

• Assumption (A1)/(A2) prevents the application to the problem of the Riemannian $L^p$ centers of mass as, in general, the function $f$ defined by (1.2) is neither necessarily quasi-convex nor differentiable in the case when $p = 1$ on the underlying Riemannian manifolds.

Our main purpose in the present paper is to deal with the more general case in which $M$ is not necessarily of curvatures bounded from below and the function $f : M \to \mathbb{R}$ is locally Lipschitz continuous on its domain (and so not necessarily continuously differentiable, or quasi-convex/convex on the whole Riemannian manifold). More precisely, we establish the global convergence result for the gradient algorithm employing more general step sizes (which includes the Armijo step sizes as a special case) under the following weaker assumption than (A1& A2):

- The generated sequence $\{x_k\}$ has a cluster point $\bar{x}$ such that $\bar{x}$ is a critical point of $f$ and $f$ is quasi-convex around $\bar{x}$.

Moreover, if the following assumption is additionally assumed, we further show that the sequence $\{x_k\}$ converges linearly to a local solution:

- The cluster point $\bar{x}$ is a local weak sharp minimizer of order 2 for problem (1.1), $f$ is convex around $\bar{x}$, and the step size sequence $\{t_k\}$ has a positive lower bound.

As explained before Theorem 3.3 and Corollary 3.2, the linear convergence result extends [35, Theorem 4.1]; while the global convergence result extends/improves particularly the corresponding ones in [23, Theorem 3.1] (and so [22, Theorem 5.3]).

As an application, the convergence results for the gradient algorithm employing the Armijo step sizes and the constant step sizes are established, respectively, for finding the Riemannian $L^p$ centers of mass for $p \in [1, +\infty)$. We note that the (linear) convergence results for the Armijo step sizes (for $p \in [1, +\infty)$) and for the constant step sizes for $p \in [1, 2)$ seem new, while the results for the constant step sizes in the case when $p \in [2, +\infty)$ extends the corresponding one in [3 Theorem 4.1] (see the explanation before Corollary 4.3).

The paper is organized as follows. As usual, some basic notions and notation on Riemannian manifolds, together with some related properties about the convexity properties of subsets and functions, are introduced in the next section. Main results, including the local/global/linear convergence properties of the gradient algorithm on general manifolds, are presented in section 3, and the application to the Riemannian $L^p$ centers of mass is provided in the last section.
2 Notation and preliminary results

Notation and terminologies used in the present paper are standard; the readers are referred to some textbooks for more details; see, e.g., [12, 25, 28].

Let $M$ be a connected and complete $n$-dimensional Riemannian manifold. We use $\nabla$ to denote the Levi-Civita connection on $M$. Let $x \in M$, and let $T_x M$ stand for the tangent space at $x$ to $M$. We denote by $\langle , \rangle_x$ the scalar product on $T_x M$ with the associated norm $\|\cdot\|_x$, where the subscript $x$ is sometimes omitted. For $y \in M$, let $\gamma : [0, 1] \to M$ be a piecewise smooth curve joining $x$ to $y$. Then, the arc-length of $\gamma$ is defined by $l(\gamma) := \int_0^1 \|\gamma'(t)\|dt$, while the Riemannian distance from $x$ to $y$ is defined by $d(x, y) := \inf_{\gamma}, l(\gamma)$, where the infimum is taken over all piecewise smooth curves $\gamma : [0, 1] \to M$ joining $x$ to $y$. The closed metric ball and the open metric ball centered at $x$ with radius $r$ are denoted by $B(x, r)$ and $U(x, r)$, respectively, that is,

$$B(x, r) := \{ y \in M : d(x, y) \leq r \} \quad \text{and} \quad U(x, r) := \{ y \in M : d(x, y) < r \}.$$

A vector field $V$ is said to be parallel along $\gamma$ if $\nabla_{\gamma'} V = 0$. In particular, for a smooth curve $\gamma$, if $\gamma'$ is parallel along itself, then $\gamma$ is called a geodesic; thus, a smooth curve $\gamma$ is a geodesic if and only if $\nabla_{\gamma'} \gamma' = 0$. A geodesic $\gamma : [0, 1] \to M$ joining $x$ to $y$ is minimal if its arc-length equals its Riemannian distance between $x$ and $y$. By the Hopf-Rinow theorem [12], $(M, d)$ is a complete metric space, and there is at least one minimal geodesic joining $x$ to $y$ for any points $x$ and $y$.

Let $Q \subseteq M$ be a subset. As usual, we use $\overline{Q}$ and $\partial Q$ to stand for the closure and the boundary of a subset $Q \subseteq M$, respectively. The distance function $d_Q(\cdot)$ associated to $Q$ and the projection $P_Q(\cdot)$ onto $Q$ are respectively defined by, for any $x \in M$,

$$d_Q(x) := \inf_{y \in Q} d(x, y) \quad \text{and} \quad P_Q(x) := \{ y \in Q : d(x, y) = d_Q(x) \}.$$

Given points $x, y \in Q$, the set of all geodesics $\gamma : [0, 1] \to M$ with $\gamma(0) = x$ and $\gamma(1) = y$ satisfying $\gamma([0, 1]) \subseteq Q$ is denoted by $\Gamma^Q_{x y}$, that is,

$$\Gamma^Q_{x y} := \{ \gamma : [0, 1] \to Q : \gamma(0) = x, \gamma(1) = y \text{ and } \nabla_{\gamma'} \gamma' = 0 \}.$$

In particular, we write $\Gamma_{x y}$ for $\Gamma^M_{x y}$. Two important structures on $M$ will be used frequently in our study: one is the exponential map $\exp_x : T_x M \to M$, and the other is the parallel transport along the geodesic $\gamma \in \Gamma_{x y}$ denoted by $P_{\gamma, y, x}$. For simplicity, we will write $P_{y, x}$ for $P_{\gamma, y, x}$ if $\gamma \in \Gamma_{x y}$ is the unique minimal geodesic and no confusion arises.

Recall two constants related to a point $x \in M$: the injectivity radius $r_{\text{inj}}(x)$ and the convexity radius $r_{\text{cvx}}(x)$ of $x$, which are defined by

$$r_{\text{inj}}(x) := \sup \{ r > 0 : \exp_x(\cdot) \text{ is a diffeomorphism on } B(0, r) \subset T_x M \}.$$
and
\[ r_{\text{cvx}}(x) := \sup \left\{ r > 0 : \text{ each ball in } B(x, r) \text{ is strongly convex} \right\}, \]  

respectively. Then, \( r_{\text{inj}}(x) \geq r_{\text{cvx}}(x) > 0 \) for any \( x \in M \); see, e.g., [25, Theorem 5.3]. In particular, \( r_{\text{inj}}(x) = r_{\text{cvx}}(x) = +\infty \) for each \( x \in M \) if \( M \) is a Hadamard manifold. Moreover, for any compact subset \( Q \subseteq M \), we have that
\[ r_{\text{inj}}(Q) := \inf \{ r_{\text{inj}}(x) : x \in Q \} > 0 \quad \text{and} \quad r_{\text{cvx}}(Q) := \inf \{ r_{\text{cvx}}(x) : x \in Q \} > 0; \]
see [25, Theorem 5.3, p. 169] or [19, Lemma 3.1].

Definition 2.1 below presents the notions of different kinds of convexities about subsets in \( M \); see e.g., [19, 30].

**Definition 2.1.** A subset \( Q \subseteq M \) is said to be

(a) weakly convex if, for any \( x, y \in Q \), there is a minimal geodesic of \( M \) joining \( x \) to \( y \) and it is in \( Q \);

(b) strongly convex if, for any \( x, y \in Q \), there is just one minimal geodesic of \( M \) joining \( x \) to \( y \) and it is in \( Q \);

(c) totally convex if, for any \( x, y \in Q \), all geodesics of \( M \) joining \( x \) to \( y \) lie in \( Q \).

Note by definition that the strongly/totally convexity implies the weakly convexity for any subset \( Q \), and note also that \( Q \) is weakly convex if and only if so is \( \overline{Q} \).

Consider now a proper real-valued function \( f : M \to \mathbb{R} := (-\infty, \infty] \) with its domain denoted by \( \mathcal{D}(f) \). Letting \( k \in \mathbb{N} \), we use \( \mathcal{D}^k(f) \) to denote the set of all points \( x \in \mathcal{D}(f) \) at which \( f \) is \( k \)th differentiable, that is,
\[ \mathcal{D}^k(f) := \{ x \in \text{int} \mathcal{D}(f) : f \text{ is } k \text{th differentiable at } x \}. \quad (2.2) \]

As usual, we say that \( f \) is \( C^k \) on \( Q \) if \( Q \subseteq \mathcal{D}^k(f) \) and its \( k \)th derivative is continuously at each point of \( Q \), and that \( f \) is \( C^2 \) around \( \bar{x} \) if it is \( C^2 \) on \( B(x, r) \) for some \( r > 0 \). The gradient (resp. the Hessian) of \( f \) at \( x \in \mathcal{D}^1(f) \) (resp. \( x \in \mathcal{D}^2(f) \)) is denoted by \( \nabla f(x) \) (resp. \( \nabla^2 f(x) \)). Recall that the gradient field \( \nabla f \) is Lipschitz continuous around \( \bar{x} \in \text{int} \mathcal{D}^1(f) \), if there exist positive constants \( \delta, L \) (with \( \delta \leq r_{\text{cvx}}(\bar{x}) \)) such that
\[ \| \nabla f(x) - P_{x,y} \nabla f(y) \| \leq L d(x, y) \quad \text{for any } x, y \in B(\bar{x}, \delta). \]

Thus, if \( f \) is \( C^2 \) around \( \bar{x} \), then \( \nabla f \) is Lipschitz continuous around \( \bar{x} \).

Item (b) in the following definition was known in [17, Definition 6.1 (b)] (for the convexity) and [23, Definition 2.2] (for the quasi-convexity in the case when \( \mathcal{D}(f) = M \)).

**Definition 2.2.** Let \( f : M \to \mathbb{R} \) be proper and let \( Q \subseteq \mathcal{D}(f) \) be weakly convex. Then, \( f \) is said to be
(a) convex (resp. quasi-convex) on $Q$ if, for any $x, y \in Q$ and any geodesic $\gamma \in \Gamma_{xy}$, the composition $f \circ \gamma : [0, 1] \to \mathbb{R}$ is convex (resp. quasi-convex) on $[0, 1]$;

(b) convex (resp. quasi-convex) if $D(f)$ is weakly convex and $f$ is convex (resp. quasi-convex) on $D(f)$.

(c) convex (resp. quasi-convex) around $x \in D(f)$ if $f$ is convex (resp. quasi-convex) on $B(x, r)$ for some $r > 0$.

It is clear that the convexity implies the quasi-convexity. The assertions in the following lemma can be proved directly by definition and are known for some special cases; see, e.g., [28, Theorems 5.1, 6.2] for assertions (i), (iii) and [21, Proposition 3.1] for assertion (ii).

**Lemma 2.1.** Let $f : M \to \mathbb{R}$ be proper. Let $Q \subseteq D(f)$ be weakly convex and let $x \in Q \cap D^1(f)$. Then, the following assertions hold.

(i) If $f$ is convex on $Q$, then it holds for any $y \in Q$ that

$$f(y) \geq f(x) + \langle \nabla f(x), \gamma'_{xy}(0) \rangle \quad \text{for all } \gamma_{xy} \in \Gamma_{xy}.$$

(ii) If $f$ is quasi-convex on $Q$, then it holds for any $y \in Q$ with $f(y) \leq f(x)$ that

$$\langle \nabla f(x), \gamma'_{xy}(0) \rangle \leq 0 \quad \text{for all } \gamma_{xy} \in \Gamma_{xy}.$$

(iii) If $f$ is $C^2$ on $Q$, then $f$ is convex on $Q$ if and only if $\nabla^2 f(x)$ is semi-positive definite for each $x \in Q$.

We show in the following lemma some inequalities, which play important roles in our study. For this purpose, we define the function $h : [0, +\infty) \to \mathbb{R}$ as in [31] by

$$h(t) := \begin{cases} \frac{\tanh t}{t} & \text{if } t \in (0, \infty), \\ 1 & \text{if } t = 0. \end{cases} \quad (2.3)$$

Note that, $h$ is continuous and decreasing monotonically on $[0, +\infty)$.

**Lemma 2.2.** Let $f : M \to \mathbb{R}$ be proper, and $Q \subseteq M$ be weakly convex such that $Q_f := D(f) \cap Q$ is weakly convex with nonempty interior (i.e., $\text{int} Q_f \neq \emptyset$) and its sectional curvatures are bounded from below by $\kappa \leq 0$. Let $t \geq 0$, $z \in \text{int} Q_f$, $x \in \text{int} Q_f \cap D^1(f)$ and $\gamma : [0, +\infty) \to M$ be the geodesic such that

$$\gamma(0) = x, \quad \gamma'(0) = -\nabla f(x) \neq 0 \quad \text{and} \quad \gamma([0, t]) \subseteq \text{int} Q_f. \quad (2.4)$$

Then, the following assertions hold:

(i) If $f$ is convex on $Q_f$, then the following inequality holds:

$$d^2(\gamma(t), z) \leq d^2(x, z) + \frac{2\sinh(\sqrt{\kappa}d(z, x))}{\sqrt{\kappa}\|
abla f(x)\|} \left( \frac{\|
abla f(x)\|^2}{2} - h\left(\sqrt{\kappa}d(x, z)\right)(f(x) - f(z)) \right). \quad (2.5)$$
(ii) If $f$ is quasi-convex on $Q_f$ and $f(z) \leq f(x)$, then the following inequalities hold:

\[
\cosh \left( \sqrt{|\kappa|}d(\gamma(t) ,z) \right) \leq \cosh \left( \sqrt{|\kappa|}d(x, z) \right) \left( 1 + \frac{|\kappa|}{2} t \|\nabla f(x)\| \sinh(t \|\nabla f(x)\|) \right); \tag{2.6}
\]

\[
d^2(\gamma(t), z) < d^2(x, z) + \frac{3t^2 \|\nabla f(x)\|^2}{2h(\sqrt{|\kappa|}d(x, z)}) \quad \text{if } \sqrt{|\kappa|}t \|\nabla f(x)\| \leq 1. \tag{2.7}
\]

Proof. (i) We note that the comparison theorem for a generalized hinge introduced in [25, p. 161, Theorem 4.2] is still true with $Q$ in place of $M$, provided that int$Q$ contains the corresponding generalized hinge. Thus, the argument for proving [32, Lemma 3.2] and [33, Lemma 3.1] remains valid. Hence, assertion (i) holds by (3.6) in [32, Lemma 3.2] (applied to corresponding generalized hinge. Thus, the argument for proving [32, Lemma 3.2] and [33, Lemma 2.3].

(ii) Suppose that $f$ is quasi-convex on $Q_f$ and $f(z) \leq f(x)$, and let $\gamma_{xz} \in \Gamma_{xz}^{Q_f}$ be a minimal geodesic joining $\{x, z\}$. Without loss of generality, we assume that $k = -1$. Then, we have from [33, (9)] (applied to $x, \gamma(t)$, $t\|\nabla f(x)\|$ in place of $x^k, x^{k+1}, t_k$) that

\[
\cosh d(\gamma(t), z) \leq \cosh d(x, z) + \cosh d(z, x) \sinh(t\|\nabla f(x)\|) \left( \frac{t\|\nabla f(x)\|}{2} - \tanh d(x, z) \cos \alpha \right),
\]

(2.8)

where $\alpha := \angle_x(\gamma, \gamma_{xz})$ is the angle between $\gamma$ and $\gamma_{xz}$ at $x$. Below, we verify that $\cos \alpha \geq 0$. Granting this, (2.6) follows immediately from (2.8). To do this, we note by Lemma 2.1(ii) (applied to $Q_f, z$ in place of $Q, y$) that $\langle \nabla f(x), \gamma_{xz}(0) \rangle \leq 0$, and so $\langle \gamma'(0), \gamma'_{xz}(0) \rangle = -\langle \nabla f(x), \gamma'_{xz}(0) \rangle \geq 0$, thanks to (2.4). Thus, by definition, $\cos \alpha = \frac{\langle \gamma'(0), \gamma'_{xz}(0) \rangle}{\|\gamma'(0)\|\|d(x, z)\|} \geq 0$ as desired to show.

To show (2.7), assume $t\|\nabla f(x)\| \leq 1$ and note that $\sinh s < \frac{3}{2}s$ holds for any $s \in (0, 1]$ (which could be easily checked by elementary calculus). Then, (2.6) implies that

\[
\cosh (d(\gamma(t), z)) \leq \cosh (d(x, z)) \left( 1 + \frac{3}{4} t^2 \|\nabla f(x)\|^2 \right).
\]

Therefore, in view of the definition of $h$ in (2.3), (2.7) is seen to hold from the following estimate (see [32, Lemma 3.1]):

\[
\cosh s_1 - \cosh s_2 \geq \frac{(s_2^2 - s_1^2) \sinh s_2}{2s_2} \quad \text{for any } s_1, s_2 \in (0, +\infty).
\]

The proof is complete. \qed

We shall use the following known lemmas in what follows; see, e.g., [31, lemma 2.3] for Lemma 2.3 and [13] for Lemma 2.4.

**Lemma 2.3.** Let $\{a_k\}, \{b_k\} \subset (0, +\infty)$ be sequences such that $\sum_{k=0}^{\infty} b_k < \infty$ and $a_{k+1} \leq a_k(1 + b_k)$ for each $k \in \mathbb{N}$. Then, $\{a_k\}$ is convergent and so it is bounded.
Lemma 2.4. Let \( \{y_k\} \subset M \) be a sequence quasi-Fejér convergent to \( S \), namely there exists a sequence \( \{\varepsilon_k\} \subset (0, +\infty) \) satisfying \( \sum_{k=1}^{\infty} \varepsilon_k < \infty \) such that \( d^2(y_{k+1}, z) \leq d^2(y_k, z) + \varepsilon_k \) for any \( k \in \mathbb{N} \) and \( z \in S \). Then, \( \{y_k\} \) is bounded. Furthermore, if \( \{y_k\} \) has a cluster point \( \bar{y} \) which belongs to \( S \), then \( \lim_{k \to \infty} y_k = \bar{y} \).

3 Gradient algorithm

As in Section 1, \( f : M \to \mathbb{R} \) is a proper locally Lipschitz continuous function. Associated to the optimization problem (1.1), let \( C_f \) denote the set of all critical points of \( f \):

\[
C_f := \{ x \in D^1(f) : \nabla f(x) = 0 \},
\]

where \( D^1(f) \) is the set defined by (2.2). We always assume for the remainder that

\[
\bar{f} := \inf_{x \in M} f(x) > -\infty \quad \text{and} \quad \text{int} D(f) \neq \emptyset. \tag{3.1}
\]

We begin with the following gradient algorithm for solving problem (1.1).

Algorithm 3.1. Give \( x_0 \in D(f) \), \( \beta \in (0, 1) \), \( R \in [1, +\infty) \) and set \( k := 0 \).

Step 1. If \( x_k \in C_f \) or \( x_k \not\in D^1(f) \), then stop; otherwise construct the geodesic \( \gamma_k \) such that

\[
\gamma_k(0) = x_k \quad \text{and} \quad \gamma_k'(0) = -\nabla f(x_k). \tag{3.2}
\]

Step 2. Select the step size \( t_k \in (0, R] \) satisfies the following inequality:

\[
f(\gamma_k(t_k)) \leq f(x_k) - \beta t_k \|\nabla f(x_k)\|^2. \tag{3.3}
\]

Step 3. Set \( x_{k+1} := \gamma_k(t_k) \), replace \( k \) by \( k + 1 \) and go to step 1.

Remark 3.1. Let \( \{x_k\} \) be a sequence generated by Algorithm 3.1 with initial point \( x_0 \in D(f) \).

Then, by Algorithm 3.1, the following inequalities hold for any \( k \in \mathbb{N} \):

\[
d(\gamma_k(t), x_k) \leq t \|\nabla f(x_k)\| \quad \text{for any} \ t \in [0, t_k]; \tag{3.4}
\]

\[
\sum_{j=0}^{k} t_j \|\nabla f(x_j)\|^2 \leq \frac{f(x_0) - f(x_{k+1})}{\beta} \leq \frac{f(x_0) - \bar{f}}{\beta} < +\infty \tag{3.5}
\]

by the blanket assumption (3.1). In particular, one has that \( t_k \|\nabla f(x_k)\| \to 0 \).

Recall that Algorithm 3.1 is said to employ the Armijo step sizes if each step size \( t_k \) in Step 2 is chosen by

\[
t_k := \max\{2^{-i} : i \in \mathbb{N}, f(\gamma_k(2^{-i})) \leq f(x_k) - \beta 2^{-i} \|\nabla f(x_k)\|^2\}; \tag{3.6}
\]
see, e.g., [14, 26]. Note that (3.6) particularly implies (3.3).

The following remark regards the well definedness and the partial convergence property of Algorithm 3.1.

**Remark 3.2.** (a) Suppose that \{x_j : 0 \leq j \leq k\} \subset \mathcal{D}(f) is generated by Algorithm 3.1 such that \(x_k \in \mathcal{D}^1(f)\) is not a critical point of \(f\). Then, using the argument as one did for proving [36, Proposition 3.1], we can check that (3.6) is well defined. Therefore, if each generated iterate \(x_k \in \mathcal{D}^1(f)\) (e.g., \(\mathcal{D}^1(f) = \mathcal{D}(f)\)), then Algorithm 3.1 employing the Armijo step sizes is well defined.

(b) Let \(\{x_k\}\) be a sequence generated by Algorithm 3.1 employing step sizes \(\{t_k\}\) with a positive lower bound or employing the Armijo step sizes. Then, any cluster point \(\bar{x}\) of \(\{x_k\}\) such that \(\nabla f\) is continuous at \(\bar{x}\) is a critical point of \(f\), that is, \(\bar{x} \in C_f\). Indeed, it is immediate from Remark 3.1 for the case when Algorithm 3.1 employs the step size \(\{t_k\}\) with a positive lower bound; while for the case when Algorithm 3.1 employs the Armijo step sizes it can be checked by the argument as one did for proving [36, Corollary 3.1].

### 3.1 Local convergence and Linear convergence

We shall consider the local convergence and the linear convergence of Algorithm 3.1 in this subsection. For this purpose, consider the following assumption:

\[
\bar{x} \in C_f \cap \text{int} \mathcal{D}^1(f), \text{ and } \nabla f \text{ is continuous at } \bar{x}. \tag{3.7}
\]

For the following key lemma, recall that \(R\) is the constant given at the beginning of Algorithm 3.1.

**Lemma 3.1.** Suppose that assumption (3.7) holds. Then, for any \(\delta > 0\), there exist \(\bar{\delta} > 0\) and \(\bar{c} \geq 3\) satisfying \(\bar{c} \bar{\delta} < \delta\) such that, for any \(x_0 \in \mathbb{B}(\bar{x}, \bar{\delta})\) and \(k \in \mathbb{N}\), if \(\{x_j : 0 \leq j \leq k+1\}\) is generated by Algorithm 3.1 to satisfy \(\{x_j : 0 \leq j \leq k\} \subset \mathbb{B}(\bar{x}, \bar{c} \bar{\delta})\), then one has the following assertions for each \(z \in \mathbb{B}(\bar{x}, \bar{c} \bar{\delta})\) satisfying \(f(z) \leq f(x_{k+1})\):

(i) If \(f\) is convex around \(\bar{x}\), then

\[
d^2(x_{k+1}, z) \leq d^2(x_k, z) + \frac{8 \sinh(\sqrt{|k|} t_k \|\nabla f(x_k)\|)}{3 \sqrt{|k|} \|\nabla f(x_k)\|} \left( \frac{1}{2\beta} - h(2\sqrt{|k|} \bar{c} \bar{\delta}) \right) (f(x_k) - f(z)). \tag{3.8}
\]

(ii) If \(f\) is quasi-convex around \(\bar{x}\), then

\[
d^2(x_{k+1}, z) \leq d^2(x_k, z) + 2R t_k \|\nabla f(x_k)\|^2 \leq d^2(x_0, z) + \delta d(x_0, z); \tag{3.9}
\]

in particular, \(x_{k+1} \in \mathbb{B}(\bar{x}, \bar{c} \bar{\delta})\) if \(f(\bar{x}) \leq f(x_{k+1})\).

**Proof.** Noting that any closed ball is compact, we have by [8, P.166] that the curvatures of the ball \(\mathbb{B}(\bar{x}, r_{cvx}(\bar{x}))\) are bounded, where \(r_{cvx}(\bar{x})\) is the convexity radius at \(\bar{x}\) defined in (2.1).
Let $\kappa \leq 0$ be a lower bound of the curvatures of $\mathcal{B}(\bar{x}, r_{cvx}(\bar{x}))$. For simplicity, we may assume, without loss of generality, that $\kappa = -1$. Thanks to assumption (3.7), there exits $\delta > 0$ such that

$$\mathcal{B}(\bar{x}, \delta) \subset \mathcal{D}^1(f), \quad \delta < \min \left\{ r_{cvx}(\bar{x}), \frac{1}{\sqrt{|\kappa|}} \right\} \quad \text{and} \quad R\sqrt{|\kappa| \norm{\nabla f(\cdot)}} \leq 1 \text{ on } \mathcal{B}(\bar{x}, \delta). \quad (3.10)$$

We further choose $0 < \delta_1 < \delta/2$ such that

$$\|\nabla f(x)\| \leq \frac{\beta \delta}{2R} \quad \text{for any } x \in \mathcal{B}(\bar{x}, \delta_1), \quad (3.11)$$

and define

$$\bar{\delta} := \frac{\delta_1^3}{\delta^2} \quad \text{and} \quad \bar{c} := \sqrt{1 + \left( \frac{\delta}{\delta_1} \right)^3}.$$ 

Then, $\bar{\delta} \leq \delta_1 \leq \frac{\bar{\delta}}{2}$ and $\bar{c}\bar{\delta} = \delta_1 \sqrt{\left( \frac{\bar{\delta}}{\delta_1} \right)^4 + \frac{\delta}{\delta_1}}$; thus one has that $\bar{c} \geq 3$ and $\bar{c}\bar{\delta} < \delta_1 < \frac{\bar{\delta}}{2}$. To proceed, let $x_0 \in \mathcal{B}(\bar{x}, \bar{\delta})$ and let $k \in \mathbb{N}$ be such that $\{x_j : 0 \leq j \leq k + 1\}$ is generated by Algorithm 3.1 and $\{x_j : 0 \leq j \leq k\} \subset \mathcal{B}(\bar{x}, \bar{c}\bar{\delta})$. Fix $j \in \{0, 1, \ldots, k\}$, and let $\gamma_j$ be the geodesic determined by (3.2). Then, by (3.11), $\|\nabla f(x_j)\| \leq \frac{\beta \delta}{2R} \quad \text{(as } \bar{c}\bar{\delta} < \delta_1 \text{), and it follows from (3.4) that, for any } t \in [0, t_j],$

$$d(\gamma_j(t), \bar{x}) \leq d(\gamma_j(t), x_j) + d(x_j, \bar{x}) < t \|\nabla f(x_j)\| + \bar{c}\bar{\delta} \leq \frac{\beta \delta}{2} + \frac{1}{2} \delta < \delta,$$

Hence, one has that

$$\gamma_j([0, t_j]) \subseteq \text{int}\mathcal{B}(\bar{x}, \delta) \subseteq \mathcal{B}(\bar{x}, r_{cvx}(\bar{x})) \cap \mathcal{D}^1(f).$$

Thus, noting that $\mathcal{B}(\bar{x}, \delta)$ is strongly convex by the second one of (3.10), Lemma 2.2 is applicable (to $\mathcal{B}(\bar{x}, \delta)$, $t_j$, $x_j$, $\gamma_j$, in place of $Q$, $t$, $x$, $\gamma$). Now let $z \in \mathcal{B}(\bar{x}, \bar{c}\bar{\delta})$ be such that $f(z) \leq f(x_{k+1})$. Then, noting that $\{f(x_j)\}$ is decreasing monotonically and using (3.3), we check that

$$f(z) \leq f(x_{k+1}) \leq f(x_{j+1}) \quad \text{and} \quad t_j\|\nabla f(x_j)\|^2 \leq \frac{f(x_j) - f(x_{j+1})}{\beta} \leq \frac{f(x_j) - f(z)}{\beta}. \quad (3.12)$$

Moreover, we have that

$$d(x_j, z) \leq d(x_j, \bar{x}) + d(z, \bar{x}) \leq 2\bar{c}\bar{\delta} \quad \text{and} \quad \sqrt{|\kappa|d(x_j, z)} < \frac{1}{2}.$$
because \( x_j, z \in \mathbb{B}(\bar{x}, \delta) \) and \( \sqrt{|\kappa|d(x_j, z) \leq \sqrt{|\kappa|\delta} \leq \frac{1}{2} \) (as \( \delta < \frac{\xi}{2} < 1 \) by (3.10)). Therefore,

\[
h \left( \sqrt{|\kappa|d(x_j, z)} \right) \geq h(2\sqrt{|\kappa|\delta}) \geq h(1) > \frac{3}{4}.
\]  

(3.13)

This, together with (3.12), implies that

\[
t_j \|\nabla f(x_j)\|^2 - h \left( \sqrt{|\kappa|d(x_j, z)} \right) (f(x_j) - f(z)) \leq \left( \frac{1}{2\beta} - h(2\sqrt{|\kappa|\delta}) \right) (f(x_j) - f(z)).
\]  

(3.14)

Thus, if \( f \) is convex around \( \bar{x} \), then we may assume, without loss of generality, that \( f \) is convex on \( \mathbb{B}(\bar{x}, \delta) \) (using a smaller \( \delta \) if necessary); hence, by using (3.13) and (3.14) (with \( j = k \)), (3.8) follows from inequality (2.5) (noting \( x_{k+1} = \gamma_k(t_k) \)), showing assertion (i).

To show assertion (ii), assume that \( f \) is quasi-convex around \( \bar{x} \). Then, \( f \) is quasi-convex on \( \mathbb{B}(\bar{x}, \delta) \) as explained earlier (using a smaller \( \delta \) if necessary). Since \( f(z) \leq f(x_j) \) and \( \sqrt{|\kappa|t_j\|\nabla f(x_j)\|} \leq \sqrt{|\kappa|R\|\nabla f(x_j)\|} \leq 1 \) by the third item of (3.10), it follows from (2.7) that

\[
d^2(\gamma_j(t_j), z) \leq d^2(x_j, z) + \frac{3t_j^2\|\nabla f(x_j)\|^2}{2h(\sqrt{|\kappa|d(x_j, z)})} \leq d^2(x_j, z) + 2Rt_j\|\nabla f(x_j)\|^2,
\]  

(3.15)

where the last inequality holds by (3.13) (recalling \( t_j \leq R \)). Summing up the inequalities in (3.15) over \( 0 \leq j \leq k \), we have that

\[
d^2(x_k, z) + t_k\|\nabla f(x_k)\|^2 \leq d^2(x_0, z) + 2R \sum_{j=0}^{k} t_j\|\nabla f(x_j)\|^2
\]

(noting that \( x_{j+1} = \gamma_j(t_j) \)). Since \( \sum_{j=0}^{k} t_j\|\nabla f(x_j)\|^2 \leq \frac{f(x_0) - f(x_{k+1})}{\beta} \leq \frac{f(x_0) - f(z)}{\beta} \) by (3.5) (recalling \( f(z) \leq f(x_{k+1}) \)), it follows that

\[
d^2(x_k, z) + t_k\|\nabla f(x_k)\|^2 \leq d^2(x_0, z) + \frac{2R}{\beta} (f(x_0) - f(z)).
\]  

(3.16)

Recalling \( x_0, z \in \mathbb{B}(\bar{x}, \delta) \subset \mathbb{B}(\bar{x}, \delta_1) \subset \mathbb{B}(\bar{x}, r_{cvx}(\bar{x})) \cap D^1(f) \), the unique minimal geodesic \( \gamma \) joining \( x_0 \) to \( \bar{x} \) is in \( \mathbb{B}(\bar{x}, \delta) \), and then we can apply the mean value theorem to choose \( \xi \in (0,1) \) such that

\[
f(x_0) - f(z) \leq \|\nabla f(\gamma(\xi))\|d(x_0, z) \leq \frac{\beta\delta}{2R}\|d(x_0, z),
\]

where the last inequality is from (3.11). This, together with (3.15) (with \( j = k \)) and (3.16), implies (3.9). Particular, if \( f(\bar{x}) < f(x_{k+1}) \), then we estimate by (3.9) (applied to \( \bar{x} \) in place
of \( z \), and noting \( d(x_0, \bar{x}) \leq \delta \) that

\[
d^2(x_{k+1}, \bar{x}) < \delta^2 + \delta \ddot{\delta} = \ddot{\delta}^2(1 + \frac{\ddot{\delta}}{\delta}) = (\dddot{c}\ddot{\delta})^2,
\]

showing that \( x_{k+1} \in \mathbb{B}(\bar{x}, \dddot{c}\ddot{\delta}) \), and so assertion (ii) is proved. The proof is complete.

**Remark 3.3.** In addition to assumption \( (3.7) \) made in Lemma \( (3.7) \) assume further that \( \dddot{x} \in M \) is a local minimizer of \( f \) and that \( f \) is quasi-convex around \( \dddot{x} \). Then, for any \( \delta > 0 \), there exist \( \dddot{\delta} > 0 \) and \( \dddot{c} \geq 3 \) satisfying \( \dddot{c}\dddot{\delta} < \delta \) such that, for any \( x_0 \in \mathbb{B}(\dddot{x}, \dddot{\delta}) \) and \( k \in \mathbb{N} \), if \( \{x_j : 0 \leq j \leq k + 1\} \) is generated by Algorithm \( (3.1) \) to satisfy \( \{x_j : 0 \leq j \leq k\} \subset \mathbb{B}(\dddot{x}, \dddot{c}\ddot{\delta}) \), then there holds that

\[
f(\dddot{x}) \leq f(x_{k+1}) \quad \text{and} \quad x_{k+1} \in \mathbb{B}(\dddot{x}, \dddot{c}\ddot{\delta}) \subseteq D^1(f). \tag{3.17}
\]

Indeed, one can choose at the beginning of the proof of Lemma \( (3.7) \) \( \delta > 0 \) and \( 0 < \delta_1 < \delta/2 \) small enough so that \( (3.10) \), \( (3.11) \) and the following condition hold:

\[
f(\dddot{x}) \leq f(x) \quad \text{for any} \quad x \in \mathbb{B}(\dddot{x}, \delta). \tag{3.18}
\]

Thus, if \( x_0 \in \mathbb{B}(\dddot{x}, \dddot{\delta}) \) and \( k \in \mathbb{N} \), and if \( \{x_j : 0 \leq j \leq k + 1\} \) is generated by Algorithm \( (3.1) \) so that \( \{x_j : 0 \leq j \leq k\} \subset \mathbb{B}(\dddot{x}, \dddot{c}\ddot{\delta}) \), then one has by \( (3.11) \) that \( t_k \|\nabla f(x_k)\| \leq R\frac{\beta\delta}{2\mathcal{R}} \leq \frac{\delta}{2} \) (as \( \beta < 1 \)) and so by \( (3.4) \) that

\[
d(\dddot{x}, x_{k+1}) \leq d(\dddot{x}, x_k) + d(x_k, x_{k+1}) \leq \dddot{c}\dddot{\delta} + t_k \|\nabla f(x_k)\| < \frac{\delta}{2} + \frac{\delta}{2} = \delta
\]

because \( \dddot{c}\dddot{\delta} < \delta_1 < \frac{\delta}{2} \) as noted in the line after \( (3.1) \), which, together with \( (3.18) \), implies that \( f(\dddot{x}) \leq f(x_{k+1}) \) and so \( x_{k+1} \in \mathbb{B}(\dddot{x}, \dddot{c}\ddot{\delta}) \) by Lemma \( (3.7)(ii) \). In particular, one can conclude by \( (3.17) \) and Remark \( (3.2)(a) \) that Algorithm \( (3.1) \) employing the Armijo step sizes with initial point \( x_0 \in \mathbb{B}(\dddot{x}, \dddot{\delta}) \) is well defined, and the generated sequence \( \{x_k\} \) satisfies

\[
\lim_{k \to +\infty} f(x_k) \geq f(\dddot{x}). \tag{3.19}
\]

For the remainder of this section, we always assume, without loss of generality, that Algorithm \( (3.1) \) does not terminate in finite steps. This particularly implies that, for each \( k \in \mathbb{N} \), \( f \) is differentiable at \( x_k \) and \( t_k \) exists to satisfy \( (3.3) \). Now, we are ready to show the first main result of this subsection.

**Theorem 3.1.** Let \( \dddot{x} \in M \) be such that assumption \( (3.7) \) holds and let \( f \) be quasi-convex around \( \dddot{x} \). Then, for any \( \delta > 0 \), there exist \( \dddot{\delta} > 0 \) and \( \dddot{c} \geq 3 \) satisfying \( \dddot{c}\dddot{\delta} < \delta \) such that, for any sequence \( \{x_k\} \) generated by Algorithm \( (3.1) \) with initial point \( x_0 \in \mathbb{B}(\dddot{x}, \dddot{\delta}) \), if it satisfies \( (3.19) \) (e.g., \( \dddot{x} \) is a local minimizer of \( f \)), then one has the following assertions:

(i) The sequence \( \{x_k\} \) stays in \( \mathbb{B}(\dddot{x}, \dddot{c}\ddot{\delta}) \) and converges to a point \( x^* \) in \( D(f) \).
(ii) If it is additionally assumed that \( \{t_k\} \) has a positive lower bound or that \( \{t_k\} \) satisfies the Armijo step sizes, then \( x^* \) is a critical point of \( f \).

**Proof.** By assumption, Lemma 3.1(ii) is applicable. For any \( \delta > 0 \), let \( \bar{\delta}, \bar{c} > 0 \) be given as in Lemma 3.1(ii) and let \( \{x_k\} \) be a sequence generated by Algorithm 3.1 with initial point \( x_0 \in B(\bar{x}, \bar{\delta}) \) which satisfies (3.19). Noting that \( f \) is quasi-convex around \( \bar{x} \), one inductively sees that \( \{x_k\} \subset B(\bar{x}, \bar{c}\bar{\delta}) \). Thus, the first conclusion of assertion (i) is shown, and the sequence \( \{x_k\} \) has at least a cluster point \( x^* \). Letting \( L_{\bar{\delta}} := \{x \in B(\bar{c}, \bar{c}\bar{\delta}) : f(x) \leq \inf_{k \in \mathbb{N}} f(x_k)\} \), one sees \( x^* \in L_{\bar{\delta}} \) since \( \{f(x_k)\} \) is decreasing and \( f \) is continuous on \( B(\bar{x}, \bar{c}\bar{\delta}) \) (choose a smaller one if necessary). Then, (3.9) holds for each \( z \in L_{\bar{\delta}} \). Thanks to \( 1 \sum_{k=1}^{\infty} t_k \|\nabla f(x_k)\|^2 < +\infty \) by (3.5), we get that \( \{x_k\} \) is quasi-Fejér convergent to \( L_{\bar{\delta}} \). Hence, recalling \( x^* \in L_{\bar{\delta}} \), we conclude by Lemma 2.4 that \( \lim_{k \to \infty} x_k = x^* \). Thus, the second conclusion of assertion (i) is seen to hold.

Assertion (ii) is a direct consequence of assertion (i) and Remark 3.2(b) (note that one can choose \( \bar{\delta}, \bar{c} > 0 \) such that \( \nabla f \) is continuous on \( B(\bar{x}, \bar{c}\bar{\delta}) \) if necessary). This completes the proof. \( \Box \)

To study the linear convergence property, we introduce in the following definition the notion of the local weak sharp minimizer of order \( q \) \( (q \geq 1) \) for problem (1.1), which is a direct extension of the corresponding one in the linear spaces to Riemannian manifolds; see, e.g., [9, 27, 34]. In particular, the notion of the local weak sharp minimizer of order 1 coincides with the local weak sharp minimizer introduced in [17] by Li et al., where some complete characterizations of which were developed on Riemannian manifolds.

**Definition 3.1.** A point \( \bar{x} \in \mathcal{D}(f) \) is said to be a local weak sharp minimizer of order \( q \geq 1 \) for problem (1.1) if there exist \( \delta, \alpha > 0 \) such that

\[
\alpha d_{\bar{S}}^q(x) \leq f(x) - f(\bar{x}) \quad \text{for any } x \in B(\bar{x}, \delta),
\]

where \( \bar{S} := \{x \in M : f(x) = f(\bar{x})\} \).

Our second main result in this subsection is on the linear convergence property of Algorithm 3.1.

**Theorem 3.2.** Suppose that \( \beta \in (\frac{1}{2}, 1) \) and that \( \inf_{k \geq 0} \{t_k\} > 0 \). Let \( \bar{x} \in M \) be such that assumption (3.7) holds,

\[
f \text{ is convex around } \bar{x}, \quad \text{and } \bar{x} \text{ is a local weak sharp minimizer of order 2 for (1.1)} \quad (3.20)
\]

Then, there exists \( \bar{\delta} > 0 \) such that any sequence \( \{x_k\} \) generated by Algorithm 3.1 with initial point \( x_0 \in B(\bar{x}, \bar{\delta}) \) converges linearly to a local minimizer \( x^* \) of \( f \), namely there exist \( \mu > 0 \) and \( \rho \in (0, 1) \) such that

\[
d(x_k, x^*) \leq \mu \rho^k \quad \text{for each } k \in \mathbb{N}. \quad (3.21)
\]
Proof. By assumption, Theorem 3.1 is applicable to getting that for any \( \delta > 0 \) there exist \( \tilde{\delta} > 0, \tilde{c} > 3 \) (satisfying \( \tilde{c} \tilde{\delta} \leq \delta \)) such that the sequence \( \{x_k\} \) generated by Algorithm 3.1 with initial point \( x_0 \in \mathbb{B}(\bar{x}, \tilde{\delta}) \) satisfies

\[
\{x_k\} \subset \mathbb{B}(\bar{x}, \tilde{c}\tilde{\delta}) \tag{3.22}
\]

and converges to a critical point \( x^* \in \mathbb{B}(\bar{x}, \tilde{c}\tilde{\delta}) \). Without loss of generality, we assume further that \( 2\tilde{c}\tilde{\delta} < r_{\text{cvx}}(\bar{x}) \), and assume by assumption that

\[
\mathbb{B}(\bar{x}, 2\tilde{c}\tilde{\delta}) \subset \mathcal{D}^1(f) \text{ and } f \text{ is convex on } \mathbb{B}(\bar{x}, 2\tilde{c}\tilde{\delta}) \tag{3.23}
\]

(and so \( x^* \) is a local minimizer of \( f \)) and there exists \( \alpha > 0 \) such that

\[
\alpha d_S^2(x) \leq f(x) - f(\bar{x}) \quad \text{for any } x \in \mathbb{B}(\bar{x}, \tilde{c}\tilde{\delta}), \tag{3.24}
\]

where \( S := \{x \in M : f(x) = f(\bar{x})\} \). Below, we show that

\[
\|\nabla f(x)\|^2 \geq \alpha(f(x) - f(\bar{x})) \quad \text{for any } x \in \mathbb{B}(\bar{x}, \tilde{c}\tilde{\delta}). \tag{3.25}
\]

To proceed, let \( x \in \mathbb{B}(\bar{x}, \tilde{c}\tilde{\delta}) \) and \( z \in P_S(x) \) (then \( z \in \mathbb{B}(\bar{x}, 2\tilde{c}\tilde{\delta}) \) as \( d(z, \bar{x}) \leq d(z, x) + d(x, \bar{x}) \leq 2\tilde{c}\tilde{\delta} \)). Recalling (3.23), we see from Lemma 2.1(ii) that

\[
f(x) - f(\bar{x}) = f(x) - f(z) \leq \langle \nabla f(x), -\exp_x^{-1}z \rangle \leq \|\nabla f(x)\|d(x, z) = \|\nabla f(x)\|d_S(x).
\]

Thus, (3.25) holds thanks to (3.24).

Now, letting \( t := \inf_{k \in \mathbb{N}} t_k > 0 \) (\( t > 0 \) is because of assumption (b)), we get that

\[
f(x_{k+1}) - f(\bar{x}) \leq f(x_k) - f(\bar{x}) - \beta t_k\|\nabla f(x)\|^2 \\
\leq (1 - \alpha\beta t_k)(f(x_k) - f(\bar{x})) \\
\leq (1 - \alpha\beta t)(f(x_k) - f(\bar{x})),
\]

where the first inequality is from (3.3) and the second inequality is because of (3.22) and (3.25). Then, there holds

\[
f(x_k) - f(\bar{x}) \leq (1 - \alpha\beta t)^k (f(x_0) - f(\bar{x})) \quad \text{for each } k \in \mathbb{N}.
\]

This, together with (3.22) and (3.24), implies that

\[
d_S(x_k) \leq \sqrt{\alpha^{-1}(f(x_0) - f(\bar{x}))} (1 - \alpha\beta t)^{\frac{1}{2}} \quad \text{for each } k \in \mathbb{N} \tag{3.26}
\]

(note that the above analysis works for all \( \beta \in (0, 1) \)).

On the other hand, by assumption \( \beta \in \left(\frac{1}{2}, 1\right) \) and the fact \( \lim_{t \to 0} h(t) = 1 \), one can choose \( \bar{\delta}, \bar{c} \) in the beginning of the proof such that they additionally satisfy \( \frac{1}{2\bar{c}} - h(2\sqrt{|\kappa|\bar{c}\bar{\delta}}) \leq 0 \).
Then, there holds from Lemma 3.1(i) that
\[ d(x_{k+1}, z) \leq d(x_k, z) \quad \text{for any } z \in L \cap \mathcal{B}(\bar{x}, 2\delta) \text{ and } k \in \mathbb{N}, \]  
(3.27)
where \( L := \{ x \in M : f(x) \leq \inf_{k \in \mathbb{N}} f(x_k) \} \). This in particular implies that \( \{ x_k \} \subset \mathcal{B}(\bar{x}, \delta) \) as \( x \) approaches \( \mathcal{B}(\bar{x}, \delta) \). Taking \( \bar{x}_k \in P_S(x_k) \), we see that \( \bar{x}_k \in S \cap \mathcal{B}(\bar{x}, 2\delta) \) as \( d(\bar{x}_k, x) \leq d(\bar{x}_k, x_k) + d(x_k, \bar{x}) \leq 2\delta \). Letting \( l, k \in \mathbb{N} \) with \( l > k \). Then, we get that
\[ d(x_l, x_k) \leq d(x_l, \bar{x}_k) + d(\bar{x}_k, x_k) \leq 2d_S(x_k) \leq 2\sqrt{\alpha^{-1}(f(x_k) - f(\bar{x}))}(1 - \alpha \beta t)^{\frac{k-K}{2}}, \]
where the second inequality is because of (3.27) and the third one is from (3.26). Letting \( l \) go to infinity, there holds that
\[ d(x_k, x^*) \leq 2\sqrt{\alpha^{-1}(f(x_k) - f(\bar{x}))}(1 - \alpha \beta t)^{\frac{k-K}{2}} \quad \text{for each } k \in \mathbb{N}, \]
which yields (3.21) with \( \mu := 2\sqrt{\alpha^{-1}(f(x_0) - f(\bar{x}))} \) and \( \rho := \sqrt{1 - \alpha \beta t} \). The proof is complete. \( \square \)

**Remark 3.4.** If the local minimizer \( \bar{x} \) in Theorem 3.2 is also isolated (namely, \( f(\cdot) > f(\bar{x}) \) on \( \mathcal{U} \setminus \{ \bar{x} \} \) for some neighbourhood \( \mathcal{U} \) of \( \bar{x} \)), then the parameter \( \beta \) can be relaxed to be in \((0,1)\). In fact, under the same assumptions made in Theorem 3.2, but relaxing \( \beta \in (0,1) \), if \( \bar{x} \) in Theorem 3.2 is an isolated local minimizer, then one can choose \( \hat{\beta} > 0 \) small enough at the beginning of the proof of Theorem 3.2 such that \( d_S(x) := d(x, \bar{x}) \) for any \( x \in \mathcal{B}(\bar{x}, \hat{\beta}) \), and \( \{ x_k \} \) satisfies (3.22) (with \( \hat{\beta} \delta \leq \delta \)) and (3.26) (as we have noted in the proof). Thus, the result is immediate from (3.26) and the fact that \( d_S(x_k) := d(x_k, \bar{x}) \) for each \( k \in \mathbb{N} \) (noting (3.22)).

The following lemma provides a sufficient condition for the step size sequence \( \{ t_k \} \) generated by the Armijo step sizes to have a positive lower bound.

**Lemma 3.2.** Let \( \{ x_k \} \) be a sequence generated by Algorithm 3.1 employing the Armijo step sizes. Suppose that \( \{ x_k \} \) converges to a point \( x^* \in \mathcal{D}^1(f) \) and \( \nabla f \) is Lipschitz continuous around \( x^* \). Then, the step size sequence \( \{ t_k \} \) has a positive lower bound: \( \inf_{k \in \mathbb{N}} t_k > 0 \).

**Proof.** By assumption, there exist \( \delta, L > 0 \) (with \( \delta \leq r_{cvx}(x^*) \)) such that \( \mathcal{B}(x^*, 3\delta) \subset \mathcal{D}^1(f) \) and
\[ \| \nabla f(x) - P_{x,y} \nabla f(y) \| \leq Ld(x, y) \quad \text{for any } x, y \in \mathcal{B}(x^*, 3\delta). \]
(3.28)
Noting that \( \lim_{k \to +\infty} x_k = x^* \), there is \( K \in \mathbb{N} \) such that
\[ x_k \in \mathcal{B}(x^*, \delta) \quad \text{for each } k \geq K. \]
Fix $k \geq K$, and assume that $t_k \leq \frac{1}{2}$. Then, by (3.28), we see that

\[ f(\gamma_k(2t_k)) - f(x_k) \geq -2\beta t_k \| \nabla f(x_k) \|^2. \]  

(3.29)

Moreover, from the mean value theorem (as $\mathbb{B}(x^*, 3\delta) \subset D^1(f)$), we get that

\[
\begin{align*}
 f(\gamma_k(2t_k)) - f(x_k) &= \langle \nabla f(\gamma_k(2\tilde{t}_k)), -2t_k P_{\gamma_k(2\tilde{t}_k), x_k} \nabla f(x_k) \rangle \\
 &= -2t_k \langle P_{\gamma_k, x_k, \gamma_k(2\tilde{t}_k)} \nabla f(\gamma_k(2\tilde{t}_k)) - \nabla f(x_k), \nabla f(x_k) \rangle - 2t_k \langle \nabla f(x_k), \nabla f(x_k) \rangle \\
 &= 2t_k \| P_{\gamma_k, x_k, \gamma_k(2\tilde{t}_k)} \nabla f(\gamma_k(2\tilde{t}_k)) - \nabla f(x_k) \| \cdot \| \nabla f(x_k) \| - 2t_k \| \nabla f(x_k) \|^2,
\end{align*}
\]

(3.30)

where $\tilde{t}_k \in (0, t_k)$ and $\gamma_k$ is the geodesic determined by (3.28). Noting that $\mathbb{B}(x^*, \delta)$ is strongly convex, one sees that $\gamma_k([0, t_k])$ is the unique minimal geodesic joining $x_k$ to $x_{k+1}$; hence

\[ d(x_{k+1}, \gamma_k(2\tilde{t}_k)) = d(\gamma_k(t_k), \gamma_k(2\tilde{t}_k)) = |t_k - 2\tilde{t}_k| \| \nabla f(x_k) \| \leq t_k \| \nabla f(x_k) \| = d(x_k, x_{k+1}). \]

(noting that the equality of (3.24) holds). It follows that

\[ d(x^*, \gamma_k(2\tilde{t}_k)) \leq d(x^*, x_{k+1}) + d(x_{k+1}, \gamma_k(2\tilde{t}_k)) \leq 3\delta. \]

Thus, we get from (3.28) that

\[ \| P_{\gamma_k, x_k, \gamma_k(2\tilde{t}_k)} \nabla f(\gamma_k(2\tilde{t}_k)) - \nabla f(x_k) \| \leq 2L t_k \| \nabla f(x_k) \| \]

noting that $d(x_k, \gamma_k(2\tilde{t}_k)) \leq t_k \| \nabla f(x_k) \|$. This, together with (3.30), implies that

\[ f(\gamma_k(2t_k)) - f(x_k) \leq 2t_k (2L t_k - 1) \| \nabla f(x_k) \|^2. \]

Combining this and (3.29), we conclude that $t_k \geq \frac{1 - \beta}{2L}$ (if $k \geq K$ and $t_k \leq \frac{1}{2}$). Thus,

\[ \inf_{k \in \mathbb{N}} t_k = \min \left\{ \frac{1}{2}, t_0, \ldots, t_{K-1}, \frac{1 - \beta}{2L} \right\} > 0 \]

as desired. \hfill \Box

In spirit of the notion of a nondegenerate critical point (in the sense that $\nabla^2 f(\bar{x})$ is positive definite; see [36] Definition 3.1), we say that a point $\bar{x} \in D^2(f)$ is a quasi-nondegenerate critical point of $f$ if

1. $f$ is convex around $\bar{x}$, and $\nabla f$ is Lipschitz continuous around $\bar{x}$;
2. $\bar{x}$ is a local weak sharp minimizer of order 2 for problem (1.1).

By definition it is clear that a nondegenerate critical point is also a quasi-nondegenerate critical point. We have the following result regarding the linear convergence of Algorithm 3.1 employing the Armijo step sizes around a quasi-nondegenerate critical point of $f$.

**Corollary 3.1.** Let $\bar{x}$ be a quasi-nondegenerate point of $f$ and let $\beta \in \left( \frac{1}{2}, 1 \right)$ or $\bar{x}$ be isolated.
Then, there exists $\delta > 0$ such that any sequence $\{x_k\}$ generated by Algorithm 3.1 employing the Armijo step sizes with initial point $x_0 \in B(\bar{x}, \delta)$ converges linearly to a local minimizer of $f$.

Proof. By assumption, we see that assumptions (3.7) and (3.20) hold. Then, Theorem 3.1(i) is applicable and so there exists $\delta > 0$ such that any sequence $\{x_k\}$ generated by Algorithm 3.1 employing the Armijo step sizes with initial point $x_0 \in B(\bar{x}, \delta)$ converges to a point $\bar{x}^* \in B(\bar{x}, \delta)$. Noting that $\nabla f$ is Lipschitz continuous around $\bar{x}$ (as $\bar{x}$ is a quasi-nondegenerate point of $f$), we get that $\nabla f$ is Lipschitz continuous around $x^*$ (choose a smaller $\delta$ if necessary). Thus, one applies Lemma 3.2 to getting $\inf \{t_k\} > 0$. Hence, Theorem 3.2 is applicable to completing the proof.

3.2 Global convergence

The following theorem regards the global convergence and the linear convergence of Algorithm 3.1. We emphasize that the convergence result as well as the linear convergence rate of Algorithm 3.1 is independent of the curvatures of $M$. In particular, in the case when the algorithm employs the Armijo step sizes, assertion (ii) extends the corresponding results in [36, Theorem 4.1], which was proven under the assumption that $\{x_k\}$ converges to a nondegenerate point $\bar{x}$ (noting that this clearly implies that (3.20) holds and $\bar{x}$ is isolated, and that $\inf_{k \geq 0} \{t_k\} > 0$ by Lemma 3.2).

Theorem 3.3. Suppose that the sequence $\{x_k\}$ generated by Algorithm 3.1 has a cluster point $\bar{x} \in \mathcal{D}(f)$ such that assumption (3.7) holds. Then, the following assertions hold:

(i) If $f$ is quasi-convex around $\bar{x}$, then $\{x_k\}$ converges to $\bar{x}$.

(ii) If $\inf_{k \geq 0} \{t_k\} > 0$ and assumption (3.20) holds, then $\{x_k\}$ converges linearly to $\bar{x}$ provided that either $\beta \in \left(\frac{1}{2}, 1\right)$ or $\bar{x}$ is isolated.

Proof. Suppose that $f$ is quasi-convex around $\bar{x}$. Noting that (3.19) is naturally satisfied as $\{f(x_k)\}$ is non-increasing monotone and $\bar{x}$ is a cluster point, we get from Theorem 3.1(i) that there exists $\delta > 0$ such that any sequence generated by Algorithm 3.1 with initial point in $B(\bar{x}, \delta)$ is convergent. Now $\bar{x}$ is a cluster point, so there exists some $k_0 \in \mathbb{N}$ such that $x_{k_0} \in B(\bar{x}, \delta)$. Thus, $\{x_k\}$ converges to some point, which in fact equals to $\bar{x}$ and assertion (i) holds.

With a similar argument that we did for assertion (i), but using Theorem 3.2 (and Remark 3.4) instead of Theorem 3.1(i), one sees that assertions (ii) holds. The proof is complete.

The following lemma provides some sufficient assumptions ensuring the boundedness of the sequence $\{x_k\}$ generated by Algorithm 3.1 (and so the existence of a cluster point). Let $L_f(c)$ denote the sub-level set of $f$ associated with constant $c \in \mathbb{R}$, that is, $L_f(c) := \{x \in M : f(x) \leq c\}$. In particular, let $L_f^0 := L_f(f(x_0))$ for simplicity.

Lemma 3.3. Let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with initial point $x_0 \in \mathcal{D}^1(f)$. Then, $\{x_k\}$ is bounded provided one of the assumptions (a) and (b) holds:
(a) $L_0^f$ is bounded.
(b) $L_1^f$ is totally convex with its curvatures being bounded from below and $f$ is quasi-convex on $L_1^f$ (e.g., $f$ is quasi-convex on $M$ and $M$ is of lower bounded curvatures).

Proof. Note that $\{x_k\} \subseteq L_0^f$ as $\{f(x_k)\}$ is non-increasing monotone. Then, $\{x_k\}$ is clearly bounded under assumption (a).

Now, suppose that assumption (b) holds. Without loss of generality, we assume that the curvatures of $L_1^f$ are bounded from below by $\kappa = -1$. To proceed, let $z \in L := \{x \in M : f(x) \leq \inf_{k \in \mathbb{N}} f(x_k)\}$. Then, we see that $\{z\} \cup \{x_k\} \subseteq L_1^0$ because $\{f(x_k)\}$ is non-increasing. Note that $f(x_k) \geq f(z)$ for each $k \in \mathbb{N}$. Then, by assumption, Lemma 2.3(ii) is applicable on $Q_f := L_1^f$ (with $x_k$, $\gamma_k$ and $t_k$ in place of $x$, $\gamma$ and $t$) to getting that for each $k \in \mathbb{N}$,

$$\cosh(d(x_{k+1}, z)) \leq \cosh(d(x_k, z)) \left(1 + \frac{1}{2} t_k \frac{\|\nabla f(x_k)\|}{\|\nabla f(x_k)\|} \sinh(t_k \|\nabla f(x_k)\|)\right).$$

(3.31)

Note further that

$$\sum_{k \in \mathbb{N}} t_k \|\nabla f(x_k)\| \sinh(t_k \|\nabla f(x_k)\|) < +\infty$$

as $\sum_{k \in \mathbb{N}} t_k^2 \|\nabla f(x_k)\|^2 < +\infty$ (by (3.33) and $\sup\{t_k\} \leq R$) and $\lim_{t \to 0} \frac{\sinh t}{t} = 1$. In view of (3.31), Lemma 2.3 is applicable (with $\{t_k \|\nabla f(x_k)\| \sinh(t_k \|\nabla f(x_k)\|)\}$ and $\{\cosh(d(x_k, z))\}$ in place of $\{b_k\}$ and $\{a_k\}$), and we get that $\{\cosh(d(x_k, z))\}$ is bounded, and so is $\{x_k\}$ as desired. The proof is complete. \qed

The following corollary is immediate from Theorem 3.3 and Lemma 3.3. Particularly, in view of Remark 3.2, the global convergence result (assertion (i)) under assumption (b) extends the corresponding one in [23, Theorem 3.1] which was established on the Riemannian manifold of nonnegative curvatures for the case when $f$ is $C^1$ and quasi-convex on $M$. As for assertion (ii), as far as we know, it is new in Riemannian manifold settings.

**Corollary 3.2.** Suppose that one of assumptions (a) and (b) in Lemma 3.3 holds. Then, any sequence $\{x_k\}$ generated by Algorithm 3.1 with initial point $x_0 \in D^1(f)$ has at least a cluster point $\bar{x}$; furthermore, if $\bar{x}$ satisfies (3.7), then assertions (i) and (ii) in Theorem 3.3 hold.

## 4 Applications to find the Riemannian $L^p$ centers of mass

Let $p \in [1, +\infty)$ and let $N$ be a positive integer such that $N \geq 2$. Let $\{y_i : 1 \leq i \leq N\} \subseteq M$ (which is always denoted by $\{y_i\}$ for short in what follows) be a data set and $\{w_i\} \subseteq (0, 1)$ be the weights satisfying $\sum_{i=1}^N w_i = 1$. In the present section, we shall apply the gradient algorithm proposed in the previous section to compute the Riemannian $L^p$ centers of mass...
of the data set \( \{y_i\} \), which are defined as solutions of the following optimization problem:

\[
\min_{x \in M} f_p(x),
\]

(4.1)

where the function \( f_p : M \to \mathbb{R} \) is defined by

\[
 f_p(x) := \frac{1}{p} \sum_{i=1}^{N} w_i d^p(x, y_i) \quad \text{for any } x \in M,
\]

(4.2)

(see, e.g., [5, Definition 2.5]). From now on, for convenience, we set

\[
 D := \bigcap_{i=1}^{N} U(y_i, r_{\text{inj}}(y_i)).
\]

Let \( D_0 \subseteq D \) be an open nonempty subset. Now consider the following optimization problem:

\[
\min_{x \in M} (f_p + \delta_{D_0})(x),
\]

(4.3)

where \( \delta_{D_0} \) is the indicator function defined by

\[
 \delta_{D_0}(x) = 0 \quad \text{if } x \in D_0 \quad \text{and} \quad \delta_{D_0}(x) = +\infty \quad \text{otherwise}.
\]

The following remark shows some properties of the function \( f_p \) defined in (4.2). For convenience, we set

\[
 I_x := \{ i \in I : x \neq y_i \}; \quad \text{see, e.g., [2].}
\]

**Remark 4.1.** The function \( f_p + \delta_D \) is \( C^1 \) on \( D \) if \( p \in (1, +\infty) \); furthermore, it is \( C^2 \) on \( D \) if \( p \in [2, +\infty) \) and on \( D \setminus \{y_i\} \) if \( p \in [1, 2) \); see, e.g., [25, p. 108-110]. Moreover, if \( f_p + \delta_D \) is differentiable at \( x \in D \), then

\[
 \nabla(f_p + \delta_D)(x) = \nabla f_p(x) = -\sum_{i \in I_x} w_i d^{p-2}(x, y_i) \exp^{-1} y_i,
\]

(4.4)

where \( I_x := \{ i \in I : x \neq y_i \} \); see, e.g., [2].

Below, we recall some results about the Riemannian centers of mass in the literature. To proceed, we fix a point \( o \in M \) and define the function \( \varrho_p : (0, +\infty) \to \mathbb{R} \) by

\[
 \varrho_p(r) := \begin{cases} 
 \frac{1}{2} \min \{r_{\text{inj}}(B(o, 2r)), \frac{\pi}{2\Delta_{B(o, 2r)}} \}, & \text{if } 1 \leq p < 2; \\
 \frac{1}{2} \min \{r_{\text{inj}}(B(o, 2r)), \frac{\pi}{\sqrt{\Delta_{B(o, 2r)}}} \}, & \text{if } 2 \leq p < +\infty,
\end{cases}
\]

for each \( r \in (0, +\infty) \),

(4.5)

where \( \Delta_{B(o, 2r)} \) is an upper bound of the sectional curvatures of \( B(o, 2r) \) (with the convention that \( \frac{\pi}{\sqrt{\Delta}} = +\infty \) for \( \Delta \leq 0 \)). Then, \( \varrho_p(\cdot) \) is non-increasing monotonically on \( (0, +\infty) \). For the remainder, let \( \rho \in (0, +\infty) \) be such that

\[
 \rho \leq \varrho_p(\rho) \quad \text{and} \quad \{y_i\} \subset U(o, \rho).
\]

(4.6)
In what follows, we need the following fact which can be found in [24] Theorem 29.

**Lemma 4.1.** Let $r > 0$ be such that $r \leq \frac{1}{2} \min \{r_{\text{inj}}(B(o, r)), \frac{\pi}{\sqrt{\Delta B(o, r)}} \}$. Then, $U(o, r)$ is strongly convex.

**Lemma 4.2.** Assume (4.6) and let $z \in \partial B(o, \rho)$. Then, the following assertions hold:

(i) If $y \in U(o, \rho)$ and $\gamma \in \Gamma_{yz}$ is minimal, then $\gamma([0, 1]) \subseteq U(o, \rho)$.

(ii) There exists $s > 0$ such that

$$\exp_z(-s \nabla f_p(z)) \in U(o, \rho) \text{ for any } s \in (0, s].$$

*(Proof.)* (i) The inclusion $B(o, \rho) \subset D$ is clear because for each $i \in I$,

$$d(x, y_i) < 2\rho \leq 2\gamma_p(\rho) \leq r_{\text{inj}}(B(o, 2\rho)) \leq r_{\text{inj}} y_i \text{ for any } x \in B(o, \rho),$$

where the third inequality is true by the definition of $\gamma_p$ (see (4.5)), while the others hold by assumption (4.6). Furthermore, the strong convexity of $U(o, \rho)$ is from Lemma 4.1 and assumption (4.6).

(ii) Let $y \in U(o, \rho)$. To show (ii), we verify below that $y$ is a weak pole of $B(o, \rho)$ in the sense that, for each $x \in B(o, \rho)$, the minimal geodesic of $M$ joining $y$ to $x$ is unique and lies in $B(o, \rho)$. Granting this, the conclusion holds by [19] Proposition 4.3 (noting that weakly convex set is locally convex). To proceed, recalling that $B(o, \rho)$ is weakly convex, one can choose a minimal geodesic $\gamma$ joining $y$ to $x$ such that $\gamma \subset B(o, \rho)$. Let $w$ be the midpoint of $\gamma$. Note that the length $l(\gamma) < 2\rho$. One sees that $y, x \in U(w, \rho)$, and $U(w, \rho) \subset U(o, 2\rho)$. By assumption (4.6), there holds that

$$\rho \leq \frac{1}{2} \min \{r_{\text{inj}}(B(o, 2\rho)), \frac{\pi}{\sqrt{\Delta B(o, 2\rho)}} \} \leq \frac{1}{2} \min \{r_{\text{inj}}(B(o, \rho)), \frac{\pi}{\sqrt{\Delta B(w, \rho)}} \}.$$

Thus, Lemma 4.1 is applicable to concluding that $B(w, \rho)$ is strongly convex and so $\gamma$ is the unique minimal geodesic joining $y$ to $x$ (noting that $x, y \in B(w, \rho)$). This shows that $y$ is a weak pole of $B(o, \rho)$ as desired, and assertion (ii) is established.

(iii) Fix $i \in I$, and write $V_i := \frac{\exp^{-1}_z y_i}{\|\exp^{-1}_z y_i\|}$. The geodesic $[0, 1] \ni t \mapsto \exp_z(t \|\exp^{-1}_z y_i\| V_i)$ is the minimal geodesic joining $z$ and $y_i$. Applying assertion (ii) just established (to $y_i$ in place of $y$), one checks that

$$\exp_z(s \|\exp^{-1}_z y_i\| V_i) \in U(o, \rho) \text{ for any } 0 < s < 1.$$
and then, by mathematical induction, that there exits \( s > 0 \) such that

\[
\exp_x s \sum_{i \in I} \lambda_i V_i \in U(o, \rho) \quad \text{for any } 0 < s \leq \bar{s},
\]

where each \( \lambda_i := w_i d^{p-1}(z, y_i) \). Taking into account that \(-\nabla f_p(z) = \sum_{i \in I} \lambda_i V_i \) by (4.4) (noting that \( I_z = I \), thanks to assumption (4.6), and \( z \in \partial B(o, \rho) \)), we conclude that (4.7) holds. The proof is complete.

For the remainder, in view of Lemma 4.2(i), we choose \( D_0 := U(o, \rho) \) for the problem (4.3) unless otherwise specified. Now we are ready to establish the following key proposition. Recall that \( \{y_i\} \) is colinear if it lies in one geodesic segment. We also need to make use of the following assumption:

\[
\min_{x \in M} f_p(x) < \min_{i \in I} f_p(y_i). \quad (4.8)
\]

**Proposition 4.1.** Assume that (4.6) holds and that \( \{y_i\} \) is not colinear if \( p = 1 \). Then,

\( \{y_i\} \) has the unique Riemannian \( L^p \) center of mass \( \bar{x}_p \), which lies in \( U(o, \rho) \) and is the unique critical point of \( f_p \) in \( B(o, \rho) \). Furthermore, the following assertions hold:

(i) \( \bar{x}_p \) is a nondegenerate critical point of \( f_p \) (and so \( \nabla f_p \) is Lipschitz continuous around \( \bar{x}_p \)) if (4.8) is additionally assumed for \( p \in [1, 2) \).

(ii) \( \bar{x}_p \) is a local weak sharp minimizer of order 2 for problem (4.3) if \( p \in (1, 2) \).

(iii) \( f_p \) is convex around \( \bar{x}_p \).

**Proof.** Note by (4.7) that \( \nabla f_p \) does not vanish on \( \partial B(o, \rho) \) thanks to assumption (4.6). Thus, by assumption, it follows from [4, Theorem 2.1 and Remark 2.5] that \( \{y_i\} \) has the unique Riemannian \( L^p \) center of mass \( \bar{x}_p \in U(o, \rho) \), which is the unique critical point of \( f_p \) in \( U(o, \rho) \).

(i) Note that the conclusion for \( p \in [2, +\infty) \) follows from [4, Theorem 2.1] (applied to \( U(o, 2\rho) \) in place of \( M \)). Thus, we assume that \( p \in [1, 2) \) and that (4.8) holds. Then, \( \bar{x}_p \in B(o, \rho) \setminus \{y_i\} \). We shall complete the proof by showing that \( \nabla^2 f_p(x) \) is positive definite for each \( x \in B(o, \rho) \setminus \{y_i\} \). To do this, let \( x \in B(o, \rho) \setminus \{y_i\} \), and let \( \gamma(\cdot) \) be a unit speed geodesic with \( \gamma(0) = x \). Then, \( \gamma([0, \epsilon]) \subseteq B(o, \rho) \setminus \{y_i\} \) for some \( \epsilon > 0 \). It suffices to verify that

\[
\frac{d^2}{dt^2} (f_p \circ \gamma)(0) = \frac{d^2}{dt^2} f_p(\gamma(t))|_{t=0} > 0 \quad (4.9)
\]

To show this, let \( i \in I \), and let \( \alpha_i \) denote the angle at \( x \) between the geodesic \( \gamma \) and the unique minimal geodesic joining \( x \) to \( y_i \). Note that

\[
d(x, y_i) < 2\rho \leq 2g_p(\rho) = \min\{r_{inj}(B(o, 2\rho)), \frac{\pi}{2\sqrt{\Delta B(o, 2\rho)}} \} \quad \text{for each } i \in I, \quad (4.10)
\]

thanks to \( \rho \leq g_p(\rho) \) (by (4.6)) and (4.5). Then, the function \( d(\gamma(\cdot), y_i) \) is analytic on \((-\epsilon, \epsilon)\),
and by the arguments for proving [21 (2.3)] and [25 p. 153-154], one has that
\[
\frac{d}{dt}d(\gamma(t), y_i)|_{t=0} = \cos \alpha_i \quad \text{and} \quad \frac{d^2}{dt^2}d(\gamma(t), y_i)|_{t=0} \geq c_{\Delta_{\beta(2\rho)}}(d(x, y_i)) \sin^2 \alpha_i, \quad (4.11)
\]
where, for any \(l > 0\), \(c_\delta(l) := \frac{1}{\sqrt{\delta}} \cot(\sqrt{\delta}l)\) if \(\delta > 0\), \(c_\delta(l) := \frac{1}{l}\) if \(\delta = 0\), and \(c_\delta(l) := \frac{1}{\sqrt{|\delta|}} \coth(\sqrt{|\delta|}l)\) otherwise. Since, for any \(t \in (-\epsilon, \epsilon)\),
\[
\frac{d^2}{dt^2}f_p(\gamma(t)) = \sum_{i \in I} w_i \left((p - 1)d^{p-2}(x, y_i) \left(\frac{d}{dt}d(\gamma(t), y_i)\right)^2 + d^{p-1}(x, y_i)\frac{d^2}{dt^2}d(\gamma(t), y_i)\right),
\]
it follows from (4.11) that
\[
\frac{d^2}{dt^2}(f_p \circ \gamma)(0) \geq \sum_{i \in I} w_i \left((p - 1)d^{p-2}(x, y_i) \cos \alpha_i + d^{p-1}(x, y_i)c_{\Delta_{\beta(2\rho)}}(d(x, y_i)) \sin^2 \alpha_i\right). \quad (4.12)
\]
Note by (4.10) that \(0 < d(x, y_i) < \frac{\pi}{2\sqrt{\Delta_{\beta(2\rho)}}}\), and then \(c_{\Delta_{\beta(2\rho)}}(d(x, y_i)) > 0\) by definition. Thus, (4.9) is clear in the case when \(p \in (1, 2)\); while, for the case when \(p = 1\), there exists an index \(i_0 \in I\) such that \(\sin \alpha_{i_0} \neq 0\) (as \(y_i\) is not colinear by assumption), and (4.9) follows from (4.12) as
\[
\frac{d^2}{dt^2}(f_p \circ \gamma)(0) \geq \sum_{i \in I} w_i c_{\Delta_{\beta(2\rho)}}(d(x, y_i)) \sin^2 \alpha_i \geq w_{i_0} c_{\Delta_{\beta(2\rho)}}(d(x, y_{i_0})) \sin^2 \alpha_{i_0} > 0.
\]
Therefore, (4.9) is valid for any \(p \in [1, 2)\), completing the proof of assertion (i).

(ii) Assume \(p \in (1, 2)\). In light of assertion (i), we only need to consider the case when (4.8) is not satisfied. Thus, we may assume that \(\bar{x}_p = y_{i_0}\) for some \(i_0 \in I\) and so \(\nabla f_p(y_{i_0}) = 0\). Consider the date set \(\{y_i : i \in \tilde{I}\}\) and the weights \(\tilde{w}_i : i \in \tilde{I}\), where \(\tilde{I} := I \setminus \{i_0\}\) and \(\tilde{w}_i := \frac{w_i}{1-w_{i_0}}\) for each \(i \in \tilde{I}\). Then, (4.6) remains true for the date set \(\{y_i : i \in \tilde{I}\}\). Let \(\tilde{f}_p\) denote the corresponding function defined by (4.2) (with \(\{y_i : i \in \tilde{I}\}, \{\tilde{w}_i : i \in \tilde{I}\}\) in place of \(\{y_i : i \in I\}, \{w_i : i \in I\}\)). Then,
\[
\tilde{f}_p(\cdot) := \frac{1}{p} \sum_{i \in \tilde{I}} \tilde{w}_i d^p(\cdot, y_i) = \frac{1}{1-w_{i_0}} \left(f_p(\cdot) - \frac{w_{i_0}}{p} d^p(\cdot, y_{i_0})\right). \quad (4.13)
\]
Hence, \(\nabla \tilde{f}_p(\bar{x}_p) = \frac{\nabla f_p(\bar{x}_p)}{1-w_{i_0}} = 0\). This means that \(\bar{x}_p\) is also the unique Riemannian \(L^p\) center of mass of \(\{y_i : i \in \tilde{I}\}\), and so (4.8) holds with \(f_p, \tilde{I}\) in place of \(f_p, I\). Thus, by assertion (i), one sees that \(\bar{x}_p\) is a nondegenerate critical point of \(\tilde{f}_p\), which in particular implies that \(\bar{x}_p\) is a local weak sharp minimizer of order 2 for problem (4.3) with \(\tilde{f}_p\) in place of \(f_p\); there exist
\(\delta, \alpha > 0\) such that
\[
\alpha d^2(x, \bar{x}_p) \leq \tilde{f}_p(x) - \tilde{f}_p(\bar{x}_p) \quad \text{for any } x \in B(\bar{x}, \delta).
\]
Since \(\tilde{f}_p(\bar{x}_p) = \frac{1}{1-w_{i0}} f_p(\bar{x}_p)\) and \(\tilde{f}_p(\cdot) \leq \frac{1}{1-w_{i0}} f_p(\cdot)\) on \(B(\bar{x}, \delta)\) by (4.13), it follows that
\[
\alpha(1-w_{i0})d^2(x, \bar{x}_p) \leq (1-w_{i0})(\tilde{f}_p(x) - \tilde{f}_p(\bar{x}_p)) \leq f_p(x) - f_p(\bar{x}_p) \quad \text{for any } x \in B(\bar{x}, \delta).
\]
Therefore \(\bar{x}_p\) is a local weak sharp minimizer of order 2 for problem (4.3), establishing assertion (ii).
(iii) It follows from assertion (i) for \(p \in [2, +\infty)\) and from [4, Theorem 2.1] for \(p \in [1, 2)\) \((f_p\) is actually convex on \(U(o, \rho)\) in the case when \(p \in [1, 2)\)). The proof is complete.

Recall from Remark 4.1 that \(f_p\) is \(C^1\) on \(D\) for \(p \in (1, +\infty)\) and \(C^1\) on \(D \setminus \{y_i\}\) for \(p = 1\). Then, we have
\[
D^1(f_p + \delta D_0) = \begin{cases} 
D_0 \setminus \{y_i\} & \text{if } p = 1, \\
D_0 & \text{if } p \in (1, +\infty). 
\end{cases}
\] (4.14)
Furthermore, it is clear that
\[
\nabla(f_p + \delta D_0) \text{ is continuous on } D^1(f_p + \delta D_0). \tag{4.15}
\]

**Theorem 4.1.** Assume that (4.6) holds and that \(\{y_i\}\) is not colinear for \(p = 1\). Let \(\{x_k\}\) be a sequence generated by Algorithm 3.1 for solving problem (4.3) with initial point \(x_0 \in U(o, \rho)\), and suppose that the step size sequence \(\{t_k\}\) has a positive lower bound: \(\inf\{t_k\} > 0\) and \(\{x_k\}\) has a cluster point \(\bar{x}_p \in D^1(f_p + \delta D_0)\). Then, \(\{x_k\}\) converges to \(\bar{x}_p\), which is the unique Riemannian \(L^p\) center of mass of \(\{y_i\}\); moreover the convergence rate is at least linear if (4.8) is additionally assumed for \(p = 1\).

**Proof.** By (4.15) and the assumption \(\inf\{t_k\} > 0\), Remark 3.2(b) is applicable and we see \(\nabla f_p(\bar{x}_p) = 0\), and so (3.7) holds (with \(\bar{x}_p\) in place of \(\bar{x}\)). Then, \(\bar{x}_p \in U(o, \rho)\) and is the unique Riemannian \(L^p\) center of mass of \(\{y_i\}\). Therefore, (3.20) is satisfied from Proposition 4.1(iii) if (4.8) is additionally assumed for \(p = 1\). Thus, Corollary 3.2 is applicable (noting \(\inf\{t_k\} > 0\)) to completing the proof.

**Corollary 4.1.** Assume that (4.6) holds and that \(\{y_i\}\) is not colinear for \(p = 1\). Let \(x_0 \in U(o, \rho)\) and suppose for \(p = 1\) that
\[
f_p(x_0) < \min_{i \in I} f_p(y_i). \tag{4.16}
\]
Then, Algorithm 3.1 for solving problem (4.3) employing the Armijo step sizes with initial point \(x_0\) is well defined, and the generated sequence \(\{x_k\}\) converges to the unique Riemann-
nian $L^p$ center of mass of \{y_i\}. Moreover, the convergence rate is at least linear if (4.8) is additionally assumed for $p \in [1, 2)$.

**Proof.** By (4.14), one sees that $D^1(f_p + \delta D_0) = D(f_p + \delta D_0) = D_0$ in the case when $p \in (1, +\infty)$; thus the first conclusion regarding the well definedness of Algorithm 3.1 follows directly from Remark 3.2(a). Below we consider the case when $p = 1$. To do this, in view of (4.16), one applies Remark 3.2(a) inductively to check that each generated point $\{x_k\}$ satisfying $\{x_k\} \subseteq L_0^0 \subset D_0 \setminus \{y_i\} = D^1(f_p + \delta D_0)$ (as $\{f_1(x_k)\}$ is decreasing), and so Algorithm 3.1 employing the Armijo step sizes is well defined, completing the proof for the first conclusion.

To show the second conclusion regarding the convergence rate, we note first that $L_0^0$ is bounded, that is, assumption (a) in Lemma 3.3 holds. Thus Corollary 3.2 is applicable to getting that $\{x_k\}$ has a cluster point, say $\bar{x}_p \in B(o, \rho)$ (noting that $D_0 := \cup(o, \rho)$). As noted before, $\{x_k\} \subseteq L_0^0$; hence $\bar{x}_p \notin \{y_i\}$ when $p = 1$. Then, one sees from (4.14) that $\bar{x}_p \in D^1(f_p + \delta D_0)$ and $\partial D_0$. Below we show that $\bar{x}_p \in D^1(f_p + \delta D_0)$. Granting this and noting that (4.8) additionally holds for $p \in [1, 2)$, we get from Proposition 4.1(i) that $\nabla f_p$ is Lipschitz continuous around $\bar{x}_p$. Therefore the corresponding step size sequence $\{t_k\}$ employing the Armijo step sizes has a positive lower bound thanks to Lemma 3.2 and then Theorem 4.1 is applicable to completing the proof.

To proceed, suppose on the contrary that $\bar{x}_p \in \partial D_0$. By assumption (4.6), it follows from Lemma 4.2(iii) that

$$\nabla f_p(\bar{x}_p) \neq 0$$

and there exists $\bar{s} > 0$ such that

$$\exp_{\bar{x}_p}[-s \nabla f_p(\bar{x}_p)] \in U(o, \rho) \quad \text{for any } 0 < s \leq \bar{s}.$$  \(4.18\)

Now, we show that there exists $\delta_0 > 0$ such that

$$\exp_x[-s \nabla f_p(x)] \in U(o, \rho) \quad \text{for any } x \in B(\bar{x}_p, \delta_0) \text{ and } 0 < s \leq \bar{s}$$  \(4.19\)

(using a smaller $\bar{s}$ if necessary). To this end, set $\bar{z} := \exp_{\bar{x}_p}[-\bar{s} \nabla f_p(\bar{x}_p)]$ and then $\bar{z} \in U(o, \rho)$ by (4.18); hence there is $\bar{\varepsilon} > 0$ such that $B(\bar{z}, \bar{\varepsilon}) \subset U(o, \rho)$. Without loss generality, we may assume

$$\bar{s} \| \nabla f_p(\bar{x}_p) \| + \bar{\varepsilon} + \bar{\delta} \leq r_{cvx}(B(\bar{x}_p, \bar{\delta}))$$  \(4.20\)

for some $\bar{\delta} > 0$. Since the mapping $x \mapsto \exp_x[-s \nabla f_p(x)]$ is continuous on $U(o, \rho)$ (as $\nabla f_p(x)$ is continuous on $U(o, \rho)$), there exists $\delta_0 \in (0, \bar{\delta})$ such that

$$\exp_x[-s \nabla f_p(x)] \in B(\bar{z}, \bar{\varepsilon}) \subset U(o, \rho) \quad \text{for any } x \in B(\bar{x}_p, \delta_0).$$

Let $x \in B(\bar{x}_p, \delta_0)$ and write $z_x := \exp_x[-s \nabla f_p(x)]$. Then, in view of (4.20), we check that

$$d(x, z_x) \leq d(x, \bar{x}_p) + d(\bar{x}_p, \bar{z}) + d(\bar{z}, z_x) \leq r_{cvx}(B(\bar{x}_p, \delta_1)) \leq r_{cvx}(x).$$
Thus, the geodesic $[0, \bar{s}] \ni s \mapsto \exp_x[-s\nabla f_p(x)]$ is the minimal geodesic joining $x$ to $z_x$, and so (4.19) holds as $\mathcal{U}(o, \rho)$ is strongly convex.

To proceed, let $\{x_k\}$ be a subsequence of $\{x\}$ converging to $\bar{x}_p$. Then, $\lim_{j \to +\infty} t_{k_j} = 0$ by Remark 3.2(i). Thus, without loss of generality, we may assume that
\[ x_{k_j} \in \mathcal{B}(\bar{x}_p, \delta_0) \quad \text{and} \quad 2t_{k_j} \leq \bar{s} \quad \text{for each} \quad j. \tag{4.21} \]

Fix $j$ and recall that the geodesic $\gamma_{k_j}$ is defined by (3.2). Then, in view of (4.19) and (4.21), we see that $\gamma_{k_j}(s) = \exp_{x_{k_j}}[-s\nabla f_p(x_{k_j})] \in \mathcal{U}(o, \rho)$ for each $s \in [0, 2t_{k_j}]$.

By using the mean value theorem, there is $\bar{t}_{k_j} \in (0, 2t_{k_j})$ such that
\[ \frac{f_p(\gamma_{k_j}(2t_{k_j})) - f_p(x_{k_j})}{-2t_{k_j}} = \langle P_{\gamma_{k_j}, x_{k_j}, \gamma_{k_j}(\bar{t}_{k_j})} \nabla f_p(\gamma_{k_j}(ar{t}_{k_j})), \nabla f_p(x_{k_j}) \rangle. \]

This, together with (3.6), implies that
\[ \langle P_{\gamma_{k_j}, x_{k_j}, \gamma_{k_j}(\bar{t}_{k_j})} \nabla f_p(\gamma(\bar{t}_{k_j})), \nabla f_p(x_{k_j}) \rangle \leq \beta \|\nabla f_p(x_{k_j})\|^2. \]

Passing to the limit as $j \to \infty$, we arrive at $\beta \geq 1$ by (4.17), which is a contradiction. Thus, the proof is complete. \qed

Below, we shall consider the gradient algorithm for solving problem (4.1) employing constant step sizes, which is stated as follows.

**Algorithm 4.1.** Give $x_0 \in \mathcal{D}(f)$, $t_0 \in (0, +\infty)$ and set $k := 0$.

Step 1. If $\nabla f(x_k) = 0$ or $x_k \notin \mathcal{D}^1(f)$, then stop; otherwise construct $\gamma_k$ as (3.2).

Step 2. Set $x_{k+1} := \gamma_k(t_0)$, replace $k$ by $k + 1$ and go to step 1.

Let $x_0 \in D_0 := \mathcal{U}(o, \rho)$, and we need the following assumption:
\[ L^0_{f_p} \subset D_0, \tag{4.22} \]

where, as done in Section 3, $L^0_{f_p} := L_{f_p}(f_p(x_0))$ is the sub-level set. Moreover, we need also the following assumption made for $p \in [1, 2)$:
\[ f_p(x_0) < \min_{i \in I} f_p(y_i). \tag{4.23} \]

Thus, under assumption (4.22), and assumption (4.23) (only for $p \in [1, 2)$), $f_p$ is $C^2$ on $L^0_{f_p}$ by Remark 4.1 and the supremum of all eigenvalues of $\nabla^2 f_p(\cdot)$ on $L^0_{f_p}$, denoted by $\lambda_p(x_0)$, is bounded (as $L^0_{f_p}$ is compact).
Corollary 4.2. Assume that (4.13) holds and that \( \{y_i\} \) is not colinear if \( p = 1 \). Let \( x_0 \in D_0 \) be such that (4.22) holds, and that (4.23) holds for \( p \in [1, 2) \). Then, Algorithm 4.1 for solving problem (4.1) with \( t_0 \in \left(0, \frac{2}{\lambda_p(x_0)}\right) \) is well defined and converges linearly to the unique \( L^p \) center of mass of \( \{y_i\} \).

Proof. As noted earlier, \( f_p \) is \( C^2 \) on \( L^0_{f_p} \), which particularly implies that \( L^0_{f_p} \subset D^1(f_p + \delta_{D_0}) \). Thus, to show the first assertion, it is sufficient to show that \( x_k \in L^0_{f_p} \) for each \( k \). Clearly, \( x_0 \in L^0_{f_p} \) by the choice of \( x_0 \). To proceed, suppose that \( x_j \in L^0_{f_p} \) for some \( j \in \mathbb{N} \). Let \( \gamma_j : [0, +\infty) \rightarrow M \) be the geodesic defined by (4.2), and set \( \bar{t} := \sup\{t : \gamma_j(s) \in L^0_{f_p} \text{ for any } 0 \leq s \leq t\} \). By the Taylor expansion and using the upper bound on the Hessian of \( f_p \) on \( L^0_{f_p} \), we check that, for each \( t \in (0, \bar{t}] \),

\[
f_p(\gamma_j(t)) = f_p(x_j) - t\|\nabla f_p(x_j)\|^2 + t^2 \int_0^1 (1 - \tau)(\nabla^2 f_p(\gamma_j(\tau t))\nabla f_p(x_j), \nabla f_p(x_j))d\tau \\
\leq f_p(x_j) - t\|\nabla f_p(x_j)\|^2 + \frac{t^2\lambda_p(x_0)}{2}\|\nabla f_p(x_j)\|^2.
\]

Noting \( f_p(x_j) \leq f_p(x_0) \), it follows that

\[
f_p(\gamma_j(t)) \leq f_p(x_0) - t(1 - \frac{t\lambda_p(x_0)}{2})\|\nabla f_p(x_j)\|^2 \quad \text{for each } t \in [0, \bar{t}]. \tag{4.24}
\]

This implies that \( \bar{t} \geq \frac{2}{\lambda_p(x_0)} > t_0 \) by definition of \( \bar{t} \) and continuity of \( f_p \). Therefore, \( x_{j+1} := \gamma_j(t_0) \in L^0_{f_p} \), and then, by mathematical induction, \( x_k \in L^0_{f_p} \) for each \( k \in \mathbb{N} \) as desired to show. Furthermore, (4.24) implies that the generated sequence \( \{x_k\} \) by Algorithm 4.1 satisfies

\[
f(\gamma_k(t_0)) \leq f(x_k) - t_0\beta\|\nabla f(x_k)\|^2 \quad \text{for each } k \in \mathbb{N},
\]

where \( \beta := 1 - \frac{t_0\lambda_p(x_0)}{2} \in (0, 1) \). This means that \( \{x_k\} \) coincides with the sequence generated by Algorithm 3.1 for solving problem (1.3) with initial point \( x_0 \) and constant step sizes \( \{t_k := t_0\} \). Note that \( \{x_k\} \subset L^0_{f_p} \) and \( L^0_{f_p} \subset D^1(f_p + \delta_{D_0}) \). Then, \( \{x_k\} \) has a cluster point in \( D^1(f_p + \delta_{D_0}) \) as \( L^0_{f_p} \) is clearly blounded. Furthermore, (4.23) particularly implies (4.8). Thus, Theorem 4.1 is applicable and \( \{x_k\} \) converges linearly to the unique Riemannian \( L^p \) center of mass of \( \{y_i\} \). The proof is complete. \( \square \)

The following corollary is new in the case when \( p \in [1, 2) \), and was proved in [5, Theorem 4.1] in the case when \( p \in [2, +\infty) \) under the assumption that \( \{y_i\} \subset \mathbb{B}(o, \frac{1}{3}r_{cx}) \) with \( r_{cx} := \frac{1}{2}\min\{r_{\text{inj}}(M), \frac{\pi}{\sqrt{\lambda_{n-1}}}\} \), which particularly implies the following assumption (4.25) with \( r_{cx} \) in place of \( \rho \).

Corollary 4.3. Assume that

\[
\rho \leq g_p(\rho) \quad \text{and} \quad \{y_i\} \subset \bigcup \left(0, \frac{1}{3}\rho\right), \tag{4.25}
\]

and \( \{y_i\} \) is not colinear if \( p = 1 \). Let \( x_0 \in \bigcup \left(0, \frac{1}{3}\rho\right) \) be such that (4.23) holds for \( p \in [1, 2) \).
Then, Algorithm 4.1 for solving problem (4.1) with initial point \( x_0 \) and \( t_0 \in \left( 0, \frac{2}{\lambda_p(x_0)} \right) \) is well defined and converges linearly to the unique Riemannian \( L^p \) center of mass of \( \{y_i\} \).

Proof. Note that (4.6) holds by (4.25). To apply Corollary 4.2, we only need to show (4.22). To do this, let \( z \in M \setminus U(o, \rho) \). Then, we have by (4.25) and the choice of \( x_0 \) that \( d(x_0, y_i) < \frac{2\rho}{\lambda} \) for each \( i \in I \); hence \( f_p(x_0) < f_p(z) \) by definition (see (1.2)). This means that \( z \notin L^0_f \), establishing (4.22) as \( z \in M \setminus U(o, \rho) \) is arbitrary. Thus, Corollary 4.2 is applicable to completing the proof.

In the special case when \( M \) is a Hadamard manifold, one checks by definition (see (4.5)) that \( \varrho_p(r) = +\infty \) for each \( r > 0 \). Then, we can choose that \( \rho := +\infty \) so that (4.6) and (4.25) hold trivially. Thus, Corollary 4.4 follows direct from Corollaries 4.1 and 4.3.

Corollary 4.4. Assume that \( M \) is a Hadamard manifold and \( \{y_i\} \) is not collinear for \( p = 1 \), and let \( x_0 \in M \). Then, the following assertions hold:

(i) If (4.23) holds for \( p = 1 \), then Algorithm 3.1 for solving problem (4.1) employing the Armijo step sizes with initial point \( x_0 \) is well defined and the generated sequence \( \{x_k\} \) converges to the unique Riemannian \( L^p \) center of mass of \( \{y_i\} \); and the convergence rate is at least linear if (4.8) is additionally assumed for \( p \in (1, 2) \).

(ii) If (4.25) holds for \( p \in [1, 2) \) and \( t_0 \in \left( 0, \frac{2}{\lambda_p(x_0)} \right) \), then Algorithm 4.1 for solving problem (4.1) with initial point \( x_0 \) is well defined and converges linearly to the unique Riemannian \( L^p \) center of mass of \( \{y_i\} \).

References

[1] P.-A. Absil, C. G. Baker, and K. A. Gallivan, Trust-Region Methods on Riemannian Manifolds, Found. Comput. Math., 7 (2007), pp. 303–330.

[2] P. A. Absil, R. Mahony, and R. Sepulchre, Optimization Algorithms on Matrix Manifolds, Princeton University Press, Princeton, 2008.

[3] R. Adler, J. P. Dedieu, J. Margulies, M. Martens, and M. Shub, Newton’s method on Riemannian manifolds and a geometric model for the human spine, IMA J. Numer. Anal., 22 (2002), pp. 359–390.

[4] B. Afsari, Riemannian \( L^p \) center of mass: existence, uniqueness, and convexity, Proc. Amer. Math. Soc., 139 (2010), pp. 655–673.

[5] B. Afsari, R. Tron, and R. Vidal, On the convergence of gradient descent for finding the Riemannian center of mass, SIAM J. Control Optim., 51 (2013), pp. 2230–2260.

[6] M. Arnaudon, C. Dombry, A. Phan, and L. Yang, Stochastic algorithms for computing means of probability measures, Stochastic Proc. Appl., 122 (2012), pp. 1437–1455.
[7] G. C. Bento, O. P. Ferreira, and P. R. Oliveira, Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds, J. Optim. Theory Appl., 154 (2012), pp. 88–107.

[8] R. L. Bishop and R. J. Crittenden, Geometry of Manifold, Academic Press, New York and London, 1964.

[9] J. V. Burke and S. Deng, Weak sharp minima revisited, I: Basic theory, Control Cybernet., 31 (2002), pp. 439–469.

[10] J. V. Burke, A. Lewis, and M. Overton, Optimal stability and eigenvalue multiplicity, Found. Comput. Math., 1 (2001), pp. 205–225.

[11] B. Chow, S. Chu, D. Glickenstein, C. Guenther, J. Isenberg, T. Ivey, D. Knopf, P. Lu, F. Luo, and L. Ni, The Ricci Flow: Techniques and Applications. Part III: Geometric-Analytic Aspects, volume 163 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2010.

[12] M. P. DoCarmo, Riemannian Geometry, Birkhäuser Boston, Boston MA, 1992.

[13] Y. M. Ermol’ev, On the method of generalized stochastic gradients and quasi-Fejér sequences, Cybernetics, 5 (1969), pp. 208–220.

[14] D. Gabay, Minimizing a differentiable function over a differential manifold, J. Optim. Theory Appl., 37 (1982), pp. 177–219.

[15] D. Groisser, Newton’s method, zeros of vector fields, and the Riemannian center of mass, Adv. Appl. Math., 33 (2004), pp. 95–135.

[16] C. Li, G. López, and V. Martín-Márquez, Monotone vector fields and the proximal point algorithm on Hadamard manifolds, J. Lond. Math. Soc., 79 (2009), pp. 663–683.

[17] C. Li, B. S. Mordukhovich, J. Wang, and J. C. Yao, Weak sharp minima on Riemannian manifolds, SIAM J. Optim., 21 (2011), pp. 1523–1560.

[18] C. Li and J. C. Yao, 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 (2012), pp. 2486–2514.

[19] S. Li, C. Li, Y. Liou, and J. Yao, Existence of solutions for variational inequalities on Riemannian manifolds, Nonlinear Anal., 71 (2009), pp. 5695–5706.

[20] D. G. Luenberger, The gradient projection method along geodesics, Management Sci., 18 (1972), pp. 620–631.
[21] S. Z. Németh, *Five kinds of monotone vector fields*, Pure Math. Appl., 9 (1998), p. 417–428.

[22] J. X. D. C. Neto, L. L. D. Lima, and P. R. Oliveira, *Geodesic algorithms in Riemannian geometry*, Balkan J. Geom. Appl., 3 (1998), pp. 89–100.

[23] E. A. Papa Quiroz, E. M. Quispe, and P. R. Oliveira, *Steepest descent method with a generalized Armijo search for quasiconvex functions on Riemannian manifolds*, J. Math. Anal. Appl., 341 (2008), pp. 467–477.

[24] P. Pertersen, *Riemannian Geometry*, Springer-Verlag, New York, 2006.

[25] T. Sakai, *Riemannian Geometry. Translations of Mathematical Monographs*, Am. Math. Soc., Providence, 1996.

[26] S. T. Smith, *Geometric Optimization Methods for Adaptive Filtering*, phd thesis, Harvard University Cambridge Massachusetts, 1994.

[27] M. Studniarski and D. E. Ward, *Weak sharp minima: characterizations and sufficient conditions*, SIAM J. Control Optim., 38 (1999), pp. 219–236.

[28] C. Udriste, *Convex Functions and Optimization Methods on Riemannian Manifolds. In: Mathematics and Its Applications*, Kluwer Academic, Dordrecht, 1994.

[29] J. H. Wang, C. Li, G. Lopez, and J. C. Yao, *Convergence analysis of inexact proximal point algorithms on Hadamard manifolds*, J Global Optim., 61 (2015), p. 553–573.

[30] J. H. Wang, G. López, V. Martín-Márquez, and C. Li, *Monotone and accretive vector fields on Riemannian manifolds*, J. Optim. Theory Appl., 146 (2010), pp. 691–708.

[31] X. M. Wang, *Subgradient algorithms on Riemannian manifolds of lower bounded curvatures*, Optimization, 67 (2018), pp. 179–194.

[32] X. M. Wang, C. Li, J. H. Wang, and J. C. Yao, *Linear convergence of subgradient algorithm for convex feasibility on Riemannian manifolds*, SIAM J. Optim., 25 (2015), pp. 2334–2358.

[33] X. M. Wang, C. Li, and J. C. Yao, *Subgradient projection algorithms for convex feasibility on Riemannian manifolds with lower bounded curvatures*, J. Optim. Theory Appl., 164 (2015), pp. 201–217.

[34] D. E. Ward, *Characterizations of strict local minima and necessary conditions for weak sharp minima*, J. Optim. Theory Appl., 80 (1994), pp. 551–571.
[35] L. Yang, *Riemannian median and its estimation*, LMS J. Comput. Math., 13 (2010), p. 461–479.

[36] Y. Yang, *Globally convergent optimization algorithms on Riemannian manifolds: uniform framework for unconstrained and constrained optimization*, J. Optim. Theory Appl., 132 (2007), pp. 245–265.

[37] S. T. Yau, *Non-existence of continuous convex functions on certain Riemannian manifolds*, Math. Ann., 207 (1974), pp. 269–270.