SIAM J. OPTIM.
Vol. 31, No. 1, pp. 172–199

CONVERGENCE ANALYSIS OF GRADIENT ALGORITHMS ON RIEMANNIAN MANIFOLDS WITHOUT CURVATURE CONSTRAINTS AND APPLICATION TO RIEMANNIAN MASS*

JINHUA WANG†, XIANGMEI WANG‡, CHONG LI§, AND 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 [B. Afsari, R. Tron, and R. Vidal, *SIAM J. Control Optim.*, 51 (2013), pp. 2230–2260; G. C. Bento et al., *J. Optim.* Theory Appl., 183 (2019), pp. 977–992].

Key words. Riemannian manifold, sectional curvature, gradient algorithm, local convergence, global convergence, linear convergence, Riemannian center of mass

AMS subject classifications. Primary, 53C20; Secondary, 53C22

DOI. 10.1137/19M1289285

1. Introduction. Let M be a Riemannian manifold and let $f : M \to \overline{\mathbb{R}} := (-\infty, \infty]$ be a proper function which is locally Lipschitz continuous on its domain. The optimization problem

$$\min_{x \in M} f(x) \tag{1.1}$$

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, 34] and the references therein), but also is a useful tool to treat some nonsmooth/nonconvex and/or constrained optimization problems appearing on the Euclidean space. As explained in [20], 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

*Received by the editors September 24, 2019; accepted for publication (in revised form) October 7, 2020; published electronically January 12, 2021.
https://doi.org/10.1137/19M1289285
Funding: The first author was supported in part by the National Natural Science Foundation of China, grant 11771397. The second author was supported in part by the National Natural Science Foundation of China, grant 11661019, and by the Guizhou Provincial Natural Science Foundation of China, grant 20161039. The third author was supported in part by the National Natural Science Foundation of China, grant 11971429, and by the Zhejiang Provincial Natural Science Foundation of China, grant LY18A010004. The fourth author was supported in part by MOST grant 108-2115-M-039-005-MY3.
†Department of Mathematics, Hangzhou Normal University, Hangzhou 311121, People's Republic of China (wangjh@hznu.edu.cn).
‡Corresponding author. College of Mathematics and Statistics, Guizhou University, Guiyang, Guizhou, 550025, People's Republic of China (xmwang2@gzu.edu.cn).
§School of Mathematical Sciences, Zhejiang University, Hangzhou, Zhejiang, 310027, People's Republic of China (cli@zju.edu.cn).
¶Center for General Education, China Medical University, Taichung, and Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, Taiwan 40402, ROC (yaojc@mail.cmu.edu.tw).

the intrinsic reduction of the dimensionality of the problem and the method's insight about the problem structure; see also [3, 5, 8, 15, 22, 23, 25, 39] 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 $\{y_i : 1 \le i \le N\} \subseteq M$, which can be formulated as a special case of problem (1.1) with the objective function f defined by

$$f(x) := \begin{cases} \frac{1}{p}\sum_{i=1}^{N} w_i \mathrm{d}^p(x, y_i), & 1 \le p < +\infty, \\ \max_{1\le i\le N} \mathrm{d}(x, y_i), & p = +\infty, \end{cases} \quad \text{for any } x \in M, \tag{1.2}$$

where $\{w_i : 1 \le i \le N\} \subseteq (0, +\infty)$ 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, 21] and the references therein). As mentioned in [4], 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 more general existence and uniqueness results for the Riemannian L^p centers of mass and some methods for locating the Riemannian centers of mass such as the gradient algorithm, the subgradient algorithm, the stochastic gradient algorithm, and Newton's method; see, e.g., [4, 5, 6, 9, 21, 45].

Related to the optimization problem (1.1), some important notions and techniques, such as weak sharp minima and variational analysis, have been developed in [23, 25], while the classical numerical methods for solving optimization problems on the Euclidean space, such as Newton's method, the trust region method, the gradient algorithm, the subgradient algorithm, the proximal point method, etc., have been extended to the Riemannian manifold setting; see, e.g., [1, 7, 20, 24, 34, 38, 40]. 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 [27], 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 [27, Theorem 1] for more details. This work was developed by Gabay in [20] with the weaker assumption that the sublevel 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 critical point at which the Hessian form of f is positive definite (see [20, (57)] for the definition of the Hessian form).

One important development in this direction is the work of Smith in [34], where he developed the gradient algorithm (together with other algorithms such as a 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 [34, Theorem 2.3]. Later, Yang studied the gradient algorithm employing the Armijo step sizes on a general Riemannian manifold and established in [46, Theorem 3.4] the global convergence result under the assumption that the generated sequence $\{x_k\}$ satisfies $\lim_{k\to+\infty} \mathrm{d}(x_k, x_{k+1}) = 0$ and has a cluster point $\bar{x}$ such that $\bar{x}$ is an isolated critical point, and in [46, 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 [30] and [31] to establish the global convergence results for the gradient algorithm (employing the Armijo step sizes) for the convex case and the quasi-convex case, respectively:

(A1) The curvatures of the Riemannian manifold M are nonnegative.[1]

(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 $\mathbb{S}^n_{++}$ endowed with the Riemannian metric defined by $\langle U, V\rangle_X := \mathrm{trace}(VX^{-1}UX^{-1})$ for any $X \in \mathbb{S}^n_{++}$, $U, V \in T_X\mathbb{S}^n_{++}$ (although it has positive curvatures when being equipped with the Bruce–Wasserstein distance proposed in [28]).
- Assumption (A2) prevents the application to some special but important Riemannian manifolds, such as compact Stiefel manifolds $\mathrm{St}(p, n)$ and Grassmann manifolds $\mathrm{Grass}(p, n)$ $(p < n)$ since there is no nontrivial (quasi-)convex function (with full domain) on a complete manifold with finite volume (see, e.g., [44, Proposition 3.4]).
- 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.

Note that the convergence properties of Newton's method and the trust region method (cf. [1, 24]) are free of the sectional curvatures of the underlying Riemannian manifolds. To remove the sectional curvature assumption (A1), Bento, Cruz Neto, and Oliveira extended in [7] the Kurdyka–Łojasiewicz condition to the Riemannian manifold setting and established convergence results (without any estimate for the convergence rate) on a general Riemannian manifold for an abstract inexact descent method (which includes the gradient algorithm as a special example) for the cost (not necessarily quasi-convex/convex) function f satisfying the Kurdyka–Łojasiewicz condition at a cluster point of the generated sequence. We remark that quasi-convex/convex functions do not satisfy the Kurdyka–Łojasiewicz condition, in general.

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 $f : M \to \overline{\mathbb{R}}$ is locally Lipschitz continuous on its domain (and so not necessarily continuously differentiable, nor quasi-convex/convex on the whole Riemannian manifold). The present paper contains two types of convergence results for the gradient algorithm for locally quasi-convex function f employing more general step sizes (which includes

[1] This assumption is relaxed to the curvatures of M having a lower bound in the recent paper [19] for establishing the convergence result (without any estimate for the convergence rate) of the gradient algorithm for differentiable and convex function f on the whole manifold M. We thank the referee for providing us this reference.

the Armijo step sizes as a special case): one is the local convergence, which means that any sequence generated with initial point close enough to a critical point converges and/or linearly converges to a critical point (see Theorems 3.2 and 3.3, which seem new in the linear space setting and/or in the case when the function f has the Kurdyka–Łojasiewicz property at the critical point), and the other is the global convergence, which means that any sequence generated with arbitrary initial point from the domain of the function f does (see Theorem 3.6 and Remark 3.4). In particular, the global convergence result is established for the gradient algorithm employing the Armijo step sizes or the step sizes having a positive lower bound (cf. Remark 3.2(b)) under the following weaker assumption than (A1) and (A2) (see Lemma 3.7):

(H) The generated sequence $\{x_k\}$ has a cluster point $\bar{x}$ 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 sizes $\{t_k\}$ have a positive lower bound.

(Note by Lemma 3.4 that the Armijo step sizes have a positive lower bound if the gradient of f is Lipschitz continuous around $\bar{x}$.)

As explained before Theorem 3.6 and Corollary 3.8, the linear convergence results in the present paper extend [46, Theorem 4.1], while the global convergence results extend/improve particularly the corresponding ones in [31, Theorem 3.1] (and so [30, Theorem 5.3]) and [19, Theorem 3.11]. The main technique used here is based on Lemmas 2.5 and 3.1 developed in the present paper, which seems new and is different from the ones in [7] and [19].

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 $p \in [2, +\infty)$ extend the corresponding one in [5, Theorem 4.1] (see the explanation before Corollary 4.7). In particular, in the special case when M is a Hadamard manifold, Corollary 4.8 extends/improves the corresponding ones in [9, Theorems 3.1 and 3.2] as we remarked before Corollary 4.8.

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., [17, 33, 37]).

Let M be a connected and complete n-dimensional Riemannian manifold. We use ∇ to denote the Levi-Civita connection on M. Let $x \in M$, and let $\mathrm{T}_x M$ 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 γ is defined by $l(\gamma) := \int_0^1 \|\gamma'(t)\| dt$, while the Riemannian distance from x to y is defined by $\mathrm{d}(x,y) := \inf_\gamma l(\gamma)$, 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 $\mathbb{B}(x,r)$ and $\mathbb{U}(x,r)$, respectively, that is,

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

A vector field V is said to be parallel along γ if $\nabla_{\gamma'} V = 0$. In particular, for a smooth curve γ, if γ' is parallel along itself, then γ is called a geodesic; thus, a smooth curve γ 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 [17], (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 ∂Q to stand for the closure and the boundary of Q, respectively. The distance function $\mathrm{d}_Q(\cdot)$ associated to Q and the projection $P_Q(\cdot)$ onto Q are respectively defined by, for any $x \in M$,

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

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 Γ^Q_{xy}, that is,

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

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

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

$$r_{\mathrm{inj}}(x) := \sup\{r > 0 : \exp_x(\cdot) \text{ is a diffeomorphism on } \mathbb{B}(0,r) \subset \mathrm{T}_x M\}$$

and

$$r_{\mathrm{cvx}}(x) := \sup\left\{ r > 0 : \begin{array}{l} \text{each ball in } \mathbb{B}(x,r) \text{ is strongly convex} \\ \text{and each geodesic in } \mathbb{B}(x,r) \text{ is minimal} \end{array} \right\}, \tag{2.1}$$

respectively. Then, $r_{\mathrm{inj}}(x) \ge r_{\mathrm{cvx}}(x) > 0$ for any $x \in M$; see, e.g., [33, Theorem 5.3]. In particular, $r_{\mathrm{inj}}(x) = r_{\mathrm{cvx}}(x) = +\infty$ for each $x \in M$ if M is a Hadamard manifold. Moreover, for any compact subset $Q \subseteq M$, we have that

$$r_{\mathrm{inj}}(Q) := \inf\{r_{\mathrm{inj}}(x) : x \in Q\} > 0 \quad \text{and} \quad r_{\mathrm{cvx}}(Q) := \inf\{r_{\mathrm{cvx}}(x) : x \in Q\} > 0;$$

see [33, Theorem 5.3, p. 169] or [26, Lemma 3.1].

Definition 2.1 below presents the notions of different kinds of convexities about subsets in M; see, e.g., [26, 39].

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 strong/total convexity implies the weak 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 $f : M \to \overline{\mathbb{R}}$ 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 kth differentiable, that is,

$$\mathcal{D}^k(f) := \{x \in \operatorname{int}\mathcal{D}(f) : f \text{ is } k\text{th differentiable at } x\}. \tag{2.2}$$

As usual, we say that f is C^k on Q if $Q \subseteq \mathcal{D}^k(f)$ and its kth derivative is continuous at each point of Q and that f is C^k around $\bar{x}$ if it is C^k on $\mathbb{B}(x, r)$ 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 ∇f is Lipschitz continuous around $\bar{x} \in \operatorname{int}\mathcal{D}^1(f)$, if there exist positive constants δ, L (with $\delta \le r_{\rm cvx}(\bar{x})$) such that

$$\|\nabla f(x) - P_{x,y}\nabla f(y)\| \le L\mathrm{d}(x, y) \quad \text{for any } x, y \in \mathbb{B}(\bar{x}, \delta).$$

Thus, if f is C^2 around $\bar{x}$, then ∇f is Lipschitz continuous around $\bar{x}$.

Item (b) in the following definition was known in [23, Definition 6.1(b)] (for the convexity) and [31, Definition 2.2] (for the quasi-convexity in the case when $\mathcal{D}(f) = M$).

Definition 2.2. *Let $f : M \to \overline{\mathbb{R}}$ be proper and let $Q \subseteq \mathcal{D}(f)$ be weakly convex. Then, f is said to be*

(a) *convex (resp., strictly convex, quasi-convex) on Q if, for any $x, y \in Q$ and any geodesic $\gamma \in \Gamma^Q_{xy}$, the composition $f \circ \gamma : [0, 1] \to \mathbb{R}$ is convex (resp., strictly convex, quasi-convex) on $[0, 1]$;*

(b) *convex (resp., strictly convex, quasi-convex) if $\mathcal{D}(f)$ is weakly convex and f is convex (resp., strictly convex, quasi-convex) on $\mathcal{D}(f)$.*

(c) *convex (resp., strictly convex, quasi-convex) around $x \in \mathcal{D}(f)$ if f is convex (resp., strictly convex, quasi-convex) on $\mathbb{B}(x, r)$ 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., [37, Theorems 5.1, 6.2] for assertions (i), (iii) and [29, Proposition 3.1] for assertion (ii).

Lemma 2.3. *Let $f : M \to \overline{\mathbb{R}}$ be proper. Let $Q \subseteq \mathcal{D}(f)$ be weakly convex and let $x \in Q \cap \mathcal{D}^1(f)$. Then, the following assertions hold.*

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

$$f(y) \ge f(x) + \langle \nabla f(x), \gamma'_{xy}(0)\rangle \quad \textit{for all } \gamma_{xy} \in \Gamma^Q_{xy}.$$

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

$$\langle \nabla f(x), \gamma'_{xy}(0)\rangle \le 0 \quad \textit{for all } \gamma_{xy} \in \Gamma^Q_{xy}.$$

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

Let $\kappa \in \mathbb{R}$. As in [33], let M^m_κ denote the m-dimensional complete simply connected Riemannian manifold of constant curvature κ. Following [33, p. 161], a generalized geodesic hinge $\Lambda(p; \gamma, \tau)$ in M is a figure consisting of a point $p \in M$ (the vertex of the hinge) and two geodesic segments γ, τ (the edges of the hinge) emanating from

p with γ being minimal. Moreover, a hinge $\Lambda(\bar{p};\bar{\gamma},\bar{\tau})$ in M_κ^m is called a comparison hinge of $\Lambda(p;\gamma,\tau)$ if it satisfies

$$l(\bar{\gamma}) = l(\gamma), \quad l(\bar{\tau}) = l(\tau), \quad \text{and} \quad \angle_{\bar{p}}(\bar{\gamma},\bar{\tau}) = \angle_p(\gamma,\tau).$$

Proposition 2.4 below is a "local" version of the Toponogov comparison theorem, i.e., [33, Theorem 4.2(2), p. 161], for a generalized hinge in a complete Riemannian manifold. Note that the crucial tool for proving [33, Theorem 4.2(2), p. 161] is [33, Theorem 2.5(2), p. 150], which, in view of the Rauch comparison theorem (see, e.g., [33, Theorem 2.3, p. 149]), remains true if $\tilde{M}$ there is the Riemannian manifold M_κ^m of constant curvature and the lower bound condition of the sectional curvatures $K_\sigma \geq \kappa$ for any plane section $\sigma \subseteq T_xM$ is assumed to hold at all points x in the involved geodesic γ, rather than in the whole M; see the proof presented in [33, pp. 150–151]. Thus [33, Lemma 4.4, p. 163] remains true in general Riemannian manifold M (which particularly shows that each closed and weakly convex nonempty subset $Q \subseteq M$ is a complete Alexandrov space; see [12] for a definition), and the argument for proving [33, Theorem 4.2(2), p. 161] works also for Proposition 2.4, and so its proof is omitted here.[2]

Proposition 2.4. *Let $\kappa < 0$ and suppose that M is an m-dimensional Riemannian manifold. Let $Q \subseteq M$ be a weakly convex set with nonempty interior (i.e., $\mathrm{int}Q \neq \emptyset$). Let $\Lambda(p;\gamma,\tau)$ be a generalized geodesic hinge in $\mathrm{int}Q$ and let $\Lambda(\bar{p};\bar{\gamma},\bar{\tau})$ be its comparison hinge in M_κ^m. Suppose that the sectional curvatures K_σ on Q are bounded from below by κ. Then $\mathrm{d}(q_\gamma,q_\tau) \leq \mathrm{d}(\bar{q}_{\bar{\gamma}},\bar{q}_{\bar{\tau}})$, where $q_\gamma, q_\tau, \bar{q}_{\bar{\gamma}}$, and $\bar{q}_{\bar{\tau}}$ denote the end points of $\gamma,\tau,\bar{\gamma}$, and $\bar{\tau}$, respectively.*

We show in the following lemma some inequalities, which play important roles in our study. For this purpose, we define the function $\hbar : [0,+\infty) \to \mathbb{R}$ as in [40] by

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

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

Lemma 2.5. *Let $f : M \to \overline{\mathbb{R}}$ be proper, and let $Q \subseteq M$ be weakly convex such that $Q_f := \mathcal{D}(f) \cap Q$ is weakly convex with nonempty interior (i.e., $\mathrm{int}Q_f \neq \emptyset$). Let $t \geq 0$, $x \in \mathrm{int}Q_f \cap \mathcal{D}^1(f)$ and $\gamma : [0,+\infty) \to M$ be the geodesic satisfying*

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

Suppose further that the sectional curvatures on Q_f are bounded from below by some $\kappa < 0$ and that f is quasi-convex on Q_f. Then the following inequalities hold for any $z \in \mathrm{int}Q_f$ satisfying $f(z) \leq f(x)$:

$$\cosh\left(\sqrt{|\kappa|}\mathrm{d}(\gamma(t),z)\right) \leq \cosh\left(\sqrt{|\kappa|}\mathrm{d}(x,z)\right)\left(1 + \frac{|\kappa|}{2}t\|\nabla f(x)\|\sinh(t\|\nabla f(x)\|)\right); \tag{2.5}$$

$$\mathrm{d}^2(\gamma(t),z) < \mathrm{d}^2(x,z) + \frac{3t^2\|\nabla f(x)\|^2}{2\hbar\left(\sqrt{|\kappa|}\mathrm{d}(x,z)\right)} \quad \text{if } \sqrt{|\kappa|}t\|\nabla f(x)\| \leq 1. \tag{2.6}$$

[2] In view of the equivalence of conditions T.C.T. (I) and (II) in [33, Theorem 4.2, pp. 161–162] (see Remark 4.3(2) there), Proposition 2.4 for the special case when $\Lambda(p;\gamma,\tau)$ is a geodesic hinge could also be derived from [12, Globalization theorem 3.2] (see also [36, Theorem 1.3 and Remark 1.5]) in a complete Alexandrov space (noting by the arguments in [12, section 2] that conditions (D) and (C, C_1) in [12] are equivalent).

Proof. Let $z \in \mathrm{int}Q_f$ be such that $f(z) \leq f(x)$, and let $\gamma_{xz} \in \Gamma_{xz}^{Q_f}$ be a minimal geodesic joining x and z. Without loss of generality, we assume that $\kappa = -1$. Then, by the argument (applied to $x, \gamma(t)$, $t\|\nabla f(x)\|$ in place of x^k, x^{k+1}, t_k) at the beginning of the proof for [42, Lemma 3.1] (using Proposition 2.4 where [42, Proposition 2.2] is used), one checks that

(2.7)
$$\begin{aligned}\cosh \mathrm{d}(\gamma(t), z) \leq \cosh \mathrm{d}(x, z) \\ + \cosh \mathrm{d}(x, z) \sinh(t\|\nabla f(x)\|) \left(\tfrac{t\|\nabla f(x)\|}{2} - \tanh \mathrm{d}(x, z) \cos\alpha\right),\end{aligned}$$

where $\alpha := \angle_x(\gamma, \gamma_{xz})$ is the angle between γ and γ_{xz} at x. Below, we verify that $\cos\alpha \geq 0$. Granting this, (2.5) follows immediately from (2.7). To do this, we note by Lemma 2.3(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, $\cos\alpha = \frac{\langle \gamma'(0), \gamma_{xz}'(0)\rangle}{\|\gamma'(0)\| \cdot \mathrm{d}(x,z)} \geq 0$ as we desired to show.

To show (2.6), 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, (2.5) implies that

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

Therefore, in view of the definition of $\hbar$ in (2.3), (2.6) is seen to hold from the following estimate (see [41, Lemma 3.1]):

$$\cosh s_1 - \cosh s_2 \geq \tfrac{(s_1^2 - s_2^2)\sinh s_2}{2s_2} \qquad \text{for any } s_1, s_2 \in (0, +\infty).$$

The proof is complete. ∎

We shall use the following known lemmas in what follows; see, e.g., [40, Lemma 2.3] for Lemma 2.6 and [18] for Lemma 2.7. Let $\mathbb{N}$ denote the set of all nonnegative integers.

LEMMA 2.6. *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 2.7. *Let $\{y_k\} \subset M$ be a sequence quasi-Fejér convergent to S, namely there exists a sequence $\{\varepsilon_k\} \subset (0, +\infty)$ satisfying $\sum_{k=0}^{\infty} \varepsilon_k < \infty$ such that $\mathrm{d}^2(y_{k+1}, z) \leq \mathrm{d}^2(y_k, z) + \varepsilon_k$ 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, $f : M \to \overline{\mathbb{R}}$ is a proper function which is locally Lipschitz continuous on its domain. Associated to the optimization problem (1.1), let C_f denote the set of all critical points of f:

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

where $\mathcal{D}^1(f)$ is the set defined by (2.2).

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

ALGORITHM 3.1. *Give $x_0 \in \mathcal{D}(f)$, $\beta \in (0, 1)$, $R \in [1, +\infty)$ and set $k := 0$.*
Step 1. *If $x_k \in C_f$ or $x_k \notin \mathcal{D}^1(f)$, then stop; otherwise construct the geodesic γ_k such that*

$$\gamma_k(0) = x_k \quad \text{and} \quad \gamma_k'(0) = -\nabla f(x_k). \tag{3.1}$$

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

$$f(\gamma_k(t_k)) \le f(x_k) - \beta t_k \|\nabla f(x_k)\|^2. \tag{3.2}$$

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

Let $\nu \in (0,1)$. Recall that Algorithm 3.1 is said to employ the (generalized) Armijo step sizes if each step size t_k in Step 2 is chosen by

$$t_k := \max\{\nu^i : i \in \mathbb{N},\ f(\gamma_k(\nu^i)) \le f(x_k) - \beta\nu^i \|\nabla f(x_k)\|^2\}; \tag{3.3}$$

see, e.g., [20, 34, 46] and [2, section 4.6.3]. Note that (3.3) particularly implies (3.2).

The following remark regards the well definedness of Algorithm 3.1.

Remark 3.1. Suppose that $\{x_j : 0 \le j \le k\} \subset \mathcal{D}(f)$ 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 [46, Proposition 3.1], we can check that (3.3) 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.

For the remainder of this section, we make the following *standing assumption*:

$$\begin{cases} \bar{f} := \inf_{x\in M} f(x) > -\infty \quad \text{and} \quad \mathrm{int}\mathcal{D}(f) \ne \emptyset; \\ \text{Algorithm 3.1 does not terminate in finite steps.} \end{cases} \tag{3.4}$$

This particularly implies that, for each $k \in \mathbb{N}$, f is differentiable at x_k and t_k exists to satisfy (3.2).

Remark 3.2. (a) Let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with initial point $x_0 \in \mathcal{D}(f)$. Then, by Algorithm 3.1, the following inequalities hold for any $k \in \mathbb{N}$:

$$\mathrm{d}(\gamma_k(t), x_k) \le t\,\|\nabla f(x_k)\| \quad \text{for any } t \in [0, t_k]; \tag{3.5}$$

$$\sum_{j=0}^{k} t_j \|\nabla f(x_j)\|^2 \le \frac{f(x_0) - f(x_{k+1})}{\beta} \le \frac{f(x_0) - \bar{f}}{\beta} < +\infty \tag{3.6}$$

by the standing assumption (3.4). In particular, one has that $t_k\|\nabla f(x_k)\| \to 0$.

(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 ∇f is continuous at $\bar{x}$ is a critical point of f, that is, $\bar{x} \in C_f$. Indeed, it is immediate from (3.6) 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 [46, 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:

$$\bar{x} \in C_f \cap \mathrm{int}\mathcal{D}^1(f),\ f \text{ is quasi-convex around } \bar{x} \text{ and } \nabla f \text{ is continuous at } \bar{x}. \tag{3.7}$$

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 $k \in \mathbb{N}$, if $\{x_j : 0 \leq j \leq k+1\}$ generated by Algorithm* 3.1 *satisfies that*

$$x_0 \in \mathbb{B}(\bar{x}, \bar{\delta}) \quad \text{and} \quad \{x_j : 1 \leq j \leq k\} \subset \mathbb{B}(\bar{x}, \bar{c}\bar{\delta}), \tag{3.8}$$

then one has that

$$\mathrm{d}^2(x_{k+1}, z) \leq \mathrm{d}^2(x_k, z) + 2Rt_k\|\nabla f(x_k)\|^2 \leq \mathrm{d}^2(x_0, z) + \delta \mathrm{d}(x_0, z) \tag{3.9}$$

if $z \in \mathbb{B}(\bar{x}, \bar{c}\bar{\delta})$ satisfies $f(z) \leq f(x_{k+1})$ and that

$$x_{k+1} \in \mathbb{B}(\bar{x}, \bar{c}\bar{\delta}) \quad \text{if } f(\bar{x}) \leq f(x_{k+1}). \tag{3.10}$$

Proof. Noting that any closed ball is compact, we have by [10, p. 166] that the curvatures of the ball $\mathbb{B}(\bar{x}, r_{\mathrm{cvx}}(\bar{x}))$ are bounded, where $r_{\mathrm{cvx}}(\bar{x})$ is the convexity radius at $\bar{x}$ defined in (2.1). Let $\kappa < 0$ be a lower bound of the curvatures of $\mathbb{B}(\bar{x}, r_{\mathrm{cvx}}(\bar{x}))$. Thanks to assumption (3.7), there exists $\delta > 0$ (using a smaller δ if necessarily) such that f is quasi-convex on $\mathbb{B}(\bar{x}, \delta)$ and that

$$\mathbb{B}(\bar{x}, \delta) \subset \mathcal{D}^1(f), \quad \delta < \min\left\{r_{\mathrm{cvx}}(\bar{x}), \frac{1}{\sqrt{|\kappa|}}\right\} \quad \text{and} \quad R\sqrt{|\kappa|}\|\nabla f(\cdot)\| \leq 1 \text{ on } \mathbb{B}(\bar{x}, \delta). \tag{3.11}$$

Now let $0 < \delta_1 < \delta/2$ be such that

$$\|\nabla f(x)\| \leq \frac{\beta\delta}{2R} \quad \text{for any } x \in \mathbb{B}(\bar{x}, \delta_1). \tag{3.12}$$

To proceed, we first choose positive numbers $\bar{\delta}$ and $\bar{c}$ to satisfy that

$$\bar{c} > 1 \quad \text{and} \quad \bar{c}\bar{\delta} \leq \delta_1. \tag{3.13}$$

Below we verify that the implication (3.8)$\Longrightarrow$(3.9) holds for any $k \in \mathbb{N}$, any $\{x_j : 0 \leq j \leq k+1\}$ generated by Algorithm 3.1, and any $z \in \mathbb{B}(\bar{x}, \bar{c}\bar{\delta})$ satisfying $f(z) \leq f(x_{k+1})$. Grant this and assume that $f(\bar{x}) < f(x_{k+1})$. Then we estimate by (3.9) (applied to $\bar{x}$ in place of z and noting $\mathrm{d}(x_0, \bar{x}) \leq \bar{\delta}$) that

$$\mathrm{d}^2(x_{k+1}, \bar{x}) \leq \mathrm{d}^2(x_0, \bar{x}) + \delta \mathrm{d}(x_0, \bar{x}) \leq (\bar{\delta} + \delta)\bar{\delta}. \tag{3.14}$$

Thus, to ensure that $x_{k+1} \in \mathbb{B}(\bar{x}, \bar{c}\bar{\delta})$, it is sufficient to choose the pair $(\bar{c}, \bar{\delta})$ to satisfy (3.13) and that $\bar{\delta} + \delta \leq \bar{c}^2\bar{\delta}$. In particular, set

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

Then, the pair $(\bar{c}, \bar{\delta})$ is as desired because one has by definition that $\bar{c} \geq 3$, $\bar{c}\bar{\delta} < \delta_1$, and $\bar{\delta} + \delta = \bar{c}^2\bar{\delta}$ (recalling $\delta_1 < \frac{\delta}{2}$).

Thus, to complete the proof, let $k \in \mathbb{N}$, and let $\{x_j : 0 \leq j \leq k+1\}$ be generated by Algorithm 3.1 to satisfy (3.8). Fix $j \in \{0, 1, \ldots, k\}$, and let γ_j be the geodesic determined by (3.1). Note by (3.13) that $\bar{c}\bar{\delta} \leq \delta_1 < \frac{1}{2}\delta$. Then, $\|\nabla f(x_j)\| \leq \frac{\beta\delta}{2R}$ by (3.12), and $\mathrm{d}(x_j, \bar{x}) \leq \delta_1$ by (3.8). Therefore it follows from (3.5) that, for any $t \in [0, t_j]$,

$$\mathrm{d}(\gamma_j(t), \bar{x}) \leq \mathrm{d}(\gamma_j(t), x_j) + \mathrm{d}(x_j, \bar{x}) < t\,\|\nabla f(x_j)\| + \delta_1 \leq \frac{\beta\delta}{2} + \frac{1}{2}\delta < \delta$$

(noting that $t \le t_j \le R$), and then one has that

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

Now let $z \in \mathbb{B}(\bar{x}, \bar{c}\bar{\delta})$ be such that $f(z) \le f(x_{k+1})$. Then, we have that

$$\mathrm{d}(x_j, z) \le \mathrm{d}(x_j, \bar{x}) + \mathrm{d}(z, \bar{x}) \le 2\bar{c}\bar{\delta} < \delta.$$

Noting that $\sqrt{|\kappa|}\delta$ by the choice of δ in (3.11), one has that

$$\hbar\left(\sqrt{|\kappa|}\mathrm{d}(x_j, z)\right) \ge \hbar(\sqrt{|\kappa|}\delta) \ge \hbar(1) > \frac{3}{4}. \tag{3.15}$$

Since $f(z) \le f(x_j)$ and $\sqrt{|\kappa|}t_j\|\nabla f(x_j)\| \le \sqrt{|\kappa|}R\|\nabla f(x_j)\| \le 1$ by the third item of (3.11), it follows from (2.6) that

$$\mathrm{d}^2(\gamma_j(t_j), z) \le \mathrm{d}^2(x_j, z) + \frac{3t_j^2\|\nabla f(x_j)\|^2}{2\hbar\left(\sqrt{|\kappa|}\mathrm{d}(x_j, z)\right)} \le \mathrm{d}^2(x_j, z) + 2Rt_j\|\nabla f(x_j)\|^2, \tag{3.16}$$

where the last inequality holds by (3.15) (recalling $t_j \le R$). Since $x_{j+1} = \gamma_j(t_j)$, it follows that

$$\mathrm{d}^2(x_{k+1}, z) \le \mathrm{d}^2(x_k, z) + 2Rt_k\|\nabla f(x_k)\|^2. \tag{3.17}$$

Noting that $\sum_{j=0}^{k} t_j\|\nabla f(x_j)\|^2 \le \frac{f(x_0)-f(x_{k+1})}{\beta} \le \frac{f(x_0)-f(z)}{\beta}$ by (3.6) (recalling $f(z) \le f(x_{k+1})$) and summing up the inequalities in (3.16) over $0 \le j \le k-1$, one concludes that

$$\mathrm{d}^2(x_k, z) + 2Rt_k\|\nabla f(x_k)\|^2 \le \mathrm{d}^2(x_0, z) + \frac{2R}{\beta}\left(f(x_0) - f(z)\right). \tag{3.18}$$

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

$$f(x_0) - f(z) \le \|\nabla f(\gamma(\xi))\|\mathrm{d}(x_0, z) \le \frac{\beta\delta}{2R}\mathrm{d}(x_0, z),$$

where the last inequality is from (3.12). This, together with (3.18), implies that

$$\mathrm{d}^2(x_k, z) + 2Rt_k\|\nabla f(x_k)\|^2 \le \mathrm{d}^2(x_0, z) + \delta\mathrm{d}(x_0, z). \tag{3.19}$$

Thus (3.9) is seen to hold by (3.17), showing the implication. The proof is complete. ∎

Remark 3.3. In addition to assumption (3.7) made in Lemma 3.1, assume further that $\bar{x} \in M$ is a local minimizer of f. Then, for any $\delta > 0$, there exist $\bar{\delta} > 0$ and $\bar{c} \ge 3$ satisfying $\bar{c}\bar{\delta} < \delta$ such that, for any $k \in \mathbb{N}$, if $\{x_j : 0 \le j \le k+1\}$ is generated by Algorithm 3.1 to satisfy (3.8), then there holds that

$$f(\bar{x}) \le f(x_{k+1}) \quad \text{and} \quad x_{k+1} \in \mathbb{B}(\bar{x}, \bar{c}\bar{\delta}) \subseteq \mathcal{D}^1(f). \tag{3.20}$$

Indeed, as at the beginning of the proof for Lemma 3.1, one chooses $\delta > 0$ and $0 < \delta_1 < \delta/2$ such that (3.11) and (3.12) hold. Without of loss of generality, we may further assume that

$$f(\bar{x}) \le f(x) \quad \text{for any } x \in \mathbb{B}(\bar{x}, \delta). \tag{3.21}$$

Let $k \in \mathbb{N}$, and assume that $\{x_j : 0 \le j \le k+1\}$ is generated by Algorithm 3.1 such that (3.8) holds. Then one has by (3.12) that $t_k\|\nabla f(x_k)\| \le R\frac{\beta\delta}{2R} \le \frac{\delta}{2}$ (as $\beta < 1$) and so by (3.5) that

$$\mathrm{d}(\bar{x}, x_{k+1}) \le \mathrm{d}(\bar{x}, x_k) + \mathrm{d}(x_k, x_{k+1}) \le \bar{c}\bar{\delta} + t_k\|\nabla f(x_k)\| < \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

because $\bar{c}\bar{\delta} < \delta_1 < \frac{\delta}{2}$ by (3.13) and the choice of δ_1, which, together with (3.21), implies that $f(\bar{x}) \le f(x_{k+1})$ and so $x_{k+1} \in \mathbb{B}(\bar{x}, \bar{c}\bar{\delta})$ by Lemma 3.1. In particular, one has by (3.20) and Remark 3.1 that Algorithm 3.1 employing the Armijo step sizes with initial point $x_0 \in \mathbb{B}(\bar{x}, \bar{\delta})$ is well defined, and the generated sequence $\{x_k\}$ satisfies

$$\lim_{k\to+\infty} f(x_k) \ge f(\bar{x}). \tag{3.22}$$

Now, we are ready to show the first main result of this subsection.

THEOREM 3.2. *Let $\bar{x} \in M$ be such that assumption* (3.7) *holds. Then, for any $\delta > 0$, there exist $\bar{\delta} > 0$ and $\bar{c} \ge 3$ satisfying $\bar{c}\bar{\delta} < \delta$ such that, for any sequence $\{x_k\}$ generated by Algorithm* 3.1 *with initial point $x_0 \in \mathbb{B}(\bar{x}, \bar{\delta})$, if it satisfies* (3.22) *(e.g., $\bar{x}$ is a local minimizer of f), then one has the following assertions:*

(i) *The sequence $\{x_k\}$ stays in $\mathbb{B}(\bar{x}, \bar{c}\bar{\delta})$ and converges to a point x^* in $\mathcal{D}(f)$.*

(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 the assumed (3.7), Lemma 3.1 is applicable. Thus, for any $\delta > 0$, there exist $\bar{\delta} > 0$ and $\bar{c} \ge 3$ satisfying $\bar{c}\bar{\delta} < \delta$ such that, for any sequence $\{x_k\}$ generated by Algorithm 3.1, if it satisfies (3.8), then (3.10) holds (for any k); hence the following implication holds for each $k \in \mathbb{N}$:

$$[(3.8) \text{ and } (3.22) \text{ hold}] \Longrightarrow x_{k+1} \in \mathbb{B}(\bar{x}, \bar{c}\bar{\delta}). \tag{3.23}$$

Now, let $\{x_k\}$ be a sequence generated by Algorithm 3.1 with initial point $x_0 \in \mathbb{B}(\bar{x}, \bar{\delta})$ such that (3.22) holds. Then one checks by (3.23) (applied to $k = 0$) that $x_1 \in \mathbb{B}(\bar{x}, \bar{c}\bar{\delta})$, and concludes by mathematical induction that $\{x_k\} \subset \mathbb{B}(\bar{x}, \bar{c}\bar{\delta})$, showing the first conclusion of assertion (i). Consequently, the sequence $\{x_k\}$ has at least one cluster point, say, x^*. Letting $L_{\bar{\delta}} := \{x \in \mathbb{B}(\bar{x}, \bar{c}\bar{\delta}) : f(x) \le \inf_{k\in\mathbb{N}} f(x_k)\}$, one sees that $x^* \in L_{\bar{\delta}}$ since $\{f(x_k)\}$ is decreasing and f is continuous on $\mathbb{B}(\bar{x}, \bar{c}\bar{\delta})$ (using a smaller δ if necessary). Then, (3.9) holds for each $z \in L_{\bar{\delta}}$. Thanks to $\sum_{k=1}^{\infty} t_k\|\nabla f(x_k)\|^2 < +\infty$ by (3.6), we get that $\{x_k\}$ is quasi-Fejér convergent to $L_{\bar{\delta}}$. Hence, we conclude by Lemma 2.7 that $\lim_{k\to\infty} x_k = x^*$ (recalling $x^* \in L_{\bar{\delta}}$). 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) (using a smaller $\delta > 0$ such that ∇f is continuous on $\mathbb{B}(\bar{x}, \delta)$ if necessary). This completes the proof. ∎

To study the linear convergence property, we consider the following condition on some ball $\mathbb{B}(\bar{x}, r)$ with some constant $\alpha > 0$:

$$\|\nabla f(x)\|^2 \ge \alpha(f(x) - f(\bar{x})) \quad \text{for each } x \in \mathbb{B}(\bar{x}, r), \tag{3.24}$$

a global version of which was also used in [37, p. 268]. Note that (3.24) implies that f has the Kurdyka–Łojasiewicz property at $\bar{x}$ with $\varphi(\cdot) := \frac{2}{\sqrt{\alpha}}(\cdot)^{\frac{1}{2}}$ (cf. [7, Definition 4.1]).

Our second main result in this subsection is on the linear convergence property of Algorithm 3.1. Note that, to guarantee the linear convergence, it is required in Theorem 3.3 that the corresponding step sizes $\{t_k\}$ have a positive lower bound, which is satisfied by the Armijo step sizes in the case when ∇f is Lipschitz continuous around $\bar{x}$; see Lemma 3.4 below.

THEOREM 3.3. *Let $\bar{x} \in M$ be such that assumption* (3.7) *holds. Suppose that*

$$\bar{x} \text{ is a local minimizer of } f, \text{ and (3.24) holds for some } \alpha > 0 \text{ and } r > 0. \tag{3.25}$$

Then, there exists $\bar{\delta} > 0$ such that, for any generated sequence $\{x_k\}$ by Algorithm 3.1 with initial point $x_0 \in \mathbb{B}(\bar{x}, \bar{\delta})$, if the corresponding step sizes $\{t_k\}$ satisfy $\underline{t} := \inf_{k \geq 0}\{t_k\} > 0$, then it converges linearly to a local minimizer x^* of f satisfying $f(x^*) = f(\bar{x})$ and satisfies

$$\mathrm{d}^2(x_k, x^*) \leq \mu(f(x_k) - f(\bar{x})) \leq \mu\rho^{2k}(f(x_0) - f(\bar{x})) \quad \textit{for each } k \in \mathbb{N}, \tag{3.26}$$

where $\mu := \frac{R}{(1-\rho)^2\beta}$ and $\rho := \sqrt{1 - \alpha\beta\underline{t}}$.

Proof. Let $0 < \delta < r$. Then, by assumption, Theorem 3.2 is applicable to getting that there exist $\bar{\delta} > 0$, $\bar{c} > 3$ satisfying $\bar{c}\bar{\delta} \leq \delta$ such that the sequence $\{x_k\}$ generated by Algorithm 3.1 with initial point $x_0 \in \mathbb{B}(\bar{x}, \bar{\delta})$ satisfies

$$\{x_k\} \subset \mathbb{B}(\bar{x}, \bar{c}\bar{\delta}) \quad \text{and} \quad x_k \to x^* \in \mathbb{B}(\bar{x}, \bar{c}\bar{\delta}) \cap C_f \tag{3.27}$$

if $\underline{t} > 0$. Below we show that $\bar{\delta}$ is as desired. In fact, let $x_0 \in \mathbb{B}(\bar{x}, \bar{\delta})$ and let $\{x_k\}$ be a sequence generated by Algorithm 3.1 employing the step sizes satisfying $\underline{t} > 0$. Then (3.27) is satisfied. Since $c\bar{\delta} \leq \delta < r$, it follows from (3.25) that

$$\|\nabla f(x_k)\|^2 \geq \alpha(f(x_k) - f(\bar{x})) \quad \text{for each } k \in \mathbb{N}. \tag{3.28}$$

Thus, for each $k \in \mathbb{N}$, one checks that

$$f(x_k) - f(\bar{x}) - \beta t_k\|\nabla f(x_k)\|^2 \leq (1 - \alpha\beta t_k)(f(x_k) - f(\bar{x})) \leq (1 - \alpha\beta\underline{t})(f(x_k) - f(\bar{x})),$$

and, in view of (3.2),

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

hence

$$f(x_{k+l}) - f(\bar{x}) \leq (1 - \alpha\beta\underline{t})^l(f(x_k) - f(\bar{x})) \quad \text{for any } k,\, l \in \mathbb{N}, \tag{3.29}$$

and so the second inequality of (3.26) follows immediately. To show the first inequality of (3.26), we note by (3.5) and (3.2) that the following relation holds for each $k, l \in \mathbb{N}$:

$$\mathrm{d}^2(x_{k+l+1}, x_{k+l}) \leq Rt_{k+l}\|\nabla f(x_{k+l})\|^2 \leq \tfrac{R(f(x_{k+l}) - f(x_{k+l+1}))}{\beta} \leq \tfrac{R(f(x_{k+l}) - f(\bar{x}))}{\beta}$$

(noting $t_{k+l} \in (0, R]$). This, together with (3.29), implies that

$$\mathrm{d}(x_{k+l+1}, x_{k+l}) \leq \rho^l\sqrt{\tfrac{R(f(x_k) - f(\bar{x}))}{\beta}} \quad \text{for any } k, l \in \mathbb{N},$$

and then

$$\mathrm{d}(x_{k+l}, x_k) \leq \textstyle\sum_{j=1}^{l} \mathrm{d}(x_{k+j}, x_{k+j-1}) \leq \tfrac{1-\rho^l}{1-\rho}\sqrt{\tfrac{R(f(x_k) - f(\bar{x}))}{\beta}}.$$

Letting l go to infinity, we have $\mathrm{d}(x_k, x^*) \leq \frac{1}{1-\rho}\sqrt{\frac{R(f(x_k) - f(\bar{x}))}{\beta}}$, showing the first inequality of (3.26). Finally, one sees that $f(x^*) = f(\bar{x})$ holds by (3.28) (as $\nabla f(x^*) = 0$ by (3.27)). The proof is complete. ∎

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 3.4. *Let $\bar{x} \in M$ be such that assumption* (3.7) *holds, and suppose that ∇f is Lipschitz continuous around $\bar{x}$. Then, there exist $\underline{t} > 0$ and $\bar{\delta} > 0$ such that, for any $x_0 \in \mathbb{B}(\bar{x}, \bar{\delta})$, if Algorithm* 3.1 *employs the Armijo step sizes and the generated sequence $\{x_k\}$ satisfies* (3.22)*, then the generated step sizes $\{t_k\}$ satisfy that $\inf_{k\in\mathbb{N}} t_k \geq \underline{t}$.*

Proof. By assumption, Theorem 3.2 is applicable to getting that, for any $\delta > 0$, there exist $\bar{\delta} > 0$ and $\bar{c} > 3$ satisfying $\bar{c}\bar{\delta} \leq \delta$ with the property stated there. Without loss of generality, we may assume further that $3\nu^{-1}\bar{c}\bar{\delta} < r_{\rm cvx}(\bar{x})$ and assume by assumption that $\mathbb{B}(\bar{x}, 3\nu^{-1}\bar{c}\bar{\delta}) \subset \mathcal{D}^1(f)$ and there exists $L > 0$ such that

$$\|\nabla f(x) - P_{x,y}\nabla f(y)\| \leq L\mathrm{d}(x, y) \quad \text{for any } x, y \in \mathbb{B}(\bar{x}, 3\nu^{-1}\bar{c}\bar{\delta}). \tag{3.30}$$

Let $\underline{t} := \min\{\nu, \frac{\nu(1-\beta)}{L}\}$. Below, we show that $\underline{t}, \bar{\delta}$ are as desired. To do this, let $x_0 \in \mathbb{B}(\bar{x}, \bar{\delta})$, and let $\{t_k\}$ and $\{x_k\}$ be the generated Armijo step sizes and the generated sequence by Algorithm 3.1 with initial point x_0, respectively. Now fix k and assume that $t_k \leq \nu$. Then, by (3.3), we see that

$$f(\gamma_k(\nu^{-1}t_k)) - f(x_k) \geq -\nu^{-1}\beta t_k\|\nabla f(x_k)\|^2. \tag{3.31}$$

Noting that $\mathbb{B}(\bar{x}, \bar{c}\bar{\delta})$ is strongly convex, one sees that $\gamma_k([0, t_k])$ is the unique minimal geodesic joining x_k to x_{k+1}. Therefore $t_k\|\nabla f(x_k)\| = \mathrm{d}(x_k, x_{k+1}))$, and it follows that

$$\mathrm{d}(x_k, \gamma_k(\nu^{-1}t_k)) \leq \nu^{-1}t_k\|\nabla f(x_k)\| = \nu^{-1}\mathrm{d}(x_k, x_{k+1}) \leq 2\nu^{-1}\bar{c}\bar{\delta}$$

(see (3.27) for the last inequality). Thus, using the triangle inequality and noting that $1 < \nu^{-1}$, one checks that $\gamma_k(\nu^{-1}t_k) \in \mathbb{B}(\bar{x}, 3\nu^{-1}\bar{c}\bar{\delta})$ because

$$\mathrm{d}(\bar{x}, \gamma_k(\nu^{-1}t_k)) \leq \mathrm{d}(\bar{x}, x_k) + \nu^{-1}\mathrm{d}(x_k, x_{k+1}) \leq 3\nu^{-1}\bar{c}\bar{\delta}.$$

Using the mean value theorem (as $\mathbb{B}(\bar{x}, 3\nu^{-1}\bar{c}\bar{\delta}) \subset \mathcal{D}^1(f)$), we can choose $\bar{t}_k \in (0, t_k)$ to satisfy that

$$f(\gamma_k(\nu^{-1}t_k)) - f(x_k) = \left\langle \nabla f\left(\gamma_k(\nu^{-1}\bar{t}_k)\right), -\nu^{-1}t_k P_{\gamma_k,\gamma_k(\nu^{-1}\bar{t}_k),x_k}\nabla f(x_k)\right\rangle. \tag{3.32}$$

Since

$$\begin{aligned}
&\left\langle \nabla f\left(\gamma_k(\nu^{-1}\bar{t}_k)\right), -P_{\gamma_k,\gamma_k(\nu^{-1}\bar{t}_k),x_k}\nabla f(x_k)\right\rangle \\
&\quad = -\left\langle P_{\gamma_k,x_k,\gamma_k(\nu^{-1}\bar{t}_k)}\nabla f\left(\gamma_k(\nu^{-1}\bar{t}_k)\right) - \nabla f(x_k), \nabla f(x_k)\right\rangle - \|\nabla f(x_k)\|^2 \\
&\quad \leq \|P_{\gamma_k,x_k,\gamma_k(\nu^{-1}\bar{t}_k)}\nabla f\left(\gamma_k(\nu^{-1}\bar{t}_k)\right) - \nabla f(x_k)\| \cdot \|\nabla f(x_k)\| - \|\nabla f(x_k)\|^2 \\
&\quad \leq (\nu^{-1}t_k L - 1)\|\nabla f(x_k)\|^2,
\end{aligned}$$

where the last inequality holds by (3.30) (as $\gamma_k(\nu^{-1}\bar{t}_k) \in \mathbb{B}(\bar{x}, 3\nu^{-1}\bar{c}\bar{\delta})$), it follows from (3.32) that

$$\begin{aligned}
f(\gamma_k(\nu^{-1}t_k)) - f(x_k) &= \nu^{-1}t_k\left\langle \nabla f\left(\gamma_k(\nu^{-1}\bar{t}_k)\right), -P_{\gamma_k,\gamma_k(\nu^{-1}\bar{t}_k),x_k}\nabla f(x_k)\right\rangle \\
&\leq \nu^{-1}t_k(\nu^{-1}Lt_k - 1)\|\nabla f(x_k)\|^2.
\end{aligned}$$

Combining this and (3.31), we conclude that $t_k \geq \frac{\nu(1-\beta)}{L}$ (in the case when $t_k \leq \nu$), and so $\inf_{k\in\mathbb{N}} t_k \geq \min\{\nu, \frac{\nu(1-\beta)}{L}\}$ as we desired to show. ∎

Remark 3.4. (a) Motivated by the corresponding notion in the linear space setting (see, e.g., [11, 13, 14, 35, 43]), we say that a point $\bar{x} \in \mathcal{D}(f)$ is a local weak sharp minimizer of order $q \geq 1$ for problem (1.1) if there exist $\alpha, r > 0$ such that

$$\alpha \mathrm{d}_{\bar{S}}^q(x) \leq f(x) - f(\bar{x}) \quad \text{for any } x \in \mathbb{B}(\bar{x}, r), \tag{3.33}$$

where $\bar{S} := \{x \in M : f(x) = f(\bar{x})\}$. In the special case when $q = 1$, this notion coincides with the one of the local weak sharp minimizer introduced in [23] by Li et al., where some complete characterizations for this notion were developed on Riemannian manifolds.

(b) Assume that f is convex around $\bar{x}$. Then that $\bar{x} \in \mathrm{int}\mathcal{D}^1(f)$ is a local weak sharp minimizer of order 2 for problem (1.1) implies condition (3.25), and the converse is also true if ∇f is additionally Lipschitz continuous around $\bar{x}$. In fact, assume that $\bar{x} \in \mathrm{int}\mathcal{D}^1(f)$ is a local weak sharp minimizer of order 2 for (1.1). Then there exist $\alpha, r > 0$ such that

$$\mathbb{B}(\bar{x}, r) \subset \mathcal{D}^1(f) \text{ and (3.33) holds with } q = 2. \tag{3.34}$$

Without loss of generality, we may assume further that $0 < r < r_{\mathrm{cvx}}(\bar{x})$ and f is convex on $\mathbb{B}(\bar{x}, r)$. Thus, for $x \in \mathbb{B}(\bar{x}, \frac{r}{2})$, one checks that

$$f(x) - f(\bar{x}) = f(x) - f(z) \leq \langle \nabla f(x), -\exp_x^{-1} z\rangle \leq \|\nabla f(x)\| \mathrm{d}(x, z) = \|\nabla f(x)\| \mathrm{d}_{\bar{S}}(x),$$

where $z \in P_{\bar{S}}(x)$ and so $z \in \mathbb{B}(\bar{x}, r)$ (as $\mathrm{d}(z, \bar{x}) \leq \mathrm{d}(z, x) + \mathrm{d}(x, \bar{x}) \leq r$), and (3.24) is seen to hold by (3.34) with $\frac{r}{2}$ in place of r. Conversely, assume that condition (3.25) holds and ∇f is additionally Lipschitz continuous around $\bar{x}$. Then, Theorem 3.3 and Lemma 3.4 are applicable. Let $r := \bar{\delta}$, the positive number given by Theorem 3.3 and Lemma 3.4 with the properties stated there. Fixing $x \in \mathbb{B}(\bar{x}, r)$, we choose an initial point $x_0 := x$ and apply Theorem 3.3 (noting by Lemma 3.4 that the corresponding step sizes $\{t_k\}$ has a positive lower bound) to conclude that there exists x^* such that $f(x^*) = f(\bar{x})$ and $\mathrm{d}^2(x_0, x^*) \leq \mu(f(x_0) - f(\bar{x}))$ (see (3.26)). This shows that (3.33) holds with 2, $\frac{1}{\mu}$ in place of q, α.

In the spirit of the notion of a nondegenerate critical point $\bar{x}$ (in the sense that $\bar{x} \in C_f$, $\nabla^2 f$ is continuous around $\bar{x}$, and $\nabla^2 f(\bar{x})$ is positive definite; see [46, Definition 3.1]), we say that a critical point $\bar{x} \in C_f$ is a quasi-nondegenerate critical point of f if

(1) f is convex around $\bar{x}$, and ∇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, which, by Remark 3.4(b), implies in turn that (3.25) holds. 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 3.5. *Let $\bar{x}$ be a quasi-nondegenerate point of f. Then, there exists $\delta > 0$ such that any sequence $\{x_k\}$ generated by Algorithm* 3.1 *employing the Armijo step sizes with initial point $x_0 \in \mathbb{B}(\bar{x}, \delta)$ converges linearly to a local minimizer of f.*

Proof. By assumption, Lemma 3.4 is applicable to getting that there exist $\underline{t} > 0$ and $\bar{\delta} > 0$ with the property stated there. Hence, Theorem 3.3 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 [46, 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 (3.25) holds and that $\inf_{k\geq 0}\{t_k\} > 0$ by Lemma 3.4).

THEOREM 3.6. *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) *$\{x_k\}$ converges to $\bar{x}$.*

(ii) *If $\inf_{k\geq 0}\{t_k\} > 0$ and assumption* (3.25) *holds, then $\{x_k\}$ converges linearly to $\bar{x}$.*

Proof. Noting that (3.22) is naturally satisfied as $\{f(x_k)\}$ is nonincreasing monotone and $\bar{x}$ is a cluster point, we get from Theorem 3.2(i) that there exists $\delta > 0$ such that any sequence generated by Algorithm 3.1 with initial point in $\mathbb{B}(\bar{x}, \delta)$ is convergent. Now $\bar{x}$ is a cluster point, so there exists some $k_0 \in \mathbb{N}$ such that $x_{k_0} \in \mathbb{B}(\bar{x}, \delta)$. Thus, $\{x_k\}$ converges to some point, which in fact equals to $\bar{x}$ and assertion (i) holds.

With a similar argument as for assertion (i), but using Theorem 3.3 instead of Theorem 3.2(i), one sees that assertion (ii) holds. The proof is complete. ∎

The following lemma provides some sufficient conditions 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 sublevel 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 3.7. *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_f^0 is bounded.*

(b) *L_f^0 is totally convex with its curvatures being bounded from below and f is quasi-convex on L_f^0 (e.g., f is quasi-convex on M and M is of lower bounded curvatures).*

Proof. Note that $\{x_k\} \subseteq L_f^0$ as $\{f(x_k)\}$ is nonincreasing monotone. Then, $\{x_k\}$ is clear bounded under assumption (a).

Now, suppose that assumption (b) holds. Without loss of generality, we assume that the curvatures of L_f^0 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 $\{z\} \cup \{x_k\} \subseteq L_f^0$ because $\{f(x_k)\}$ is nonincreasing. Note that $f(x_k) \geq f(z)$ for each $k \in \mathbb{N}$. Then, by assumption, Lemma 2.5 is applicable on $Q_f := L_f^0$ (with x_k, γ_k, and t_k in place of x, γ, and t) to getting that for each $k \in \mathbb{N}$,

$$\cosh\left(\mathrm{d}(x_{k+1}, z)\right) \leq \cosh\left(\mathrm{d}(x_k, z)\right)\left(1 + \frac{1}{2}t_k\|\nabla f(x_k)\|\sinh(t_k\|\nabla f(x_k)\|)\right). \tag{3.35}$$

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 (3.6) and $\sup\{t_k\} \le R$) and $\lim_{t\to 0}\frac{\sinh t}{t} = 1$. In view of (3.35), Lemma 2.6 is applicable (with $\{\frac{1}{2}t_k\|\nabla f(x_k)\|\sinh(t_k\|\nabla f(x_k)\|\}$ and $\{\cosh(\mathrm{d}(x_k,z))\}$ in place of $\{b_k\}$ and $\{a_k\}$), and we get that $\{\cosh(\mathrm{d}(x_k,z))\}$ is bounded, and so is $\{x_k\}$ as desired. The proof is complete. ∎

The following corollary is immediate from Theorem 3.6 and Lemma 3.7. Particularly, the global convergence result (assertion (i)) under assumption (b) in Lemma 3.7 extends the corresponding one in [31, Theorem 3.1] and [19, Theorem 3.11] which were established respectively for the case when f is C^1 and quasi-convex on the Riemannian manifold M of nonnegative curvatures and the case when f is C^1 and convex on the Riemannian manifold M of lower bound curvatures (noting that, in each of both cases, any cluster point $\bar{x}$ of a sequence generated by Algorithm 3.1 employing the Armijo step sizes satisfies (3.7) by Remark 3.2(b)). As for assertion (ii), as far as we know, it is new in the Riemannain manifold setting.

Corollary 3.8. *Suppose that one of assumptions* (a) *and* (b) *in Lemma* 3.7 *holds. Then, any sequence* $\{x_k\}$ *generated by Algorithm* 3.1 *with initial point* $x_0 \in \mathcal{D}^1(f)$ *has at least a cluster point* $\bar{x}$*; furthermore, if* $\bar{x}$ *satisfies* (3.7)*, then assertions* (i) *and* (ii) *in Theorem* 3.6 *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 $N \ge 2$. Let $\{y_i : 1 \le i \le N\} \subset M$ (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:

$$\min_{x\in M} f_p(x), \tag{4.1}$$

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

$$f_p(x) := \frac{1}{p}\sum_{i=1}^N w_i \mathrm{d}^p(x,y_i) \quad \text{for any } x \in M \tag{4.2}$$

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

$$D := \bigcap_{i=1}^N \mathbb{U}(y_i, r_{\mathrm{inj}}(y_i)).$$

The following remark shows some properties of the function f_p defined in (4.2). As usual, let δ_C be the indicator function associated with the subset $C \subset M$, which is defined by $\delta_C(x) = 0$ if $x \in C$ and $\delta_C(x) = +\infty$ otherwise, and set $I := \{1,2,\ldots,N\}$ for the sake of brevity.

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., [33, pp. 108–110]. Moreover, if $f_p + \delta_D$ is differentiable at $x \in D$, then

$$\nabla(f_p + \delta_D)(x) = \nabla f_p(x) = -\sum_{i\in I_x} w_i \mathrm{d}^{p-2}(x,y_i)\exp_x^{-1} y_i, \tag{4.3}$$

where $I_x := \{i \in I : x \ne y_i\}$; see, e.g., [4].

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 $\varrho_p : (0, +\infty) \to \overline{\mathbb{R}}$ by

$$\varrho_p(r) := \begin{cases} \frac{1}{2}\min\{r_{\rm inj}(\mathbb{B}(o,2r)), \frac{\pi}{2\sqrt{\Delta_{\mathbb{B}(o,2r)}}}\} & \text{if } 1 \le p < 2; \\ \frac{1}{2}\min\{r_{\rm inj}(\mathbb{B}(o,2r)), \frac{\pi}{\sqrt{\Delta_{\mathbb{B}(o,2r)}}}\} & \text{if } 2 \le p < +\infty, \end{cases} \tag{4.4}$$

for each $r \in (0, +\infty)$, where $\Delta_{\mathbb{B}(o,2r)}$ is an upper bound of the sectional curvatures of $\mathbb{B}(o, 2r)$ (with the convention that $\frac{1}{\sqrt{\Delta}} = +\infty$ for $\Delta \le 0$). Then, $\varrho_p(\cdot)$ is nonincreasing monotonically on $(0, +\infty)$.

In what follows, we need two lemmas: the first one is known in [32, Theorem 29].

LEMMA 4.1. *Let $r > 0$ be such that*

$$r \le \frac{1}{2}\min\left\{r_{\rm inj}(\mathbb{B}(o,r)), \frac{\pi}{\sqrt{\Delta_{\mathbb{B}(o,r)}}}\right\}. \tag{4.5}$$

Then, $\mathbb{U}(o,r)$ is strongly convex.

LEMMA 4.2. *Let $r > 0$ be such that $\{y_i\} \subset \mathbb{U}(o,r)$ and*

$$r \le \varrho_p(r) \quad or \quad r < \frac{1}{2}\min\left\{r_{\rm inj}(\mathbb{B}(o,r)), \frac{\pi}{\sqrt{\Delta_{\mathbb{B}(o,r)}}}\right\}. \tag{4.6}$$

Then, $\mathbb{B}(o,r) \subset D$, and the following assertions hold:

(i) *$\mathbb{U}(o,r)$ is strongly convex (so $\mathbb{B}(o,r)$ is weakly convex), and $\mathbb{B}(o,r)$ is strongly convex if $r < \frac{1}{2}\min\{r_{\rm inj}(\mathbb{B}(o,r)), \frac{\pi}{\sqrt{\Delta_{\mathbb{B}(o,r)}}}\}$.*

(ii) *For any $y \in \mathbb{U}(o,r)$ and $z \in \mathbb{B}(o,r)$, if $\gamma \in \Gamma_{yz}$ is minimal, then $\gamma([0,1)) \subseteq \mathbb{U}(o,r)$.*

(iii) *For any $z \in \mathbb{B}(o,r)$, there exists $\bar{s} > 0$ such that*

$$\exp_z(-s\nabla f_p(z)) \in \mathbb{U}(o,r) \quad \text{for any } s \in (0, \bar{s}]; \tag{4.7}$$

in particular, $\nabla f_p(z) \neq 0$ if $z \in \partial\mathbb{B}(o,r)$.

Proof. Note by assumption (4.6) that (4.5) holds. Thus the inclusion $\mathbb{B}(o,r) \subset D$ is clear because, for each $i \in I$, $y_i \in \mathbb{U}(o,r)$ (and so $\mathrm{d}(o,y_i) < r$), and

$$\mathrm{d}(x,y_i) < 2r \le r_{\rm inj}(\mathbb{B}(o,r)) \le r_{\rm inj} y_i \quad \text{for any } x \in \mathbb{B}(o,r),$$

where the second inequality is true by (4.5).

(i) Since (4.5) holds as noted earlier, the strongly convexity of $\mathbb{U}(o,r)$ follows from Lemma 4.1. Furthermore, assume that $r < \frac{1}{2}\min\{r_{\rm inj}(\mathbb{B}(o,r)), \frac{\pi}{\sqrt{\Delta_{\mathbb{B}(o,r)}}}\}$. Then there exists $r' > r$ such that (4.5) holds with r' in place of r. Thus $\mathbb{B}(o,r) \subseteq \mathbb{U}(o,r')$, and $\mathbb{U}(o,r')$ is strongly convex by Lemma 4.1. Noting that $\mathbb{B}(o,r)$ is weakly convex, one checks by definition that $\mathbb{B}(o,r)$ is strongly convex.

(ii) Let $y \in \mathbb{U}(o,r)$. It is sufficient to show that y is a weak pole of $\mathbb{B}(o,r)$ in the sense that, for each $x \in \mathbb{B}(o,r)$, the minimal geodesic of M joining y to x is unique and lies in $\mathbb{B}(o,r)$. Granting this, the conclusion holds by [26, Proposition 4.3] (noting that the weakly convex set is locally convex). In the case when $r < \frac{1}{2}\min\{r_{\rm inj}(\mathbb{B}(o,r)), \frac{\pi}{\sqrt{\Delta_{\mathbb{B}(o,r)}}}\}$, we see from assertion (i) that $\mathbb{B}(o,r)$ is strongly convex;

hence y is clearly a weak pole of $\mathbb{B}(o,r)$. Thus we only need to consider the case when $r \le \varrho_p(r)$. To do this, let $x \in \mathbb{B}(o,r)$. Noting by assertion (i) that $\mathbb{B}(o,r)$ is weakly convex, one can choose a minimal geodesic γ joining y to x such that $\gamma \subset \mathbb{B}(o,r)$. Let w be the midpoint of γ. Note that the length $l(\gamma) < 2r$. One sees that $x, y \in \mathbb{U}(w,r)$, and $\mathbb{U}(w,r) \subset \mathbb{U}(o,2r)$. Since $r \le \varrho_p(r)$, it follows that

$$r \le \frac{1}{2}\min\left\{ r_{\rm inj}(\mathbb{B}(o,2r)), \frac{\pi}{\sqrt{\Delta_{\mathbb{B}(o,2r)}}} \right\} \le \frac{1}{2}\min\left\{ r_{\rm inj}(\mathbb{B}(w,r)), \frac{\pi}{\sqrt{\Delta_{\mathbb{B}(w,r)}}} \right\}.$$

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

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

$$\exp_z(s\|\exp_z^{-1} y_i\| V_i) \in \mathbb{U}(o,r) \quad \text{for any } 0 < s < 1.$$

Thus, applying [16, Lemma H.18] to $\{V_1, V_2\}$ (with $\mathbb{B}(o,r)$ in place of C), one can conclude that there exists $s_1 > 0$ such that

$$\exp_z s(\lambda_1 V_1 + \lambda_2 V_2) \in \mathbb{U}(o,r) \quad \text{for } 0 < s \le s_1$$

and then, by mathematical induction, that there exists $\bar{s} > 0$ such that

$$\exp_z s \sum_{i\in I} \lambda_i V_i \in \mathbb{U}(o,r) \quad \text{for any } 0 < s \le \bar{s},$$

where each $\lambda_i := w_i \mathrm{d}^{p-1}(z, y_i)$. Taking into account that $-\nabla f_p(z) = \sum_{i\in I} \lambda_i V_i$ by (4.3) (noting that $I_z = I$, thanks to inclusion $\{y_i\} \subset \mathbb{U}(o,r)$ and $z \in \partial\mathbb{B}(o,r)$), we conclude that (4.7) holds. This particularly implies $\nabla f_p(z) \ne 0$, completing the proof. ∎

For the remainder, let $\rho \in (0, +\infty]$ be such that

$$\rho \le \varrho_p(\rho) \quad \text{and} \quad \{y_i\} \subset \mathbb{U}(o,\rho). \tag{4.8}$$

For convenience, we set $D_\rho := \mathbb{U}(o,\rho)$ and consider the following optimization problem:

$$\min_{x\in M} (f_p + \delta_{D_\rho})(x). \tag{4.9}$$

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:

$$\min_{x\in M} f_p(x) < \min_{i\in I} f_p(y_i). \tag{4.10}$$

Proposition 4.3. *Assume that* (4.8) *holds and that* $\{y_i\}$ *is not colinear if* $p = 1$. *Then, the following statements hold:*

(i) $\{y_i\}$ *has the unique Riemannian* L^p *center of mass* $\bar{x}_p$*, which lies in* $\mathbb{U}(o,\rho)$ *and is the unique critical point*[3] *of* f_p *in* $\mathbb{U}(o,\rho)$*.*

(ii) ∇f_p *does not vanish on* $\partial\mathbb{U}(o,\rho)$*.*

(iii) *If* $\bar{x}_p \in \mathbb{U}(o,\rho)$ *is a critical point of* f_p*, then it has the following properties (and so* (3.25) *holds with* $\bar{x}_p$ *in place of* $\bar{x}$*):*

(iii-a) $\bar{x}_p$ *is a nondegenerate critical point of* f_p *if* (4.10) *is additionally assumed for* $p \in [1,2)$*.*

(iii-b) $\bar{x}_p$ *is a local weak sharp minimizer of order* 2 *for problem* (4.9) *if* $p \in (1,2)$*.*

(iii-c) f_p *is convex around* $\bar{x}_p$ *(and indeed strictly convex on* $\mathbb{U}(o,\rho)$ *if* $p \in [1,2)$*).*

Proof. (i) By assumption (4.8), [4, Theorem 2.1 and Remark 2.5] is applicable and $\{y_i\}$ has the unique Riemannian L^p center of mass $\bar{x}_p$, which lies in $\mathbb{U}(o,\rho)$, and is the unique critical point of f_p in $\mathbb{U}(o,\rho)$ (noting that f_1 is strictly convex on $\mathbb{U}(o,\rho)$ as $\{y_i\}$ is not colinear; see [4, Theorem 2.1 and Remark 2.5]).

(ii) With ρ in place of r, one sees that (4.6) holds by assumption (4.8). The assertion is immediate from Lemma 4.2(iii) (with $\bar{x}_p$ in place of z).

(iii) To show assertion (iii), we assume that $\bar{x}_p \in \mathbb{U}(o,\rho)$ is a critical point of f_p.

(iii-a) In the case when $p \in [2,+\infty)$, we see from Remark 4.1 that f_p is C^2 on D; thus it is sufficient to check that $\nabla^2 f_p(\bar{x}_p)$ is positive definite, which is implicitly contained in the proof of [4, Theorem 2.1]; see [4, (2.19)] and the argument in the paragraph from page 663, line 3, to page 665, line 5, there. Thus, we only need to consider the case when $p \in [1,2)$. To proceed, assume that $p \in [1,2)$ and that (4.10) holds. Then, $\bar{x}_p \in \mathbb{B}(o,\rho)\setminus\{y_i\}$. By Remark 4.1, f_p is C^2 on $D\setminus\{y_i\} \supseteq \mathbb{B}(o,\rho)\setminus\{y_i\}$. Thus we shall complete the proof by showing that $\nabla^2 f_p(x)$ is positive definite for each $x \in \mathbb{B}(o,\rho)\setminus\{y_i\}$. To do this, let $x \in \mathbb{B}(o,\rho)\setminus\{y_i\}$, and let $\gamma(\cdot)$ be a unit speed geodesic with $\gamma(0) = x$. Then, $\gamma([-\epsilon,\epsilon]) \subseteq \mathbb{B}(o,\rho)\setminus\{y_i\}$ for some $\epsilon > 0$. It suffices to verify that

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}(f_p\circ\gamma)(0) = \frac{\mathrm{d}^2}{\mathrm{d}t^2} f_p(\gamma(t))|_{t=0} > 0. \tag{4.11}$$

To show this, let $i \in I$, and let α_i denote the angle at x between the geodesic γ and the unique minimal geodesic joining x to y_i. Since

$$\rho \le \varrho_p(\rho) \le \frac{1}{2}\min\left\{r_{\mathrm{inj}}(\mathbb{B}(o,\rho)), \frac{\pi}{\sqrt{\Delta_{\mathbb{B}(o,\rho)}}}\right\}$$

by assumption (4.8) and the definition of ϱ_p, it follows that

$$\mathrm{d}(x,y_i) < 2\rho \le \min\left\{r_{\mathrm{inj}}(\mathbb{B}(o,\rho)), \frac{\pi}{\sqrt{\Delta_{\mathbb{B}(o,\rho)}}}\right\}. \tag{4.12}$$

Then, by the argument used for proving [4, (2.3)] (see also [33, pp. 153–154], applied to $\mathbb{U}(o,\rho)$ in place of M), one has that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{d}(\gamma(t),y_i)|_{t=0} = \cos\alpha_i \quad \text{and} \quad \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathrm{d}(\gamma(t),y_i)|_{t=0} \ge c_{\Delta_{\mathbb{B}(o,2\rho)}}(\mathrm{d}(x,y_i))\sin^2\alpha_i, \tag{4.13}$$

[3] In the case when $p = 1$ and $\bar{x}_p = y_i$ for some i, $\bar{x}_p$ being a critical point means that $0 \in \partial(f_p + \delta_{D_\rho})(\bar{x}_p)$, the subdifferentials of f_p in the sense of convex analysis; see, e.g., [25].

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 $t \in (-\epsilon, \epsilon)$,

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} f_p(\gamma(t)) = \sum_{i \in I} w_i \left((p-1)\mathrm{d}^{p-2}(x, y_i) \left(\frac{\mathrm{d}}{\mathrm{d}t} \mathrm{d}(\gamma(t), y_i) \right)^2 + \mathrm{d}^{p-1}(x, y_i) \frac{\mathrm{d}^2}{\mathrm{d}t^2} \mathrm{d}(\gamma(t), y_i) \right),$$

it follows from (4.13) that

$$\begin{aligned} &\frac{\mathrm{d}^2}{\mathrm{d}t^2}(f_p \circ \gamma)(0) \\ &\qquad \geq \sum_{i \in I} w_i \left((p-1)\mathrm{d}^{p-2}(x, y_i) \cos^2 \alpha_i + \mathrm{d}^{p-1}(x, y_i) c_{\Delta_{\mathbb{B}(o,2\rho)}}(\mathrm{d}(x, y_i)) \sin^2 \alpha_i \right). \end{aligned} \tag{4.14}$$

Note by (4.12) that $0 < \mathrm{d}(x, y_i) < \frac{\pi}{2\sqrt{\Delta_{\mathbb{B}(o,2\rho)}}}$, and then $c_{\Delta_{\mathbb{B}(o,2\rho)}}(\mathrm{d}(x, y_i)) > 0$ by definition. Thus, (4.11) 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 (4.11) follows from (4.14) as

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}(f_p \circ \gamma)(0) \geq \sum_{i \in I} w_i c_{\Delta_{\mathbb{B}(o,2\rho)}}(\mathrm{d}(x, y_i)) \sin^2 \alpha_i \geq w_{i_0} c_{\Delta_{\mathbb{B}(o,2\rho)}}(\mathrm{d}(x, y_{i_0})) \sin^2 \alpha_{i_0} > 0.$$

Therefore, (4.11) is valid for any $p \in [1, 2)$, completing the proof of assertion (iii-a).

(iii-b) Assume $p \in (1, 2)$. In light of assertion (iii-a), we only need to consider the case when (4.10) 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 data 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, (4.8) remains true for the data 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,

$$\tilde{f}_p(\cdot) := \frac{1}{p} \sum_{i \in \tilde{I}} \tilde{w}_i \mathrm{d}^p(\cdot, y_i) = \frac{1}{1 - w_{i_0}} \left(f_p(\cdot) - \frac{w_{i_0}}{p} \mathrm{d}^p(\cdot, y_{i_0}) \right). \tag{4.15}$$

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 (4.10) holds with $\tilde{f}_p$, $\tilde{I}$ 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 (4.9) with $\tilde{f}_p$ in place of f_p: there exist $\delta, \alpha > 0$ such that

$$\alpha \mathrm{d}^2(x, \bar{x}_p) \leq \tilde{f}_p(x) - \tilde{f}_p(\bar{x}_p) \quad \text{for any } x \in \mathbb{B}(\bar{x}, \delta).$$

Since $\tilde{f}_p(\bar{x}_p) = \frac{1}{1 - w_{i_0}} f_p(\bar{x}_p)$ and $\tilde{f}_p(\cdot) \leq \frac{1}{1 - w_{i_0}} f_p(\cdot)$ on $\mathbb{B}(\bar{x}, \delta)$ by (4.15), it follows that

$$\alpha(1 - w_{i_0})\mathrm{d}^2(x, \bar{x}_p) \leq (1 - w_{i_0})(\tilde{f}_p(x) - \tilde{f}_p(\bar{x}_p)) \leq f_p(x) - f_p(\bar{x}_p) \text{ for any } x \in \mathbb{B}(\bar{x}_p, \delta).$$

Therefore $\bar{x}_p$ is a local weak sharp minimizer of order 2 for problem (4.9), establishing assertion (iii-b).

(iii-c) It follows from assertion (iii-a) for $p \in [2, +\infty)$ and from [4, Theorem 2.1, Remark 2.5] for $p \in [1, 2)$. The proof is complete. ∎

Remark 4.2. Note that there may be multiple such balls $\mathbb{U}(o,\rho)$ with different center o and radius ρ satisfying (4.8). Since any local minimizer $\bar{x}_p$ of f_p in $\mathbb{U}(o,\rho)$ is a critical point, it follows from Proposition 4.3(i) that $\bar{x}_p$ is the unique Riemannian L^p center of mass. This implies that f_p has the unique local minimizer in the union of all such balls.

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

$$\mathcal{D}^1(f_p+\delta_{D_\rho})=\begin{cases}\mathbb{U}(o,\rho)\setminus\{y_i\} & \text{if } p=1,\\ \mathbb{U}(o,\rho) & \text{if } p\in(1,+\infty)\end{cases} \tag{4.16}$$

(recalling that $D_\rho=\mathbb{U}(o,\rho)$). Furthermore, it is clear that

$$\nabla f_p \text{ is locally Lipschitz continuous on } \mathcal{D}^1(f_p+\delta_{D_\rho}). \tag{4.17}$$

THEOREM 4.4. *Assume that* (4.8) *holds and that* $\{y_i\}$ *is not colinear and satisfies* (4.10) *for* $p=1$. *Let* $\{x_k\}$ *be a sequence generated by Algorithm* 3.1 *for solving problem* (4.9) *with initial point* $x_0\in\mathbb{U}(o,\rho)$, *and suppose that the step size sequence* $\{t_k\}$ *has a positive lower bound:* $\inf\{t_k\}>0$. *Then,* $\{x_k\}$ *converges to a point* $\bar{x}_p$, *which is the unique Riemannian* L^p *center of mass of* $\{y_i\}$; *moreover the convergence rate is at least linear.*

Proof. In view of Algorithm 3.1, $\{x_k\}$ is contained in $\mathbb{U}(o,\rho)$. Then $\nabla(f_p+\delta_{D_\rho})(x_k)=\nabla f_p(x_k)$ for each k. Moreover, $\{x_k\}$ has at least a cluster point, say, $\bar{x}_p$, and $\bar{x}_p\in\mathbb{B}(o,\rho)$. Recalling from Remark 3.2(a) that $t_k\|\nabla f_p(x_k)\|=t_k\|\nabla(f_p+\delta_{D_\rho})(x_k)\|\to 0$, we see that $\lim_{k\to+\infty}\nabla f_p(x_k)=0$ as $\inf\{t_k\}>0$ by assumption. We claim that $\bar{x}_p\in\mathbb{U}(o,\rho)$. In fact, otherwise, $\bar{x}_p\in\partial\mathbb{U}(o,\rho)$. Then, $\bar{x}_p\in D\setminus\{y_i\}$, and ∇f_p is continuous at $\bar{x}_p$ by Remark 4.1 and so $\nabla f_p(\bar{x}_p)=0$, which is a contradiction by Proposition 4.3(i), and the claim stands. Thus, if $p\in(1,+\infty)$ or $\bar{x}_p\notin\{y_i : i\in I\}$, one has that $\nabla f_p(\bar{x}_p)=0$, thanks to (4.16) and (4.17); this shows that $\bar{x}_p$ is a critical point of f_p. Now, consider the case when $p=1$ and $\bar{x}_p=y_i$ for some i. Note by Proposition 4.3(iii-c) that $f_1+\delta_{D_\rho}$ is convex, and note also that $\partial(f_1+\delta_{D_\rho})$ is upper Kuratowski semicontinuous (see [25, Proposition 6.2]). One checks that $0\in\partial(f_1+\delta_{D_\rho})(\bar{x}_1)$. This means that $\bar{x}_p$ is also a critical point of f_p in this case. Thus, applying Proposition 4.3(i), we get that $\bar{x}_p$ is the unique Riemannian L^p center of mass of $\{y_i\}$. This, together with (4.10), implies that $\bar{x}_p\notin\{y_i\}$ in the case when for $p=1$, and then ∇f_p is continuous at $\bar{x}_p$ (see (4.17)). Moreover, f_p is convex around $\bar{x}_p$ by Proposition 4.3(iii-c). Thus (3.7) is checked to hold with $\bar{x}_p$ in place of $\bar{x}$. Hence Theorem 3.6(i) is applicable to concluding that $\{x_k\}$ converges to $\bar{x}_p$. Furthermore, one sees that (3.25) holds too (with $\bar{x}_p$ in place of $\bar{x}$) by Proposition 4.3(iii) (noting that each nondegenerate critical point of f_p is a local weak sharp minimizer of order 2 for problem (4.9)). Thus, the convergence rate is at least linear by Theorem 3.6(ii) (noting $\inf\{t_k\}>0$ as assumed). The proof is complete. ∎

COROLLARY 4.5. *Assume that* (4.8) *holds and that* $\{y_i\}$ *is not colinear for* $p=1$. *Let* $x_0\in\mathbb{U}(o,\rho)$ *and suppose for* $p=1$ *that*

$$f_p(x_0)<\min_{i\in I} f_p(y_i). \tag{4.18}$$

Then, Algorithm 3.1 *for solving problem* (4.9) *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* $\{y_i\}$. *Moreover, the convergence rate is at least linear if* (4.10) *is additionally assumed for* $p\in[1,2)$.

Proof. By (4.16), one sees that $\mathcal{D}^1(f_p + \delta_{D_\rho}) = \mathcal{D}(f_p + \delta_{D_\rho}) = \mathbb{U}(o,\rho)$ in the case when $p \in (1, +\infty)$; thus the first conclusion regarding the well definedness of Algorithm 3.1 follows directly from Remark 3.1. Below we consider the case when $p = 1$. To do this, in view of (4.18) and (4.16), one applies Remark 3.1 inductively to check that each generated point x_k satisfies

$$x_k \in L^0_{f_1+\delta_{D_\rho}} \subset \mathbb{U}(o,\rho) \setminus \{y_i\} = \mathcal{D}^1(f_1 + \delta_{D_\rho}) \tag{4.19}$$

(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 other conclusions regarding the convergence and the convergence rate, we note first that $\{x_k\}$ has a cluster point, say, $\bar{x}_p \in \mathbb{B}(o,\rho)$ (as $\{x_k\} \subseteq \mathbb{U}(o,\rho)$). Note by (4.19) that $\bar{x}_p \notin \{y_i\}$ in the special case when $p = 1$. One sees that $\bar{x}_p \in \mathcal{D}^1(f_p + \delta_{D_\rho}) \cup \partial\mathbb{U}(o,\rho)$ (for all $p \in [1, +\infty)$). Below we show that $\bar{x}_p \in \mathcal{D}^1(f_p + \delta_{D_\rho})$. Granting this, we have by (4.17) that $\nabla(f_p + \delta_{D_\rho})$ is continuous at $\bar{x}_p$. This, together with Remark 3.2(b), implies that $\nabla(f_p + \delta_{D_\rho})(\bar{x}_p) = 0$, and so $\nabla(f_p)(\bar{x}_p) = 0$ (as $\bar{x}_p \in \mathcal{D}^1(f_p + \delta_{D_\rho}) \subseteq \mathbb{U}(o,\rho)$). Hence, by Proposition 4.3(iii-c), (3.7) is seen to hold with $\bar{x}_p$ in place of $\bar{x}$, and Theorem 3.6(i) is applicable to showing that $\{x_k\}$ converges to $\bar{x}_p$. Moreover it follows from Proposition 4.3(i) that $\bar{x}_p$ is the unique Riemannian L^p center of mass of $\{y_i\}$.

Now, assume that (4.10) holds additionally for $p \in [1, 2)$. To show the linear convergence property, we grant that $\bar{x}_p \in \mathcal{D}^1(f_p+\delta_{D_\rho})$ again. Then, as noted in (4.17), $\nabla(f_p + \delta_{D_\rho})$ is Lipschitz continuous around $\bar{x}_p$. Since (3.7) holds with $\bar{x}_p$ in place of $\bar{x}$, it follows from Lemma 3.4 (applied to $\bar{x}_p$ in place of $\bar{x}$) that the corresponding step size sequence $\{t_k\}$ has a positive lower bound, and so $\{x_k\}$ converges linearly to $\bar{x}_p$ thanks to Theorem 4.4.

Thus, to complete the proof, we only need to verify that $\bar{x}_p \in \mathcal{D}^1(f_p + \delta_{D_\rho})$. To do this, suppose on the contrary that $\bar{x}_p \in \partial\mathbb{U}(o,\rho)$. Then, by Proposition 4.3(ii), suppose that

$$\nabla f_p(\bar{x}_p) \neq 0, \tag{4.20}$$

and it follows from Lemma 4.2(iii) (applied to $\bar{x}_p$, ρ in place of z, r) that there exists $\bar{s} > 0$ such that

$$\exp_{\bar{x}_p}[-s\nabla f_p(\bar{x}_p)] \in \mathbb{U}(o,\rho) \quad \text{for any } 0 < s \leq \bar{s}. \tag{4.21}$$

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

$$\exp_x[-s\nabla f_p(x)] \in \mathbb{U}(o,\rho) \quad \text{for any } x \in \mathbb{B}(\bar{x}_p, \delta_0) \text{ and } 0 < s \leq \bar{s} \tag{4.22}$$

(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 $\bar{z} \in \mathbb{U}(o,\rho)$ by (4.21); hence there is $\bar{\varepsilon} > 0$ such that $\mathbb{B}(\bar{z}, \bar{\varepsilon}) \subset \mathbb{U}(o,\rho)$. Without loss of generality, we may assume

$$\bar{s}\|\nabla f_p(\bar{x}_p)\| + \bar{\varepsilon} + \bar{\delta} \leq r_{\rm cvx}(\mathbb{B}(\bar{x}_p, \bar{\delta})) \tag{4.23}$$

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

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

Let $x \in \mathbb{B}(\bar{x}_p, \delta_0)$ and write $z_x := \exp_x[-\bar{s}\nabla f_p(x)]$. Then, in view of (4.23), we check that

$$\mathrm{d}(x, z_x) \leq \mathrm{d}(x, \bar{x}_p) + \mathrm{d}(\bar{x}_p, \bar{z}) + \mathrm{d}(\bar{z}, z_x) \leq r_{\mathrm{cvx}}(\mathbb{B}(\bar{x}_p, \delta_1)) \leq r_{\mathrm{cvx}}(x).$$

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 (4.22) holds as $\mathbb{U}(o, \rho)$ is strongly convex.

To proceed, let $\{x_{k_j}\}$ be a subsequence of $\{x_k\}$ converging to $\bar{x}_p$. Then, $\lim_{j\to+\infty} t_{k_j} = 0$ by (3.6). Thus, without loss of generality, we may assume that

$$x_{k_j} \in \mathbb{B}(\bar{x}_p, \delta_0) \quad \text{and} \quad 2t_{k_j} \leq \bar{s} \quad \text{for each } j. \tag{4.24}$$

Fix j and recall that the geodesic γ_{k_j} is defined by (3.1). Then, in view of (4.22) and (4.24), we see that

$$\gamma_{k_j}(s) = \exp_{x_{k_j}}[-s\nabla f_p(x_{k_j})] \in \mathbb{U}(o, \rho) \quad \text{for each } s \in [0, 2t_{k_j}].$$

By using the mean value theorem, there is $\bar{t}_{k_j} \in (0, 2t_{k_j})$ such that

$$\frac{f_p(\gamma_{k_j}(2t_{k_j})) - f_p(x_{k_j})}{-2t_{k_j}} = \langle P_{\gamma_k, x_{k_j}, \gamma_{k_j}(\bar{t}_{k_j})} \nabla f_p(\gamma_{k_j}(\bar{t}_{k_j})), \nabla f_p(x_{k_j})\rangle.$$

This, together with (3.3), implies that

$$\langle P_{\gamma_{k_j}, x_k, \gamma_{k_j}(\bar{t}_{k_j})} \nabla f_p(\gamma(\bar{t}_{k_j})), \nabla f_p(x_{k_j})\rangle \leq \beta \|\nabla f_p(x_{k_j})\|^2.$$

Passing to the limit as $j \to \infty$, we arrive at $\beta \geq 1$ by (4.20), which is a contradiction. Thus, the proof is complete. ∎

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

ALGORITHM 4.1. *Give $x_0 \in M$, $t_0 \in (0, +\infty)$ and set $k := 0$.*
Step 1. *If $\nabla f_p(x_k) = 0$ or $x_k \notin \mathcal{D}^1(f_p)$, then stop; otherwise construct γ_k as* (3.1).
Step 2. *Set $x_{k+1} := \gamma_k(t_0)$, replace k by $k+1$, and go to Step* 1.

Let $x_0 \in \mathbb{U}(o, \rho)$, and we need the following assumption:

$$L^0_{f_p} \subset \mathbb{U}(o, \rho), \tag{4.25}$$

where, as done in section 3, $L^0_{f_p} := L_{f_p}(f_p(x_0))$ is the sublevel set. Moreover, we need also the following assumption made for $p \in [1, 2)$:

$$f_p(x_0) < \min_{i \in I} f_p(y_i). \tag{4.26}$$

Thus, under assumption (4.25), and assumption (4.26) (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 finite (as $L^0_{f_p}$ is compact). Recall that $D_\rho = \mathbb{U}(o, \rho)$.

COROLLARY 4.6. *Assume that* (4.8) *holds and that $\{y_i\}$ is not colinear if $p = 1$. Let $x_0 \in \mathbb{U}(o, \rho)$ be such that* (4.25) *holds and that* (4.26) *holds for $p \in [1, 2)$. Then, Algorithm* 4.1 *for solving problem* (4.1) *with $t_0 \in (0, \frac{2}{\lambda_p(x_0)})$ 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 $L^0_{f_p} \subset \mathcal{D}^1(f_p + \delta_{D_\rho})$. 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 $\gamma_j : [0, +\infty) \to M$ be the geodesic defined by (3.1), and set $\bar{t} := \sup\{t : \gamma_j(s) \in L^0_{f_p}$ for any $0 \le s \le t\}$. 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 $t \in (0, \bar{t}]$,

$$\begin{aligned} f_p(\gamma_j(t)) &= f_p(x_j) - t\|\nabla f_p(x_j)\|^2 + t^2 \textstyle\int_0^1 (1-\tau)\langle \nabla^2 f_p(\gamma_j(\tau t))\nabla f_p(x_j), \nabla f_p(x_j)\rangle d\tau \\ &\le f_p(x_j) - t\|\nabla f_p(x_j)\|^2 + \tfrac{t^2\lambda_p(x_0)}{2}\|\nabla f_p(x_j)\|^2. \end{aligned}$$

Noting $f_p(x_j) \le f_p(x_0)$, it follows that

$$f_p(\gamma_j(t)) \le f_p(x_0) - t\left(1 - \frac{t\lambda_p(x_0)}{2}\right)\|\nabla f_p(x_j)\|^2 \quad \text{for each } t \in [0, \bar{t}]. \tag{4.27}$$

This implies that $\bar{t} \ge \frac{2}{\lambda_p(x_0)} > t_0$ 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 we desired to show. Furthermore, (4.27) implies that the generated sequence $\{x_k\}$ by Algorithm 4.1 satisfies

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

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 (4.9) with initial point x_0 and constant step sizes $\{t_k := t_0\}$. Note that $\{x_k\} \subset L^0_{f_p}$ and $L^0_{f_p} \subset \mathcal{D}^1(f_p + \delta_{D_\rho})$. Furthermore, (4.26) particularly implies (4.10). Thus, Theorem 4.4 is applicable and $\{x_k\}$ converges linearly to the unique Riemannian L^p center of mass of $\{y_i\}$. The proof is complete. ∎

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 $r_{cx} := \frac{1}{2}\min\{r_{\text{inj}}(M), \frac{\pi}{\sqrt{\Delta_M}}\}$, which particularly implies the following assumption (4.28) with r_{cx} in place of ρ.

COROLLARY 4.7. *Assume that*

$$\rho \le \varrho_p(\rho) \quad \textit{and} \quad \{y_i\} \subset \mathbb{U}\left(o, \frac{1}{3}\rho\right), \tag{4.28}$$

and $\{y_i\}$ *is not colinear if* $p = 1$. *Let* $x_0 \in \mathbb{U}(o, \frac{1}{3}\rho)$ *be such that* (4.26) *holds for* $p \in [1,2)$. *Then, Algorithm* 4.1 *for solving problem* (4.1) *with initial point* x_0 *and* $t_0 \in (0, \frac{2}{\lambda_p(x_0)})$ *is well defined and converges linearly to the unique Riemannian* L^p *center of mass of* $\{y_i\}$.

Proof. Note that (4.8) holds by (4.28). To apply Corollary 4.6, we only need to show (4.25). To do this, let $z \in M \setminus \mathbb{U}(o, \rho)$. Then, we have by (4.28) and the choice of x_0 that $\mathrm{d}(x_0, y_i) < \frac{2\rho}{3} < \mathrm{d}(z, y_i)$ for each $i \in I$; hence $f_p(x_0) < f_p(z)$ by definition (see (4.2)). This means that $z \notin L^0_{f_p}$, establishing (4.25) as $z \in M \setminus \mathbb{U}(o, \rho)$ is arbitrary. Thus, Corollary 4.6 is applicable to completing the proof. ∎

In the special case when M is a Hadamard manifold, one checks by definition (see (4.4)) that $\varrho_p(r) = +\infty$ for each $r > 0$. Then, we can choose that $\rho := +\infty$ so that

(4.8) and (4.28) hold trivially. Thus, Corollary 4.8 follows directly from Corollaries 4.5 and 4.7. In particular, Corollary 4.8 extends/improves the corresponding ones in [9, Theorems 3.1 and 3.2], where only the convergence property of the generated sequence $\{x_k\}$ is established for the special cases when $p = 1$ and when $p = 2$.

COROLLARY 4.8. *Assume that M is a Hadamard manifold and $\{y_i\}$ is not colinear for $p = 1$, and let $x_0 \in M$. Then, the following assertions hold:*

(i) *If* (4.26) *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* (4.10) *is additionally assumed for $p \in (1, 2)$.*

(ii) *If* (4.26) *holds for $p \in [1, 2)$ and $t_0 \in (0, \frac{2}{\lambda_p(x_0)})$, 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 $\{y_i\}$.*

We end this section with the following remark regarding the discussion on why to use $\mathbb{B}(o, 2r)$ and the possibility of using $\mathbb{B}(o, r)$ in the definition of $\varrho_p(r)$ in (4.4). Consider

$$\varrho_p(r) := \begin{cases} \frac{1}{2}\min\{r_{\mathrm{inj}}(\mathbb{B}(o,r)), \frac{\pi}{2\sqrt{\Delta_{\mathbb{B}(o,r)}}}\} & \text{if } 1 \le p < 2; \\ \frac{1}{2}\min\{r_{\mathrm{inj}}(\mathbb{B}(o,r)), \frac{\pi}{\sqrt{\Delta_{\mathbb{B}(o,r)}}}\} & \text{if } 2 \le p < +\infty, \end{cases} \tag{4.29}$$

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

Remark 4.3. If one uses the function $\varrho_p(\cdot)$ defined by (4.29), then, under the assumption made in Proposition 4.3, claim (iii) there is still true (noting that the comparison theorem on $\mathbb{B}(o, r)$ is used in the proof), but whether claim (i) remains true is an open question as remarked in [4, Remark 2.5] (as the comparison theorem on $\mathbb{B}(o, 2r)$ is required for the proof); moreover, we also don't know whether claim (ii) remains true. Thus, in this case, we don't know whether the assumption made in Proposition 4.3 guarantees the following assertions:

(a) Any critical point in $\mathbb{U}(o, \rho)$ of f_p is a Riemannian center of mass of $\{y_i\}$.

(b) $\{x_k\}$ has a cluster point $\bar{x}_p$ in $\mathbb{U}(o, \rho)$ satisfying $\nabla f_p(\bar{x}_p) = 0$.

In particular, assertion (b) plays a crucial role in establishing convergence results of Algorithm 3.1. Therefore, in the case when the function $\varrho_p(\cdot)$ is defined by (4.29), we don't know whether our results (i.e., Theorem 4.4 and Corollaries 4.5–4.7) remain true. However, if the strict inequality of (4.8) holds, $\rho < \varrho_p(\rho)$, and assumption (4.10) is replaced by (4.18) when $p = 1$, then our arguments (using (4.18) instead of (4.10) when $p = 1$) for analyzing the convergence properties do work for the case when the function $\varrho_p(\cdot)$ is defined by (4.29) (noting that assertion (b) above is true thanks to Lemma 4.2), and so Theorem 4.4 and Corollaries 4.5–4.7 remain true in this case with $\bar{x}_p$ being a critical point of f_p, rather than a Riemannian center of mass.

Acknowledgments. The authors are indebted to the handling editor and two anonymous referees for their very helpful remarks and comments, which allowed us to improve the original presentation.

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, NJ, 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 Process Appl., 122 (2012), pp. 1437–1455.

[7] G. C. Bento, J. X. Cruz Neto, and P. R. Oliveira, *Convergence of Inexact Descent Methods for Nonconvex Optimization on Riemannian Manifolds*, arXiv:1103.4828v1, 2011.

[8] 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.

[9] G. C. Bento, S. D. B. Bitar, J. X. Cruz Neto, P. R. Oliveira, and J. C. O. Souza, *Computing Riemannian center of mass on Hadamard manifolds*, J. Optim. Theory Appl., 183 (2019), pp. 977–992.

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

[11] J. F. Bonnans and A. Ioffe, *Second-order sufficiency and quadratic growth for nonisolated minima*, Math. Oper. Res., 20 (1995), pp. 801–817.

[12] Yu. Burago, M. Gromov, and G. Perel'man, *A.D. Alexandrov spaces with curvature bounded below*, Uspeckhi Mat. Nank., 47 (1992), pp. 3–51.

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

[14] J. V. Burke and M. C. Ferris, *Weak sharp minima in mathematical programming*, SIAM J. Control Optim., 31 (1993), pp. 1340–1359.

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

[16] 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*, Math. Surveys Monographs 163, AMS, Providence, RI, 2010.

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

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

[19] O. P. Ferreira, M. S. Louzeiro, and L. F. Prudente, *Gradient method for optimization on Riemannian manifolds with lower bounded curvature*, SIAM J. Optim., 29 (2019), pp. 2517–2541.

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

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

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

[23] 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.

[24] C. Li and J. Wang, *Newton's method on Riemannian manifolds: Smale's point estimate theory under the γ-condition*, IMA J. Numer Anal., 26 (2006), pp. 228–251.

[25] 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.

[26] 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.

[27] D. G. Luenberger, *The gradient projection method along geodesics*, Management Sci., 18 (1972), pp. 620–631.

[28] E. Massart and P. A. Absil, *Quotient geometry with simple geodesics for the manifold of fixed-rank positive-semidefinite matrices*, SIAM J. Matrix Anal. Appl., 41 (2020), pp. 171–198.

[29] S. Z. Németh, *Five kinds of monotone vector fields*, Pure Math. Appl., 9 (1998), pp. 417–428.

[30] 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.

[31] 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.

[32] P. Pertersen, *Riemannian Geometry*, Springer-Verlag, New York, 2006.
[33] T. Sakai, *Riemannian Geometry*, Transl. Math. Monogr., AMS, Providence, 1996.
[34] S. T. Smith, *Geometric Optimization Methods for Adaptive Filtering*, Ph.D. thesis, Harvard University, Cambridge, MA, 1994.
[35] M. Studniarski and D. E. Ward, *Weak sharp minima: characterizations and sufficient conditions*, SIAM J. Control Optim., 38 (1999), pp. 219–236.
[36] X. Su, H. Sun, and Y. Wang, *Pythagorean Theorem & Curvature with Lower or Upper Bound*, https://arxiv.org/pdf/1911.01019.pdf, 2019.
[37] C. Udriste, *Convex Functions and Optimization Methods on Riemannian Manifolds*, Math. Appl., Kluwer, Dordrecht, 1994.
[38] 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), pp. 553–573.
[39] 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.
[40] X. M. Wang, *Subgradient algorithms on Riemannian manifolds of lower bounded curvatures*, Optimization, 67 (2018), pp. 179–194.
[41] X. M. Wang, C. Li, J. H. Wang, and J. C. Yao, *Linear convergence of subgradient algorithm for convex feasibility on Reimannian manifolds*, SIAM J. Optim., 25 (2015), pp. 2334–2358.
[42] 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.
[43] D. E. Ward, *Characterizations of strict local minima and necessary conditions for weak sharp minima*, J. Optim. Theory Appl., 80 (1994), pp. 551–571.
[44] T. Yamaguchi, *Locally geodesically quasiconvex functions on complete Riemannian manifolds*, Trans. Amer. Math. Soc., 298 (1986), pp. 307–330.
[45] L. Yang, *Riemannian median and its estimation*, LMS J. Comput. Math., 13 (2010), pp. 461–479.
[46] 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.