Nonconvex Weak Sharp Minima on Riemannian Manifolds

Mohammad Mahdi Karkhaneei · Nezam Mahdavi-Amiri

Received: 18 August 2018 / Accepted: 8 May 2019 / Published online: 25 June 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
We establish some necessary conditions (of the primal and dual types) for the set of weak sharp minima of a nonconvex optimization problem on a Riemannian manifold. Here, we provide a generalization of some characterizations of weak sharp minima for convex problems on Riemannian manifold introduced by Li et al. (SIAM J Optim 21(4):1523–1560, 2011) for nonconvex problems. We use the theory of the Fréchet and limiting subdifferentials on Riemannian manifold to give some necessary conditions of the dual type. We also consider a theory of contingent directional derivative and a notion of contingent cone on Riemannian manifold to give some necessary conditions of the primal type. Several definitions have been provided for the contingent cone on Riemannian manifold. We show that these definitions, under some modifications, are equivalent. We establish a lemma about the local behavior of a distance function. We use this lemma to establish some necessary conditions by expressing the Fréchet sub-differential (contingent directional derivative) of a distance function on a Riemannian manifold in terms of normal cones (contingent cones). As an application, we show how one can use weak sharp minima property to model a Cheeger-type constant of a graph as an optimization problem on a Stiefel manifold.

Keywords  Weak sharp minima · Riemannian manifolds · Distance functions · Nonconvex functions · Generalized differentiation · Graph clustering

Mathematics Subject Classification 49J52 · 90C26 · 05C40

Communicated by Alexandru Kristály.

Nezam Mahdavi-Amiri
nezamm@sharif.edu
Mohammad Mahdi Karkhaneei
karkhaneei@gmail.com

1 Sharif University of Technology, Tehran, Iran
1 Introduction

In recent decades, extensive research has been carried out on optimization on manifolds. These studies scatter in various contexts such as convex [1], smooth [2–4], nonsmooth [5], and over special manifolds such as the manifold of positive definite matrices [6], Stiefel manifolds [7], etc. Particularly, special attention has been devoted to nonconvex analysis and optimization on Riemannian manifolds. Several problems in machine learning, pattern recognition, and computer vision can be modeled as nonconvex optimization problems on Riemannian manifolds [8–12]. Moreover, various numerical procedures for solving nonconvex optimization problems on Riemannian manifolds have been designed, such as line search and trust region methods along with Newton-like methods in smooth cases (see [2,3,13–15]), and subgradient descent, gradient sampling, and proximal algorithms in nonsmooth cases (see [16,17]).

Distance functions appear in various optimization methods, such as proximal point methods, penalty methods, etc. Generalized differential properties of distance functions play remarkable roles in variational analysis, optimization, and their applications. The authors of [18,19] investigated properties of generalized derivatives of distance functions in the linear space setting. Properties of distance functions on a Riemannian manifold are not trivially obtained by a generalization of the corresponding properties in the linear space setting. Li et al. [20] gave a relation for the subdifferential of a convex distance function in terms of a normal cone to the corresponding set in the Riemannian manifold setting with the nonpositive sectional curvature. They used a comparison result for geodesic triangles in a Riemannian manifold with the nonpositive sectional curvature as a global property. Subdifferentials and normal cones are local notions, and we do not require any global condition on a Riemannian manifold for investigating local properties of a distance function. Here, we establish a lemma about the local behavior of a distance function on a manifold (called local distance lemma). Using this lemma, we express the Fréchet subdifferential (contingent directional derivative) of a distance function on a Riemannian manifold in terms of normal cones (contingent cones).

Then, we investigate the notion of weak sharp minima for nonconvex optimization problems on Riemannian manifolds. We provide a generalization of some characterizations of weak sharp minima of [20] to the nonconvex optimization problems on a Riemannian manifold modeled in a Hilbert space. To the best of our knowledge, our work here is the first study concerning the notion of weak sharp minima for nonconvex optimization problems on Riemannian manifolds.

The concept of sharp minima was introduced by Polyak [21] in the case of finite-dimensional Euclidean space for sensitivity analysis of optimization problems and convergence analysis of some numerical algorithms. Next, Ferris [22] extended the notion of weak sharp minima to the situation where the optimization problem has multiple solutions. The concept was expanded by many authors for convex and nonconvex optimization problems over finite- and infinite-dimensional linear spaces (see [23–29]).

The primary goal of our work here is to present some primal and dual necessary conditions for the set of weak sharp minima of a nonconvex optimization problem on a Riemannian manifold. The key ingredient of deriving our necessary conditions
is the representation of a generalized derivative of a distance function in terms of a cone. We will use the Fréchet and limiting subdifferentials along with the Fréchet and limiting normal cones on Riemannian manifold to state some necessary conditions of the dual type for the set of weak sharp minima of a nonconvex optimization problem. To state some necessary conditions of a primal type for the set of weak sharp minima of a nonconvex optimization problem, we use a contingent directional derivative and a contingent cone on Riemannian manifold. Our introduced contingent directional derivative is closely related to the Fréchet subdifferential. Several definitions have been provided for the contingent cone on Riemannian manifold; see [30,31]. We will show that these definitions, with some modifications, are equivalent.

The remainder of our work is organized as follows: In Sect. 2, some needed preliminaries on linear spaces and metric spaces are given. Also, we recall some fundamental notions of variational analysis in the linear space setting. In Sect. 3, we first introduce some basic notions of Riemannian manifolds. Then, definitions and properties of the (limiting) Fréchet subdifferential and (limiting) normal cone on Riemannian manifold are provided. Also, we introduce contingent directional derivative on Riemannian manifold and recall some definitions of contingent cones on Riemannian manifold and show that these definitions, with some modifications, are equivalent.

In Sect. 4, we establish a local distance lemma and use it to attain a formula for the Fréchet subdifferential (the directional derivative) of a distance function on a Riemannian manifold in terms of the normal cone (the continent cone). In Sect. 5, we establish some necessary conditions for weak sharp minima on Riemannian manifold in the nonconvex case. In Sect. 6, we give an application of the results of the previous sections. In Sect. 7, we provide our concluding remarks.

2 Linear and Metric Spaces

We provide some useful definitions and symbols required by variational analysis and topology; the reader is referred to [32] for more details. We denote \( \mathbb{R} := \mathbb{R} \cup \{ \infty \} \) as the extended real numbers. Suppose that \( E \) is a Banach space. Then, \( E^*, B_E, \) and \( I_E \), respectively, denote dual space, closed unit ball, and identity function on \( E \). Let \( (M, d) \) be a metric space and \( A \subset M \). Recall that \( \text{cl} A \) stands for the closure of \( A \), and the distance function for the subset \( A \) is defined by \( \text{dist}(p; A) := \inf_{u \in M} d(p, u) \), for all \( p \in M \). The closed ball with center \( p \) and radius \( r > 0 \) is denoted by \( B(p, r) := \{ q \in M : d(p, q) \leq r \} \). Moreover, if \( f : M \to \mathbb{R} \) is a function on \( M \), then we denote \( \text{dom}(f) := \{ p \in M : f(p) < \infty \} \). We say that \( f \) is proper, if \( \text{dom}(f) \neq \emptyset \). Also, we say that \( f \) is lower semicontinuous (l.s.c.) at \( p \in M \), if \( f(p) \leq \lim \inf_{u \to p} f(u) \). We say that \( f \) is locally Lipschitz at \( p \in M \) with rate \( r \), if there exists a neighborhood of \( p \), say \( U \), such that \( |f(u) - f(p)| \leq rd(u, p) \), for all \( u \in U \). Furthermore, we say that \( f \) is a Lipschitzian function with rate \( r \) (around \( q \in M \)), if the earlier inequality holds for all \( p, u \in M \) (for all \( p \) and \( u \) in a neighborhood of \( q \)). Let \( \Omega \subset M \). The indicator function of \( \Omega \) is defined as \( \delta_{\Omega}(u) := 0 \), for \( u \in \Omega \), and \( \delta_{\Omega}(u) := \infty \), for \( u \notin \Omega \). Now, we recall definition of weak sharp minima on a metric space.
**Definition 2.1 (Weak sharp minima)** Let \( f : M \to \overline{\mathbb{R}} \) be a proper function on a metric space \( M \) and \( S \subset M \). We say that

(i) \( p \in \Omega \) (where \( \Omega := \text{argmin}_S f \)) is a local weak sharp minimizer for the problem \( \min_{u \in S} f(u) \) with modulus \( \alpha > 0 \), if there is \( \epsilon > 0 \) such that for all \( u \in S \cap B(p, \epsilon) \), we have

\[
f(u) \geq f(p) + \alpha \text{dist}(u; \Omega).
\]

(1)

(ii) \( \Omega := \text{argmin}_S f \) is the set of (global) weak sharp minima for the problem \( \min_{u \in S} f(u) \) with the modulus \( \alpha > 0 \), if (1) holds for all \( p \in \Omega \) and \( u \in S \).

In the sequel, we recall definitions of some basic notions of variational analysis in the linear space setting. Suppose that \( f : E \to \overline{\mathbb{R}} \) is a function on a Banach space \( E \) and is finite at \( p \in E \). The Fréchet subdifferential of \( f \) at \( p \) is defined to be

\[
\hat{\partial} f(p) := \left\{ x^* \in E^* : \liminf_{u \to p} \frac{f(u) - f(p) - \langle x^*, u - p \rangle}{\|u - p\|} \geq 0 \right\}.
\]

The limiting subdifferential of \( f \) at \( p \) is the set \( \partial f(p) \) of all \( x^* \in E^* \) such that there is a sequence of covectors \( x^*_i \in \hat{\partial} f(p_i) \) such that \( f(p_i) \to f(p) \), and \( x^*_i = x^* \). The contingent directional derivative of \( f : E \to \overline{\mathbb{R}} \) at \( p \) in the direction \( v \in E \) is defined to be

\[
f^{-}(p; v) := \liminf_{w \to v, t \downarrow 0} \frac{f(p + tw) - f(p)}{t}.
\]

(2)

Suppose that \( \Omega \) is a subset of the Banach space \( E \) and \( p \in \text{cl} \, \Omega \). The Fréchet normal cone of \( \Omega \) at \( p \) is defined to be

\[
\hat{N}_\Omega(p) := \left\{ x^* \in E^* : \limsup_{u \to p, u \to \Omega} \frac{\langle x^*, u - p \rangle}{\|u - p\|} \leq 0 \right\}.
\]

The limiting normal cone of \( \Omega \) at \( p \) is the set \( N_\Omega(p) \) of limits of sequences \( \{x^*_i\} \subset E^* \) for which there is a sequence \( \{p_i\} \subset \Omega \) converging to \( p \) such that \( x^*_i \in \hat{N}_\Omega(p_i) \), for all \( i \). The contingent cone of \( \Omega \) at \( p \) is defined to be

\[
\hat{T}_\Omega(p) := \{ v \in E : \exists v_i \in E, t_i \downarrow 0 \text{ such that } v_i \to v, p + t_i v_i \in \Omega \}.
\]

**Proposition 2.1** For a subset \( \Omega \) of \( E \), with \( E \) being a Banach space, its distance function \( \text{dist}(\cdot; \Omega) \), and \( p \in \text{cl} \, \Omega \), one has

\[
\hat{\partial} \text{dist}(p; \Omega) = \hat{N}_\Omega(p) \cap \mathbb{B}_{E^*}.
\]

(3)

The following theorem relates directional derivative of the distance function to contingent cone.

**Proposition 2.2** (Burke et al. [18, Theorem 4]) Suppose that \( E \) is a Banach space and \( p \in \Omega \subset E \). Then, for all \( v \in E \), we have

\[
\hat{\partial} \text{dist}(p; \Omega) = \hat{N}_\Omega(p) \cap \mathbb{B}_{E^*}.
\]
\[ d^-_\Omega(p; v) \leq \text{dist}\left(v; \hat{T}_\Omega(p)\right), \tag{4} \]

where \( d^-_\Omega(p; v) \) is the contingent directional derivative of \( \text{dist}(\cdot; \Omega) \) at \( p \) in direction \( v \). If it is further assumed that \( E \) is finite-dimensional, then equality holds in (4).

3 Variational Analysis on Riemannian Manifolds

Here, we recall some definitions and results about variational analysis on Riemannian manifolds which will be useful later on; see, e.g., [31,33] for more details. We will be dealing with functions defined on Riemannian manifolds (either finite- or infinite-dimensional). A Riemannian manifold \((\mathcal{M}, g)\) is a \( C^\infty \) smooth manifold \( \mathcal{M} \) modeled on some Hilbert space \( \mathbb{H} \) (possibly infinite-dimensional), such that for every \( p \in \mathcal{M} \) we are given a scalar product \( g(p) = g_p := \langle \cdot, \cdot \rangle_p \) on the tangent space \( T_p\mathcal{M} \cong \mathbb{H} \) so that \( \|x\|_p = \langle x, x \rangle_p^{1/2} \) defines an equivalent norm on \( T_p\mathcal{M} \) for each \( p \in \mathcal{M} \) and in such a way that the mapping \( p \in \mathcal{M} \rightarrow g_p \in S^2(\mathcal{M}) \) is a \( C^\infty \) section of the bundle of symmetric bilinear forms. For each \( p \in \mathcal{M} \), the metric \( g \) induces a natural isomorphism between \( T_p\mathcal{M} \) and \( T^*_p\mathcal{M} \). So, we define the norm on \( T^*_p\mathcal{M} \) as \( \|v^*\|^2_p = g_p(v, v) \). If a function \( f : \mathcal{M} \rightarrow \mathbb{R} \) is (Fréchet) differentiable at \( p \in \mathcal{M} \), then norm of the differential \( df(p) \in T^*_p\mathcal{M} \) at the point \( p \) is defined by \( \|df(p)\|_p := \sup\{df(p)(v): v \in T_p\mathcal{M}, \|v\|_p \leq 1\} \). Given two points \( p, q \in \mathcal{M} \), the Riemannian distance from \( p \) to \( q \) is denoted by \( d_{\mathcal{M}}(p, q) \). Throughout our work here, \( \mathcal{M} \) is a Riemannian manifold modeled on a Hilbert space.

Since Fréchet/limiting subdifferential and Fréchet/limiting normal cone are local notions in the linear space setting, we can define these notions on Riemannian manifold, using a local chart. These definitions are independent of the chosen chart; see, e.g., [31,33,34] for more details. Here, we define these concepts on Riemannian manifold by exponential charts. Suppose that \( f : \mathcal{M} \rightarrow \mathbb{R} \) is a function on a Riemannian manifold \( \mathcal{M} \) and is finite at \( p \in \mathcal{M} \). The Fréchet subdifferential and the limiting subdifferential of \( f \) at \( p \), respectively, are defined to be \( \hat{\partial}_\mathcal{M}f(p) := \hat{\partial}(f \circ \exp_p)(0) \) and \( \partial f(p) := \partial(f \circ \exp_p)(0) \), where \( \exp_p : U \rightarrow \mathcal{M} \) is the exponential map of \( \mathcal{M} \) defined on \( U \), which is a sufficiently small neighborhood of \( 0 \) in \( T_p\mathcal{M} \). Similarly, the contingent directional derivative of \( f : \mathcal{M} \rightarrow \mathbb{R} \) at \( p \) in the direction \( v \in T_p\mathcal{M} \) is defined to be \( f^-(p; v) := (f \circ \exp_p)^- (0; v) \). Let \( \delta \Omega \subset \mathcal{M} \) and \( p \in \text{cl } \delta \Omega \). The Fréchet normal cone and the limiting normal cone of \( \Omega \) at \( p \), respectively, are defined to be \( \hat{N}_{\mathcal{M}}^\mathcal{M}(p) := \hat{N}_{\exp_p^{-1} \delta \Omega}(0) \) and \( N_{\mathcal{M}}^\mathcal{M}(p) := N_{\exp_p^{-1} \delta \Omega}(0) \), where \( \delta \Omega \) is the intersection of \( \Omega \) with an arbitrarily small neighborhood of \( p \) on which \( \exp_p^{-1} \) is defined. Recall that the Fréchet normal cone of \( \Omega \) at point \( p \) is equal to the Fréchet subdifferential of the indicator function of \( \Omega \) at \( p \), that is,

\[ \hat{N}_{\mathcal{M}}^\mathcal{M}(p) = \hat{\partial}_\mathcal{M}\delta\Omega(p). \tag{5} \]

Similarly, for the limiting subdifferential and normal cone, \( N_{\mathcal{M}}^\mathcal{M}(p) = \partial\mathcal{M}\delta\Omega(p) \). Suppose that \( \mathcal{M} \) is a submanifold of a Euclidean space \( E \), \( \Omega \subset \mathcal{M} \) and \( p \in \text{cl } \Omega \). Then, from the definition, we have
\[ \hat{\mathcal{N}}^\mathcal{M}_\Omega(p) = \hat{\mathcal{N}}^E\Omega(p) \cap T^*_p\mathcal{M}. \] (6)

The following proposition states that the Fréchet subdifferential on Riemannian manifold has the homotone property. The proof is directly obtained using the definition.

**Proposition 3.1** (Homotone property of subdifferential) Consider functions \( f, g : \mathcal{M} \rightarrow \mathbb{R} \) on a Riemannian manifold \( \mathcal{M} \). Suppose that \( f \) and \( g \) are finite at \( p \in \mathcal{M} \), \( g \leq f \) and \( f(p) = g(p) \). Then, we have \( \hat{\partial}_\mathcal{M}g(p) \subset \hat{\partial}_\mathcal{M}f(p) \).

### 3.1 Contingent Cone

In the sequel, we define contingent cone on a Riemannian manifold. Similarly, one can define contingent cone on a Riemannian manifold by means of the exponential function and the corresponding definition in the linear space setting. But, at first, for a general subset \( \Omega \) of a Riemannian manifold \( \mathcal{M} \) and a point \( p \in \text{cl}\, \Omega \), the notion of contingent cone was introduced by Ledyaev and Zhu [31, Definition 3.8], as all tangent vectors \( v \in T_p\mathcal{M} \) such that

\[
\text{there are a sequence } t_i \downarrow 0 \text{ and } v_i \in T_p\mathcal{M} \text{ so that } v_i \rightarrow v \text{ and } c_{v_i}(t_i) \in \Omega,
\] (7)

where \( c_{v_i} \) is an integral curve on \( \mathcal{M} \) with \( c_{v_i}(0) = p \) and \( c'_{v_i}(0) = v_i \). Li et al. [20, Remark 3.6] pointed out that this definition is incomplete and introduced some additional restrictions on the sequence \( \{c_{v_i}\} \), which were in fact used by Ledyaev and Zhu [31, Proposition 3.9]; the restrictions are

\[
\lim_i c_{v_i}(t_i) = p \text{ and } \lim_i c'_{v_i}(t_i) = v.
\] (8)

But, it appears that these additional conditions are still incomplete. We should add a condition, so that the conclusions of [20, Remark 3.6] and [31, Proposition 3.9] hold true. That condition is

\[
\text{uniform convergence of } c'_{v_i} \text{ to } c'_v,
\] (9)

in the sense that for every scalar function \( g \in C^1(\mathcal{M}) \), \( d(g \circ c_{v_i}) \) uniformly converges to \( d(g \circ c_v) \) on a neighborhood of 0. With these additional conditions, we can show that the definition of contingent cone as given in [31, Definition 3.8] reduces to the following definition (see Remark 3.1 below).

**Definition 3.1** (Contingent cone) Suppose that \( \Omega \) is a subset of the Riemannian manifold \( \mathcal{M} \) and \( p \in \text{cl}\, \Omega \). The contingent cone of \( \Omega \) at \( p \) is defined to be

\[
\hat{T}^\mathcal{M}_\Omega(p) := \{ v \in T_p\mathcal{M} : \exists v_i \in T_p\mathcal{M}, t_i \downarrow 0 \text{ such that } v_i \rightarrow v, \exp_p(t_i v_i) \in \Omega \}.
\]

It is worthwhile to mention that Definition 3.1 has been used by Hosseini and Pouryayevali [30].
Remark 3.1  Here, we show that Definition 3.1 is equivalent to [31, Definition 3.8] when we add additional restrictions (8) and (9) on a sequence \( \{c_{v_i}\} \) for a vector \( v_i \in T_pM \) therein. Our proof makes use of a technique of [20, Proposition 3.5]. Let \( T_\Omega(p) \) be the set of all tangent vectors \( v \in T_pM \) for which conditions (7), (8), and (9) hold true. We want to show that \( \hat{T}_\Omega^M(p) = T_\Omega(p) \). Obviously, \( \hat{T}_\Omega^M(p) \subset T_\Omega(p) \). We will show the reverse of the inclusion. Let \( v \in T_\Omega(p) \). By (7), there exist a sequence \( t_i \downarrow 0 \) and \( v_i \in T_pM \) such that \( v_i \rightarrow v \) and \( c_{v_i}(t_i) \in \Omega \). Denote \( w_i := \exp_p^{-1} c_{v_i}(t_i) \) for sufficiently small \( t_i \). Then, we show

\[
\begin{equation}
    v = \lim_{t_i \to 0} \frac{w_i}{t_i},
\end{equation}
\]

which clearly shows \( v \in \hat{T}_\Omega^M(p) \). To establish (10), take \( f \in C^1(M) \) and, by its smoothness, obtain

\[
    f(c_{v_i}(t_i)) = f(\exp_p w_i) = f(p) + \langle df(p), w_i \rangle + o(\|w_i\|),
\]

which in turn implies

\[
    \langle df(p), v \rangle = \lim_{t_i \downarrow 0} \frac{f(c_{v_i}(t_i)) - f(p)}{t_i} = \lim_{t_i \downarrow 0} \frac{w_i}{t_i} + \lim_{t_i \downarrow 0} \frac{o(\|w_i\|)}{t_i}. \tag{11}
\]

Since \( c'_{v_i} \) uniformly converges to \( c'_{v} \), for a constant \( L > 0 \), we have

\[
    \|w_i\| = d(c_{v_i}(t_i), p) \leq \int_0^{t_i} \|c'_{v_i}(t)\| dt \leq L t_i.
\]

We get that \( \frac{\|w_i\|}{t_i} \) is bounded as \( t_i \downarrow 0 \). The latter, together with \( \lim_{i} w_i = 0 \) and (11), implies that (10) holds, because \( f \in C^1(M) \) was chosen arbitrarily. Thus, the proof is complete. 

4 Generalized Derivatives of a Distance Function

Generalized differential properties of distance functions play remarkable roles in variational analysis, optimization, and their applications. Generalized derivatives of distance functions have fundamental roles in the analysis of optimization algorithms, such as proximal point methods in both linear spaces and Riemannian manifolds. The authors of [18,19] investigated the properties of generalized derivatives of distance functions in the linear space setting. Properties of distance functions on a Riemannian manifold are not trivially obtained by a generalization of the corresponding properties in the linear spaces setting.

The following statement shows a relation between the distance of two points on a Riemannian manifold and the distance of the image of two points under a chart.
Proposition 4.1 ([35]) For any point $p \in M$ and chart $(U', \psi)$ around $p$ there exist a $U \subseteq U'$ and a constant $C \geq 1$ such that for all $p, x \in U$, we have
\[
\frac{1}{C} \| \psi(p) - \psi(x) \| \leq d(p, x) \leq C \| \psi(p) - \psi(x) \|. \tag{12}
\]

Now, suppose that $\Omega \subset M$ and $p \in \text{cl} \; \Omega$. For every $r > 0$, denote $\Omega_r := \Omega \cap B(p, r)$. In Proposition 4.1, by setting $h = \exp_p^{-1}$ and taking supremum over $x \in \Omega_r$, we get
\[
\frac{1}{C} \text{dist}(\exp_p^{-1}(u); \exp_p^{-1}(\Omega)) \leq d(u; \Omega_r) \leq C \text{dist}(\exp_p^{-1}(u); \exp_p^{-1}(\Omega_r)), \tag{13}
\]
which is a local estimate of a distance function on a Riemannian manifold in terms of normal coordinates. By these estimates, one can obtain some properties of a distance function on a Riemannian manifold from the corresponding results in the linear space setting. For example, by using homotone property of the Fréchet subdifferential and a well-known result in the linear space setting (Proposition 2.1), we get
\[
\frac{1}{C} \hat{N}_M^M(p) \cap B_{T_pM} \leq \hat{d}_M \text{dist}(p; \Omega) \leq C \hat{N}_M^M(p) \cap B_{T_pM}. \tag{14}
\]

Li et al. [20] established (14) with $C = 1$ for a convex subset of a Riemannian manifold with the nonpositive sectional curvature. In their proof, there were two fundamental points: First, since the authors used the convex calculus on manifolds, it was necessary to assume that the manifolds were Hadamard manifolds and $\Omega$ was a geodesic convex set so that the distance function $\text{dist}(\cdot; \Omega)$ be a convex function; second, they used a comparison result for geodesic triangles in a Hadamard manifold, which was a global property. Since we use the nonsmooth calculus on manifolds, we do not need the distance function to be convex. On the other hand, since the Fréchet subdifferential and normal cone are local notions, we can establish (14) with $C = 1$ for an arbitrary subset of an arbitrary Riemannian manifold with a better local estimate of a distance function than (13).

Lemma 4.1 (Local distance lemma) Let $\Omega$ be a subset of $M$ and $p \in \text{cl} \; \Omega$. For sufficiently small $r > 0$, denote $\Omega_r := \Omega \cap B(p, r)$. For each $u \in B(p, r)$, we have
\[
\frac{\text{dist}(\exp_p^{-1}u; \exp_p^{-1}\Omega_r)}{1 + |o(r)|} \leq \text{dist}(u; \Omega_r) \leq (1 + |o(r)|) \text{dist}(\exp_p^{-1} u; \exp_p^{-1} \Omega_r). \tag{15}
\]

Proof Let $r$ be sufficiently small such that $\exp_p^{-1}$ is defined on $B(p, r)$. It is enough to show that for every $u_1, u_2 \in B(p, r)$, we have
\[
(1 - |o(r)|)d(v_1, v_2) \leq d(u_1, u_2) \leq (1 + |o(r)|)d(v_1, v_2), \tag{15}
\]

\[1 \text{ See: } https://mathoverflow.net/q/301064.\]
where \( v_1 := \exp_p^{-1} u_1 \) and \( v_2 = \exp_p^{-1} u_2 \). Suppose that \( \gamma(t) = v_1 + t(v_2 - v_1), \)
\( 0 \leq t \leq 1 \), is the straight line segment joining \( v_1 \) and \( v_2 \).

Denote \( c(t) := \exp_p \gamma(t), 0 \leq t \leq 1 \). Let \( f = \exp_p^* (g) \) be the pull back of the
metric \( g \) of \( M \) in \( \exp_p^{-1} \mathbb{B}(p, r) \). We have \( \dot{c}(t) = (D \exp_p \gamma(t)(\dot{\gamma}(t)) \)
and

\[
d(u_1, u_2) \leq \int_0^1 \sqrt{g_p(\dot{\gamma}, \dot{\gamma}) + q_\gamma(t)(\dot{\gamma}, \dot{\gamma}) + h_\gamma(t)(\dot{\gamma}, \dot{\gamma}) dt}
\]

\[
= \int_0^1 \sqrt{1 + \frac{1}{3} \left( \frac{R_p(\gamma, \gamma, \gamma)}{g_p(\dot{\gamma}, \dot{\gamma}) g_p(\gamma, \gamma)} \right) g_p(\gamma, \gamma) + \frac{h_\gamma(t)(\dot{\gamma}, \dot{\gamma})}{g_p(\dot{\gamma}, \dot{\gamma})}} dt
\]

\[
\leq \left(1 + \frac{\|R_p\|}{6} r^2 + O(r^3)\right) d(v_1, v_2),
\]

where \( \|R_p\| \) is the supremum of \( R_p(w_1, w_2, w_3, w_4) \) over all \( w_i \in T_pM \), for \( i = 1, \ldots, 4 \), with norms equal to 1. So, the second inequality of (15) is established. To establish the first inequality of (15), we consider an arbitrary curve \( c \) joining \( u_1 \) and \( u_2 \) in \( \mathbb{B}(p, r) \). Then, using a similar technique as above, we can show that the length of \( c \) is at least \( (1 - \frac{\|R_p\|}{6} r^2 + O(r^3))d(v_1, v_2) \). So, by taking the infimum, the desired inequality is at hand. Therefore, the proof is complete. \( \square \)

**Remark 4.1** The proof of Lemma 4.1 gives a tighter estimate of \( o(r) \). Indeed, we have
\[
\text{dist}(u; \Omega_r) = \left(1 \pm \frac{\|R_p\|}{6} r^2 + O(r^3)\right) \text{dist}(\exp_p^{-1} u; \exp_p^{-1} \Omega_r).
\]

Now, we can prove (14) with \( C = 1 \) in a general setting, which plays an essential role in stating some necessary conditions of the dual type for the set of weak sharp minima of a nonconvex optimization problem in Sect. 5.

Note that results of this type, relating subdifferential of a distance function and a normal to the corresponding set, are known for various subdifferentials in general nonconvex settings of Banach spaces and are of great importance for many aspects of variational analysis; e.g., see [32,37].

**Theorem 4.1** With \( \Omega \) a subset of \( M \), for the distance function \( \text{dist}(\cdot; \Omega) \), and \( p \in \text{cl} \Omega \), one has

\[
\hat{\partial}_M \text{dist}(p; \Omega) = \hat{N}_\Omega^M(p) \cap \mathbb{B}_{T^*_pM}.
\]

\( \square \) Springer
Proof By local distance lemma (Lemma 4.1), for sufficiently small values of \( r \), we have \( \text{dist}(u; \Omega) = \text{dist}(u; \Omega_r) \leq (1 + |o(r)|) \text{dist}(\exp_{p}^{-1} u; \exp_{p}^{-1} \Omega_r) \), for each \( u \in B(0, r/2) \). Now, the homotone property of subdifferential and the similar property in the linear space setting (Proposition 2.1) imply
\[
\hat{\partial}_{M} \text{dist}(p; \Omega) \leq (1 + |o(r)|) \hat{N}_{\exp_{p}^{-1} \Omega_r}(0) \cap B_{T_p^*M}.
\]
By \( r \to 0 \), we have \( \hat{\partial}_{M} \text{dist}(p; \Omega) \leq \hat{N}^{M}_{\Omega}(p) \cap B_{T_p^*M} \). The reverse of the inequality can be established similarly, to complete the proof.

Immediately, we have the following corollary.

**Corollary 4.1** With the notation of Theorem 4.1, we have
\[
\hat{N}^{M}_{\Omega}(p) = \text{cone} \hat{\partial}_{M} \text{dist}(p; \Omega).
\]

The following theorem relates a directional derivative of a distance function to a contingent cone. It immediately follows from a corresponding property in the linear space setting by using the local distance lemma. We use this property to obtain a necessary condition for weak sharp minima.

**Theorem 4.2** Suppose that \( M \) is a Riemannian manifold and \( p \in \Omega \subset M \). Then, for all \( v \in T_pM \), we have
\[
\frac{\text{dist}(\exp_{p} tw; \Omega) - \text{dist}(p; \Omega)}{t} \leq \frac{|o(r)| \text{dist}(tw; \exp_{p}^{-1} \Omega_r) - 0}{t}.
\]

By taking \( \text{lim inf} \) as \( t \downarrow 0 \) and \( w \to v \), and by using a similar property in the linear space setting (Proposition 2.2), we have
\[
\frac{\text{dist}(\exp_{p} tw; \Omega) - \text{dist}(p; \Omega)}{t} \leq (1 + |o(r)|) \text{dist}(\exp_{p}^{-1} u; \exp_{p}^{-1} \Omega_r).
\]

By taking \( \text{lim inf} \) as \( t \downarrow 0 \) and \( w \to v \), and by using a similar property in the linear space setting (Proposition 2.2), we have
\[
d_{\Omega}(p; v) \leq \text{dist}(v; \hat{T}_{\exp_{p}^{-1} \Omega_r}(0)) = \text{dist}(v; \hat{T}^{M}_{\Omega}(p)).
\]

Similarly, one can prove that if it is further assumed that \( M \) is finite-dimensional, then equality holds in (17).

The following corollary is now at hand.
Corollary 4.2 With the notation of Theorem 4.2, if \( p \in \Omega \subset \mathcal{M} \), then
\[
\hat{T}_\Omega^\mathcal{M}(p) \subset \{ v : d_\Omega(p; v) \leq 0 \},
\]
and equality holds when \( \mathcal{M} \) is finite-dimensional.

Remark 4.2 To prove Theorems 4.1 and 4.2, we do not need the underlying Riemannian manifold to be a Hadamard manifold. But, for the development of some optimization techniques on Riemannian manifolds, such as variational inequalities [38] and equilibrium problems [39], properties of Hadamard manifolds are indeed essential; see Kristály [39, Remark 5.1].

5 Nonconvex Weak Sharp Minima on a Riemannian Manifold

Here, we give some necessary conditions of the primal and dual types for the set of weak sharp minima of an optimization problem on a (possibly infinite-dimensional) Riemannian manifold. Formerly, Li at al. [20] provided some characterizations of weak sharp minima in the case of convex problems on finite-dimensional Riemannian manifold.

Note that \( p \in \Omega \) (where \( \Omega := \text{argmin}_S f \)) being a local weak sharp minimizer for the problem \( \min_{u \in S} f(u) \) with modulus \( \alpha > 0 \), is equivalent to \( p \in \Omega \) being a local minimizer of the following perturbed problem:
\[
\min_{u \in S} (f(u) - \alpha \text{dist}(u; \Omega)).
\]
(18)

Similarly, \( \Omega \) being the set of weak sharp minima for the problem \( \min_{u \in S} f(u) \) with the modulus \( \alpha > 0 \), is equivalent to each \( p \in \Omega \) being a global minimizer of the perturbed problem (18). So, the set of weak sharp minima of an optimization problem is equivalent to the set of minimizers of a difference optimization problem.

Remark 5.1 The properties of weak sharp minima on Riemannian manifold could not trivially be advocated as the properties of weak sharp minima on the linear space setting. Indeed, if we substitute \( u \) by \( \exp_p w \) in (18), then the objective function is converted to \( (f \circ \exp_p)(w) - \alpha \text{dist}(\exp_p w; \Omega) \). But, the local minimum of the converted problem is not trivially related to the weak sharp minima of an optimization problem on a linear space.

The homotone property of the Fréchet subdifferential (Proposition 3.1) admits a necessary optimality condition for a local minimum of a function on a Riemannian manifold.

Proposition 5.1 Let \( f : \mathcal{M} \to \mathbb{R} \) be a function on a Riemannian manifold \( \mathcal{M} \). Suppose that the value of \( f \) is finite at \( p \in \mathcal{M} \). If \( p \) is a local minimizer of \( f \), then \( 0 \in \hat{\partial}_\mathcal{M} f(p) \).

Next, we state a simple rule about the Fréchet subdifferential of the sum of two functions, directly deduced from the definition.
Proposition 5.2 Consider functions $f_1, f_2: \mathcal{M} \to \mathbb{R}$ on a Riemannian manifold $\mathcal{M}$. Suppose $p \in \mathcal{M}$ and $f_1(p), f_2(p) < \infty$. Then, we have

$$\hat{\partial} \mathcal{M} f_1(p) + \hat{\partial} \mathcal{M} f_2(p) \subset \hat{\partial} \mathcal{M} (f_1 + f_2)(p).$$

(19)

Moreover, if one of the $f_i$ is Fréchet differentiable, then we have equality in (19).

Now, we state some necessary conditions of the primal type and the dual type for a local weak sharp minimizer of an unconstrained optimization problem.

Theorem 5.1 (Necessary conditions for a local weak sharp minimizer of an unconstrained problem on a Riemannian manifold) Suppose $S := \mathcal{M}$ and $p \in \Omega := \text{argmin}_S f$ is a local weak sharp minimizer for the problem $\min_{u \in S} f(u)$ with modulus $\alpha > 0$. Then, the following hold.

(i) We have

$$\alpha \mathbb{B}_{T_p \mathcal{M}} \cap \hat{N}_{\Omega}^\mathcal{M}(p) \subset \hat{\partial} \mathcal{M} f(p).$$

(20)

(ii) For all $v \in T_p \mathcal{M}$, we have

$$f^-(p; v) \geq \alpha \text{dist}(v; \hat{T}_{\Omega}^\mathcal{M}(p)).$$

(21)

Proof By definition, there exists $\epsilon > 0$ such that

$$f(u) \geq f(p) + \alpha \text{dist}(u; \Omega), \quad \forall u \in \mathbb{B}(p, \epsilon).$$

Since the Fréchet subdifferential has homotone property (Proposition 3.1), we have $\hat{\partial} \mathcal{M} \alpha \text{dist}(p; \Omega) \subset \hat{\partial} \mathcal{M} f(p)$. Theorem 4.1 implies

$$\alpha \hat{N}_{\Omega}^\mathcal{M}(p) \cap \mathbb{B}_{T_p \mathcal{M}} \subset \hat{\partial} \mathcal{M} f(p).$$

So, (i) is proved. Next, we prove (ii). Let $v \in T_p \mathcal{M}$. The hypothesis guarantees that for all $w$ sufficiently close to $v$ and for all sufficiently small $t > 0$, we have

$$f(\exp_p t w) - f(p) \geq \alpha \text{dist}(\exp_p t w; \Omega),$$

which implies

$$\frac{f(\exp_p t w) - f(p)}{t} \geq \frac{\alpha \text{dist}(\exp_p t w; \Omega) - \text{dist}(p; \Omega)}{t}.$$

By taking lim inf of both sides of the latter inequality, as $w \to v$ and $t \downarrow 0$, and applying Theorem 4.2, (ii) is obtained. \qed

Since every element of the set of global weak sharp minima is a local weak sharp minimizer, we immediately have the following corollary.

Corollary 5.1 (Necessary conditions for unconstrained weak sharp minima on a Riemannian manifold) Suppose $S := \mathcal{M}$ and $\Omega := \text{argmin}_S f$ is the set of weak sharp minima for the problem $\min_{u \in S} f(u)$ with modulus $\alpha > 0$. Then,

(i) for every $p \in \Omega$, we have the inclusion (20);

\begin{center}
\includegraphics[width=\textwidth]{Springer}
\end{center}
(ii) for every $p \in \Omega$ and $v \in T_p M$, we have the inequality (21).

Remark 5.2 Similar to Ward [28], with some modifications of the definition of the contingent directional derivative on Riemannian manifold, one can state some necessary conditions for weak sharp minima of higher orders.

Remark 5.3 Similar to the approach of Studniarski and Ward [27] in linear space, one can state some sufficient conditions for weak sharp minima on Riemannian manifold based on a generalization of the contingent directional derivative.

5.1 Constrained Weak Sharp Minima on Riemannian Manifolds

In the sequel, we give some necessary conditions for the weak sharp minima of a constrained optimization problem on a Riemannian manifold. As pointed out at the beginning of this section, the set of weak sharp minima is equivalent to the set of minimizers of a difference optimization problem (18). So, we first state some necessary conditions for minimizers of a difference optimization problem, i.e., an optimization problem whose cost function is given in a difference form. To present these necessary conditions, we use the Fréchet and limiting subdifferentials on a Riemannian manifold. Consider the difference optimization problem with a geometric constraint as follows:

$$\min_{u \in S} f(u), \quad (22)$$

where $f : M \to \bar{\mathbb{R}}$ may be represented as $f = f_1 - f_2$. With the help of indicator function on the set $S$, the constrained optimization problem (22) can be rewritten as the following unconstrained problem:

$$\min_{u \in M} f(u) + \delta_S(u). \quad (23)$$

To present some necessary optimality conditions for the unconstrained optimization problem, we need to decompose the subdifferential of the objective function of problem (23). We say that $f$ is Fréchet decomposable at $p \in S$ on $S$, when

$$(f + \delta_S)(p) \subset \delta_M f(p) + \hat{N}_S^M(p). \quad (24)$$

Note that this definition is a Riemannian manifold counterpart of the one used by Mordukhovich et al. [25] in the linear space setting. If $f$ is Fréchet differentiable at $p$, then Proposition 5.2 implies that $f$ is Fréchet decomposable at $p \in S$ on $S$. Moreover, by [20, Proposition 4.3], for a convex function $f$ with dom $f$ having nonempty interior and a nonempty convex set $S$ such that $S \cap \text{dom } f$ is convex, we have that $f$ is Fréchet decomposable on $S$ at every point in int(dom $f$) $\cap$ $S$. According to the attractive form of calculus for the limiting subdifferential of a Lipschitzian function, the decomposable condition (24) for the limiting subdifferential is at hand. In the first assertion of the following theorem, we impose the decomposition assumption for the Fréchet subdifferential on a Riemannian manifold, while the second part is justified
without this assumption via the limiting subdifferential on a Riemannian manifold and its attractive forms in the Lipschitzian case, i.e., if \( f_1: \mathcal{M} \to \overline{\mathbb{R}} \) is Lipschitzian around \( p \in \mathcal{M} \) and \( f_2: \mathcal{M} \to \overline{\mathbb{R}} \) is an l.s.c. function and is finite at \( p \), then

\[
\partial_{\mathcal{M}}(f_1 + f_2)(p) \subset \partial_{\mathcal{M}} f_1(p) + \partial_{\mathcal{M}} f_2(p).
\] (25)

**Theorem 5.2** (Necessary conditions for difference problems with geometric constraints) Suppose that \( p \) is a local solution of (22), and \( f \) is given by \( f = f_1 - f_2 \), where \( f_1: \mathcal{M} \to \overline{\mathbb{R}} \) is finite at \( p \). Then, the following hold.

(i) If \( f_1 \) is Fréchet decomposable at \( p \in S \) on \( S \), then we have

\[
\hat{\partial}_{\mathcal{M}} f_2(p) \subset \hat{\partial}_{\mathcal{M}} f_1(p) + \hat{N}_{\mathcal{M}}^S(p).
\] (26)

Particularly, when \( f \equiv -f_2 \), we have \( \hat{\partial}_{\mathcal{M}} f_2(p) \subset \hat{N}_{\mathcal{M}}^S(p) \).

(ii) If \( f_1 \) is Lipschitzian around \( p \) and \( S \) is closed, then we have

\[
\hat{\partial}_{\mathcal{M}} f_2(p) \subset \partial_{\mathcal{M}} f_1(p) + N_{\mathcal{M}}^S(p).
\]

**Proof** Problem (22) can be written as the following unconstrained difference problem:

\[
\min_{x \in \mathcal{M}} [f_1(x) + \delta_S(x) - f_2(x)].
\]

By Corollary 5.1, we have \( 0 \in \hat{\partial}_{\mathcal{M}}(f_1(x) + \delta_S(x) - f_2(x))(p) \). Proposition 5.2 implies

\[
\hat{\partial}_{\mathcal{M}} f_2(p) \subset \hat{\partial}_{\mathcal{M}}(f_1 + \delta_S)(p).
\] (27)

If \( f_2 \) is Fréchet decomposable at \( p \in S \) on \( S \), then we have

\[
\hat{\partial}_{\mathcal{M}} f_2(p) \subset \hat{\partial}_{\mathcal{M}} f_1(p) + \hat{N}_{\mathcal{M}}^S(p).
\]

So, (i) is proved. Now, we prove part (ii). Since \( S \) is assumed to be closed, \( \delta_S \) is l.s.c. and inclusion (27) with the sum rule for the limiting subdifferential (25) implies

\[
\hat{\partial}_{\mathcal{M}} f_2(p) \subset \hat{\partial}_{\mathcal{M}}(f_1 + \delta_S)(p) \subset \partial_{\mathcal{M}}(f_1 + \delta_S)(p)
\]

\[
\subset \partial_{\mathcal{M}} f_1(p) + \partial_{\mathcal{M}} \delta_S(p) = \partial_{\mathcal{M}} f_1(p) + N_{\mathcal{M}}^S(p),
\]

and the proof of (ii) is complete. \( \square \)

In the following, as a corollary of Theorem 5.2, we can state some necessary conditions for the set of weak sharp minima of a constrained optimization problem. The result can be established using Theorem 5.2 by reducing weak sharp minima to minimizers of a constrained difference optimization problem.

**Corollary 5.2** (Necessary conditions for weak sharp minima under geometric constraints on Riemannian manifolds) Let \( \Omega := \text{argmin}_S f \). Suppose \( \Omega \) is the set of weak sharp minima for the problem \( \min_S f \) with modulus \( \alpha > 0 \). Then, the following hold.
(i) If \( p \in \Omega \) and \( f \) is Fréchet decomposable on \( S \) at the point \( p \), then
\[
\alpha B_{T_p^* M} \cap \hat{N}_\Omega^M(p) \subset \partial_M f(p) + \hat{N}_S^M(p).
\]

(ii) If \( f \) is Lipschitzian around \( p \in S \) and \( S \) is closed, then we have
\[
\alpha B_{T_p^* M} \cap \hat{N}_\Omega^M(p) \subset \partial_M f(p) + N_S^M(p).
\]

6 An Application

Here, we give an application of the nonconvex weak sharp minima on Riemannian manifolds. We will show how the notion of weak sharp minima and the results of the previous sections can be used to model a discrete optimization problem as an unconstrained optimization problem on a Stiefel manifold. Recall that for integers \( 0 < k \leq n \), the Stiefel manifold \( St(n, k) \) is defined as the set of all matrices \( U \in \mathbb{R}^{n \times k} \), with \( U^T U = I_k \), where \( U^T \) denotes the transpose of \( U \) and \( I_k \) denotes the \( k \times k \) identity matrix. For each \( P \in St(n, k) \), we have
\[
T_P St(n, k) = \{ X \in \mathbb{R}^{n \times k} : X^T P + P^T X = 0 \},
\]
and we endow \( T_P St(n, k) \) with the induced metric of the Euclidean space \( \mathbb{R}^{n \times k} \); \( \langle X, Y \rangle := \text{tr}(X^T Y) \), for \( X, Y \in \mathbb{R}^{n \times k} \); also, we identify \( T_P^* St(n, k) \) with \( T_P St(n, k) \), by the natural isomorphism induced by this Riemannian metric. For more details on properties of the Stiefel manifold, see [3]. For each matrix \( A = [a_{ij}] \), denote
\[
\| A \|_\beta := (\sum_{i,j} |a_{ij}|^\beta)^{1/\beta} \quad \text{and} \quad A^- := \max\{-a_{ij}, 0\}.
\]

**Example 6.1** Let \( G \) be a graph with vertices \( V(G) := \{ v_1, \ldots, v_n \} \) and edges \( E(G) \). Fix integer \( k > 0 \). Denote
\[
D_k(G) := \{(A_1, \ldots, A_k) : \emptyset \neq A_i \subset V(G), A_i \cap A_j = \emptyset, \text{ for all } i \neq j\}
\]
as the set of all \( k \) subpartitions of \( V(G) \). Define \( \partial A \), for a subset \( A \) of \( V(G) \), as the set of all edges of \( G \) whose only one endpoint belongs to \( A \). Now, consider the following Cheeger-type constant of \( G \):
\[
\gamma_k(G) := \min \sum_{i=1}^k \frac{|\partial A_i|}{\sqrt{|A_i|}}, \quad (28)
\]
where the minimum runs over all \( (A_1, \ldots, A_k) \in D_k(G) \). The constant \( \gamma_k(G) \) indicates how well \( G \) can be partitioned into \( k \) clusters; see [40–42] for more information on the Cheeger constant of graphs and its applications to clustering. For each \( \alpha := (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n \), define
\[
\| \nabla \alpha \|_1 := \sum_{\{v_i, v_j\} \in E(G)} |\alpha_i - \alpha_j|.
\]
Using a technique similar to that of Rothaus [43], one can show\(^2\) that the discrete minimization problem (28) is equivalent to the following continuous optimization problem on a manifold with nonnegativity constraints:

\[
\min_{U \in \text{unconstrained optimization problem on a Stiefel manifold:}} \sum_{i=1}^{k} \| \nabla u_i \|_1,
\]

where the minimum runs over all \(U := (u_1, \ldots, u_k) \in \text{St}(n, k)\) with nonnegative entries (which hereafter is denoted by \(\text{St}_+(n, k)\)). The nonnegativity constraints provide a combinatorial nature to the feasible set of problem (29). But, since the cost function of (29) is Lipschitzian on \(\text{St}(n, k)\) with a rate \(C\), the penalization lemma (see [32, Proposition 1.121]) implies that problem (29) is equivalent to the following unconstrained optimization problem on a Stiefel manifold:

\[
\min_{U \in \text{St}(n, k)} \sum_{i=1}^{k} \| \nabla u_i \|_1 + C \text{dist}(U; \text{St}_+(n, k)).
\]

Although computing the distance term in (30) is not a simple task, a simple observation shows that one can replace the distance term with any upper estimate \(h(U)\) such that 0 is the minimum value of \(h\), the set of minimizers of \(h\) is \(\text{St}_+(n, k)\), and for every \(U \in \text{St}(n, k),\)

\[
\text{dist}(U; \text{St}_+(n, k)) \leq h(U).
\]

In other words, \(\text{St}_+(n, k)\) should be the set of weak sharp minima of the problem \(\min_{\text{St}(n, k)} h(U)\) with 0 as the optimal value. Since the distance of a matrix \(U \in \text{St}(n, k)\) from the nonnegative matrices in the Euclidean space \(\mathbb{R}^{n \times k}\) is a function of the negative part of \(U\), it seems that a natural candidate for \(h(U)\) is a function of \(U^-\). So, we investigate the following question.

Is \(\text{St}_+(n, k)\) the set of weak sharp minima for the problem of minimizing \(f_\beta(\tilde{U}) := \| U^- \|_\beta^\beta\) over the Stiefel manifold \(\text{St}(n, k)\), for some \(\beta > 0\)?

According to Corollary 5.1, a necessary condition for provision of an affirmative response to this question is that there exists \(\alpha > 0\) such that, for every \(P \in \text{St}_+(n, k),\)

\[
\alpha \mathbb{B}_{T_P \text{St}(n, k)} \cap \overline{\mathcal{N}}_{\text{St}_+(n, k)}(P) \subset \partial \text{St}(n, k) f_\beta(P).
\]

Fix \(P := [p_{ij}] \in \text{St}_+(n, k)\) and, for simplicity, suppose that only the first \(t\) rows of \(P\) are nonzero (for some \(t \geq 0\)). For each \(n \times k\) matrix \(A\), denote \(\tilde{A}\) by the matrix whose rows are the first \(t\) rows of \(A\) and \(\hat{A}\) by the matrix whose rows are the last \(n - t\) rows of \(A\). According to (6), we have \(X \in \overline{\mathcal{N}}_{\text{St}_+(n, k)}(P)\), if and only if \(X \in T_P \text{St}(n, k)\) and \(\limsup (X, U - P) / \|U - P\| \leq 0\), where the lim sup runs over all \(U\) approaching \(P\) in \(\text{St}_+(n, k)\). By approaching \(P\), from some suitable curves in \(\text{St}_+(n, k)\), one can see

\(^2\) The key point is to show that for each vector \(u := (u_1, \ldots, u_n) \in \mathbb{R}^n\) with \(\|u\|_2 = 1\), there is a nonempty subset \(A \subset \{v_i : u_i \neq 0\}\) such that \(\frac{|A|}{\sqrt{|A|}} < \|\nabla u\|_1\).
where \( \circ \) is the entry-wise product, \( O_{t \times k} \) is the \( t \times k \) zero matrix, and \( \widehat{X} \leq 0 \) means that \( \widehat{X} \) is a nonpositive matrix. Next, we will investigate when the necessary condition (32) holds. First, note that if \( \beta > 1 \), then \( f_{\beta} \) is smooth. So, the subdifferential \( \partial_{\beta} f_{\beta}(P) \) has only one element. But, there are some \( P \in St_{+) (n, k) \) such that the left-hand side of (32) has infinitely many elements. So, for \( \beta > 1 \), the necessary condition (32) does not hold.

Next, we will show that the necessary condition (32) holds, for each \( \beta < 1 \), with \( \alpha = 1 \). Consider an arbitrary element \( X \in B_{TPSt(n,k)} \cap \hat{N}_{St_{+}(n,k)}(P) \). We will show \( X \in \partial_{\beta} f_{\beta}(P) \). Define the smooth function \( \tilde{g}(U) = \langle X, \exp_{P}^{-1} U \rangle \) on a sufficiently small neighborhood of \( P \). For simplicity, denote \( dU = \exp_{P}^{-1} U \). Note that since \( X \in B_{TPSt(n,k)} \), we have \( \text{tr}(X^T X) \leq 1 \). So, the absolute value of the entries of \( \widehat{X} \) are at most equal to 1. Moreover, \( \widehat{X} \leq 0 \). Thus, \( \langle \widehat{X}, dU \rangle \leq \| dU \|_1 \). Therefore, by the Cauchy–Schwarz inequality, we have

\[
\tilde{g}(U) = \langle \widehat{X}, \widehat{dU} \circ \widehat{IP} \rangle + \langle \widehat{X}, \widehat{dU} \rangle \leq \| \widehat{dU} \circ \widehat{IP} \|_2 + \| \widehat{dU} \|_1, \tag{34}
\]

where \( IP \) is the \( n \times k \) matrix whose \((i, j)\)th entry is 1 if \( p_{ij} = 0 \) and is 0 otherwise.

One can see that the function \( \| \cdot \|_1 \) is a norm on

\[
S := \{ V \in \mathbb{R}^{t \times k} : \widehat{P}^T V + V^T \widehat{P} = 0 \}.
\]

Indeed, the function \( \| \cdot \|_1 \) satisfies the triangle inequality, since for each real \( a, b \) we have \((a + b)^{\beta} \leq a^{\beta} + b^{\beta}\). So, it is enough to show that \( V = 0 \), whenever \( V \in S \) and \( \| V \|_1 = 0 \). This follows from nonnegativity of the entries of \( P \) and that \( P \) has no zero row. Now, since \( dU \in TPSt(n,k) \), we have \( \widehat{dU} \in S \) and \( \widehat{dU} \circ \widehat{IP} \in S \). By the equivalence of norms on a finite-dimensional linear space, we have \( \| dU \circ \widehat{IP} \|_2 \leq C_{P} \| (dU \circ \widehat{IP}) \|_1 \), for some \( C_{P} > 1 \). Note that \( \widehat{dU} = \widehat{dU} \circ \widehat{IP} \). So, from (34), we have

\[
\tilde{g}(U) \leq C_{P} \| (dU \circ \widehat{IP}) \|_1 + \| (dU \circ \widehat{IP}) \|_1 = C_{P} \| (dU \circ IP) \|_1. \tag{35}
\]

Now, consider the function \( g(U) := \tilde{g}(U) + C_{P} \| (U \circ IP) \|_1 - C_{P} \| (dU \circ IP) \|_1 \). Since \( \| (U \circ IP) \|_1 - \| (dU \circ IP) \|_1 \) is of order \( o(\| U - P \|) \) (note that the function \( \| \cdot \|_1 \) satisfies the triangle inequality, \( dU = (U - P) + o(\| U - P \|) \)), and \( P \circ IP = O_{n \times k} \), we have \( dg(P) = d\tilde{g}(P) = X \). Therefore, by use of (35), we have

\[
g(U) \leq C_{P} \| (U \circ IP) \|_1 \leq C_{P} \| U \|_1 - C_{P} \| (dU \circ IP) \|_1 \leq C_{P} \| U \|_1^{\beta} (C_{P} \| U \|_1^{1-\beta}) \\
\leq \| U \|_1^{\beta} \leq \| U \|_1^{\beta}, \tag{36}
\]

on a sufficiently small neighborhood of \( P \), in which \( C_{P} \| U \|_1^{\beta} \leq 1 \). (The last inequality of (36) follows from the fact that for each \( a, b \geq 0 \) and \( 0 < \beta < 1 \), we have \((a + b)^{\beta} \leq a^{\beta} + b^{\beta}\). Thus, \( g \leq f_{\beta} \) on a neighborhood of \( P \). Since \( g(P) = f_{\beta}(P) \)
and \( \partial g(P) = X \), as the homotone property of the Fréchet subdifferential, we have \( X \in \partial_{\text{St}(n,k)} f_{\beta}(P) \).

\[\square\]

7 Conclusions

We presented a lemma and used it to derive some local properties of a distance function on a Riemannian manifold using the corresponding properties in linear space. In this regard, we established a relation between the Fréchet subdifferential (directional derivative) of a distance function and a normal cone (contingent cone) on a Riemannian manifold. Then, we established some primal and dual necessary conditions for the set of weak sharp minima of nonconvex optimization problems on Riemannian manifold. As an application, we showed how the notion of weak sharp minima and our stated results could be used to model a Cheeger-type constant of a graph as an unconstrained optimization problem on a Stiefel manifold.

Acknowledgements The authors thank Sharif University of Technology for supporting this work. The first author is grateful to Amir Daneshgar, Mostafa Einollahzadeh, Mehdi Shaeiri, and Mostafa Kiyae for constructive discussions.

References

1. Udriște, C.: Convex Functions and Optimization Methods on Riemannian Manifolds, Mathematics and Its Applications, vol. 297. Kluwer Academic Publishers Group, Dordrecht (1994)
2. Absil, P.A., Baker, C.G., Gallivan, K.A.: Trust-region methods on Riemannian manifolds. Found. Comput. Math. 7(3), 303–330 (2007)
3. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton (2007)
4. Smith, S.T.: Optimization techniques on Riemannian manifolds. In: Bloch, A. (ed.) Hamiltonian and Gradient Flows, Algorithms and Control, Fields Institute Communications, vol. 3, pp. 113–136. American Mathematical Society, Providence (1994)
5. Grohs, P., Hosseini, S.: Nonsmooth trust region algorithms for locally Lipschitz functions on Riemannian manifolds.IMA J. Numer. Anal. 36(3), 1167–1192 (2016)
6. Hosseini, R., Sra, S.: An alternative to EM for Gaussian mixture models: batch and stochastic Riemannian optimization. Math. Program. (2019). https://doi.org/10.1007/s10107-019-01381-4
7. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. 20(2), 303–353 (1999)
8. Dong, X., Frossard, P., Vandergheynst, P., Nefedov, N.: Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds. IEEE Trans. Signal Process. 62(4), 905–918 (2014)
9. Journée, M., Bach, F., Absil, P.A., Sepulchre, R.: Low-rank optimization on the cone of positive semidefinite matrices. SIAM J. Optim. 20(5), 2327–2351 (2010)
10. Shalit, U., Weinshall, D., Chechik, G.: Online learning in the embedded manifold of low-rank matrices. J. Mach. Learn. Res. 13, 429–458 (2012)
11. Sun, J., Qu, Q., Wright, J.: Complete dictionary recovery over the sphere II: recovery by Riemannian trust-region method. IEEE Trans. Inf. Theory 63(2), 885–914 (2017)
12. Vandereycken, B.: Low-rank matrix completion by Riemannian optimization. SIAM J. Optim. 23(2), 1214–1236 (2013)
13. Adler, R.L., Dedieu, J.P., Margulies, J.Y., Martens, M., Shub, M.: Newton’s method on Riemannian manifolds and a geometric model for the human spine.IMA J. Numer. Anal. 22(3), 359–390 (2002)
14. Ishteva, M., Absil, P.A., Van Huffel, S., De Lathauwer, L.: Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme. SIAM J. Matrix Anal. Appl. 32(1), 115–135 (2011)
15. Mishra, B., Meyer, G., Bach, F., Sepulchre, R.: Low-rank optimization with trace norm penalty. SIAM J. Optim. 23(4), 2124–2149 (2013)
16. Bento, G.C., Ferreira, O.P., Oliveira, P.R.: Proximal point method for a special class of nonconvex functions on Hadamard manifolds. Optimization 64(2), 289–319 (2015)
17. Hosseini, S., Uschmajew, A.: A Riemannian gradient sampling algorithm for nonsmooth optimization on manifolds. SIAM J. Optim. 27(1), 173–189 (2017)
18. Burke, J.V., Ferris, M.C., Qian, M.: On the Clarke subdifferential of the distance function of a closed set. J. Math. Anal. Appl. 166(1), 199–213 (1992)
19. Mordukhovich, B.S., Nam, N.M.: Subgradient of distance functions with applications to Lipschitzian stability. Math. Program. 104(2–3, Ser. B), 635–668 (2005)
20. Li, C., Mordukhovich, B.S., Wang, J., Yao, J.C.: Weak sharp minima on Riemannian manifolds. SIAM J. Optim. 21(4), 1523–1560 (2011)
21. Polyak, B.T.: Sharp minima, institute of control sciences lecture notes, Moscow, USSR, 1979; Presented at the IIASA Workshop on Generalized Lagrangians and Their Applications. IIASA, Laxenburg, Austria (1979)
22. Ferris, M.C.: Weak sharp minima and penalty functions in mathematical programming. Ph.D. thesis, University of Cambridge, Cambridge (1988)
23. Burke, J.V., Deng, S.: Weak sharp minima revisited. II. Application to linear regularity and error bounds. Math. Program. 104(2–3, Ser. B), 235–261 (2005)
24. Burke, J.V., Ferris, M.C.: Weak sharp minima in mathematical programming. SIAM J. Control Optim. 31(5), 1340–1359 (1993)
25. Mordukhovich, B.S., Nam, N.M., Yen, N.D.: Fréchet subdifferential calculus and optimality conditions in nondifferentiable programming. Optimization 55(5–6), 685–708 (2006)
26. Ng, K.F., Zheng, X.Y.: Global weak sharp minima on Banach spaces. SIAM J. Control Optim. 41(6), 1868–1885 (2003)
27. Studniarski, M., Ward, D.E.: Weak sharp minima: characterizations and sufficient conditions. SIAM J. Control Optim. 38(1), 219–236 (1999)
28. Ward, D.E.: Characterizations of strict local minima and necessary conditions for weak sharp minima. J. Optim. Theory Appl. 80(3), 551–571 (1994)
29. Zhou, J., Mordukhovich, B.S., Xiu, N.: Complete characterizations of local weak sharp minima with applications to semi-infinite optimization and complementarity. Nonlinear Anal. 75(3), 1700–1718 (2012)
30. Hosseini, S., Pouryayevali, M.R.: On the metric projection onto prox-regular subsets of Riemannian manifolds. Proc. Am. Math. Soc. 141(1), 233–244 (2013)
31. Ledyjaev, Y.S., Zhu, Q.J.: Nonsmooth analysis on smooth manifolds. Trans. Am. Math. Soc. 359(8), 3687–3732 (2007)
32. Penot, J.P.: Calculus Without Derivatives, Graduate Texts in Mathematics, vol. 266. Springer, New York (2013)
33. Azagra, D., Ferrera, J., López-Mesas, F.: Nonsmooth analysis and Hamilton–Jacobi equations on Riemannian manifolds. J. Funct. Anal. 220(2), 304–361 (2005)
34. Motreanu, D., Pavel, N.: Quasi-tangent vectors in flow-invariance and optimization problems on Banach manifolds. J. Math. Anal. Appl. 88(1), 116–132 (1982)
35. Grigor’yan, A.: Heat Kernel and Analysis on Manifolds, AMS/IP Studies in Advanced Mathematics, vol. 47. American Mathematical Society, Providence (2009)
36. Lang, S.: Differential and Riemannian Manifolds, Graduate Texts in Mathematics, vol. 160, 3rd edn. Springer, New York (1995)
37. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation. I, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 330. Springer, Berlin (2006)
38. Németh, S.Z.: Variational inequalities on Hadamard manifolds. Nonlinear Anal. 52(5), 1491–1498 (2003)
39. Kristály, Á.: Nash-type equilibria on Riemannian manifolds: a variational approach. J. Math. Pures Appl. (9) 101(5), 660–688 (2014)
40. Daneshgar, A., Hajiairolhassan, H., Javadi, R.: On the isoperimetric spectrum of graphs and its approximations. J. Comb. Theory Ser. B 100(4), 390–412 (2010)
41. Tudisco, F., Hein, M.: A nodal domain theorem and a higher-order Cheeger inequality for the graph p-Laplacian. J. Spectr. Theory 8(3), 883–908 (2018)
42. von Luxburg, U.: A tutorial on spectral clustering. Stat. Comput. 17(4), 395–416 (2007)
43. Rothaus, O.S.: Analytic inequalities, isoperimetric inequalities and logarithmic Sobolev inequalities. J. Funct. Anal. 64(2), 296–313 (1985)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.