OPTIMALITY CONDITIONS FOR GLOBAL MINIMA OF NONCONVEX FUNCTIONS ON RIEMANNIAN MANIFOLDS

S. HOSSEINI

Abstract. A version of Lagrange multipliers rule for locally Lipschitz functions is presented. Using Lagrange multipliers, a sufficient condition for x to be a global minimizer of a locally Lipschitz function defined on a Riemannian manifold, is proved. Then, a necessary and sufficient condition for feasible point x to be a global minimizer of a concave function on a Riemannian manifold is obtained.

1. introduction

Nonsmooth optimization refers to the general problem of minimizing (or maximizing) functions that are typically not differentiable at their minimizers (maximizers). Since the classical theory of optimization presumes certain differentiability and strong regularity assumptions upon the functions to be optimized, it cannot be directly utilized. However, due to the complexity of the real world, functions involved in practical applications are often nonsmooth. That is, they are not necessarily differentiable.

Nonsmooth optimization problems, in general, are difficult to solve, especially when they are constrained. In last decades global optimization problems were studied intensively, [12, 14, 17], because there exists a large number of real-life applications where it is necessary to solve such problems; see [8, 9].

Optimization on nonlinear spaces finds also a lot of applications, such as in computer vision, signal processing, motion and structure estimation; see [1, 2, 24]. A manifold, in general, does not have a linear structure, hence the usual techniques, which are often used to study optimization problems on linear spaces, cannot be applied. Therefore, new techniques are needed for dealing with optimization problems posed on manifolds. Tools from Riemannian geometry have been used in mathematical programming to obtain both theoretical results and practical algorithms; see [5, 6, 7, 11, 25]. In considering optimization problems with nonsmooth objective functions on Riemannian manifolds, generalization of the concepts of nonsmooth analysis on Riemannian manifolds are of essential importance. In the past few years, a number of results have been obtained on numerous aspects of nonsmooth analysis on Riemannian manifolds; [3, 4, 18, 20, 21].

The goal of this paper is to get optimality conditions for global minima of nonconvex functions on Riemannian manifolds. First, we present Lagrange multipliers

Key words and phrases. Riemannian manifolds, Lipschitz function, Clarke subdifferential, convex function.

Seminar for Applied Mathematics, ETH, Zürich, Switzerland.
seyedehsomayeh.hosseini@math.ethz.ch .

rule for the following optimization problem with equality and inequality constraints;

$$\min g_0(x), \tag{1}$$

$$\text{subject to } x \in C, g_i(x) \leq 0,\ h_j(x) = 0,\ i \in I, j \in J,$$

where $g_i : M \to \mathbb{R}$ and $h_j : M \to \mathbb{R}$ are locally Lipschitz functions on a complete Riemannian manifold M, $I = \{1, ..., n\}$, $J = \{1, ..., m\}$, $m,\ n \in \mathbb{N}$ and C is a nonempty closed convex subset of M. This Lagrange multipliers rule is a generalization of the one in [19], which requires that the objective function and the inequality constraints be Fréchet differentiable and the equality constraints be continuously differentiable. In this paper, multipliers rule is generalized in the direction of replacing the usual gradient by certain generalized gradients under Lipschitz assumptions. Using Lagrange multipliers, we provide a sufficient condition for a locally Lipschitz function to have a general minimum on a closed convex subset of a complete Riemannian manifold. It is worthwhile to mention that a necessary optimality condition for global minimization of a locally Lipschitz function defined on a Riemannian manifold was obtained in [19]. Moreover, we prove a necessary and sufficient condition for a global minimum of a concave function on a Riemannian manifold. It is worth pointing out that our key tool is Ekeland's variational principle, hence we shall work only with complete Riemannian manifolds.

2. PRELIMINARIES

In this paper, we use the standard notations and known results of Riemannian manifolds, see, e.g. [13, 23, 22]. Throughout this paper, M is an n-dimensional complete manifold endowed with a Riemannian metric $\langle ., . \rangle$ on the tangent space T_xM. As usual we denote by $B_\delta(x)$ the open ball centered at x with radius δ, by $\mathrm{int}N(\mathrm{cl}N)$ the interior (closure) of the set N, by $\mathrm{conv}N$ the convex hull of the set N. Also, let S be a nonempty closed subset of a Riemannian manifold M, we define $d_S : M \longrightarrow \mathbb{R}$ by

$$d_S(x) := \inf\{d(x, s) : s \in S\ \},$$

where d is the Riemannian distance on M. Recall that the set S in a Riemannian manifold M is called convex if every two points $p_1,\ p_2 \in S$ can be joined by a unique geodesic whose image belongs to S. For the point $x \in M$, $\exp_x : U_x \to M$ will stand for the exponential function at x, where U_x is an open subset of T_xM. Recall that $\exp_x$ maps straight lines of the tangent space T_xM passing through $0_x \in T_xM$ into geodesics of M passing through x.

In the present paper, we are concerned with the minimization of locally Lipschitz functions which we now define.

Definition 2.1 (Lipschitz Condition)**.** *Recall that a real valued function f defined on a Riemannian manifold M is said to satisfy a Lipschitz condition of rank k on a given subset S of M if $\mid f(x) - f(y) \mid \leq kd(x, y)$ for every $x, y \in S$, where d is the Riemannian distance on M. A function f is said to be Lipschitz near $x \in M$ if it satisfies the Lipschitz condition of some rank on an open neighborhood of x . A function f is said to be locally Lipschitz on M if f is Lipschitz near x, for every $x \in M$.*

Let us continue with the definition of the Clarke generalized directional derivative for locally Lipschitz functions on Riemannian manifolds; see [18, 21].

Definition 2.2 (Clarke generalized directional derivative)**.** *Suppose $f : M \to \mathbb{R}$ is a locally Lipschitz function on a Riemannian manifold M. Let $\phi_x : U_x \to T_xM$ be an exponential chart at x. Given another point $y \in U_x$, consider $\sigma_{y,v}(t) := \phi_y^{-1}(tw)$, a geodesic passing through y with derivative w, where (ϕ_y, y) is an exponential chart around y and $D(\phi_x o \phi_y^{-1})(0_y)(w) = v$. Then, the generalized directional derivative of f at $x \in M$ in the direction $v \in T_xM$, denoted by $f^\circ(x; v)$, is defined as*

$$f^\circ(x; v) = \limsup_{y \to x,\ t \downarrow 0} \frac{f(\sigma_{y,v}(t)) - f(y)}{t}.$$

If f is differentiable in $x \in M$, we define the gradient of f at x as the unique vector $\operatorname{grad} f(x) \in T_xM$ which satisfies

$$\langle \operatorname{grad} f(x), \xi \rangle = df(x)(\xi) \quad \text{for all } \xi \in T_xM.$$

Using the previous definition of a Riemannian Clarke derivative we can also generalize the notion of subdifferential to a Riemannian context.

Definition 2.3 (Subdifferential)**.** *We define the subdifferential of f, denoted by $\partial f(x)$, as the subset of T_xM whose support function is $f^\circ(x; .)$. It can be proved* [18] *that*

$$\partial f(x) = \mathrm{conv}\{\lim_{\mathrm{i} \to \infty} \mathrm{grad\, f(x_i)} :\ \{\mathrm{x_i}\} \subseteq \Omega_\mathrm{f},\ \mathrm{x_i} \to \mathrm{x}\},$$

where Ω_f is a dense subset of M on which f is differentiable.

It is worthwhile to mention that $\lim \operatorname{grad} f(x_i)$ in the previous definition is obtained as follows. Let $\xi_i \in T_{x_i}M$, $i = 1, 2, ...$ be a sequence of tangent vectors of M and $\xi \in T_xM$. We say ξ_i converges to ξ, denoted by $\lim \xi_i = \xi$, provided that $x_i \to x$ and, for any smooth vector field X, $\langle \xi_i, X(x_i) \rangle \to \langle \xi, X(x) \rangle$.

Note that function $f : M \to \mathbb{R}$ is convex if and only if, for any geodesic segment γ, the composition $f \circ \gamma$ is convex (in the usual sense). Given $x \in M$, a vector $s \in T_xM$ is said to be a subgradient of a convex function f at x, iff for any geodesic segment γ with $\gamma(0) = x$ and every t in domain γ,

$$(f \circ \gamma)(t) \geq f(x) + t\langle s, \gamma^\circ(0) \rangle.$$

The set of all subgradients of f at x is called the subdifferential of f at x.

Definition 2.4 (Regular function)**.** *Let $f : M \to \mathbb{R}$ be a locally Lipschitz function defined on a Riemannian manifold. If the directional derivative of f at x in the direction $v \in T_xM$, defined by*

$$f'(x; v) = \lim_{t \downarrow 0} \frac{f \circ \exp_x(tv) - f(x)}{t},$$

exists and $f^\circ(x; v) = f'(x; v)$ for every $v \in T_xM$, then f is regular.

Now, we recall definitions of the tangent and normal cones to a closed subset of a Riemannian manifold.

Definition 2.5 (Clarke tangent Cone)**.** *Let S be a nonempty closed subset of a Riemannian manifold M, $x \in S$ and (φ, U) be a chart of M at x. Then the (Clarke) tangent cone to S at x, denoted by $T_S(x)$ is defined as follows;*

$$T_S(x) := D\varphi(x)^{-1}[T_{\varphi(S \cap U)}(\varphi(x))],$$

where $T_{\varphi(S \cap U)}(\varphi(x))$ is the tangent cone to $\varphi(S \cap U)$ as a subset of $\mathbb{R}^n$.

Obviously, $0_x \in T_S(x)$ and $T_S(x)$ is closed and convex.

Remark 2.6. The definition of $T_S(x)$ does not depend on the choice of the chart φ at x, [18]. Hence, for any normal neighborhood U of x, we have that

$$T_S(x) = T_{\exp_x^{-1}(S\cap U)}(0_x). \tag{2}$$

In the case of submanifolds of $\mathbb{R}^n$, the tangent space and the normal space are orthogonal to one another. In an analogous manner, for a closed subset S of a Riemannian manifold, the normal cone to S at x, denoted $N_S(x)$, is defined as the (negative) polar of the tangent cone $T_S(x)$, i.e.

$$N_S(x) := T_S(x)^\circ := \{\xi \in T_xM : \langle \xi, z\rangle \le 0 \quad \forall z \in T_S(x)\}.$$

Let S be a closed convex subset of a Riemannian manifold M, the normal cone to S at $x \in S$, denoted by $N_S(x)$, is as follows;

$$N_S(x) = \{\xi \in T_xM : \langle \xi, \exp_x^{-1}(y)\rangle \le 0 \text{ for every } y \in S\}.$$

Reader can refer to [18, 20, 21] for more details about the normal cone and the tangent cone.

Lemma 2.7. [15] *Let M be a complete Riemannian manifold. If $d_p : M \to \mathbb{R}$ is defined by $d_p(q) = d(p,q)$, then*

$$\partial d_p(p) = B,$$

where B is the closed unit ball of T_pM.

The differentiability of d^2 on all $M \times M$ is equivalent to the condition that any two points of M have a unique minimal geodesic connection and this is equivalent to the condition "points have a unique geodesic connection". Examples for manifolds of this type are simply connected complete Riemannian manifolds with nonpositive sectional curvature, the so-called Hadamard manifolds. But it is not true that all manifolds with unique geodesic connection for any two given points, are Hadamard manifolds. There exist examples of complete manifolds with unique geodesic connection for any two given points, where the sectional curvature changes sign; see [16, 26]. Note that if C is a convex closed subset of a complete Riemannian manifold, then for every $p \in C$, $d_p^2 : C \to \mathbb{R}$ is differentiable and

$$\operatorname{grad} \frac{1}{2} d_p^2(q) = -\exp_q^{-1}(p).$$

Moreover, $d_p : C \setminus \{p\} \to \mathbb{R}$ is differentiable and

$$\operatorname{grad} d_p(q) = -\frac{\exp_q^{-1}(p)}{d(p,q)}.$$

We finish this section with Ekeland's variational principle on complete Riemannian manifolds; see [3].

Theorem 2.8. *Let M be a complete Riemannian manifold, and let $f : M \to \mathbb{R} \cup \{-\infty\}$ be a proper upper semicontinuous function, which is bounded above. Let $\epsilon > 0$ and $x_0 \in M$ such that $f(x_0) > \sup\{f(x) : \ x \in M \ \} - \epsilon$. Then, for every $\lambda > 0$ there exists a point $z \in \operatorname{dom}(f) = \{s \in M : \ f(s) > -\infty\}$ such that*
(i) $\frac{\epsilon}{\lambda} d(z, x_0) \le f(z) - f(x_0)$
(ii) $d(z, x_0) \le \lambda$
(iii) $\frac{\epsilon}{\lambda} d(x, z) + f(z) > f(x)$ *whenever* $x \neq z$.

3. MAIN RESULTS

The classical Lagrange multipliers rule on linear spaces usually needs that the objective function and the inequality constraints be Fréchet differentiable and the equality constraints be continuously differentiable. Most extensions of the classical Lagrange multipliers on linear spaces are obtained under two different assumptions: differentiability and Lipschitz continuity. On one hand, the classical multipliers rule was given in the direction of eliminating the smoothness assumption while keeping the differentiability assumption. On the other hand, the classical multipliers rule was generalized with replacing the usual gradient by certain generalized gradients under Lipschitz assumptions such as in [10].

In [19], a generalization of the classical Lagrange multipliers rule, which requires differentiability of the objective function and constraints on a Riemannian manifold M was presented. In this paper, we show a generalization of the classical Lagrange multipliers rule by replacing the usual gradient by subdifferential of Lipschitz functions.

Let C be a closed and nonempty subset of a complete Riemannian manifold M. For given nonnegative integers n and m, we denote $I = \{1, 2, ..., n\}$ and $J = \{1, 2, ..., m\}$. We assume that the following locally Lipschitz functions are given;

$$g_i : M \to \mathbb{R},\ i \in I \cup \{0\},$$

$$h_j : M \to \mathbb{R},\ j \in J.$$

Now, we consider the following problem with constraints on a complete Riemannian manifold M;

$$\min g_0(x), \tag{3}$$

$$\text{subject to } x \in C,\ g_i(x) \leq 0,\ h_j(x) = 0,\ i \in I, j \in J.$$

Theorem 3.1. *If x solves (3) locally, then there exist numbers r_0, r_i, s_j, $i \in I, j \in J$ not all zero and a vector $\xi \in T_xM$ such that*
(a) $r_0 \geq 0$, $r_i \geq 0$, $i \in I$.
(b) $r_i g_i(x) = 0$, $i \in I$.
(c) $\xi \in r_0 \partial g_0(x) + \Sigma_{i \in I} r_i \partial g_i(x) + \Sigma_{j \in J} s_j \partial h_j(x)$.
(d) $-\xi \in N_C(x)$.

Proof. Assume that $\varepsilon > 0$ is so small that $\exp_x : B_\varepsilon(0_x) \to B_\varepsilon(x)$ is diffeomorphism and x is a solution of (3) on $B_\varepsilon(x) \cap C$. We consider the following minimization problem;

$$\min g_0 \circ \exp_x(v), \tag{4}$$

$$\text{subject to } v \in \exp_x^{-1}(C \cap B_\varepsilon(x)),$$

$$g_i \circ \exp_x(v) \leq 0,\ h_j \circ \exp_x(v) = 0,\ i \in I, j \in J.$$

Hence, x solves (3) locally if and only if 0_x solves (4) locally. By Lagrange multipliers rule for locally Lipschitz functions defined on linear spaces; see[10], there exist numbers r_0, r_i, s_j, $i \in I$, $j \in J$ not all zero and a vector $\xi \in T_xM$ such that
(a') $r_0 \geq 0$, $r_i \geq 0$, $i \in I$.
(b') $r_i g_i \circ \exp_x(0_x) = 0$, $i \in I$.
(c') $\xi \in r_0 \partial(g_0 \circ \exp_x)(0_x) + \Sigma_{i \in I} r_i \partial(g_i \circ \exp_x)(0_x) + \Sigma_{j \in J} s_j \partial(h_j \circ \exp_x)(0_x)$.

(d') $-\xi \in N_{\exp_x^{-1}(C\cap B_\varepsilon(x))}(0_x)$,
and the proof is complete. $\square$

Example 3.2. Let us consider the set $\mathrm{Pos}_2(\mathbb{R})$ of symmetric positive definite 2×2 matrices and the set $\mathrm{Sym}_2(\mathbb{R})$ of symmetric 2×2 matrices endowed with the Frobenius metric $\langle U,V\rangle_X = \mathrm{tr}(X^{-1}UX^{-1}V)$, where $X \in \mathrm{Pos}_2(\mathbb{R})$ and $U,V \in T_X(\mathrm{Pos}_2\mathbb{R}) = \mathrm{Sym}_2(\mathbb{R})$. The set $\mathrm{Pos}_2(\mathbb{R})$ is a Hadamard manifold; see[22]. Consider the following problem on $\mathrm{Pos}_2(\mathbb{R})$;

$$
\begin{array}{lll}
(\mathbf{P_1}) & \min & g_0(X) = |x_1 - 1|, \\
 & \text{s.t.} & g(X) = |x_2| + |x_3| - 7 \le 0 \\
 & & h(X) = -x_1 + 1 \le 0 \\
 & & C = \{A \in \mathrm{Pos}_2(\mathbb{R}) :\ \det(A) = 1\}. \\
 & & X = \begin{bmatrix} x_1 & x_2 \\ x_2 & x_3 \end{bmatrix} \in \mathrm{Pos}_2(\mathbb{R}),
\end{array}
$$

It is easy to check that

$$\bar{X} = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix},$$

is a global optimal solution for $(\mathbf{P_1})$. Moreover, we have

$$\partial g_0(\bar{X}) = \{\begin{bmatrix} t & 0 \\ 0 & 0 \end{bmatrix},\ t \in [-1,1]\}.$$

Note that

$$\operatorname{grad} h(\bar{X}) = \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix}.$$

It is trivial that for $r_0 = 1$, $r_1 = 0$, $s_1 = 1$, we have $0 \in \partial g_0(\bar{X}) + r_1\partial g(\bar{X}) + s_1\partial h(\bar{X})$, moreover $0 \in N_C(\bar{X})$.

Now by means of the notion of the subdifferential and Lagrange multipliers, we present a sufficient condition of optimality for the problem;

$$\min_{x\in C} f(x),$$

where $f : M \to \mathbb{R}$ is a locally Lipschitz function and C is an arbitrary nonempty closed convex subset of a Riemannian manifold M. It is worthwhile to mention that a necessary condition for a feasible point x to be a global minimizer of a locally Lipschitz function on a Riemannian manifold was presented in [19].

Theorem 3.3 (**Sufficient condition for a global minimum**). *Consider the following minimization problem;*

$$\min f(x), \tag{5}$$

$$\textit{subject to } x \in C,$$

where $f : M \to \mathbb{R}$ is a locally Lipschitz function, C is a nonempty closed convex subset of a complete Riemannian manifold M and $f^{-1}(c) = \{y \in C :\ f(y) = c\}$. If

$$\partial f(y) \cap N_C(y) = \emptyset \quad \forall\, y \in f^{-1}(f(z)), \tag{6}$$

and

$$-\partial f(y) \subset N_C(y) \quad \forall\ y \in f^{-1}(f(z)),$$

then $z \in C$ is a global solution to Problem (5).

Proof. Assume that z is not a solution to Problem (5). Hence, there exists $u \in C$ such that $f(u) < f(z)$. We define a closed subset D of M, by $D := \{x \in C :\ f(x) \geq f(z)\}$. Note that $u \notin D$, and $\varepsilon = \dfrac{1}{2}d_D(u) > 0$. By Ekeland's variational principle, we obtain a point $y \in D$ such that

$$g(y) \leq g(x) + \varepsilon d(x, y) \qquad \forall x \in D,$$

where $g(x) = \dfrac{1}{2}d(x, u)^2$. Thus, y is a solution for the following minimization problem;

$$\min \frac{1}{2}d(x, u)^2 + \varepsilon d(y, x), \tag{7}$$

$$\text{subject to } x \in C,$$

$$-f(x) \leq -f(z).$$

By Theorem 3.1, there exist numbers $r_0, r_1 \geq 0$ and a vector $\xi \in T_yM$ such that

$$r_1(f(y) - f(z)) = 0, \tag{8}$$

and

$$\xi \in r_0\partial(\frac{1}{2}d(., u)^2 + \varepsilon d(y, .))(y) + r_1\partial(-f)(y), \tag{9}$$

$$-\xi \in N_C(y). \tag{10}$$

Hence, Proposition 3.1 in [18] and Lemma 2.7 imply the existence of $\eta_1 = -\exp_y^{-1}(u)$, $\eta_2 \in \partial d_y(y) = B$ and $\eta_3 \in \partial f(y)$ such that

$$\xi = r_0(\eta_1 + \varepsilon\eta_2) - r_1\eta_3.$$

If $r_0 = 0$, $r_1 > 0$, then by (8), $f(y) = f(z)$. Therefore, $\eta_3 = -\frac{1}{r_1}\xi \in \partial f(y) \cap N_C(y)$, which contradicts our assumption. Hence, $r_0 > 0$. Assuming that $r_1 = 0$, since $-\xi \in N_C(y)$, we have $\langle \xi, -\exp_y^{-1}(u)\rangle \leq 0$. Moreover,

$$\begin{aligned} 0 \geq &\langle \xi, -\exp_y^{-1}(u)\rangle \\ &= r_0(d(y, u)^2 + \varepsilon\langle \eta_2, -\exp_y^{-1}(u)\rangle) \\ &\geq r_0(d(y, u)^2 - \varepsilon\|\eta_2\|d(u, y)) \\ &\geq r_0 d(u, y)(d_D(u) - \varepsilon) \\ &> 0, \end{aligned} \tag{11}$$

as a contradiction. Whence, we must have $r_1 > 0$ and $f(y) = f(z)$. We consider

$$\eta_3 = \frac{1}{r_1}(r_0(\eta_1 + \varepsilon\eta_2) - \xi).$$

Then,

$$\begin{aligned}\langle \eta_3, -\exp_y^{-1}(u)\rangle =& \frac{1}{r_1}\langle r_0(\eta_1+\varepsilon\eta_2)-\xi, -\exp_y^{-1}(u)\rangle\\ \geq& \frac{r_0}{r_1}(d(u,y)^2+\varepsilon\langle \eta_2, -\exp_y^{-1}(u)\rangle)\\ \geq& \frac{r_0}{r_1}(d(u,y)^2-\varepsilon\|\eta_2\| d(u,y))\\ \geq& \frac{r_0}{r_1}d(u,y)(d(u,y)-\varepsilon)>0,\end{aligned} \tag{12}$$

which is another contradiction and the proof is complete. □

Theorem 3.4 (**Necessary and Sufficient condition for a global maximum of a convex function).** *Consider the following maximization problem;*

$$\max f(x), \tag{13}$$

$$\textit{subject to } x \in C,$$

where $f : M \to \mathbb{R}$ is a convex function, C is a nonempty closed convex subset of a complete Riemannian manifold M and $f^{-1}(c) = \{y \in C : \ f(y) = c\}$. Consider a point $\bar{x} \in C$ such that

$$-\infty \leq \inf_C f < f(\bar{x}).$$

Then, $\bar{x}$ is a global maximum of f on C if and only if

$$\partial f(x) \subset N_C(x) \qquad \textit{for all } x \in f^{-1}(f(\bar{x})). \tag{14}$$

Proof. Let $\bar{x}$ be a global maximum of f on C and $\xi \in \partial f(x)$, where $x \in f^{-1}(f(\bar{x}))$. Hence, for the unique minimal geodesic $\gamma(t) = \exp_x(t\exp_x^{-1}(x'))$ connecting x and an arbitrary point $x' \in C$,

$$\langle \xi, \gamma^\circ(0)\rangle \leq f(x') - f(x).$$

Note that $\gamma^\circ(0) = \exp_x^{-1}(x')$ and $f(x) = f(\bar{x})$, hence $f(x') \leq f(x)$. Consequently, for every $x' \in C$,

$$\langle \xi, \exp_x^{-1}(x')\rangle \leq 0,$$

which implies $\xi \in N_C(x)$.

We claim that (14) is equivalent to saying for every $x \in f^{-1}(f(\bar{x}))$ and $c \in C$,

$$f'(x, \exp_x^{-1}(c)) \leq 0.$$

Since if $\xi \in \partial f(x) \subset N_C(x)$,

$$\langle \xi, \exp_x^{-1}(c)\rangle \leq 0 \qquad \forall c \in C,$$

then by the definition of support function,

$$\sup_{\xi \in \partial f(x)} \langle \xi, \exp_x^{-1}(c)\rangle = f'(x, \exp_x^{-1}(c)) \leq 0.$$

For the converse, assume that for every $c \in C$, $f'(x, \exp_x^{-1}(c)) \leq 0$. Suppose that the contrary holds; let $\xi \in \partial f(x) \setminus N_C(x)$, then there exists $c \in C$ such that

$$0 < \langle \xi, \exp_x^{-1}(c)\rangle \leq f'(x, \exp_x^{-1}(c)) \leq 0,$$

which is a contradiction and the claim is proved.

Now, we prove the sufficient condition. If $\bar{x}$ is not a maximum of f on C. Then, there exists $z \in C$ such that

$$f(z) > f(\bar{x}).$$

Since $\inf_C f < f(\bar{x})$, there exists $y \in C$ such that $f(y) < f(\bar{x})$. Assume that γ is the unique minimal geodesic connecting y and z, continuity of $f \circ \gamma$ implies the existence of t_0 such that $f \circ \gamma(t_0) = f(\bar{x})$. Convexity of f implies that f is increasing along γ . Therefore, $f'(\gamma(t_0), \exp^{-1}_{\gamma(t_0)}(z)) > 0$, which is a contradiction. □

Following theorem gives a necessary and sufficient condition for a global maximum of a regular function.

Theorem 3.5 (**Necessary and Sufficient condition for a global maximum of a regular function).** *Consider the following maximization problem;*

$$\max f(x), \tag{15}$$

subject to $x \in C$,

where $f : M \to \mathbb{R}$ *is a regular function,* C *is a nonempty closed convex subset of a complete Riemannian manifold* M *and* $f^{-1}(c) = \{y \in C : \ f(y) = c\}$. *Consider a point* $\bar{x} \in C$ *such that for all* $x \in f^{-1}(f(\bar{x}))$, *there exists* $c_x \in C$,

$$f'(x, \exp_x^{-1}(c_x)) < 0. \tag{16}$$

Then, $\bar{x}$ *is a global maximum of* f *on* C *if and only if*

$$\partial f(x) \subset N_C(x) \qquad \textit{for all } x \in f^{-1}(f(\bar{x})). \tag{17}$$

Proof. Let $\bar{x}$ be a global maximum of f and $x \in f^{-1}(f(\bar{x}))$. For each $c \in C$, assume that γ_c is the unique minimal geodesic connecting x and c. Then $f(\gamma_c(t)) \leq f(x)$, hence

$$f'(x, \gamma_c^{\circ}(0)) = \lim_{t \downarrow 0} \frac{f(\gamma_c(t)) - f(x)}{t} \leq 0.$$

Therefore, for every $\xi \in \partial f(x)$, we have $\langle \xi, \gamma_c^{\circ}(0) \rangle \leq 0$, which implies $\partial f(x) \subset N_C(x)$.

For the converse, assume that $\bar{x}$ is not a global maximum of f. Hence, there exists $\tilde{x}$ such that $f(\tilde{x}) > f(\bar{x})$. We define $D := \{x \in C : f(x) \leq f(\bar{x})\}$. By Ekeland's variational principle, for every $\varepsilon > 0$, we obtain a point $y_\varepsilon \in D$ such that

$$g(y_\varepsilon) \leq g(x) + \varepsilon d(x, y_\varepsilon) \qquad \forall x \in D,$$

where $g(x) = \dfrac{1}{2} d(x, \tilde{x})^2$. First, we prove that for ε small enough with $d_D(\tilde{x}) \geq \varepsilon > 0$ $f(y_\varepsilon) = f(\bar{x})$.

Assume on the contrary that there exists a sequence $\varepsilon_k \downarrow 0$ such that $y_k = y_{\varepsilon_k}$ and $f(y_k) < f(\bar{x})$. Let γ_k be a minimal geodesic connecting y_k and $\tilde{x}$. By convexity of C, $\gamma_k \in C$ and for t small enough

$$f(\gamma_k(t)) \leq f(\bar{x}).$$

Consequently, there exists t_k such that for all positive $t \leq t_k$, $\gamma_k(t) \in D$. Moreover,

$$\frac{1}{2} d(\gamma_k(t), \tilde{x})^2 + \varepsilon_k d(\gamma_k(t), y_k) \geq \frac{1}{2} d(y_k, \tilde{x})^2 \ \text{ for all } 0 < t \leq t_k.$$

Hence,

$$\frac{1}{2}(1-t)^2 d(y_k, \tilde{x})^2 + t\varepsilon_k d(\tilde{x}, y_k) \geq \frac{1}{2} d(y_k, \tilde{x})^2 \ \text{ for all } 0 < t \leq t_k.$$

Therefore,

$$d(y_k, \tilde{x}) \leq \varepsilon_k + \frac{t}{2} d(y_k, \tilde{x}).$$

By getting limit for $t \downarrow 0$, we have $d(y_k, \tilde{x}) \leq \varepsilon_k$ and $y_k \to \tilde{x}$, as k goes to ∞. By continuity of f, $f(\tilde{x}) \leq f(\bar{x})$, which is a contradiction and the claim is proved.

Now assume ε is small enough that $f(y_\varepsilon) = f(\bar{x})$. By the assumption

$$\partial f(y_\varepsilon) \subset N_C(y_\varepsilon).$$

We assume that $y_\varepsilon \in D$ is such that

$$g(y_\varepsilon) \leq g(x) + \varepsilon d(x, y_\varepsilon) \quad \forall x \in D,$$

where $g(x) = \frac{1}{2} d(x, \tilde{x})^2$. Thus, y_ε is a solution to the following minimization problem;

$$\min \frac{1}{2} d(x, \tilde{x})^2 + \varepsilon d(y_\varepsilon, x), \tag{18}$$

$$\text{subject to } x \in C,$$

$$f(x) \leq f(\bar{x}).$$

By Theorem 3.1, there exist numbers $r_0, r_1 \geq 0$ and a vector $\xi \in T_{y_\varepsilon} M$ such that

$$r_1(f(y_\varepsilon) - f(\bar{x})) = 0, \tag{19}$$

and

$$\xi \in r_0 \partial(1/2 d(., \tilde{x})^2 + \varepsilon d(y_\varepsilon, .))(y_\varepsilon) + r_1 \partial f(y_\varepsilon), \tag{20}$$

$$-\xi \in N_C(y_\varepsilon). \tag{21}$$

Hence, Proposition 3.1 in [18] and Lemma 2.7 imply the existence of $\eta_1 = -\exp_{y_\varepsilon}^{-1}(\tilde{x})$, $\eta_2 \in \partial d_{y_\varepsilon}(y_\varepsilon) = B$ and $\eta_3 \in \partial f(y_\varepsilon)$ such that

$$\xi = r_0(\eta_1 + \varepsilon \eta_2) + r_1 \eta_3.$$

If $r_0 = 0$, $r_1 > 0$, then by (19) $f(y_\varepsilon) = f(\bar{x})$. Therefore, $\eta_3 = \frac{1}{r_1}\xi \in \partial f(y_\varepsilon) \subset N_C(y_\varepsilon)$, which means ξ and $-\xi$ are in $N_C(y_\varepsilon)$ and for every $c \in C$, $\langle \xi, \exp_{y_\varepsilon}^{-1}(c) \rangle = 0$. Hence, we have $f'(y_\varepsilon, \exp_{y_\varepsilon}^{-1}(c)) \geq 0$ for every $c \in C$, which contradicts our assumption. Consequently, $r_0 > 0$. Now assume that $r_1 = 0$, since $-\xi \in N_C(y_\varepsilon)$, we have $\langle \xi, -\exp_{y_\varepsilon}^{-1}(\tilde{x}) \rangle \leq 0$. Moreover,

$$\begin{aligned} 0 \geq & \langle \xi, -\exp_{y_\varepsilon}^{-1}(\tilde{x}) \rangle \\ & = r_0(d(y_\varepsilon, \tilde{x})^2 + \varepsilon \langle \eta_2, -\exp_{y_\varepsilon}^{-1}(\tilde{x}) \rangle) \\ & \geq r_0(d(y_\varepsilon, \tilde{x})^2 - \varepsilon \|\eta_2\| d(\tilde{x}, y_\varepsilon)) \\ & \geq r_0 d(\tilde{x}, y_\varepsilon)(d_D(\tilde{x}) - \varepsilon) \\ & > 0, \end{aligned} \tag{22}$$

as a contradiction. Therefore, we must have $r_1 > 0$ and $f(y_\varepsilon) = f(\tilde{x})$. We consider

$$-\eta_3 = \frac{1}{r_1}(r_0(\eta_1 + \varepsilon \eta_2) - \xi).$$

Then,

$$\begin{aligned}\langle \eta_3, \exp_{y_\varepsilon}^{-1}(\tilde{x})\rangle =& \frac{1}{r_1}\langle r_0(\eta_1+\varepsilon\eta_2)-\xi, -\exp_{y_\varepsilon}^{-1}(\tilde{x})\rangle \\ &\geq \frac{r_0}{r_1}(d(\tilde{x},y_\varepsilon)^2+\varepsilon\langle \eta_2, -\exp_{y_\varepsilon}^{-1}(\tilde{x})\rangle) \\ &\geq \frac{r_0}{r_1}(d(\tilde{x},y_\varepsilon)^2-\varepsilon\|\eta_2\|d(\tilde{x},y_\varepsilon)) \\ &\geq \frac{r_0}{r_1}d(\tilde{x},y_\varepsilon)(d(\tilde{x},y_\varepsilon)-\varepsilon)>0,\end{aligned} \tag{23}$$

but $\eta_3 \in \partial f(y_\varepsilon) \subset N_C(y_\varepsilon)$ and $\langle \eta_3, \exp_{y_\varepsilon}^{-1}(\tilde{x})\rangle \leq 0$ as another contradiction. □

Following lemma proves that the qualification condition in (16) and (6) are equivalent.

Lemma 3.6. *Let $f: M \to \mathbb{R}$ be a locally Lipschitz function on a complete Riemannian manifold M, and C be a nonempty closed convex subset of M and $f^{-1}(c)=\{y \in C: \ f(y)=c\}$. If $\bar{x} \in C$, then for all $x \in f^{-1}(f(\bar{x}))$, there exists $c_x \in C$,*

$$(-f)^\circ(x, \exp_x^{-1}(c_x)) < 0. \tag{24}$$

if and only if for all $x \in f^{-1}(f(\bar{x}))$,

$$N_C(x) \cap \partial f(x) = \emptyset. \tag{25}$$

Proof. $\Rightarrow$) Suppose for contradiction, there exist $x \in f^{-1}(f(\bar{x}))$ and $\xi \in \partial f(x) \cap N_C(x)$. Therefore, for every $c \in C$, $\langle \xi, -\exp_x^{-1}(c)\rangle \geq 0$, by the definition of support function and Proposition 2.4 in [18] $f^\circ(x, -\exp_x^{-1}(c)) = (-f)^\circ(x, \exp_x^{-1}(c)) \geq 0$, which is a contradiction.

$\Leftarrow$) Suppose for contradiction, there exists $x \in f^{-1}(f(\bar{x}))$ such that for every $c \in C$, $f^\circ(x, -\exp_x^{-1}(c)) \geq 0$, so there exists $\xi \in \partial f(x)$ such that $\langle \xi, -\exp_x^{-1}(c)\rangle \geq 0$, which implies $\xi \in N_C(x)$, as a contradiction. □

References

[1] P. A. Absil, R. Mahony, R. Sepulchre, Optimization Algorithm on Matrix Manifolds, Princeton University Press, 2008.

[2] R. L. Adler, J. P. Dedieu, J. Y. Margulies, M. Martens, M. Shub, Newton's method on Riemannian manifolds and a geometric model for the human spine, IMA J. Numer. Anal., **22** (2002), 359-390.

[3] D. Azagra, J. Ferrera, F. López-Mesas, Nonsmooth analysis and Hamilton-Jacobi equations on Riemannian manifolds, J. Funct. Anal., **220** (2005), 304-361.

[4] D. Azagra, J. Ferrera, Applications of proximal calculus to fixed point theory on Riemannian manifolds, Nonlinear. Anal., **67** (2007), 154-174.

[5] G. C. Bento, O. P Ferreira, P. R. Oliveira, Local convergence of the proximal point method for a special class of nonconvex functions on Hadamard manifolds, Nonlinear. Anal., **73** (2010), 564 - 572.

[6] G. C. Bento, O. P. Ferreira, P. R. Oliveira, Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds, J. Optim. Theory Appl., **154** (2012), 88 - 107.

[7] G. C. Bento, J. G. Melo, A subgradient method for convex feasibility on Riemannian manifolds, J. Optim. Theory Appl., **152** (2012), 773 - 785.

[8] S. H. Brooks, Discussion of random methods for locating surface maxima,Operation Research, **6** (1958), 244 - 251.

[9] J. Calvin, A. Zilinskas, On the convergence of the P-algorithm for one-dimensional global optimization of smooth functions, JOTA., **102** (1999), 479 - 495.
[10] F. H. Clarke, Optimization and Nonsmooth Analysis, Classics in Applied Mathematics, SIAM, Philadelphia, 1990.
[11] J. X. da Cruz Neto, L. L. Lima, P. R. Oliveira, Geodesic Algorithm in Riemannian Manifolds, Balkan J. Geom. Appl., **2** (1998), 89 - 100.
[12] M. Dür, R. Horst, M. Locatelli, Necessary and sufficient global optimality conditions for convex maximization revisited, J. Math. Anal. App., **217** (1998), 637 - 649.
[13] M. P. Do Carmo, Riemannian Geometry, Birkhäuser, Boston, 1992.
[14] P. Georgiev, A. Chinchuluun, P. Pardalos. Optimality conditions of first order for global minima of locally Lipschitz functions, Optimization., **60** (2011), 277 - 282.
[15] P. Grohs, S. Hosseini, Optimization algorithm of locally Lipschitz functions on Riemannian manifolds, Submitted.
[16] R. Gulliver, On the variety of manifolds without conjugate points, Trans. Amer. Math. Soc., **210** (1975), 185–200.
[17] J. B. Hiriart-Urruty, Y. S. Ledyaev, A note on the characterization of the global maxima of a (Tangentially) convex function over a convex set, J. Convex. Anal., **3** (1996), No. 1, 55–61.
[18] S. Hosseini, M. R. Pouryayevali, Generalized gradients and characterization of epi-Lipschitz sets in Riemannian manifolds, Nonlinear Anal., **74** (2011), 3884–3895.
[19] S. Hosseini, M. R. Pouryayevali, Nonsmooth optimization techniques on Riemannian Manifolds, JOTA., **158** (2013), 328–342.
[20] S. Hosseini, M. R. Pouryayevali, Euler characterization of epi-Lipschitz subsets of Riemannian manifolds, J. Convex. Anal., **20** (2013), No. 1, 67–91.
[21] S. Hosseini, M. R. Pouryayevali, On the metric projection onto prox-regular subsets of Riemannian manifolds, Proc. Amer. Math. Soc., **141** (2013), 233–244.
[22] S. Lang, Fundamentals of Differential Geometry, Graduate Texts in Mathematics, Vol. **191**, Springer, New York, 1999.
[23] J. M. Lee, Riemannian Manifolds, An Introuction to Curvature, Springer, 1997.
[24] P. Y. Lee, Geometric Optimization for Computer Vision, PhD thesis, Australian National University, 2005.
[25] C. Udriste, Convex Functions and Optimization Methods on Riemannian Manifolds, Kluwer Acadmic Publishers, Dordrecht, Netherlands, 1994.
[26] F. E. Wolter, Distance function and cut loci on a complete Riemannian manifold, Arch. Math., (Basel) **32** (1979) 92–96.