Institut für Numerische Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn
Wegelerstraße 6 • 53115 Bonn • Germany
phone +49 228 73-3427 • fax +49 228 73-7527
www.ins.uni-bonn.de

S. Hosseini

Convergence of nonsmooth descent methods via Kurdyka-Lojasiewicz inequality on Riemannian manifolds

INS Preprint No. 1523

Nov 2015
revised version, July 2017

CONVERGENCE OF NONSMOOTH DESCENT METHODS VIA KURDYKA-LOJASIEWICZ INEQUALITY ON RIEMANNIAN MANIFOLDS

S. HOSSEINI*

Abstract. We develop a subgradient-oriented descent method in nonsmooth optimization on Riemannian manifolds and prove convergence of the method in the sense of subsequences for nonsmooth functions whose standard models are strict. Moreover, we present a nonsmooth version of the Kurdyka-Lojasiewicz inequality and show that a locally Lipschitz $\mathcal{C}$-function defined on an analytic manifold satisfies this inequality. Finally, we prove that if the objective function satisfies the Kurdyka-Lojasiewicz inequality and its standard model is strict, then the sequence of iterates of the subgradient-oriented descent algorithm converges to a singular critical point.

1. introduction

We consider the optimization problem

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

where M is a complete Riemannian manifold of dimension n and $f : M \to \mathbb{R}$ is locally Lipschitz on M. In this paper, we develop subgradient-oriented descent methods for solving (1.1). Much attention has been paid over centuries to understanding and solving the problem of minimization of functions. Compared to linear programming and nonlinear unconstrained optimization problems, nonlinear constrained optimization problems are much more difficult. Since the procedure of finding an optimizer is a search based on the local information of the constraints and the objective function, it is very important to develop techniques using geometric properties of the constraints and the objective function. In fact, differential geometry provides a powerful tool to characterize and analyze these geometric properties. Thus, there is clearly a link between the techniques of optimization on manifolds and standard constrained optimization approaches. Furthermore, there are manifolds that are not defined as constrained sets in $\mathbb{R}^n$; an important example is a Grassmann manifold. Hence, to solve optimization problems on these spaces, intrinsic methods are used.

Unconstrained smooth optimization algorithms on linear spaces can be classified into two main categories: line-search descent methods and trust-region methods. The classical convergence results established for these two classes of methods show

Key words and phrases. Riemannian manifolds, Lipschitz functions, Descent directions, Clarke subdifferential.

AMS Subject Classifications: 49J52, 65K05, 58C05.

* *Hausdorff Center for Mathematics and Institute for Numerical Simulation, University of Bonn, 53115 Bonn, Germany (hosseini@ins.uni-bonn.de).*

that accumulation points of the sequence of iterates are critical points of the objective function. But the convergence of the whole sequence to a single limit-point is not guaranteed. Though if it is known that a point x^* is an accumulation point of the sequence of iterates, then in order to have convergence of the whole sequence to x^*, it is sufficient to require that the so-called Kurdyka-Lojasiewicz inequality holds in a neighborhood of x^*; see [2].

In nonsmooth optimization problems on linear spaces, several numerical algorithms have been so far proposed and their convergence results have been studied. In particular, it has been proved in [18] that convergence of a class of nonsmooth algorithms for locally Lipschitz objective functions, even in the case of subsequences, only happens if the objective function has a strict standard model. To prove convergence of the whole sequence to a single critical point, nonsmooth generalizations of Kurdyka-Lojasiewicz were needed to be exploited; see [5, 6].

There are a number of problems that can be expressed as a minimization of a function over a smooth Riemannian manifold. Applications range from linear algebra to the analysis of shape spaces; see [1] and references therein. Therefore, there have been some attempts to adapt standard optimization methods to problems on manifolds. Line-search and trust region techniques were proposed and analyzed on manifolds by several authors; see, e.g., [1, 19, 20, 21]. For instance; in [19] line-search optimization methods on Riemannian manifolds are introduced. Moreover, the requirements on the search direction, focusing on the key property of the angle between the search direction and the negative of the Riemannian gradient are discussed. Similar to linear cases for proving convergence of the whole sequence of iterations to a single critical point, an extension of Kurdyka-Lojasiewicz inequality is required. Lageman [15] extended the Kurdyka-Lojasiewicz inequality for analytic manifolds and differentiable $\mathcal{C}$-functions in an analytic-geometric category (satisfying a certain descent condition, namely, angle and Wolfe-Powell conditions) and established an abstract result of convergence of the descent method, see [15, Theorem 2.1.22]. It is also worth pointing out [4] which presents an abstract convergence analysis of inexact descent methods in Riemannian context for functions satisfying Kurdyka-Lojasiewicz inequality. In particular, without any restrictive assumption about the sign of the sectional curvature of the manifold, it obtains full convergence of a bounded sequence generated by the proximal point method, in the case that the objective function is nonsmooth and nonconvex, and the subproblems are determined by a quasi distance which does not necessarily coincide with the Riemannian distance.

Although, the theory and algorithms for optimization of smooth functions on a Riemannian manifold is a well established topic, the case of nonsmooth function is not so developed, because results are not so simple. Papers [8, 9, 11, 12] are among the first papers on numerical algorithms for minimization of nonsmooth functions on Riemannian manifolds. However, the convergence results established in those papers show that accumulation points of the sequence of iterates are critical points of the objective function. But the convergence of the whole sequence to a single limit-point is not guaranteed.

Our main contributions in this paper are twofold. First, we innovate a nonsmooth descent optimization algorithm for locally Lipschitz functions on Riemannian manifolds. Second, we fill the gap among the convergence results obtained in previous nonsmooth optimization algorithms on Riemannian manifolds and present

some requirements to prove the convergence of a nonsmooth descent method to a single limit-point. To this goal, we extend the Kurdyka-Lojasiewicz inequality for nonsmooth functions on Riemannian manifolds and prove that a locally Lipschitz $\mathcal{C}$-function defined on an analytic manifold satisfies this inequality. Then, we consider a nonsmooth generalization of the Taylor expansion to define a first order model for a locally Lipschitz function defined on a Riemannian manifold. It is worthwhile to mention that the generalized directional derivative defines a first order model which is called a standard model. Moreover, we present a definition of a strict first order model for a locally Lipschitz function defined on a Riemannian manifold. Finally, we prove convergence of our proposed descent method to a single limit-point for a locally Lipschitz objective function defined on a Riemannian manifold satisfying the Kurdyka-Lojasiewicz inequality with a strict standard model.

2. Preliminaries

In this paper, we use the standard notations and known results of Riemannian manifolds, see, e.g. [16]. 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 $B(x,\delta)$ the open ball centered at x with radius δ, by $\operatorname{int} N(\operatorname{cl} N)$ the interior (closure) of the set N.

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.

We will also use the parallel transport of vectors along geodesics. Recall that, for a given curve $\gamma : I \to M$, number $t_0 \in I$, and a vector $V_0 \in T_{\gamma(t_0)}M$, there exists a unique parallel vector field $V(t)$ along $\gamma(t)$ such that $V(t_0) = V_0$. Moreover, the map defined by $V_0 \mapsto V(t_1)$ is a linear isometry between the tangent spaces $T_{\gamma(t_0)}M$ and $T_{\gamma(t_1)}M$, for each $t_1 \in I$. In the case when γ is a minimizing geodesic and $\gamma(t_0) = x, \gamma(t_1) = y$, we will denote this map by L_{xy}, and we will call it the parallel transport from T_xM to T_yM along the curve γ. Note that, L_{xy} is well defined when the minimizing geodesic which connects x to y, is unique. For example, the parallel transport L_{xy} is well defined when x and y are contained in a convex neighborhood. In what follows, L_{xy} will be used wherever it is well defined. We use of a class of mappings called retractions:

Definition 2.1 (Retraction)**.** *A retraction on a manifold M is a smooth map $R : TM \to M$ with the following properties. Let R_x denote the restriction of R to T_xM.*

- *$R_x(0_x) = x$, where 0_x denotes the zero element of T_xM.*
- *With the canonical identification $T_{0_x}T_xM \simeq T_xM$, $DR_x(0_x) = id_{T_xM}$, where id_{T_xM} denotes the identity map on T_xM.*

By the inverse function Theorem, we have that R_x is a local diffeomorphism. For example, the exponential function defined by $\exp : TM \to M$, $v \in T_xM \to \exp_x v$, $\exp_x(v) = \gamma(1)$, where γ is a geodesic starting at x with $\gamma'(0) = v$, is a retraction; see [1].

To prove our results, the retractions in this paper must satisfy the following condition: for all $x \in M$ and $g \in T_xM$, there exist $m_1 > 0$ and $m_2 > 0$ such that

$$m_1\|g\| \leq \operatorname{dist}(x, R_x(g)) \leq m_2\|g\|,$$

where dist is the Riemannian distance on M.

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

Definition 2.2 (Lipschitz condition)**.** *Recall that a real valued function f defined on a Riemannian manifold M is said to satisfy a Lipschitz condition of constant k on a given subset S of M if $|\ f(x)-f(y)\ | \leq k\,\mathrm{dist}(x,y)$ for every $x, y \in S$, where* dist *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 constant 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 [10, 13].

Definition 2.3 (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 $D(\phi_x \circ \phi_y^{-1})(0_y)(w) = v$. Then, the Clarke 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*

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

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

Definition 2.4 (Subdifferential)**.** *We define the subdifferential of f at x, denoted by $\partial f(x)$, as the subset of T_xM with support function given by $f^\circ(x; .)$, i.e., for every $v \in T_xM$,*

$$f^\circ(x; v) = \sup\{\langle \xi, v\rangle :\ \xi \in \partial f(x)\}.$$

Every element of $\partial f(x)$ is called a subgradient of f at x. If x is a solution of the problem (1.1), then $0 \in \partial f(x)$. Moreover, if $0 \in \partial f(x)$, then x is called a critical point for f. We say p is a descent direction at x, if there exists $\alpha > 0$ such that for every $t \in (0, \alpha)$, we have

$$f(R_x(tp)) - f(x) < 0.$$

It is obvious that if $f^\circ(x; p) < 0$, then p is a descent direction at x.

We know that the search direction for a smooth optimization problem often has the form $p = -P\,\mathrm{grad}\, f(x)$, where $\mathrm{grad}\, f(x)$ denotes the Riemannian gradient of f at x and P is a symmetric non-singular linear map. Therefore, it is not far from expectation to use elements of the subdifferential of f at x in Definition 2.5 and produce a subgradient-oriented descent sequence in nonsmooth problems.

Definition 2.5 (Subgradient-oriented descent sequence)**.** *A sequence $\{p_k\}$ of normalized descent directions is called subgradient-oriented if there exist a sequence of subgradients $\{g_k\}$ and a sequence of positive definite linear maps $\{P_k : T_{x_k}M \to T_{x_k}M\}$ satisfying*

$$\lambda \leq \lambda_{min}(P_k) \leq \lambda_{max}(P_k) \leq \Lambda \quad \text{for some } 0 < \lambda < \Lambda < \infty \text{ and all } k \in \mathbb{N},$$

where $\lambda_{min}(P_k)$ *and* $\lambda_{max}(P_k)$ *denote respectively the smallest and largest eigenvalues of* P_k*, such that* $p_k = \dfrac{-P_k g_k}{\|P_k g_k\|}$.

3. Nonsmooth Kurdyka-Lojasiewicz inequality on Riemannian manifolds

In this section, we present a nonsmooth version of the Kurdyka-Lojasiewicz inequality. Then, we prove that a locally Lipschitz $\mathcal{C}$-function defined on an analytic manifold satisfies the Kurdyka-Lojasiewicz inequality at every point of its domain.

Definition 3.1 (The Kurdyka-Lojasiewicz inequality)**.** *A locally Lipschitz function* $f : M \to \mathbb{R}$ *satisfies the Kurdyka-Lojasiewicz inequality at* $x \in M$ *iff there exist* $\eta \in (0, \infty)$*, a neighborhood* U *of* x*, and a continuous concave function* $\kappa : [0, \eta] \to [0, \infty)$ *such that*

- $\kappa(0) = 0$,
- κ *is of class* C^1 *on* $(0, \eta)$,
- $\kappa' > 0$ *on* $(0, \eta)$,
- *For every* $y \in U$ *with* $f(x) < f(y) < f(x) + \eta$*, we have*

$$\kappa'(f(y) - f(x)) \operatorname{dist}(0, \partial f(y)) \geq 1,$$

where $\operatorname{dist}(0, \partial f(y)) = \inf\{\|v\| : \quad v \in \partial f(y)\}$.

The following lemma states that a locally Lipschitz function f defined on a Riemannian manifold M satisfies the Kurdyka-Lojasiewicz inequality at any noncritical point x.

Lemma 3.2. *Let* $f : M \to \mathbb{R}$ *be a locally Lipschitz function defined on a Riemannian manifold* M *and* $0 \notin \partial f(x)$*. Then,* f *satisfies the Kurdyka-Lojasiewicz inequality at* x*.*

Proof. First we claim that there exist a neighborhood $U(x)$ and a scalar $\delta > 0$ such that for every $y \in U(x)$, $\operatorname{dist}(0, \partial f(y)) > \delta$. We prove the claim by contradiction, assume that there exist sequences $y_i \in B(x, \frac{1}{i}) \subset \operatorname{cl} B(x, 1)$ and $v_i \in \partial f(y_i)$ such that $\|v_i\| \leq \dfrac{1}{i}$. Since f is Lipschitz on $\operatorname{cl} B(x, 1)$, we conclude from Theorem 2.9 in [10] that $L_{y_i x}(v_i)$ is a bounded sequence in $T_x M$ and has a convergent subsequence to zero. Moreover, $\{y_i\}$ has a subsequence converging to x, hence Theorem 2.9 in [10] implies that $0 \in \partial f(x)$ as a contradiction. We consider $\kappa(t) := \dfrac{t}{\delta}$, and $\eta := \dfrac{\delta}{2}$. It is obvious that for every $y \in U(x)$,

$$\kappa'(f(y) - f(x)) \operatorname{dist}(0, \partial f(y)) = \frac{\operatorname{dist}(0, \partial f(y))}{\delta} > 1,$$

which completes the proof. □

Now we aim to present a class of locally Lipschitz functions satisfying the Kurdyka-Lojasiewicz inequality on their domains. First we need to recall some definitions referring to o-minimal structures on $(\mathbb{R}, +, .)$ and analytic geometric categories; see [6].

Definition 3.3 (o-minimal structure)**.** *Let* $\mathcal{O} := \{\mathcal{O}_n\}_{n \in \mathbb{N}}$ *be a sequence such that every* $\mathcal{O}_n$ *is a collection of subsets of* $\mathbb{R}^n$*.* $\mathcal{O}$ *is said to be an o-minimal structure on the real field* $(\mathbb{R}, +, .)$ *if for every* $n \in \mathbb{N}$ *the following conditions are satisfied:*

- $\mathcal{O}_n$ *is a Boolean algebra.*
- *If* $A \in \mathcal{O}_n$*, then* $A \times \mathbb{R} \in \mathcal{O}_{n+1}$ *and* $\mathbb{R} \times A \in \mathcal{O}_{n+1}$*.*
- *If* $A \in \mathcal{O}_{n+1}$*, then* $\pi_n(A) \in \mathcal{O}_n$*, where* π_n *is the projection on the first* n *coordinates.*
- $\mathcal{O}_n$ *contains the family of algebraic subsets of* $\mathbb{R}^n$*.*
- $\mathcal{O}_1$ *consists of all finite unions of points and open intervals.*

We say the elements of $\mathcal{O}$ are definable in $\mathcal{O}$. Moreover, a function $f : \mathbb{R}^n \to \mathbb{R}$ is called definable in $\mathcal{O}$ if its graph belongs to $\mathcal{O}_{n+1}$.

Definition 3.4 (An analytic-geometric category)**.** *An analytic-geometric category* $\mathcal{C}$ *assigns to each real analytic manifold* M *a collection of sets* $\mathcal{C}(M)$ *such that for all real analytic manifolds* M *and* N *the following conditions are satisfied:*

- $\mathcal{C}(M)$ *is a Boolean algebra of subsets of* M*, with* $M \in \mathcal{C}(M)$*.*
- *If* $A \in \mathcal{C}(M)$*, then* $A \times \mathbb{R} \in \mathcal{C}(M \times \mathbb{R})$*.*
- *If* $f : M \to N$ *is a proper analytic map and* $A \in \mathcal{C}(M)$*, then* $f(A) \in \mathcal{C}(N)$*.*
- *If* $A \subset M$ *and* $\{U_i : \quad i \in \Lambda\}$ *is an open covering of* M*, then* $A \in \mathcal{C}(M)$ *if and only if* $A \cap U_i \in \mathcal{C}(U_i)$*, for all* $i \in \Lambda$*.*
- *For every bounded set* $A \in \mathcal{C}(\mathbb{R})$*, the topological boundary* ∂A *consists of a finite number of points.*

The elements of $\mathcal{C}(M)$ are called $\mathcal{C}$-sets. If the graph of a continuous function $f : A \to B$ with $A \in \mathcal{C}(M)$ and $B \in \mathcal{C}(N)$ is contained in $\mathcal{C}(M \times N)$, then f is called a $\mathcal{C}$-function. The following theorem proves that a locally Lipschitz $\mathcal{C}$-function defined on an analytic manifold satisfies the Kurdyka-Lojasiewicz inequality at every point of its domain.

Theorem 3.5. *Let* $f : M \to \mathbb{R}$ *be a locally Lipschitz* $\mathcal{C}$*-function defined on an analytic Riemannian manifold* M*. Then,* f *satisfies the Kurdyka-Lojasiewicz inequality at every* $\bar{x} \in M$*.*

Proof. Using Lemma 3.2, it is enough to prove that f satisfies the Kurdyka-Lojasiewicz inequality at every critical point $\bar{x}$. Assume that (Φ, V) is a local analytic chart of M around $\bar{x}$. We can suppose that V is bounded, therefore by the Lipschitzness of f we conclude that $f(V)$ is also bounded. From [15, Proposition 1.1.5], we deduce that $f \circ \Phi^{-1}$ is definable. Moreover, using Theorem 11 of [6] and Theorem 4.1 of [3], we result that the Kurdyka-Lojasiewicz inequality for $f \circ \Phi^{-1}$ holds at $\bar{y} := \Phi(\bar{x})$. Therefore, there exist $\eta \in (0, \infty)$ and a concave function $\kappa : [0, \eta] \to [0, \infty)$ such that

- $\kappa(0) = 0$,
- κ is of class C^1 on $(0, \eta)$,
- $\kappa' > 0$ on $(0, \eta)$,
- For every $y \in \Phi(V) = U$ with $f \circ \Phi^{-1}(\bar{y}) < f \circ \Phi^{-1}(y) < f \circ \Phi^{-1}(\bar{y}) + \eta$ we have

$$\kappa'(f \circ \Phi^{-1}(y) - f \circ \Phi^{-1}(\bar{y}))\, \mathrm{dist}(0, \partial(f \circ \Phi^{-1})(y)) \geq 1.$$

Since Φ is analytic on V, we have $D\Phi$ is continuous on V and therefore for every compact subset K in V, there exists C_K such that $C_K := \sup_{y \in K} \|D\Phi(y)\|$, where $\|.\|$ denotes the operator norm. Now we prove that f satisfies the Kurdyka-Lojasiewicz inequality at $\bar{x}$. We assume that V' is an open set containing $\bar{x}$ in V

such that $K := \operatorname{cl} V' \subset \operatorname{int} V$ is compact, then we define $\tilde{\kappa} := C_K \kappa$. It is clear that for every $x \in V'$ with $f(\bar{x}) < f(x) < f(\bar{x}) + \eta$, we have

$$\tilde{\kappa}'(f(x) - f(\bar{x})) \operatorname{dist}(0, \partial f(x)) \geq 1.$$

□

4. Subgradient-oriented descent methods

In this section, our aim is to present a subgradient-oriented descent optimization algorithm on a Riemannian manifold and to study the convergence analysis of the proposed algorithm. Our approach is based on the concept of a first order model function of the objective function f in a neighborhood of the current iterate, which can be considered as a nonsmooth generalization of the Taylor expansion. First, we present the notion of a first order model for a locally Lipschitz function defined on a Riemannian manifold. The generalized directional derivative defines a first order model which is called a standard model. Moreover, we present the definition of a strict first order model for a locally Lipschitz function defined on a Riemannian manifold. Using an approximation of the standard first order model, we develop a subgradient-oriented descent optimization algorithm on a Riemannian manifold. Finally, we prove convergence of our proposed descent algorithm to a single limit-point for a locally Lipschitz objective function defined on a Riemannian manifold satisfying the Kurdyka-Lojasiewicz inequality with a strict standard model.

Definition 4.1 (First order model function)**.** *Let $f : M \to \mathbb{R}$ be a locally Lipschitz function on a Riemannian manifold M. A function $\Phi : TM \to \mathbb{R}$ is called a first order model of f if the following conditions are satisfied:*

- *$\Phi|_{T_x M}$ for every $x \in M$ is convex.*
- *$\Phi(0_x) = f(x)$ and $\partial\Phi(0_x) \subset \partial f(x)$ for every $x \in M$.*
- *Φ is upper semicontinuous.*
- *For every x and $\epsilon > 0$, there exists $\delta > 0$ such that $B(x, \delta)$ is convex and $f(y) \leq \Phi(\exp_x^{-1}(y)) + \epsilon \operatorname{dist}(x, y)$ whenever $\operatorname{dist}(x, y) \leq \delta$, where* dist *denotes the Riemannian distance on M.*

Definition 4.2 (Standard first order model)**.** *Every locally Lipschitz function f has a first order local model, called standard model, defined as follows:*

$$\Phi(v_x) := f(x) + f^\circ(x; v_x).$$

Definition 4.3 (Strict first order model)**.** *A first oder model $\Phi : TM \to \mathbb{R}$ is called strict at $\bar{x} \in M$ if the following condition is satisfied:*

- *For every $\epsilon > 0$, there exists $\delta > 0$ such that $B(\bar{x}, \delta)$ is convex and $f(y) \leq \Phi(\exp_x^{-1}(y)) + \epsilon \operatorname{dist}(x, y)$ whenever $x, y \in B(\bar{x}, \delta)$.*

For a convex function $f : M \to \mathbb{R}$, we can define a strict model function $\Phi(v_x) := f \circ \exp_x(v_x)$. For a concave function defined on a Riemannian manifold, the standard model is strict.

Remark 4.4. A locally Lipschitz function $f : M \to \mathbb{R}$ is called prox-regular at $\bar{x}$ with respect to $\bar{v} \in \partial f(\bar{x})$ if there exist $\varepsilon > 0$ and $r > 0$ such that $B(\bar{x}, \varepsilon)$ is convex and

$$f(y) > f(x) + \langle v, \exp_x^{-1}(y) \rangle - \frac{r}{2} \operatorname{dist}(y, x)^2,$$

whenever $\text{dist}(y,\bar{x}) < \varepsilon$ and $\text{dist}(x,\bar{x}) < \varepsilon$ with $y \neq x$ and $|f(x) - f(\bar{x})| < \varepsilon$, while $\|L_{x\bar{x}}v - \bar{v}\| < \varepsilon$ with $v \in \partial f(x)$; see [13]. Assume that f is a locally Lipschitz function and $-f$ is prox-regular at $\bar{x}$. Then the standard model of f is strict on a neighborhood of $\bar{x}$.

As our focus in this paper is on subgradient-oriented methods, we use the generalized directional derivative to define our first order model. What is important in our approach is that we do not use the generalized directional derivative directly to generate the sequence of directions, because this may be too costly. Instead we build a so-called working model which we define as follows.

Definition 4.5 (First order working model)**.** *A function $\Phi^x : T_xM \to \mathbb{R}$ is called a first order working model of f around x if*

- *Φ^x is convex,*
- *$\Phi^x(v_x) \leq f^\circ(x; v_x) + f(x)$ for all $v_x \in T_xM$,*
- *$\Phi^x(0_x) = f(x)$ and $\partial\Phi^x(0_x) \subset \partial f(x)$.*

If $\mathcal{G}$ is a set containing a finite number of subgradients of f around x, then we can define $\Phi^x(v_x) := f(x) + \max_{g\in\mathcal{G}}\langle g, v_x\rangle$.

4.1. **Descent step finding algorithm.** Our minimization algorithm contains an inner loop called descent step finding algorithm, in which we find a descent direction at the current iterate of the minimization algorithm by using first order working models. Here, we will explain our approach in the descent step finding algorithm:

The idea is that after some iterations of this algorithm, the obtained descent direction will improve the current x and define the new iterate x^+. Indeed, at counter k of Algorithm 1, we produce a descent direction d_k^* by solving the following tangent program with proximity control

$$d_k^* = \text{argmin}\{Q_k^x(d) = \Phi_k^x(d) + \frac{1}{2t_k}\langle Pd, d\rangle : \ d \in T_xM\}, \tag{4.1}$$

where $1/t_k > 0$ is the proximity control parameter and the first order working model $\Phi_k^x : T_xM \to \mathbb{R}$ is defined by

$$\Phi_k^x(d) = f(x) + \max_{g\in\mathcal{G}_k}\langle g, d\rangle$$

where $\mathcal{G}_k$ is a finite subset of $\partial f(x)$. If $d_k^* = 0_x$, then $0_x \in \partial\Phi_k^x(0_x) \subset \partial f(x)$ and therefore x is a critical point of f. If the solution $d_k^* \neq 0_x$ gives sufficient decrease in f, it defines by using the retraction the new iterate $x^+ = R_x(d_k^*)$. If d_k^* is not satisfactory, we keep x and use the information transmitted by d_k^* to improve the first order working model. In order to decide whether d_k^* is satisfactory or not, we introduce a constant $c_1 \in (0,1)$, and compute the quotient

$$\rho_k = \frac{f(x) - f(R_x(d_k^*))}{f(x) - \Phi_k^x(d_k^*)},$$

which reflects the agreement between f and the first order working model. Indeed, if the working model is close to the true f, we expect ρ_k to be close to one. We say that agreement between f and the first order working model is acceptable if $\rho_k > c_1$. Note that in this case, d_k^* defines a descent direction, because

$$f(x) - \Phi_k^x(d_k^*) = \Phi_k^x(0_x) - \Phi_k^x(d_k^*) \geq Q_k^x(0_x) - Q_k^x(d_k^*) > 0.$$

Therefore, $f(x) - f(R_x(d_k^*)) > 0$.

Algorithm 1 Descent step finding by backtracking; $(x^+, t^+, \rho) = Descent(x, P, t)$

1: **Require**: A Riemannian manifold M, a locally Lipschitz function $f : M \to \mathbb{R}$ and a retraction R from TM to M.
2: **Parameters**: Scalars $c_1 \in (0,1)$, $c_2 \in (c_1, 1)$, $0 < \theta < \Theta < 1$.
3: **Input**: Current iterate x, a positive definite matrix $P \in \mathbb{R}^{n\times n}$ and $t > 0$.
4: **Output**: New iterate x^+, step size t^+ and ρ.
5: **Initialize**: $k = 1$, $t_1 = t$, $g_0 \in \partial f(x)$ and $\mathcal{G}_1 = \{g_0\}$.
6: find

$$d_k^* = \operatorname{argmin}\{Q_k^x(d) = \Phi_k^x(d) + \frac{1}{2t_k}\langle Pd, d\rangle : \ d \in T_xM\}. \tag{4.2}$$

where $\Phi_k^x : T_xM \to \mathbb{R}$ is defined by $\Phi_k^x(d) = f(x) + \max_{g\in\mathcal{G}_k}\langle g, d\rangle$.
7: Compute

$$\rho_k = \frac{f(x) - f(R_x(d_k^*))}{f(x) - \Phi_k^x(d_k^*)}.$$

8: **if** $\rho_k \geq c_1$ **then** $x^+ = R_x(d_k^*)$, $t^+ = t_k$ and $\rho = \rho_k$ and stop.
9: **end if**
10: Pick $g_k \in \partial f(x)$ such that $f^\circ(x; d_k^*) = \langle g_k, d_k^*\rangle$. Include g_k into the new $\mathcal{G}_{k+1}$. Moreover, include $g_k^* = \dfrac{-1}{t_k} P d_k^*$ into the new $\mathcal{G}_{k+1}$.
11: Compute the test quotient

$$\tilde{\rho}_k = \frac{-f^\circ(x, d_k^*)}{f(x) - \Phi_k^x(d_k^*)}.$$

12: **if** $\tilde{\rho}_k \geq c_2$ **then** select $t_{k+1} \in [\theta t_k, \Theta t_k]$, $k = k + 1$ and go to line 6.
13: **else** $t_{k+1} = t_k$, $k = k + 1$ and go to line 6.
14: **end if**

Suppose that d_k^* is not satisfactory. Then the first order working model was not entirely useful, and we need to improve it. Since the quadratic term $\langle Pd, d\rangle$ in (4.1) remains unchanged during the step finding algorithm, we have to improve the first order working model, which is done through adding elements to the set $\mathcal{G}_k$. As the first order working model is an approximation of the standard first order model, we need to make sure if the agreement between the first order working model and $\Phi(d_k^*) = f^\circ(x; d_k^*) + f(x)$ is good. In order to decide, we introduce another control parameter

$$\tilde{\rho}_k = \frac{f(x) - \Phi(d_k^*)}{f(x) - \Phi_k^x(d_k^*)},$$

which reflects the agreement between the first order working model and the standard first order model. We fix a constant c_2 with $1 > c_2 > c_1$, which plays a role similar to c_1. We say that the first order working model is far from the standard first order model if $\tilde{\rho}_k < c_2$.

Suppose that $\rho < c_1$ and $\tilde{\rho}_k \geq c_2$, it means that the first order working model is not close to f, however the agreement between the first order model and first order working model is good. Adding new elements to $\mathcal{G}_k$ would make the first order working model closer to the standard first order model, but would not suffice to make the first order working model close to f. Therefore, we have to tighten proximity control, which means decreasing t_k.

It is left to explain how the first order working model is built. We start with one subgradient g_0 and define

$$\Phi_1^x(d) = f(x) + \max_{g\in\mathcal{G}_1=\{g_0\}} \langle g, d\rangle.$$

As along as the found direction is not satisfactory, two subgradients will be added to improve the first order working model. At counter k of Algorithm 1, we pick $g_k \in \partial f(x)$ such that

$$f^\circ(x; d_k^*) = \langle g_k, d_k^*\rangle$$

and include g_k into the new $\mathcal{G}_{k+1}$. Moreover, as we have

$$0 \in \partial\Phi_k^x(d_k^*) + t_k^{-1} P d_k^*,$$

therefore $g_k^* := -t_k^{-1} P d_k^* \in \partial\Phi_k^x(d_k^*) \subset \partial f(x)$. We include

$$g_k^* = \frac{-1}{t_k} P d_k^*$$

into the new $\mathcal{G}_{k+1}$.

Remark 4.6. Note that if $\mathcal{G}_0 = \partial f(x)$, then $\tilde{\rho}_k$ is always equal to one and therefore we always reduce the step size in case of non-satisfactory directions. Therefore, $d_k^* = -t_k P^{-1} g$, where $g \in \partial f(x)$ is the projection of 0_x onto $\partial f(x)$ with respect to $\|.\|_{P^{-1}} = \langle ., P^{-1}.\rangle^{1/2}$; see[11, 12].

The following theorem proves that Algorithm 1 terminates after a finite number of iterations. It can be proved along the same lines as Theorem 3.1 of [17].

Theorem 4.7. *Let f be a locally Lipschitz function on a Riemannian manifold M and $0 \notin \partial f(x)$. Then after a finite number of iterations k the descent step finding algorithm finds a subgradient $g_k^* \in \partial f(x)$ and a step size $t_k > 0$ such that $x^+ = R_x(-t_k P^{-1} g_k^*)$ satisfies the descent condition $\rho_k \geq c_1$.*

4.2. **Minimization algorithm.** Now, we present the main algorithm and prove the convergence result. This algorithm contains all subgradient-oriented algorithms. Moreover, it is beneficial in practical situations, where the full subdifferential is inaccessible. In the minimization algorithm, we deal with updating the step size t_j^* and the matrix P_j. Recall that in Algorithm 1 we had a constant $c_1 \in (0, 1)$, and computed the quotient

$$\rho_k = \frac{f(x) - f(R_x(d_k^*))}{f(x) - \Phi_k^x(d_k^*)},$$

which reflects the agreement between f and the first order working model at every point x. Indeed, if the working model is close to the true f around x, we expect ρ_k to be close to one. We mentioned that agreement between f and the first order working model around x is acceptable if $\rho_k \geq c_1$. In the following algorithm, we introduce another constant $\Gamma \in (c_1, 1)$ and we say that agreement between f and the first order working model around x is good if the quotient is bigger than Γ.

As in Algorithm 1 the step size is never increased, we increase the step size if the agreement between f and the first order working model around the current iteration is good. If the agreement is only acceptable, then we memorize the last step size. As we wish to avoid too small step sizes in our convergence analysis, we put a lower bound T on the step sizes. This part is not obligatory in practice though.

The sequence of matrices in the minimization algorithm can be constant or even can be considered equal to identity. But we can also use the BFGS strategy presented in [12] to update this sequence.

Algorithm 2 Minimization algorithm

Require: A Riemannian manifold M, a locally Lipschitz function $f : M \to \mathbb{R}$ and a retraction R from TM to M.
Parameters: $0 < \theta < \Theta < 1$, $0 < c_1 < \Gamma < 1$, $0 < \lambda < \Lambda < \infty$, $T \geq 0$.
Initialize: $j = 1$, $t_1^* > 0$, $x_1 \in M$ and P_1 is a positive definite matrix such that $\lambda\|.\| \leq \|.\|_1 \leq \Lambda\|.\|$, with $\|x\|_1^2 := \langle x, Px\rangle$.
for $j = 1, 2, \ldots$ **do**
    if $0 \in \partial f(x_j)$ **then** Stop
    end if
    $(x_{j+1}, t_{j+1}^*, \rho) = Descent(x_j, P_j, t_j^*)$ and choose a new P_{j+1} such that $\lambda\|.\| \leq \|.\|_{j+1} \leq \Lambda\|.\|$.
    if $\rho \geq \Gamma$ **then** $t_{j+1}^* = \max\{T, \theta^{-1} t_{j+1}^*\}$.
    end if
end for

The following theorem proves the convergence in the sense of subsequence of Algorithm 2. Moreover, convergence to a single critical point can also be proved if the Kurdyka-Lojasiewicz inequality is satisfied.

Theorem 4.8. *Let $f : M \to \mathbb{R}$ be locally Lipschitz on a Riemannian manifold M and $L = \{x \in M : \ f(x) \leq f(x_1)\}$ be bounded. Assume that x_j is the sequence generated by Algorithm 2.*

- *If the standard model of f is strict and $T > 0$, then every accumulation point of x_j is critical.*
- *If the standard model of f is strict and $T = 0$, then there exists at least one accumulation point of x_j which is critical.*
- *If the standard model is strict and f satisfies the Kurdyka-Lojasiewicz inequality, then x_j converges to a single critical point. (In this case, $T \geq 0$.)*

Proof. Assuming $0 \notin \partial f(x_j)$, then by Theorem 4.7 we deduce that after a finite number of iterations k_j the descent step finding algorithm finds a subgradient $g_{k_j}^* \in \partial f(x_j)$ and a step size $t_{k_j} > 0$ such that $x_{j+1} = R_{x_j}(-t_{k_j} P_j^{-1} g_{k_j}^*)$ satisfies the descent condition $\rho \geq c_1$. Set $d_{k_j}^* = -t_{k_j} P_j^{-1} g_{k_j}^*$. Therefore,

$$f(x_j) - f(x_{j+1}) \geq c_1\big(f(x_j) - \Phi_{k_j}^{x_j}(d_{k_j}^*)\big). \tag{4.3}$$

Since $d_{k_j}^* = \operatorname{argmin}\{\Phi_{k_j}^{x_j}(d) + \dfrac{1}{2t_{k_j}}\langle P_j d, d\rangle : \ d \in T_{x_j}M\}$, we have $g_{k_j}^* = \dfrac{-1}{t_{k_j}} P_j d_{k_j}^* \in \partial\Phi_{k_j}^{x_j}(d_{k_j}^*)$, hence the subgradient inequality gives

$$\langle g_{k_j}^*, -d_{k_j}^*\rangle \leq \Phi_{k_j}^{x_j}(0) - \Phi_{k_j}^{x_j}(d_{k_j}^*) = f(x_j) - \Phi_{k_j}^{x_j}(d_{k_j}^*).$$

Consequently,

$$\frac{1}{t_{k_j}}\|d_{k_j}^*\|_j^2 \leq \frac{1}{c_1}\big(f(x_j) - f(x_{j+1})\big), \tag{4.4}$$

where $\|d_{k_j}^*\|_j^2 = \langle P_j d_{k_j}^*, d_{k_j}^*\rangle$. Now summing (4.4) over $j = 1, ..., J-1$ on both sides implies

$$\sum_{j=1}^{J-1} \frac{1}{t_{k_j}} \|d_{k_j}^*\|_j^2 \leq \frac{1}{c_1}\big(f(x_1) - f(x_J)\big).$$

Since $d_{k_j}^*$ is a descent direction, the sequence $\{f(x_j)\}$ is decreasing and $\{x_j\} \subset L$ is bounded. Moreover, since f is locally Lipschitz on the compact set L, it can be proved that it is Lipschitz of some constant K on L which implies that

$$\sum_{j=1}^{J-1} \frac{1}{t_{k_j}} \|d_{k_j}^*\|_j^2 \leq \frac{1}{c_1}\big(f(x_1) - f(x_J)\big) \leq \frac{K}{c_1} d(x_1, x_J).$$

Therefore, the series $\sum_j \frac{1}{t_{k_j}} \|d_{k_j}^*\|_j^2$ is summable and $\frac{1}{t_{k_j}} \|d_{k_j}^*\|_j^2 \to 0$. Since the norms $\|\cdot\|_j$ are uniformly equivalent, we conclude that $\frac{1}{t_{k_j}} \|d_{k_j}^*\|^2 \to 0$. Now we shall have to deal with two cases;

i)An infinite subsequence $\{g_{k_j}^*\}_{j\in\mathcal{N}}$ converging to zero. We claim that every accumulation point of $\{x_j\}_{j\in\mathcal{N}}$ is critical. Let x^* be an accumulation point of $\{x_j\}_{j\in\mathcal{N}}$, without loss of generality, we assume that $x_j \to x^*$, $j \in \mathcal{N}$. Since $g_{k_j}^* \in \partial f(x_j)$, by Theorem 2.9 in [10], $0 \in \partial f(x^*)$.

ii) Let $\{g_{k_j}^*\}_{j\in\mathcal{J}}$ be an infinite subsequence of $g_{k_j}^*$ with $\|g_{k_j}^*\| \geq \eta > 0$ for all $j \in \mathcal{J}$ and some positive number η. We first prove that under this assumption, $\{t_{k_j}\}_{j\in\mathcal{J}}$ converges to zero. To prove the claim, we assume on the contrary that there exists $\tau > 0$ such that $t_{k_j} \geq \tau > 0$. Moreover, we assume that x^* is an accumulation point of $\{x_j\}_{j\in\mathcal{J}}$, then there exist subsequences $\{L_{x_j x^*}(P_j)\}_{j\in\mathcal{J}'}$, $\{L_{x_j x^*}(d_{k_j}^*)\}_{j\in\mathcal{J}'}$ and $\{\frac{1}{t_{k_j}}\}_{j\in\mathcal{J}'}$ converging to P, δx, $\frac{1}{t}$, respectively. Consequently, $\frac{1}{t}\|P\delta x\| \geq \eta$. But we proved that $\frac{1}{t_{k_j}} \|d_{k_j}^*\|^2 \to 0$, which means $\frac{1}{t}\|\delta x\|^2 = 0$. Hence, either $\delta x = 0$ or $\frac{1}{t} = 0$, which is a contradiction. Therefore, we conclude that $\{t_{k_j}\}_{j\in\mathcal{J}}$ converges to zero.

Now we divide $\mathcal{J}$ into two classes;

$$\mathcal{J}_1 \subset \{j \in \mathbb{N} : \ t_j^* > t_{k_j}\}, \qquad \mathcal{J}_2 \subset \{j \in \mathbb{N} : \ t_j^* = t_{k_j}\}.$$

Let $\hat{x}$ be an accumulation point of a subsequence in $\mathcal{J}_1$, we show that $\hat{x}$ is a critical point. Without loss of generality, we may assume that $x_j \to \hat{x}$, for $j \in \mathcal{J}_1$. Suppose that for $j \in \mathcal{J}_1$ the backtracking rule was applied for the last time at step $k_j - \nu_j$ with $\nu_j \geq 1$. Consequently,

$$\rho_{k_j-\nu_j} = \frac{f(x_j) - f(R_{x_j}(d_{k_j-\nu_j}^*))}{f(x_j) - \Phi_{k_j-\nu_j}^{x_j}(d_{k_j-\nu_j}^*)} < c_1,$$

and

$$\tilde{\rho}_{k_j-\nu_j} = \frac{f(x_j) - \Phi(d_{k_j-\nu_j}^*)}{f(x_j) - \Phi_{k_j-\nu_j}^{x_j}(d_{k_j-\nu_j}^*)} \geq c_2.$$

Moreover, $t_{k_j} = \theta_{k_j-\nu_j} t_{k_j-\nu_j}$ for uniformly bounded $\theta_{k_j-\nu_j}$ and $\tilde{g}_j^* = -\theta_{k_j-\nu_j} t_{k_j}^{-1} P_j d_{k_j-\nu_j}^* \in \partial\Phi_{k_j-\nu_j}^{x_j}(d_{k_j-\nu_j}^*)$. We prove that $\tilde{g}_j^* \to 0$ and therefore $0 \in \partial f(\hat{x})$. First, it is clear

that that $\{\tilde{g}_j^* : j \in \mathcal{J}_1\}$ is bounded. Moreover, $\|d^*_{k_j-\nu_j}\|$ is convergent to zero. Now if $\tilde{g}_j^*$ does not converge to zero, there exists $\theta > 0$ such that $\|\tilde{g}_j^*\| \geq \theta$ for all $j \in \mathcal{J}_1$. Note that

$$\langle \tilde{g}_j^*, -d^*_{k_j-\nu_j} \rangle \leq f(x_j) - \Phi^{x_j}_{k_j-\nu_j}(d^*_{k_j-\nu_j}).$$

Therefore, the left hand side behaves asymptotically like $c\|\tilde{g}_j^*\|\|d^*_{k_j-\nu_j}\|$ for some $c > 0$. Hence, we have

$$c\theta\|d^*_{k_j-\nu_j}\| \leq f(x_j) - \Phi^{x_j}_{k_j-\nu_j}(d^*_{k_j-\nu_j}).$$

Now using the fact that f has a strict standard model, there exists $\epsilon_j \to 0$ such that

$$f(\exp_{x_j}(d^*_{k_j-\nu_j})) - \Phi(d^*_{k_j-\nu_j}) \leq \epsilon_j\|d^*_{k_j-\nu_j}\|.$$

Hence, using the fact that any retraction is a first order approximation for the exponential map on the manifold, we have that

$$\begin{aligned}
\tilde{\rho}_{k_j-\nu_j} &= \rho_{k_j-\nu_j} + \frac{f(R_{x_j}(d^*_{k_j-\nu_j})) - \Phi(d^*_{k_j-\nu_j})}{f(x_j) - \Phi^{x_j}_{k_j-\nu_j}(d^*_{k_j-\nu_j})} \\
&\leq \rho_{k_j-\nu_j} + \frac{f(\exp_{x_j}(d^*_{k_j-\nu_j})) - \Phi(d^*_{k_j-\nu_j})}{f(x_j) - \Phi^{x_j}_{k_j-\nu_j}(d^*_{k_j-\nu_j})} + \frac{f(R_{x_j}(d^*_{k_j-\nu_j})) - f(\exp_{x_j}(d^*_{k_j-\nu_j}))}{f(x_j) - \Phi^{x_j}_{k_j-\nu_j}(d^*_{k_j-\nu_j})} \\
&\leq \rho_{k_j-\nu_j} + \frac{\epsilon_j}{c\theta} + \frac{K\,\mathrm{dist}(R_{x_j}(d^*_{k_j-\nu_j}), \exp_{x_j}(d^*_{k_j-\nu_j}))}{c\theta\|d^*_{k_j-\nu_j}\|},
\end{aligned}$$

which shows that $\limsup_{j\to\infty} \tilde{\rho}_{k_j-\nu_j} \leq \limsup_{j\to\infty} \rho_{k_j-\nu_j} \leq c_1 < c_2$, contradicting the fact that $\tilde{\rho}_{k_j-\nu_j} \geq c_2$ for every $j \in \mathcal{J}_1$. This shows that $\tilde{g}_j^*$ converges to zero.

Now assume that $\hat{x}$ is an accumulation point of a subsequence of $\{x_j\}_{\mathcal{J}_2}$. Therefore, $t_j^* = t_{k_j}$. But this cannot happen. Since t_{k_j} is convergent to zero and $t_j^* > T$.

To prove the second part of the theorem; we proved that if $g^*_{k_j}$ has an infinite subsequence $\{g^*_{k_j}\}_{j\in\mathcal{N}}$ converging to zero, then every accumulation point of $\{x_j\}_{j\in\mathcal{N}}$ is critical. If $g^*_{k_j}$ has an infinite subsequence which is bounded below, then as before we define $\mathcal{J}_1$ and $\mathcal{J}_2$, since t_{k_j} converges to zero in this case, hence there exists an infinite subsequence of $\{x_j\}_{j\in\mathcal{J}_1}$ converging to some $\hat{x}$. Using the same argument as in the first part of the theorem, we can prove $\hat{x}$ is critical.

To prove the third part of the theorem, assume that f satisfies the Kurdyka-Lojasiewicz inequality. We prove that x_j converges to a single critical point x^*. We have shown that the sequence x_j has at least one accumulation point x^*, which is critical. Assume that L' is the set of all accumulation points of x_j, it is obvious that L' is closed. Since $f(x_j)$ is decreasing, f is constant on the set L'. Using the Kurdyka-Lojasiewicz inequality for every $x \in L'$, we may find a neighborhood $U(x)$ of x and a continuous concave function $\kappa_x : [0, \eta_x] \to [0, \infty)$ of class C^1 on $(0, \eta_x)$ with $\kappa_x(0) = 0$, $\kappa_x' > 0$ on $(0, \eta_x)$, such that

$$\kappa_x'(f(x') - f(x))\,\mathrm{dist}(0, \partial f(x')) \geq 1, \quad x' \in U(x) \quad \text{with} \quad f(x) < f(x') < f(x) + \eta_x.$$

By compactness of L', we find finite points $x_1, ..., x_r \in L'$ such that $U(x_1), ..., U(x_r)$ cover L'. Then, we choose $\epsilon > 0$ such that $V := \{x \in M : \mathrm{dist}(x, L') < \epsilon\} \subset \bigcup_{i=1}^r U(x_i)$. Set $\eta = \min_{i=1,...,r} \eta_{x_i}$, $\kappa'(t) = \max_{i=1,...,r} \kappa'_{x_i}(t)$ and $\kappa(t) = \int_0^t \kappa'(\tau)d\tau$. We claim that for every $x \in L'$ and $x' \in V$ with $f(x) < f(x') < f(x)+\eta$, we have

$$\kappa'(f(x') - f(x))\,\mathrm{dist}(0, \partial f(x')) \geq 1.$$

To prove the claim, we find x_i such that $x' \in U(x_i)$, then

$$\kappa'(f(x') - f(x)) \operatorname{dist}(0, \partial f(x')) \geq \kappa'_{x_i}(f(x') - f(x_i)) \operatorname{dist}(0, \partial f(x')) \geq 1,$$

which proves our claim. We assume without loss of generality that f is zero on L'. We know that

$$\frac{1}{t_{k_j}} \|d^*_{k_j}\|_j^2 \leq \frac{1}{c_1}\big(f(x_j) - f(x_{j+1})\big).$$

By concavity of κ, we have

$$\kappa(f(x_j)) - \kappa(f(x_{j+1})) \geq \kappa'(f(x_j))(f(x_j) - f(x_{j+1})) \geq c_1 \kappa'(f(x_j)) \frac{1}{t_{k_j}} \|d^*_{k_j}\|_j^2,$$

whenever $0 < f(x_j) < \eta$, $0 < f(x_{j+1}) < \eta$. By the Kurdyka-Lojasiewicz inequality, we conclude that $\kappa'(f(x_j)) \geq \|g\|^{-1}$ for every Clarke subgradient $g \in \partial f(x_j)$. Therefore, $\kappa'(f(x_j)) \geq \|g^*_{k_j}\|^{-1}$, which implies that

$$\kappa(f(x_j)) - \kappa(f(x_{j+1})) \geq c_1 \frac{{t_{k_j}}^{-1} \|d^*_{k_j}\|_j^2}{{t_{k_j}}^{-1} \|P_j d^*_{k_j}\|} \geq \lambda' \|d^*_{k_j}\|,$$

for some constant $\lambda' > 0$. This proves the summability of $\|d^*_{k_j}\|$, hence $d^*_{k_j} \to 0$ and $\operatorname{dist}(R_{x_j}(d^*_{k_j}), x_j) \to 0$. Therefore, x_j is a Cauchy sequence converging to x^* and $L' = \{x^*\}$. Since L' has at least one critical point of f, we deduce that x^* is critical and the proof is complete.

□

References

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

[2] P. A. Absil, R. Mahony, B. Andrews,*Convergence of the iterates of descent methods for analytic cost functions*, SIAM J. Optim., 6 (2005), pp. 531-547.

[3] H. Attouch, P. Redont, J. Bolte, A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems, An approach based on the Kurdyka-Lojasiewicz inequality,* Math. Oper. Res., 35(2) (2010), pp. 438-457.

[4] G. C. Bento, J. X. da Cruz Neto, P. R. Oliveira, *Convergence of inexact descent methods for nonconvex optimization on Riemannian manifolds*, Submitted.

[5] J. Bolte, J. A. Daniilidis, A. Lewis, *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., 17(4) (2006), pp. 1205-1223.

[6] J. Bolte, J. A. Daniilidis, A. Lewis, M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM J. Optim., 18(2) (2007), pp. 556-572.

[7] E. Bierstone, P. D. Milman, *Semianalytic and subanalytic sets*, Inst. Hautes Études Sci. Publ. Math., 67 (1988), pp. 5-42.

[8] P. Grohs, S. Hosseini, *ε-subgradient algorithms for locally Lipschitz functions on Riemannian manifolds,* Adv. Comput. Math., 42(2)(2016), pp. 333-360.

[9] P. Grohs, S. Hosseini, *Nonsmooth trust region algorithms for locally Lipschitz functions on Riemannian manifolds*, IMA J. Numer. Anal., 36(3) (2016), pp. 1167-1192.

[10] S. Hosseini, M. R. Pouryayevali, *Generalized gradients and characterization of epi-Lipschitz sets in Riemannian manifolds*, Nonlinear Anal., 74 (2011), pp. 3884-3895.

[11] S. Hosseini, A. Uschmajew, *A Riemannian gradient sampling algorithm for nonsmooth optimization on manifolds*, SIAM J. Optim., 27(1) (2017), pp. 173-189.

[12] S. Hosseini, W. Huang, R. Yousefpour, *Line search algorithms for locally Lipschitz functions on Riemannian manifolds*, INS Preprint No. 1626.

[13] S. Hosseini, M. R. Pouryayevali, *On the metric projection onto prox-regular subsets of Riemannian manifolds*, Proc. Amer. Math. Soc., 141 (2013), pp. 233-244.

[14] K. Kurdyka, *On gradients of functions definable in o-minimal structures,* Ann. Inst. Fourier., 48 (1998), pp. 769-783.
[15] C. Lageman, *Convergence of gradient-like dynamical systems and optimization algorithms*, PhD thesis, www.opus-bayern.de/uni- wuerzburg/volltexte/2007/2394/pdf/diss.pdf.
[16] S. Lang, *Fundamentals of Differential Geometry,* Graduate Texts in Mathematics, Vol. 191, Springer, New York, 1999.
[17] D. Noll, *Convergence of non-smooth descent methods using the Kurdyka-Lojasiewicz Inequality,* J. Optim. Theory Appl., 160(2014), pp. 553 -572.
[18] D. Noll, O. Prot, A. Rondepierre, *A proximity control algorithm to minimize nonsmooth and nonconvex functions,* Pac. J. Optim., 4(2008), pp. 569-602.
[19] W. Ring, B. Wirth, *Optimization methods on Riemannian manifolds and their application to shape space*, SIAM J. Optim., 22(2) (2012), pp. 596-627.
[20] S. T. Smith, *Optimization techniques on Riemannian manifolds*, Fields Institute Communications, 3 (1994), pp. 113-146.
[21] C. Udriste, *Convex Functions and Optimization Methods on Riemannian Manifolds*, Kluwer Academic Publishers, Dordrecht, Netherlands, 1994.