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-Łojasiewicz inequality on Riemannian manifolds

INS Preprint No. 1523

November 2015

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

S. HOSSEINI*

Abstract. We examine convergence of subgradient-oriented descent methods in nonsmooth optimization on Riemannian manifolds. We prove convergence in the sense of subsequences for nonsmooth functions whose standard model is strict. Adding the Kurdyka-Lojasiewicz condition, we demonstrate convergence to a singular critical point.

1. introduction

This paper is concerned with subgradient-oriented descent methods in nonsmooth nonconvex optimization problems on Riemannian manifolds. 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.

In smooth optimization algorithms on linear spaces for choosing the search direction p_j at x_j, we need that the angle θ_j, defined below, is bounded away from $90°$;

$$\cos\theta_j = \frac{-\langle \operatorname{grad} f(x_j), p_j\rangle}{\|\operatorname{grad} f(x_j)\|\|p_j\|}.$$

Then convergence of methods is obtained by the Armijo condition along with a safeguard against too small step sizes; see [14]. Indeed, classical convergence results establish that accumulation points of the sequence of iterates are critical points of the objective function f and 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, then in order to have convergence of the whole sequence to x^* it is sufficient to

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).*

require that the so-called Kurdyka-Lojasiewicz inequality holds in a neighborhood of x^*; see [2]. The Kurdyka-Lojasiewicz inequality was introduced for differentiable functions definable in an o-minimal structure defined in $\mathbb{R}^n$; see [11].

The question is that whether similar results are correct in nonsmooth optimization problems? We know that in convex optimization problems on linear spaces, the steepest descent method converges only if the step sizes t_j satisfy $\sum_j t_j = \infty$, $\sum_j t_j^2 < \infty$. As this condition is essential and cannot be verified algorithmically in nonconvex cases, therefore the nonsmooth situation seems more complicated. It has been proved in [16] that convergence of subgradient-oriented methods 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 are needed to be exploited; for extensions of Kurdyka-Lojasiewicz inequality to subanalytic nonsmooth functions see [5] and for a more general extension for Clarke subdifferential see [6]. In [15] it is proved that if the locally Lipschitz objective function f has a strict model function and satisfies a nonsmooth version of the Kurdyka-Lojasiewicz inequality, then convergence to a single critical point may be guaranteed.

There have been some attempts to adapt standard optimization methods to problems on manifolds. Line-search techniques were proposed and analyzed on manifolds by several authors; see, e.g., [1, 17, 18, 19]. For instance; in [17] 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 gradient $-\operatorname{grad} f(x_j)$, are discussed. Similar to linear case for proving convergence of the whole sequence of iterations to a single critical point, an extension of Kurdyka-Lojasiewicz inequality is required. Lageman [12] 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 [12, 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.

In this paper, we prove convergence of subgradient-oriented descent methods to a single limit-point for locally Lipschitz functions on Riemannian manifolds with a strict standard model function satisfying the Kurdyka-Lojasiewicz inequality.

2. Preliminaries

In this paper, we use the standard notations and known results of Riemannian manifolds, see, e.g. [13]. Throughout this paper, M is an n-dimensional complete manifold endowed with a Riemannian metric $\langle ., . \rangle$ on the tangent space T_xM. 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 $\gamma^\circ(0) = v$, 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 $m > 0$ and $M > 0$ such that

$$m\|g\| \leq \mathrm{dist}(x, R_x(g)) \leq M\|g\|.$$

By using retractions, we extend the concepts of nonsmooth analysis on Riemannian manifolds.

Let $f : M \to \mathbb{R}$ be a locally Lipschitz function on a Riemannian manifold. For $x \in M$, we let $\hat{f}_x = f \circ R_x$ denote the restriction of the pullback $\hat{f} = f \circ R$ to T_xM. The Clarke generalized directional derivative of f at x in the direction $p \in T_xM$, denoted by $f^\circ(x;p)$, is defined by $f^\circ(x;p) = \hat{f}_x^\circ(0_x;p)$, where $\hat{f}_x^\circ(0_x;p)$ denotes the Clarke generalized directional derivative of $\hat{f}_x : T_xM \to \mathbb{R}$ at 0_x in the direction $p \in T_xM$. Therefore, the generalized subdifferential of f at x, denoted by $\partial f(x)$, is defined by $\partial f(x) = \partial \hat{f}_x(0_x)$. A point x is a stationary point of f if $0 \in \partial f(x)$. A necessary condition that f achieve a local minimum at x is that x is a stationary point of f; see [8, 9]. The following theorem can be proved along the same lines as [9, Theorem 2.9].

Theorem 2.2. *Let M be a Riemannian manifold, $x \in M$ and $f : M \to \mathbb{R}$ be a Lipschitz function of rank K near x. Then*
(a) $\partial f(x)$ is a nonempty, convex, compact subset of T_xM, and $\|\xi\| \leq K$ for every $\xi \in \partial f(x)$.
(b) for every v in T_xM, we have

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

(c) if $\{x_i\}$ and $\{\xi_i\}$ are sequences in M and TM such that $\xi_i \in \partial f(x_i)$ for each i, and if $\{x_i\}$ converges to x and ξ is a cluster point of the sequence $\{\xi_i\}$, then we have $\xi \in \partial f(x)$.
(d) ∂f is upper semicontinuous at x.

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 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*
$$\kappa'(f(y) - f(x))dist(0, \partial f(y)) \geq 1,$$
where $dist(0, \partial f(y)) = \inf\{\|v\| : \quad v \in \partial f(y)\}$.

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 $\delta > 0$ such that for every $y \in U(x)$, $\text{dist}(0, \partial f(y)) > \delta$. Now we prove the claim by contradiction, assume that there exists a sequence $y_i \in B(x, \frac{1}{i}) \subset \text{cl}B(x, 1)$ such that $v_i \in \partial f(y_i)$ and $\|v_i\| \leq \dfrac{1}{i}$. Since f is Lipschitz on $\text{cl}B(x, 1)$, we conclude from Theorem 2.2 that $L_{y_i x}(v_i)$ is a bounded sequence in $T_x M$ which has a convergent subsequence to zero. Moreover $\{y_i\}$ has a subsequence converging to x, hence Theorem 2.2 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 $y \in U(x)$,
$$\kappa'(f(y) - f(x))\text{dist}(0, \partial f(y)) = \frac{\text{dist}(0, \partial f(y))}{\delta} > 1,$$
which completes the proof. $\square$

Now we aim to present a class of locally Lipschitz functions satisfying the Kurdyka-Lojasiewicz inequality at every point of their domains. First we need to recall some definitions referring to o-minimal structures on $(\mathbb{R}, +, .)$ and analytic geometric categories; see [6]. O-minimal structures on $(\mathbb{R}, +, .)$ are a generalization of semialgebraic sets, which are determined by a finite number of polynomial inequalities and equations.

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 subset 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 $\mathcal{O}_{n+1}$. On analytic manifolds the analogue of semialgebraic sets in $\mathbb{R}^n$ are the semianalytic and subanalytic sets. The semianalytic sets are locally described by a finite number of analytic equations and inequalities. The subanalytic sets are locally projections of relatively compact semianalytic sets; for more information see [7]. The analyticgeometric categories are the analogue generalization of the semianalytic and subanalytic sets.

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 $A \in \mathcal{C}(M)$, 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 $A \in \mathcal{C}(M)$, $B \in \mathcal{C}(N)$ 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 [12, Proposition 1.1.5], we deduce that $f \circ \Phi^{-1}$ is definable in $\mathcal{O}(\mathcal{C})$. 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

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

Since Φ is analytic on V, we have $D\Phi^{-1}$ is continuous on U and therefore for every compact subset K in U, there exists C_K such that $C_K := \sup_{y \in K} \|D\Phi^{-1}(y)\|$, where $\|.\|$ denotes the norm operator. Now we prove that f satisfies the Kurdyka-Lojasiewicz inequality at $\bar{x}$. We assume that U' is an open set containing $\bar{x}$ in U such that $K := \text{cl}U' \subset \text{int}U$ is compact, then we define $\tilde{\kappa} := C_K \kappa$. It is clear that for every $x \in U'$ with $f(\bar{x}) < f(x) < f(\bar{x}) + \eta$, we have

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

□

Finally, we present the following corollary to state that locally Lipschitz definable functions on submanifolds of $\mathbb{R}^n$ satisfy the Kurdyka-Lojasiewicz inequality at every point of their domains.

Corollary 3.6. *Let $f : M \subset \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz definable function in an o-minimal structure $\mathcal{O}$ and M be endowed with the induced metric of $\mathbb{R}^n$. Then, f satisfies the Kurdyka-Lojasiewicz inequality at every $\bar{x} \in M$.*

4. Nonsmooth descent methods on Riemannian manifolds

In nonsmooth problems the angle condition does not propose a proper set of search directions, because the descent directions form a cone with opening angle less than 180°, which means a direction p with $\angle(p, -g) < 90°$, where g is the steepest ascent subgradient, is not necessarily descent. The following definition is equivalent to gradient-orientedness carried over nonsmooth problems; see [15].

Definition 4.1 (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 symmetric linear maps $\{P_k : T_{x_k}M \to T_{x_k}M\}$ satisfying*

$$0 < \lambda \leq \lambda_{min}(P_k) \leq \lambda_{max}(P_k) \leq \Lambda < \infty,$$

for $0 < \lambda < \Lambda < \infty$ and all $k \in \mathbb{N}$, such that $p_k = \dfrac{-P_k g_k}{\|P_k g_k\|}$.

Note that $\lambda_{min}(P_k)$ and $\lambda_{max}(P_k)$ denote respectively the smallest and largest eigenvalues of P_k.

Our purpose is to study the conditions presented in [15] on Riemannian manifolds. Therefore, our approach is based on the concept of a local 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.

Definition 4.2 (A 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;*

- *$\Phi|_{T_xM}$ for every $x \in M$ is convex.*
- *$\Phi(x, 0) = f(x)$ and $\partial\Phi(x, 0) \subset \partial f(x)$ 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 $f(y) \leq \Phi(x, \exp_x^{-1}(y)) + \epsilon dis(x, y)$ whenever $dist(x, y) \leq \delta$, where dist denotes the Riemannian distance on M.*

Note that every locally Lipschitz function f has a first order local model, called standard model, defined as follows;

$$\Phi(x, v) = f(x) + f^\circ(x, v).$$

A local 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 $f(y) \leq \Phi(x, \exp_x^{-1}(y)) + \epsilon \mathrm{dist}(x, y)$ whenever $x, y \in B(\bar{x}, \delta)$.

A local model $\Phi : TM \to \mathbb{R}$ is called strong at $\bar{x} \in M$ if the following condition is satisfied;

- There exist $\delta > 0$ and $L > 0$ such that $B(\bar{x}, \delta)$ is convex and $f(y) \leq \Phi(x, \exp_x^{-1}(y)) + L\mathrm{dist}(x, y)^2$ whenever $x, y \in B(\bar{x}, \delta)$.

Remark 4.3. 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

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

whenever $\mathrm{dist}(y, \bar{x}) < \varepsilon$ and $\mathrm{dist}(x, \bar{x}) < \varepsilon$ 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 [10]. Assume that f is a locally Lipschitz function and $-f$ is prox regular at $\bar{x}$. Then the standard model of f is strong on a neighborhood of $\bar{x}$.

As we are working with subgradient-oriented methods, therefore we choose the standard model to build the descent directions. The following theorem can be

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

1: **Require**: Riemannian manifold M, a locally Lipschitz function $f : M \to \mathbb{R}$, 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, $P \in S(n)$ 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

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

where $\Phi_k : T_xM \to \mathbb{R}$ 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 $g_k^* = \dfrac{-1}{t_k}Pd_k^*$.
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 Step 6.
13: **else** $t_{k+1} = t_k, k = k + 1$ and go to Step 6.
14: **end if**

proved along the same lines as Theorem 3.1 of [15] due to the fact that the use of retractions yields Algorithm 1 be an expression of Algorithm 1 of [15] in Euclidean space T_xM.

Theorem 4.4. *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 $g_k \in \partial f(x)$ and a step size $t_k > 0$ such that $x^+ = R_x(-t_k P^{-1} g_k)$ satisfies the descent condition $\rho_k \geq c_1$.*

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. Note that the

Algorithm 2 Subgradient-oriented descent method

1: **Require**: Riemannian manifold M, a locally Lipschitz function $f: M \to \mathbb{R}$, retraction R from TM to M.
2: **Parameters**: scalars $c_1 \in (0,1)$, $c_2 \in (c_1, 1)$, $0 < \theta < \Theta < 1$, $0 < c_1 < \Gamma < 1$, $0 \le t_- < t^- \le i(L)$, where $i(L)$ is the R-injectivity radius of L.
3: **Initialize**: $j = 1$, $t_1^* > 0$, $x_1 \in M$ and $P_1 \in S(n)$.
4: At counter j, stop if $0 \in \partial f(x_j)$. Otherwise $(x_{j+1}, t_{j+1}^*, \rho) = Descent(x_j, P_j, t_j^*)$ and update P_j.
5: **if** $\rho \ge \Gamma$ **then** $t_{j+1}^* = \theta^{-1} t_{j+1}^*$.
6: **end if**
7: **Small step size safeguard rule** $t_{j+1}^* = \max\{t_-, t_{j+1}^*\}$.
8: **Large step size safeguard rule** $t_{j+1}^* = \min\{t^-, t_{j+1}^*\}$ and go to Step 4.

R-injectivity radius at a point x of a Riemannian manifold is the largest radius for which the retraction R at x is a diffeomorphism. The R-injectivity radius of L is the infimum of the injectivity radii at all points of L. 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 4.5. *Let $f: M \to \mathbb{R}$ be locally Lipschitz on a Riemannian manifold M and $L = \{x \in M : \ f(x) \le f(x_1)\}$ be bounded. Assume that x_j is the sequence generated by Algorithm 2. Then the following are satisfied:*

- *If the standard model of f is strict, then x_j has at least one accumulation point 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.*

Proof. Assuming $0 \notin \partial f(x_j)$, then by Theorem 4.4 we deduce that after a finite number of iterations k_j the descent step finding algorithm finds a $g_{k_j} \in \partial f(x_j)$ 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 $\rho \ge c_1$. Set $d_{k_j}^* = -t_{k_j} P_j^{-1} g_{k_j}$. Therefore,

$$f(x_j) - f(x_{j+1}) \ge c_1\big(f(x_j) - \Phi_{k_j}(x_j, d_{k_j}^*)\big). \tag{4.2}$$

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

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

Consequently,

$$\frac{1}{t_{k_j}}\|d_{k_j}^*\|_j^2 \le \frac{1}{c_1}\big(f(x_j) - f(x_{j+1})\big), \tag{4.3}$$

where $\|d^*_{k_j}\|_j^2 = \langle P_j d^*_{k_j}, d^*_{k_j}\rangle$. Now summing (4.3) 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 $\frac{1}{t_{k_j}} \|d^*_{k_j}\|_j^2 \to 0$, and 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) there exists an infinite subsequence $j \in \mathcal{N}$ of $\mathbb{N}$ where $g^*_{k_j} \to 0$, $j \in \mathcal{N}$. 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 $x_j \to x^*$, $j \in \mathcal{N}$. By the subgradient inequality, we have

$$\langle g^*_{k_j}, h\rangle \leq \Phi_{k_j}(x_j, d^*_{k_j} + h) - \Phi_{k_j}(x_j, d^*_{k_j}) \leq \Phi(x_j, d^*_{k_j} + h) - \Phi_{k_j}(x_j, d^*_{k_j}),$$

for every $h \in T_{x_j}M$. By (4.2), we have

$$\langle g^*_{k_j}, h\rangle \leq \Phi(x_j, d^*_{k_j} + h) - f(x_j) + \frac{1}{c_1}\big(f(x_j) - f(x_{j+1})\big).$$

Setting $h = -d^*_{k_j} + L_{x^* x_j}(h')$, where h' is an arbitrary vector in $T_{x^*}M$, we obtain

$$\langle g^*_{k_j}, -d^*_{k_j}\rangle + \langle g^*_{k_j}, L_{x^* x_j}(h')\rangle \leq \Phi(x_j, L_{x^* x_j}(h')) - f(x_j) + \frac{1}{c_1}\big(f(x_j) - f(x_{j+1})\big).$$

Note that for every $x_j \in L$, $\Phi_{k_j} : T_{x_j}M \to \mathbb{R}$ is Lipschitz with constant K, because

$$\begin{aligned}\Phi_{k_j}(x_j, v) - \Phi_{k_j}(x_j, w) &= \max_{g \in \mathcal{G}_k}\langle g, v\rangle - \max_{g \in \mathcal{G}_k}\langle g, w\rangle \\ &\leq \langle g_v, v\rangle - \langle g_v, w\rangle \leq \langle g_v, v - w\rangle \leq \|g_v\|\|v - w\| \\ &\leq K\|v - w\|,\end{aligned}$$

where v, w are two arbitrary vectors in $T_{x_j}M$ and $g_v := \operatorname{argmax}_{g \in \mathcal{G}_k}\langle g, v\rangle$. We know that t_{k_j} is bounded and for every $g^*_{k_j}$, $\|g^*_{k_j}\| \leq K$, therefore we have $\{L_{x_{k_j} x^*}(d^*_{k_j})\}$ is bounded. By limiting on $j \in \mathcal{N}$, the upper semicontinuity of Φ implies that

$$\begin{aligned}0 &\leq \lim_{j \to \infty, j \in \mathcal{N}} \Phi(x_j, L_{x^* x_j}(h')) - f(x_j) + \frac{1}{c_1}\big(f(x_j) - f(x_{j+1})\big) \\ &\leq \Phi(x^*, h') - f(x^*) = \Phi(x^*, h') - \Phi(x^*, 0).\end{aligned}$$

Since h' is arbitrary in $T_{x^*}M$, this shows that $0 \in \partial\Phi(x^*, 0) \subset \partial f(x^*)$, which proves the claim.

ii) Infinite subsequences $\mathcal{J}$ of $\mathbb{N}$ satisfy $\|g^*_{k_j}\| \geq \eta > 0$ for all $j \in \mathcal{J}$. 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, 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 the set of infinite subsequences $\mathcal{J}$ of $\mathbb{N}$ satisfying $\|g^*_{k_j}\| \geq \eta > 0$ for all $j \in \mathcal{J}$ 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 sequence 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 $j \in \mathcal{J}_1$. First we claim that $B := \{(x_j, R_{x_j}(d^*_k)) :\ k \leq k_j,\ j \in \mathcal{J}_1, t_k \leq \mathcal{V}\}$ is bounded for fixed $\mathcal{V}$. It is enough to prove that $\{R_{x_j}(d^*_k) :\ k \leq k_j,\ j \in \mathcal{J}_1, t_k \leq \mathcal{V}\}$ is bounded. Let $g^*_k = \frac{-1}{t_k}P_j d^*_k \in \partial\Phi_k(x_j, d^*_k)$ and $g_{j_0} \in \partial f(x_j)$ be the first subgradient pick, then

$$\langle\frac{-1}{t_k}P_j d^*_k, -d^*_k\rangle \leq \Phi_k(x_j, 0) - \Phi_k(x_j, d^*_k) \leq \langle g_{j_0}, -d^*_k\rangle \leq K\|d^*_k\|,$$

which implies

$$\frac{\lambda}{t_k}\|d^*_k\|^2 \leq \frac{1}{t_k}\|d^*_k\|^2_j \leq K\|d^*_k\|,$$

where $\lambda\|.\| \leq \|.\|_j$ for all $j \in \mathbb{N}$, and hence

$$\frac{\lambda}{t_k}\|d^*_k\| \leq K,$$

consequently,

$$\|d^*_k\| \leq \frac{t_k}{\lambda}K \leq \frac{\mathcal{V}}{\lambda}K,$$

for all $j \in \mathcal{J}_1$ and $1 \leq k \leq k_j$ with $t_k \leq \mathcal{V}$. Since for all $x \in M$ and $g \in T_xM$, there exist $m > 0$ and $M > 0$ such that

$$m\|g\| \leq \operatorname{dist}(x, R_x(g)) \leq M\|g\|,$$

the proof of the claim is complete. 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

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

Moreover, $t_{k_j} = \theta_{k_j-\nu_j}t_{k_j-\nu_j}$ and $\tilde{g}_j = -\theta_{k_j-\nu_j}t^{-1}_{k_j}P_j d^*_{k_j-\nu_j} \in \partial\Phi_{k_j-\nu_j}(x_j, d^*_{k_j-\nu_j})$. We prove that $\tilde{g}_j \to 0$ and therefore $0 \in \partial f(\hat{x})$. First, we prove that $\{\tilde{g}_j :\ j \in \mathcal{J}_1\}$ is bounded. Note that

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

Therefore, the left hand side behave asymptotically like $\lambda\|\tilde{g}_j\|\|d^*_{k_j-\nu_j}\|$. Let $\tilde{g}_{0_j} \in \partial f(x_j)$ be the first subgradient pick, then

$$\lambda\|\tilde{g}_j\|\|d^*_{k_j-\nu_j}\| \leq \|\tilde{g}_{0_j}\|\|d^*_{k_j-\nu_j}\| \leq K\|d^*_{k_j-\nu_j}\|.$$

which proves that $\{\tilde{g}_j : j \in \mathcal{J}_1\}$ is bounded. Now if $\tilde{g}_j$ does not converge to zero, there exists $\theta > 0$ such that $\|\tilde{g}_j\| \geq \theta$ for all $j \in \mathcal{J}_1$. Therefore, by subgradient inequality and the asymptotic behavior of $\langle \tilde{g}_j, -d^*_{k_j-\nu_j}\rangle$, we have

$$\lambda\theta\|d^*_{k_j-\nu_j}\| \leq f(x_j) - \Phi_{k_j-\nu_j}(x_j, d^*_{k_j-\nu_j}).$$

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

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

Hence,

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

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. Therefore, $0 \in \partial f(\hat{x})$, which completes the proof of the first part of the theorem.

To prove the second 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

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

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 $V := \{x \in M : \operatorname{dist}(x, L') < \epsilon\} \subset \bigcup_{i=1}^r U(x_i)$. 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

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

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 $\mathrm{dist}(R_{x_j}(d^*_{k_j}), x_j) \to 0$. 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,* to appear in Adv. Comput. Math.

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

[10] 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.

[11] K. Kurdyka, *On gradients of functions definable in o-minimal structures,* Ann. Inst. Fourier., 48 (1998), pp. 769-783.

[12] 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.

[13] S. Lang, *Fundamentals of Differential Geometry,* Graduate Texts in Mathematics, Vol. 191, Springer, New York, 1999.

[14] J. Nocedal, S. J. Wright,*Numerical Optimization,* Springer, 1999.

[15] D. Noll, *Convergence of non-smooth descent methods using the Kurdyka-Lojasiewicz Inequality,* J. Optim. Theory Appl., 160(2014), pp. 553 -572.

[16] D. Noll, O. Prot, A. Rondepierre, *A proximity control algorithm to minimize nonsmooth and nonconvex functions,* Pac. J. Optim., 4(2008), pp. 569-602.

[17] W. Ring, B. Wirth, *Optimization methods on Riemannian manifolds and their application to shape space*, SIAM J. Optim., 22(2) (2012), pp. 596-627.

[18] S. T. Smith, *Optimization techniques on Riemannian manifolds*, Fields Institute Communications, 3 (1994), pp. 113-146.

[19] C. Udriste, *Convex Functions and Optimization Methods on Riemannian Manifolds*, Kluwer Academic Publishers, Dordrecht, Netherlands, 1994.