NONSMOOTH TRUST REGION ALGORITHMS FOR LOCALLY LIPSCHITZ FUNCTIONS ON RIEMANNIAN MANIFOLDS

P. GROHS*, S. HOSSEINI*

ABSTRACT. This paper presents a Riemannian trust region algorithm for unconstrained optimization problems with locally Lipschitz objective functions defined on complete Riemannian manifolds. To this end we define a function $\Phi : TM \to \mathbb{R}$ on the tangent bundle TM, and at k-th iteration, using the restricted function $\Phi|_{T_{x_k}M}$ where $T_{x_k}M$ is the tangent space at x_k, a local model function Q_k that carries both first and second order information for the locally Lipschitz objective function $f : M \to \mathbb{R}$ on a Riemannian manifold M, is defined and minimized over a trust region. We establish the global convergence of the proposed algorithm. Moreover, using the Riemannian ε-subdifferential, a suitable model function is defined. Numerical experiments illustrate our results.

1. INTRODUCTION

Most classical problems considered in optimization are formulated in Banach spaces, where the linear structure plays an important role. However, many problems in computer vision, robotics, signal processing and geometric mechanics, to name but a few, are more conveniently expressed as optimization problems on Riemannian manifolds [4, 38, 34, 45, 48, 51]. Therefore it is of eminent interest to develop useful computational and theoretical tools of optimization on manifolds. This paper is concerned with the numerical solution of optimization problems defined on Riemannian manifolds where the objective function may be nonsmooth.

Many algorithms for solving the following unconstrained optimization problem,

$$\min_{x\in\mathbb{R}^n} f(x)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable, have been proposed. Trust region methods are an important class of iterative methods due to their strong global convergence and fast local convergence; see [17] . In this class of iterative methods, a step to the $k+1$-th iterate is obtained by minimizing a model function Q_k defined by

$$Q_k(x_k, d) = f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T B_k d,$$

over a restricted region centered at the current iterate. It is worth pointing out that in this model function, B_k is adequately selected and the model function preserves the first and second order information of the objective function f. The

Key words and phrases. Riemannian manifolds, Lipschitz function, Trust region method, Clarke subdifferential.

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

*ETH Zürich, Seminar for Applied Mathematics, Rämistrasse 101, 8092 Zürich, Switzerland (pgrohs@sam.math.ethz.ch, seyedehsomayeh.hosseini@math.ethz.ch).

so-called trust region ratio evaluates an agreement between the model and the actual objective reductions along the computed step. Considering the trust region ratio, one can decide whether the step is accepted or rejected. After that the trust region radius is updated and a new point is obtained.

The classical trust region methods for smooth problems cannot be used for nonsmooth ones since in general the gradient of the objective function at the current iterate does not exist. Several trust region methods for minimizing a nonsmooth objective function defined on a linear space have been presented and applied to the nonlinear equations problem, the nonlinear fitting problem, and the constrained optimization problems; see [42, 18, 3] and references therein. The most well-known nonsmooth trust region methods begin from a starting point x_1 which may not be close to the minimum of the objective function $f : \mathbb{R}^n \to \mathbb{R}$. They use a function $\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ to build at each iteration a model Q_k defined by

$$Q_k(x_k, d) = f(x_k) + \Phi(x_k, d) + \frac{1}{2} d^T B_k d,$$

which must be an approximation of $f(x_k + d)$ for small d. Therefore, they should impose some conditions on the function $\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ and the sequence of the symmetric matrices B_k. For instance, [42] is among the first works on trust region methods for unconstrained optimization problems with locally Lipschitz objective functions; in that article several conditions on B_k and $\Phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ are proposed, which insure that Q_k is an approximation for $f(x_k + d)$ for small d and the algorithm is convergent.

Nonsmooth trust region algorithms approximately solve the subproblem

$$\min_{\{d \in \mathbb{R}^n :\ \|d\| \leq \delta_k\}} Q_k(x_k, d)$$

to obtain d_k. Using the trust region ratio, either the step is accepted or rejected.

The extension of nonsmooth trust region algorithms and their (global) convergence properties to Riemannian manifolds are the subject of the present paper. 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 and new techniques need to be developed.

There is clearly a link between the techniques of optimization on manifolds and standard constrained optimization approaches. However, there are manifolds that are not defined as constrained sets in $\mathbb{R}^n$; important examples are Grassmann manifolds [6, 14] or symmetric positive definite matrices [9, 40, 54]. To solve optimization problems on these spaces, intrinsic methods are a popular method of choice.

A manifold is, per definition, locally isomorphic to a linear space via chart maps. For this reason one might wonder whether it suffices to simply work in a chart domain and use classical linear algorithms. Unfortunately, such an approach does in general not lead to useful algorithms: First of all symmetries of the underlying Riemannian manifold will in general not be respected by such algorithms. Moreover, there often does not exist a canonical useful representation for charts. But more fundamentally, localizing to a chart inevitably leads to distortions in the metric which leads to much slower convergence. Finally, we are interested in establishing global convergence of the algorithms which we study. Such a property is clearly out of reach by working in a chart domain, which is a local procedure (it should now also be clear that the mathematical analysis of global convergence requires global arguments from Riemannian geometry and cannot be deduced from corresponding

linear results). For all those reasons, together with the evident need for efficient and reliable Riemannian optimization algorithms, the construction and study of intrinsic algorithms has become a thriving area of research in the past few years.

The development of Riemannian nonsmooth optimization algorithms is primarily motivated by large-scale applications that have gained much popularity in recent years regarding the framework of l^1 optimization, stochastic optimization, and statistical learning. Those applications include robust, sparse, structured principal component analysis, statistics on manifolds (e.g. median calculation of positive semidefnite tensors), and low-rank optimization (matrix completion, collaborative filtering, source separation); see [32, 52, 53, 51]. Furthermore, these algorithms have many applications in image processing, computer vision, nonsmooth constrained optimization problems on linear spaces; [5, 15, 21].

Previous Work. For the optimization of smooth objective functions many classical methods for unconstrained minimization, such as Newton-type and trust-region methods, have been successfully generalized to problems on Riemannian manifolds; see [1, 4, 19, 38, 44, 49, 50, 55]. The recent monograph [2] by Absil, Mahony and Sepulchre discusses, in a systematic way, the framework and many numerical first-order and second-order manifold-based algorithms for minimization problems on Riemannian manifolds with an emphasis on applications to numerical linear algebra; see [2].

As we discussed above, one of the most important methods in the unconstrained optimization of smooth functions is the trust region method due to its strong global convergence and fast local convergence. In [1], a Riemannian trust region method for smooth functions on Riemannian manifolds was introduced. Similar to Euclidean trust region methods, the Riemannian trust region method ensures global convergence properties while allowing superlinear local convergence. The trick in [1] was to define a retraction R on a Riemannian manifold M that defines for any $x \in M$, a one-to-one correspondence R_x between a neighborhood of x in M and a neighborhood of 0_x in the tangent space T_xM. Using this retraction, the objective function f on M is lifted to an objective function $\overline{f_x} = f \circ R_x$ on T_xM. Then a quadratic model of $\overline{f_x}$ is defined, and a classical method on the Euclidean space T_xM is used to compute a minimizer of the model within a trust region around $0_x \in T_xM$. Afterward, the minimizer is lifted back to M to be a new candidate for the next iterate. The most noticeable point in the mentioned trust region method is that it does not deal with a unique objective function since in any iteration the retraction is changed. However, the authors proved that under some conditions, the nice properties of the classical trust region method are preserved in their Riemannian generalizations.

In considering optimization problems with nonsmooth objective functions on Riemannian manifolds, it is necessary to generalize concepts of nonsmooth analysis to Riemannian manifolds. In the past few years a number of results have been obtained on numerous aspects of nonsmooth analysis on Riemannian manifolds; see [7, 8, 27, 28, 29, 37].

Recently, some mathematicians have started developing nonsmooth optimization algorithms for manifold settings although their attempts are limited to generalizing some subgradient based and proximal point algorithms. In [10, 13, 22], constrained minimization problems on Hadamard manifolds are solved using a generalization of the proximal point method. There have also been some studies by Ferreira,

Bento and Oliveira and their colleagues who generalized subgradient-type methods for convex and quasiconvex functions defined on Riemannian manifolds; see [11, 12, 20, 23, 41]. Finally, it is worth mentioning paper [21], which presents a survey on Riemannian geometry methods for smooth and nonsmooth constrained optimization as well as gradient and subgradient descent algorithms on a Riemannian manifold. In that paper, the methods are illustrated by applications from robotics and multi antenna communication.

Contributions. Our main contributions are twofold. First, we impose some conditions on a function $\Phi : TM \to \mathbb{R}$, where TM is the tangent bundle of a Riemannian manifold M, to insure that our model function

$$Q_k(x_k, \cdot) = f(x_k) + \Phi(x_k, \cdot) + \frac{1}{2}\langle B_k \cdot, \cdot\rangle : T_{x_k}M \to \mathbb{R}$$

preserves the first order information of the locally Lipschitz objective function $f : M \to \mathbb{R}$. Our proposed conditions are generalizations of the conditions in [42]. Our first main result, Theorem 3.6, states that, provided these assumptions are satisfied, then the Riemannian trust region scheme which iteratively (approximately) minimizes the model function Q_k on a ball in $T_{x_k}M$ converges globally to a critical point of f.

As the second main contribution of this paper, in Section 4 we propose several choices for a suitable function Φ, the simplest one being the Clarke generalized directional derivative of f as a function $TM \to \mathbb{R}$. However, this function is only practical if the subdifferential of f can be computed explicitly, which might not be the case. Therefore, based on an approximation scheme for the so-called ε-subdifferential first introduced in [25] for the Riemannian case, we also present a numerical approximation of the subdifferential which is efficiently computable and which still yields a globally convergent algorithms. This is carried out in Section 4.

To the best of our knowledge, the resulting algorithm is the first practical nonsmooth trust region algorithm for locally Lipschitz functions defined on Riemannian manifolds. In Section 5 we present some numerical experiments. Our proposed algorithm is implemented in MATLAB and applied to some nonsmooth problems with locally Lipschitz objective functions. Numerical results show that the proposed algorithm has a far better performance in some problems in comparison with existing first order methods, such as subgradient descent.

2. PRELIMINARIES

Throughout this paper, M is an n-dimensional complete manifold endowed with a Riemannian metric $\langle ., . \rangle$ on the tangent space T_xM; see [33]. As usual $B(x, \delta)$ denotes the open ball with respect to the Riemannian distance centered at x with radius δ. For the point $x \in M$, $\exp_x : U_x \to M$ defines the exponential function at x, where U_x is an open subset of T_xM. For a minimizing geodesic $\gamma : [0, l] \to M$ connecting x to y in M, and for a vector $v \in T_xM$ there is a unique parallel vector field P along γ such that $P(0) = v$, this is called the parallel translation of v along γ. The mapping $T_xM \ni v \mapsto P(l) \in T_yM$ is a linear isometry from T_xM onto T_yM. This map is denoted by L_{xy}. An easy consequence of the definition of the parallel translation along a curve as a solution to an ordinary linear differential equation implies that the mapping

$$C : TM \to T_{x_0}M,\ C(x, \xi) = L_{xx_0}(\xi), \tag{2.1}$$

when x is in a neighborhood U of x_0, is well defined and continuous at (x_0, ξ_0); that is, if $(x_n, \xi_n) \to (x_0, \xi_0)$ in TM then $L_{x_n x_0}(\xi_n) \to L_{x_0 x_0}(\xi_0) = \xi_0$, for every $(x_0, \xi_0) \in TM$; see [7, Remark 6.11].

Note that $i_M(x)$ denotes the injectivity radius of M at x; that is the supremum of the radius r of all balls $B(0_x, r)$ in T_xM for which $\exp_x$ is a diffeomorphism from $B(0_x, r)$ onto $B(x, r)$. If U is a compact subset of a Riemannian manifold M and $i(U) := \inf\{i_M(x) : \ x \in U\}$, then $0 < i(U)$; see [30].

A set S in a Riemannian manifold M is said to be convex if every two points $p_1,\ p_2 \in S$ can be joined by a unique geodesic whose image belongs to S.

A real valued function f is said to be locally Lipschitz on M if f is Lipschitz near x, for every $x \in M$; that is for every $x \in M$, there exit an open neighborhood $B(x, \delta)$ and a real number $k > 0$ such that $\mid f(z) - f(y) \mid \leq k\text{dist}(z, y)$ for every $z, y \in B(x, \delta)$, where dist is the Riemannian distance on M.

It is easy to see that the metric on TM can be defined as follows; see [16]. Assume that $u,\ v \in TM$, $\pi : TM \to M$ is the projection map, and let γ be a piecewise smooth path from $x = \pi(u)$ to $y = \pi(v)$, whose derivative is never zero. Let L^{γ}_{xy} be the parallel translation along γ. We define the square of the distance from u to v as the infimum over all paths γ from $\pi(u)$ to $\pi(v)$ of

$$\text{dist}_{\gamma}(u, v)^2 = \|L^{\gamma}_{xy}(u) - v\|^2 + \text{length}(\gamma)^2.$$

To justify the above assertion, note that over a small piece of γ, parallel translation gives us a canonical identification of any two tangent spaces, and hence the tangent bundle is metrically the product of an interval and $\mathbb{R}^n$.

If x and y are two points in the same convex neighbourhood in M, let γ be the geodesic joining them and L_{xy} be the parallel translation along γ from the tangent space at x to the tangent space at y. Then

$$\text{dist}_{TM}(u, v)^2 \leq \|L_{xy}(u) - v\|^2 + \text{dist}(x, y)^2.$$

Moreover,

$$\text{dist}(x, y) \leq \text{dist}_{TM}(u, v).$$

3. A nonsmooth trust region method on Riemannian manifolds

Assume that $f : M \to \mathbb{R}$ is a locally Lipschitz function on a complete Riemannian manifold M.

The trust region algorithm proposed in the present section, which is based on [42], relies on the choice of a function $\Phi : TM \to \mathbb{R}$ (modeling the derivative of f) and a sequence $\{B_k : \ k = 1, 2, ..\}$ of $n \times n$ symmetric matrices (modeling the Hessian of f) from which we build a sequence of model functions

$$Q_k : \left\{ \begin{array}{ccc} T_{x_k}M & \to & \mathbb{R} \\ (x_k, d) & \mapsto & f(x_k) + \Phi(x_k, d) + \frac{1}{2}\langle B_k d, d\rangle \end{array} \right.$$

analogous to a second order Taylor expansion in the Euclidean case.

Then the proposed trust region scheme iteratively computes approximative minima of these model functions over a trust region as follows:

Let

(3.1)

$$d_k^* = \text{argmin}\{Q_k(x_k, d_k) = f(x_k) + \Phi(x_k, d_k) + 1/2\langle B_k d_k, d_k\rangle : \ d_k \in T_{x_k}M,\ \|d_k\| \leq \delta_k\}.$$

Then in each iteration step we compute d_k^* or, if this turns out to be not practical, an approximation $\bar{d}_k$ of d_k^* in the sense that

$$f(x_k) - Q_k(x_k, \bar{d}_k) \geq c_0[f(x_k) - Q_k(x_k, d_k^*)] \quad \text{and} \quad \|\bar{d}_k\| \leq \delta_k, \tag{3.2}$$

for a fixed constant $0 < c_0 \leq 1$.

Then, depending on the trust ratio in each step we either reduce the trust region by reducing the value δ_k and solving (3.1), or we update

$$x_{k+1} := \exp_{x_k}(\bar{d}_k), \tag{3.3}$$

see Algorithm 1.

Remark 3.1. *A natural choice for the function Φ would be the Clarke generalized directional derivative of f as defined below in (4.1).*

In case the Clarke generalized directional derivative of f is not given explicitly (as is the case for instance in [21, Sections 3.2 and 3.3] *and* [31]*), in Section 4 we also propose a numerical approximation which only solves (3.1). The choice of matrices $\{B_k\}$ is quite general; in our results they simply need to be uniformly bounded but even this property can be weakened; see for instance* [42]*. In our algorithms we have used a Riemannian generalization of the BFGS method to iteratively update the matrices B_k; see also* [43].

Remark 3.2. *Instead of using the exponential map to update x_k, we can choose a retraction $R : TM \to M$. The notion of retraction on a manifold, includes all first-order approximations to the Riemannian exponential; see* [1]*. The retraction can be used to take a step in the direction of a tangent vector. Using a good retraction amounts to finding an approximation of the exponential mapping that can be computed with low computational cost while not adversely affecting the behavior of the optimization algorithm.*

In the remainder of this section we formulate assumptions on the Φ to ensure global convergence of the sequence $(x_k)_k$ to a critical point of f, in the following sense.

Definition 3.3. *With $f : M \to \mathbb{R}$ and $\Phi : TM \to \mathbb{R}$, define*

$$\psi(x, \delta) = \sup\{-\Phi(x, d) : \ d \in T_xM, \ \|d\| \leq \delta\}. \tag{3.5}$$

The point $x \in M$ is called a critical point with respect to Φ of the objective function f if there exists $\delta > 0$ such that $\psi(x, \delta) = 0$.

In the Euclidean case the notion of critical point as defined in Definition 3.3 is common in the context of nonsmooth trust region methods; see [42, 18]. Note that the upper Dini directional derivative of f at x in the direction $d \in T_xM$ denoted by $f^+(x; d)$ is defined as follows;

$$f^+(x; d) := \limsup_{t \downarrow 0} \frac{f(\exp_x(td)) - f(x)}{t}.$$

A point x is called a Dini stationary point if for all $d \in T_xM$, $f^+(x; d) \geq 0$. Under weak assumptions a critical point corresponds to a so-called 'Dini stationary point' [42]. Indeed, we may impose the following assumption on Φ in order to be able to prove that any critical point of f is also a Dini stationary point of f.

Algorithm 1 A nonsmooth trust region algorithm on Riemannian manifolds

Data: An n-dimensional complete Riemannian manifold (M, g); a real valued locally Lipschitz function f on M.
Parameters: $\delta_0 > 0,\ \delta_0 > \delta_1 > 0,\ c_0, c_1, c_2, c_3, c_4 > 0,\ c_2 < c_1 < 1,\ c_0 \leq 1.$
Input: initial iterate $x_1 \in M$, and $B_1 \in S(n)$, where $S(n)$ denotes the space of symmetric $n \times n$-matrices.
Output: sequence of iterates $\{x_k\}$.
for $k = 1, 2, ...$ **do** find

(3.4)
$$d_k^* = \operatorname{argmin}\{Q_k(x_k, d_k) = f(x_k) + \Phi(x_k, d_k) + 1/2\langle B_k d_k, d_k\rangle :\ d_k \in T_{x_k}M,\ \|d_k\| \leq \delta_k\}.$$

where $\Phi : TM \to \mathbb{R}$ is a given function.

Assume $\bar{d}_k$ is an inexact solution of 3.4 in the sense that

$$f(x_k) - Q_k(x_k, \bar{d}_k) \geq c_0[f(x_k) - Q_k(x_k, d_k^*)]$$

and $\|\bar{d}_k\| \leq \delta_k$.

if $\bar{d}_k = 0$ **then**, Stop.
else

$$r_k = \frac{f(x_k) - f(\exp_{x_k}(\bar{d}_k))}{f(x_k) - Q_k(x_k, \bar{d}_k)},$$

if $c_2 < r_k$, **then** $x_{k+1} = \exp_{x_k}(\bar{d}_k)$ and update B_k.
end if
if $r_k \leq c_2$, **then** $x_{k+1} = x_k$, $\delta_{k+1} = c_3\delta_k$.
else
if $c_2 < r_k \leq c_1$, **then** $\delta_{k+1} = \delta_k$.
else $\delta_{k+1} = \min\{c_4\delta_k, \delta_0\}$.
end if
end if
end if
end for

Assumption 3.1. *Assume that D is a bounded open convex set containing $N := \{x \in M :\ f(x) \leq f(x_0)\}$ and for all $x \in D$ and $d \in T_xM$ it holds that*

$$\liminf_{t \downarrow 0} \frac{\Phi(x, td)}{t} \leq f^+(x; d),$$

where $f^+(x; d)$ is the upper Dini directional derivative of f at x in the directional $d \in T_xM$.

It is obvious that if x is a critical point of f in the sense of Definition 3.3 then for t small enough $\Phi(x, td) \geq 0$ and therefore $\liminf_{t\downarrow 0} \frac{\Phi(x,td)}{t} \geq 0$. Hence, using Assumption 3.1, we have that for all $d \in T_xM$, $f^+(x; d) \geq 0$. One can also show that a local minimizer x of a locally Lipschitz function $f : M \to \mathbb{R}$ is always a critical point, provided that the function Φ satisfies some natural assumption. This will be done in Lemma 3.4 below.

3.1. **Convergence Conditions.** Let $\Phi : TM \to \mathbb{R}$ and $I = \{1, 2, ...\}$, $N = \{x \in M :\ f(x) \leq f(x_1)\}$ where x_1 is a starting point for Algorithm 1. Let

$\{B_k : k = 1, 2, ..\}$ be a sequence of $n \times n$ symmetric matrices. In this section, we make some assumptions on Φ and B_k and show that these assumptions ensure global convergence of Algorithm 1 to a critical point of f in the sense of Definition 3.3. We extend results of [42] to the Riemannian case.

We start with the following assumptions on the function Φ.

Assumption 3.2. *Let $\Phi : TM \to \mathbb{R}$. Assume that*

$$\Phi(x, 0_x) = 0 \qquad \forall x \in M, \tag{3.6}$$

$$\Phi(x, \alpha d) \leq \alpha \Phi(x, d), \quad \forall (x, d) \in TM, 0 \leq \alpha \leq 1, \tag{3.7}$$

$$\text{for all } x \in M, \Phi|_{T_x M} \text{ is lower semi continuous}, \tag{3.8}$$

for any $(x, d) \in TM$ it holds that

$$f(\exp_x(d)) - f(x) \leq \Phi(d) + o(\|d\|), \tag{3.9}$$

and there exists δ_* such that

$$\text{for all } \delta < \delta_* \text{ the function } \psi(., \delta) \text{ is lower semi continuous}, \tag{3.10}$$

where ψ is defined in (3.5) and the implicit constant in the o-term is uniform over compact sets.

First we show that, provided Φ satisfies Assumption 3.2, every local minimum of a locally Lipschitz function $f : M \to \mathbb{R}$ is a critical point in the sense of Definition 3.3.

Lemma 3.4. *Suppose that $f : M \to \mathbb{R}$ and $\Phi : TM \to \mathbb{R}$ such that Assumption 3.2 holds. Then every local minimizer of f is a critical point in the sense of Definition 3.3.*

In order to establish this result we utilize the following simple lemma which is a straightforward extension of [42, Lemma 3.1].

Lemma 3.5. *Suppose that Φ satisfies Assumption 3.2 and let ψ be defined by (3.5). Then for any $x \in M$, the function $d \mapsto \psi(x, d)$, $d \in T_x M$, is nonnegative and nondecreasing and for any $\alpha \in [0, 1]$,*

$$\psi(x, \alpha\delta) \geq \alpha \psi(x, \delta),$$

and for any $\delta > 0$,

$$\psi(x, \delta) = 0 \text{ if and only if } \psi(x, 1) = 0.$$

We can now proceed to the

Proof of Lemma 3.4. Suppose that $\bar{x}$ is a local minimizer of f on some neighborhood $B(\bar{x}, \delta)$ which is not a critical point, then Lemma 3.5 implies that $\psi(\bar{x}, 1) > 0$. Therefore, there exists $d_{\bar{x}}$ with $\|d_{\bar{x}}\| \leq 1$ such that $\Phi(\bar{x}, d_{\bar{x}}) < 0$. Assume that $t < 1$ is small enough, hence by Assumptions (3.9) and (3.7)

$$f(\exp_{\bar{x}}(t d_{\bar{x}})) - f(\bar{x}) \leq \Phi(\bar{x}, t d_{\bar{x}}) + o(\|t d_{\bar{x}}\|) \leq t \Phi(\bar{x}, d_{\bar{x}}) + o(\|t d_{\bar{x}}\|).$$

Therefore, if t is small enough

$$\frac{f(\exp_{\bar{x}}(t d_{\bar{x}})) - f(\bar{x})}{t} \leq \Phi(\bar{x}, d_{\bar{x}})/2 < 0,$$

which means that $\bar{x}$ is not a local minimizer of f and this is a contradiction. □

Now we want to establish conditions for the convergence of Algorithm 1 to a critical point of f.

We need to make more assumptions on the starting point and the sequence $(B_k)_k$.

Assumption 3.3. *Recall $N = \{x \in M : \ f(x) \leq f(x_1)\}$ where x_1 is the starting point of Algorithm 1. Assume that N is bounded. Furthermore assume that there exists $C > 0$ such that $\|B_k\| \leq C$, for all $k = 1, 2, ...$*

The main theorem of this section reads as follows.

Theorem 3.6. *Suppose that Φ and $(B_k)_k$ are such that Assumptions 3.2 and 3.3 hold true. If $\bar{x}$ is an accumulation point of $\{x_k\}$, generated by Algorithm 1, then $\bar{x}$ is a critical point of f in the sense of Definition 3.3.*

In order to establish Theorem 3.6 we require the following result which can be proved exactly in the same way as Lemma 3.2 in [42].

Lemma 3.7. *Let $x \in M$ and $\delta > 0, 0 < c_0 \leq 1$. Define*

$$d^* := \operatorname{argmin}\{Q_x(d) := f(x) + \Phi(x, d) + 1/2\langle Bd, d\rangle : \ d \in T_xM, \ \|d\| \leq \delta\}$$

and let $\bar{d}$ be an approximate solution of the above problem in the sense that

$$f(x) - Q_x(\bar{d}) \geq c_0[f(x) - Q_x(d^*)] \quad \text{and} \quad \|\bar{d}\| \leq \delta.$$

Then, for all $\nu \geq \delta$ it holds that

$$f(x) - Q_x(\bar{d}) \geq \frac{c_0}{2\nu}\psi(x, \nu) \min\{\delta, \psi(x, \nu)/(\|B\|\nu)\},$$

where the second term in the min notation is understood as ∞ if $B = 0$.

The following lemma proves that if Algorithm 1 generates a sequence $\{x_k\}$ with $x_k = \bar{x}$ for all large k, then $\bar{x}$ is a critical point of f.

Lemma 3.8. *Suppose that $\bar{x}$ is an accumulation point of $\{x_k\}$ which is not a critical point. Then there exist $\epsilon > 0$ and $\beta > 0$ such that for all k satisfying*

$$dist(x_k, \bar{x}) < \epsilon, \ 0 < \delta_k < \beta, \quad \|B_k\| \leq C, \tag{3.11}$$

we have

$$r_k = \frac{f(x_k) - f(\exp_{x_k}(\bar{d_k}))}{f(x_k) - Q_k(x_k, \bar{d_k})} > c_2,$$

where $x_k, \ \delta_k, c_2$ are the same as in algorithm 1.

Proof. Since $\bar{x}$ is not a critical point, Lemma 3.5 implies that $\psi(\bar{x}, \bar{\delta}) > 0$ for every $\bar{\delta} > 0$. Using Assumption (3.10), let δ_* be such that for all $0 \leq \delta \leq \delta_*$, $\psi(., \delta)$ is lower semi continuous. Therefore, there exist $\theta > 0$ and $\epsilon > 0$ such that $\psi(\bar{x}, \delta_*) = 2\theta > 0$ and

$$\psi(x_k, \delta_*) > \theta \text{ provided } \operatorname{dist}(x_k, \bar{x}) < \epsilon.$$

Assuming that $\delta_k < \beta < \delta_*$, then by Assumption (3.9) and Lemma 3.7

(3.12)
$$\begin{aligned}\frac{f(\exp_{x_k}(\bar{d}_k)) - Q_k(x_k,\bar{d}_k)}{f(x_k) - Q_k(x_k,\bar{d}_k)} &\leq \frac{o(\|\bar{d}_k\|) - 1/2\langle B_k\bar{d}_k,\bar{d}_k\rangle}{(c_0/2\delta_*)\psi(x_k,\delta_*)\min\{\delta_k,\psi(x_k,\delta_*)/(\|B_k\|\delta_*)\}}\\ &\leq \frac{o(\|\bar{d}_k\|) + 1/2\|B_k\|\|\bar{d}_k\|^2}{(c_0/2\delta_*)\psi(x_k,\delta_*)\min\{\delta_k,\psi(x_k,\delta_*)/(\|B_k\|\delta_*)\}}\\ &\leq \frac{o(\|\bar{d}_k\|) + 1/2C\|\bar{d}_k\|\delta_*}{(c_0/2\delta_*)\psi(x_k,\delta_*)\min\{\delta_k,\psi(x_k,\delta_*)/(\|B_k\|\delta_*)\}}\\ &\leq \frac{o(\delta_k)}{(c_0/2\delta_*)\theta\min\{\delta_k,\theta/(C\delta_*)\}}.\end{aligned}$$

Assuming that β is small enough, then

$$\frac{f(\exp_{x_k}(\bar{d}_k)) - Q_k(x_k,\bar{d}_k)}{f(x_k) - Q_k(x_k,\bar{d}_k)} < 1 - c_2,$$

which completes the proof.

□

We can now proceed to the

Proof of Theorem 3.6. Note that M is a complete Riemannian manifold and N is compact. We assume that an infinite subsequence $\{x_k : k \in I_*\}$ converges to some $\bar{x}$ which is not a critical point. Let $I_0 = \{k : r_k > c_2\}$ and $I_* \subset I_0$. Therefore, by Lemma 3.7

$$\begin{aligned}f(x_k) - f(x_{k+1}) &\geq c_2(f(x_k) - Q_k(x_k,d_k))\\ &\geq \frac{c_0c_2}{2\delta_0}\psi(x_k,\delta_0)\min\{\delta_k,\psi(x_k,\delta_0)/(\|B_k\|\delta_0)\}\\ &\geq \frac{c_0c_2}{2\delta_0}\psi(x_k,\delta_0)\min\{\delta_k,\psi(x_k,\delta_0)/(C\delta_0)\}.\end{aligned}$$

Therefore,

$$+\infty > f(x_0) - f(\bar{x}) \geq \frac{c_0c_2}{2\delta_0}\sum_{k\in I_0}\psi(x_k,\delta_0)\min\{\delta_k,\psi(x_k,\delta_0)/(C\delta_0)\}. \tag{3.13}$$

By Lemma 3.5, we may find β and ϵ_0 such that for all $0 < \epsilon \leq \epsilon_0$ and $x_k \in \mathrm{cl}B(\bar{x},\epsilon)$,

$$\psi(x_k,\delta_0) \geq \psi(x_k,\delta_*) \geq \beta > 0. \tag{3.14}$$

Set $I_1 = \{k \in I_0 : \delta_k \geq \dfrac{\psi(x_k,\delta_0)}{C\delta_0}\}$ and $I_2 = I_0 \setminus I_1$. Then (3.13) shows that

$$\sum_{k\in I_1}\big(\psi(x_k,\delta_0)\big)^2 < +\infty, \tag{3.15}$$

and

$$\sum_{k\in I_2}\psi(x_k,\delta_0)\delta_k < +\infty. \tag{3.16}$$

Hence, we may assume that there is $N(\epsilon) > 0$ such that for all $k \geq N(\epsilon)$ and $k \in I_*$, we have

$$\mathrm{dist}(x_k,\bar{x}) \leq \frac{\epsilon}{2} \text{ and } \sum_{k\in I_2,\ k\geq N(\epsilon)}\psi(x_k,\delta_0)\delta_k \leq \frac{\epsilon}{2}. \tag{3.17}$$

By (3.17) and (3.14), we can prove that for all k with $k_0 < k \leq k_1$, where k_0 is an index in I_* with $k_0 \geq N(\epsilon)$ and k_1 is the first index in I_1 greater than k_0,

$$\begin{aligned}
\operatorname{dist}(x_k, \bar{x}) &\leq \operatorname{dist}(x_{k-1}, x_k) + \operatorname{dist}(x_{k-1}, \bar{x}) \\
&\leq \operatorname{dist}(x_{k_0}, \bar{x}) + \sum_{k_0 \leq m < k, m \in I_2} \operatorname{dist}(x_{m+1}, x_m) \\
&\leq \operatorname{dist}(x_{k_0}, \bar{x}) + \sum_{k_0 \leq m < k, m \in I_2} \operatorname{dist}(\exp_{x_m}(d_m), x_m) \\
&\leq \frac{1}{2}\epsilon + \sum_{k_0 \leq m < k, m \in I_2} \delta_m \\
&\leq \epsilon,
\end{aligned}$$

and therefore

$$\psi(x_k, \delta_0) \geq \psi(x_k, \delta_*) \geq \beta > 0. \tag{3.18}$$

This means that I_1 must be finite since otherwise $\sum_{k \in I_1} \big(\psi(x_k, \delta_0)\big)^2$ is not finite. Therefore, we have $\lim_{k \to \infty} x_k = \bar{x}$ for $k \in I_0$ and $\sum_{k \in I_0} \delta_k < \infty$. This contradicts the fact that $\delta_{k+1} \geq \delta_k$ for all large $k \in I_0$ and the proof is complete. □

4. A suitable model function

In the previous Section 3, we have developed a general trust region method and established global convergence, provided that some criteria on the function Φ and the sequence $(B_k)_k$ hold true. The present section presents several choices for Φ which lead to convergent algorithms.

4.1. The Clarke generalized directional derivative.

If f were a smooth function, then clearly the function Φ would simply be the directional derivative of f. It is clear that since the function is not necessarily differentiable, we cannot use the differential of the objective function. However we might be able to use generalized directional derivatives instead. Let us continue with the definition of the Clarke generalized directional derivative for locally Lipschitz functions on Riemannian manifolds; see [27, 29].

Definition 4.1 (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*

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

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

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

Definition 4.2 (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* [27] *that*

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

where Ω_f is a dense subset of M on which f is differentiable.[1]

Therefore, we have a good candidate for the function $\Phi : TM \to \mathbb{R}$, which can be defined by

$$\Phi(x,d) := f^\circ(x,d) = \sup\{\langle \xi, d\rangle : \ \xi \in \partial f(x)\}, \tag{4.1}$$

see [27].

The resulting model function leads to a convergent algorithm as the following result shows.

Theorem 4.3. *Let $f : M \to \mathbb{R}$ be locally Lipschitz and define Φ as in (4.1). Then the function Φ satisfies Assumption 3.2. In particular, Algorithm 1 converges globally to a critical point of f.*

Proof. The fact that Φ satisfies Assumption 3.2 follows directly from Theorem 2.4 and Theorem 2.9 of [27]. Therefore, by Theorem 3.6 the resulting trust region algorithm converges. □

4.2. **The ε-subdifferential.** A crucial observation is that the computation of Φ as in (4.1) can be impractical in case that no explicit expression for the subdifferential $\partial f(x)$ is available. Using an approximation of the Clarke subdifferential, we overcome this problem and define a local model that is practically and efficiently implementable.

The following definition presents an approximation of the subdifferential which can be computed approximately; see [25].

Definition 4.4 (ε-subdifferential)**.** *Let $f : M \to \mathbb{R}$ be a locally Lipschitz function on a Riemannian manifold M, $\varepsilon < i_M(x)$. We define the ε-subdifferential of f at x denoted by $\partial_\varepsilon f(x)$ as follows;*

$$\partial_\varepsilon f(x) = \operatorname{conv}\{d\exp_x^{-1}(y)(\partial f(y)) : y \in \operatorname{cl}B(x,\varepsilon)\}.$$

The following result has been proven in [24].

Lemma 4.5. *Let U be a compact subset of M and $\varepsilon < i(U)$; then for every open neighborhood W in U, the set valued mapping $\partial_\varepsilon f : W \to TM$ is upper semi continuous.*

Now we construct a suitable model function for the following unconstrained optimization problem,

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

where $f : M \to \mathbb{R}$ is a locally Lipschitz function. Assume that D is a bounded open subset of M and $\varepsilon < i(\operatorname{cl}(D))$. We define $\Phi : TD \to \mathbb{R}$ by

$$\Phi(x,d) := \sup\{\langle \xi, d\rangle : \ \xi \in \partial_\varepsilon f(x)\}, \text{ for every } x \in D. \tag{4.2}$$

We now show that this function Φ satisfies Assumption 3.2.

[1]Note that $\lim \operatorname{grad} f(x_i)$ in this 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, $\langle \xi_i, X(x_i)\rangle \to \langle \xi, X(x)\rangle$.

Theorem 4.6. *The function $\Phi : TD \to \mathbb{R}$ defined by (4.2) satisfies Assumption 3.2.*

Proof. It is easy to prove that Φ satisfies Assumptions (3.6) and (3.8). Moreover, $\Phi|_{T_xM}$ is upper semi continuous, to show this for every $\epsilon > 0$, we set $\delta < \epsilon/K$ where K is the Lipschitz constant of f on a neighborhood of x. Now if $\|d - w\| < \delta$, then

$$\Phi(x, w) \leq \Phi(x, w - d) + \Phi(x, d) \leq K\|w - d\| + \Phi(x, d) < \epsilon + \Phi(x, d),$$

therefore the claim is proved.

To prove that Φ satisfies Assumption (3.9), note that for $(x, d) \in TM$,

$$f(\exp_x(d)) - f(x) \leq f^\circ(x, d) + o(\|d\|) \leq \Phi(x, d) + o(\|d\|).$$

Now we prove that Φ satisfies Assumption (3.10). To this end, we first prove that: For each $\epsilon > 0$, there exists $\delta > 0$ such that for $x \in B(x_0, \delta) \subset D$

$$L_{xx_0}(\partial_\varepsilon f(x)) \subseteq \partial_\varepsilon f(x_0) + \epsilon B_{T_{x_0}M},$$

where $B_{T_{x_0}M}$ is the unit ball of $T_{x_0}M$. To see this; note that for $\epsilon > 0$ the set $\partial_\varepsilon f(x_0) + \epsilon B_{T_{x_0}M}$ is an open neighborhood of $L_{x_0x_0}(\partial_\varepsilon f(x_0)) = \partial_\varepsilon f(x_0)$. It follows from the continuity of (2.1) that there exists an open neighborhood $V \subset TM$ of $\partial_\varepsilon f(x_0)$ such that

$$C(V) \subseteq \partial_\varepsilon f(x_0) + \epsilon B_{T_{x_0}M}.$$

By the upper semi continuity of $\partial_\varepsilon f$, there exists a neighborhood V' of x_0 such that for each $x \in V'$, we have $\partial_\varepsilon f(x) \subseteq V$. Now let $x \in V'$, then

$$L_{xx_0}(\partial_\varepsilon f(x)) \subseteq \partial_\varepsilon f(x_0) + \epsilon B_{T_{x_0}M},$$

as required. Now we claim that there exists $\delta_* > 0$ such that for all $x \in D$ and $d_x \in T_xM$ with $\|d_x\| < \delta_*$, we have Φ is upper semi continuous at d_x.

To prove the claim; since $l_x(y) := d\exp_x^{-1}(y)$ is a smooth function with respect to y, we have that it is bounded on $\mathrm{cl}B(x, \varepsilon)$ by some $m_x \geq 0$, from the Lipschitzness of f on $\mathrm{cl}B(x, \varepsilon)$, Theorem 2.9 of [27] implies that for every $\xi \in \partial_\varepsilon f(x)$, $\|\xi\| \leq m_x K_x$. Since $\mathrm{cl}(D)$ is compact, there exists a finite number of neighborhoods $B(x_i, \varepsilon)$ such that $D \subset \bigcup_{i=1}^n B(x_i, \varepsilon)$. Assume that $m_1K := \min\{m_{x_i}K_{x_i} : i = 1, \ldots, n\}$, $\delta_* := m_1K$. Let x and d_x be, respectively, arbitrary elements of D and T_xM with $\|d_x\| < \delta_*$.

We prove that for each $\epsilon > 0$, there exists $\delta > 0$ such that

$$\Phi(y, d_y) < \Phi(x, d_x) + \epsilon, \text{ provided } \mathrm{dist}_{TM}(d_x, d_y) < \delta.$$

Assume that $0 < \epsilon < 1$, then there exists $\delta_1 > 0$ such that for each $y \in B(x, \delta_1) \subset D$,

$$L_{yx}(\partial_\varepsilon f(y)) \subseteq \partial_\varepsilon f(x) + \frac{\epsilon}{3Km_1} B_{T_xM}.$$

Since $\Phi|_{T_xM}$ is upper semi continuous and $\Phi(x, 0_x) = 0$, hence there exists $\delta_3 > 0$ such that $\|w_x\| < \delta_3$ implies that $\Phi(x, w_x) < \epsilon/3$. Assume that $\sigma := \min\{\delta_3, \frac{m_1K\epsilon}{3}\}$, then the continuity of (2.1) implies that there exists $\delta_2 > 0$,

$$\|L_{yx}(d_y) - d_x\| < \sigma, \text{ provided } \mathrm{dist}_{TM}(d_x, d_y) < \delta_2.$$

Let $\delta < \min\{\delta_1, \delta_2, \delta_3\}$ be such that $B(x,\delta)$ is convex. We suppose that $\mathrm{dist}_{TM}(d_x, d_y) < \delta$, then $\mathrm{dist}(x,y) < \delta$ and $y \in B(x,\delta) \subset D$. Therefore

$$\begin{aligned}
\Phi(y, d_y) &= \sup_{\xi}\{\langle \xi, d_y\rangle :\ \xi \in \partial_\varepsilon f(y)\} \\
&= \sup\{\langle L_{yx}\xi, L_{yx} d_y\rangle :\ \xi \in \partial_\varepsilon f(y)\} \\
&\le \sup_{\eta}\{\langle \eta, L_{yx} d_y\rangle :\ \eta \in \partial_\varepsilon f(x) + \frac{\epsilon}{3Km_1} B_{T_xM}\} \\
&\le \sup_{\theta, v}\{\langle \theta + v, d_x + w_x\rangle :\ \theta \in \partial_\varepsilon f(x),\ v \in \frac{\epsilon}{3Km_1} B_{T_xM}, \|w_x\| < \sigma\} \qquad (4.3) \\
&\le \sup_{\theta}\{\langle \theta, d_x\rangle + \langle \theta, w_x\rangle :\ \theta \in \partial_\varepsilon f(x), \|w_x\| < \sigma\} + \frac{\epsilon^2}{9} + \frac{\epsilon}{3} \\
&\le \Phi(x, d_x) + \frac{\epsilon^2}{9} + \frac{\epsilon}{3} + \Phi(x, w_x) \le \Phi(x, d_x) + \epsilon,
\end{aligned}$$

□

It is worth mentioning that every critical point with respect to Φ defined by (4.2) is an ε-stationary point; i.e. there is y in $\mathrm{cl}B(x,\varepsilon)$ such that $0 \in \partial f(y)$. A key property of the ε-subdifferential is that it can be approximated efficiently. In our implementations, we substitute the ε-subdifferential of the objective function f with its approximation presented in [24]. Indeed, to approximate the ε-subdifferential at x_k, we start with the gradient of an arbitrary point nearby x_k and move the gradient to the tangent space in x_k via the derivative of the logarithm mapping, and in every subsequent iteration, the gradient of a new point nearby x_k is computed and moved to the tangent space in x_k to add to the working set to improve the approximation of $\partial_\varepsilon f(x_k)$. Indeed, we do not want to provide a description of the entire ε-subdifferential set at each iteration; what we do is approximate $\partial_\varepsilon f(x_k)$ by the convex hull of its elements. In this way, let $W_l := \{v_1, ..., v_l\} \subseteq \partial_\varepsilon f(x_k)$; then we define

$$w_l := \operatorname*{argmin}_{v \in \mathrm{conv}W_l} \|v\|.$$

Now if we have

$$f(\exp_{x_k}(\varepsilon g_l)) - f(x_k) \le -c\varepsilon\|w_l\|,\ c \in (0,1) \qquad (4.4)$$

where $g_l = -\frac{w_l}{\|w_l\|}$, then we can say $\mathrm{conv}W_l$ is an acceptable approximation for $\partial_\varepsilon f(x_k)$. Otherwise, we add a new element of $\partial_\varepsilon f(x_k) \setminus \mathrm{conv}W_l$ to W_l. Indeed, having (4.4) implies that the set $\mathrm{conv}W_l$ contains a vector w_l such that $g_l = -\frac{w_l}{\|w_l\|}$ is a good approximation of the steepest descent direction. See [24] for further details on how to algorithmically realize this approximation procedure. The following lemma proves that if W_l is not an acceptable approximation for $\partial_\varepsilon f(x)$, then there exists $v_{l+1} \in \partial_\varepsilon f(x)$ such that $\langle v_{l+1}, g_l\rangle \ge -c\|w_l\| > -\|w_l\|$, therefore $v_{l+1} \in \partial_\varepsilon f(x) \setminus \mathrm{conv}W_l$; for a proof see [24].

Lemma 4.7. *Let $W_l = \{v_1, ..., v_l\} \subset \partial_\varepsilon f(x)$, $0 \notin \mathrm{conv}W_l$ and*

$$w_l = \mathrm{argmin}\{\|v\| :\ v \in \mathrm{conv}W_l\}.$$

If we have $f(\exp_x(\varepsilon g_l)) - f(x) > -c\varepsilon\|w_l\|$, where $g_l = \frac{-w_l}{\|w_l\|}$, then there exist $\theta_0 \in (0, \varepsilon]$ and $\bar{v}_{l+1} \in \partial f(\exp_x(\theta_0 g_l))$ such that

$$\langle d\exp_x^{-1}(\exp_x(\theta_0 g_l))(\bar{v}_{l+1}), g_l\rangle \ge -c\|w_l\|,$$

and $v_{l+1} := d\exp_x^{-1}(\exp_x(\theta_0 g_l))(\bar{v}_{l+1}) \notin \text{conv}W_l$.

Algorithm 2 is used to find a vector $v_{l+1} \in \partial_\varepsilon f(x)$ which can be added to the set W_l in order to improve the approximation of $\partial_\varepsilon f(x)$. It is easy to prove by Proposition 3.2 and Proposition 3.3 of [39] that this algorithm terminates after finitely many iterations. Therefore, using an approximation of $\partial_\varepsilon f(x_k)$ we define a

Algorithm 2 An h-increasing point algorithm; $v = Increasing(x, g, a, b)$.

Input $x \in M, g \in T_xM, a, b \in \mathbb{R}$.
Let $t = b$.
repeat
  select $v \in \partial f(\exp_x(tg))$ such that $\langle v, d\exp_x(tg)(g)\rangle + c\|w\| \in \partial h(t)$
  if $\langle v, d\exp_x(tg)(g)\rangle + c\|w\| < 0$ **then**
    $t = \frac{a+b}{2}$
    if $h(b) > h(t)$ **then**
      $a = t$
    else
      $b = t$
    end if
  end if
until $\langle v, d\exp_x(tg)(g)\rangle + c\|w\| \geq 0$

function $\Phi(x_k, d) := \max\{\langle \xi, d\rangle : \ \xi \in \text{conv}W_l\}$ which approximately satisfies our assumptions and is easily computable at every $d \in T_{x_k}M$. Indeed, if we assume that $i \in \{1, ..., l\}$ is such that for a fixed $d \in T_{x_k}M$ we have $\langle v_j, d\rangle \leq \langle v_i, d\rangle$ for every $j \in \{1, ..., l\}$, then for every $\xi \in \text{conv}W_l$, we have $\xi := \sum_{s=1}^{l} \alpha_s v_s$ such that $\sum_{s=1}^{l} \alpha_s = 1$ and therefore $\langle \xi, d\rangle \leq \langle v_i, d\rangle$.

5. Numerical Experiments

The Riemannian nonsmooth trust region algorithm presented in the previous section was implemented in Matlab. To the best of our knowledge, our algorithm is the first practical nonsmooth trust region algorithm for locally Lipschitz functions defined on Riemannian manifolds. In our implementation, we update the sequence of matrices $\{B_k\}$ by using the BFGS method; see [43]. Moreover, we compare the nonsmooth trust region algorithm with Riemannian subgradient descent algorithms presented in [15, 25, 21]. The number of function evaluations is used as a measure efficiency for the algorithms. The parameters are initialized similar to the smooth version of the classical trust method. We set the parameters as $c_1 = 0.75$, $c_2 = 0$, $c_3 = 0.5$, $c_4 = 2$, $c = 10^{-4}$, $\varepsilon = 10^{-6}$, $\delta = 10^{-1}$, $\delta_1 = 10^{-1}$.

The unit sphere S^2 is the smooth compact manifold

$$S^2 = \{x \in \mathbb{R}^3 : \ \|x\| = 1\},$$

and the global coordinates on S^2 are naturally given by this embedding into $\mathbb{R}^3$. The tangent space at a point $x \in S^2$ is

$$T_xS^2 = \{v \in \mathbb{R}^3 : \ \langle x, v\rangle = 0\}.$$

The inner product on T_xS^2 is defined by

$$\langle v, w\rangle_{T_xS^2} = \langle v, w\rangle_{\mathbb{R}^3}.$$

The exponential map

$$\exp_x : T_xS^2 \to S^2$$

is defined by

$$\exp_x(v) = \cos(\|v\|)x + \sin(\|v\|)\frac{v}{\|v\|}.$$

Moreover, if $x \in S^2$, then

$$\exp_x^{-1} : S^2 \to T_xS^2$$

is defined by

$$\exp_x^{-1}(y) = \frac{\theta}{\sin(\theta)}(y - x\cos(\theta)),$$

where $\theta = \arccos\langle x, y\rangle$. The Riemannian distance between two points x, y in S^2 is given by

$$\mathrm{dist}(x, y) = \arccos\langle x, y\rangle.$$

Let $t \to \gamma(t)$ be a geodesic on S^2, and let $u = \frac{\gamma^\circ(0)}{\|\gamma^\circ(0)\|}$. The parallel translation of a vector $v \in T_{\gamma(0)}S^2$, along the geodesic γ, is given by [2]

$$L_{\gamma(0)\gamma(t)}(v) = -\gamma(0)\sin(\|\gamma^\circ(0)\|t)u'v + u\cos(\|\gamma^\circ(0)\|t)u'v + (I - uu')v.$$

5.1. **Denoising on a sphere.** First, we are going to solve the one dimensional total variation problem for functions which map into a two dimensional sphere S^2. Assuming that M is a manifold, consider the minimization problem

$$\min_{u\in BV([0,1];M)} \{F(u) := \mathrm{dist}_2(f, u)^2 + \lambda\|\nabla u\|_1\} \tag{5.1}$$

where $f : [0, 1] \to M$ is the given (noisy) function, u is a function of bounded variation from $[0, 1]$ to M, dist_2 is the distance on the function space $L^2([0, 1]; M)$ and $\lambda > 0$ is a Lagrangian parameter, [46]. Note that for every $w \in [0, 1]$, $\nabla u(w) : \mathbb{R} \to T_{u(w)}M$ and $\|\nabla u\|_1 = \int_{[0,1]} \|\nabla u(w)\|dw$. Now we can formulate a discrete version of the problem (5.1) by restricting the space of functions to V_h^M which is the space of all geodesic finite element functions for M associated with a regular grid on $[0, 1]$; see [47, 24]. We refer to [47] for the definition of geodesic finite element spaces V_h^M.

Using the nodal evaluation operator $\varepsilon : V_h^M \to M^n$, $(\varepsilon(v_h))_i = v_h(x_i)$, where x_i is the i-th vertex of the simplicial grid on $[0, 1]$, one can find an equivalent problem defined on M^n as follows,

$$\min_{u\in M^n} \{F_*(u) := \mathrm{dist}_*(\varepsilon(f), u)^2 + \lambda\|\nabla(\varepsilon^{-1}(u))\|_1\} \tag{5.2}$$

where dist_* is the Riemannian distance on M^n.

Let $\varepsilon(f) = (p_1, ..., p_n)$, then $F_* : M^n \to \mathbb{R}$ can be defined by

$$F_*(u_1, ..., u_n) = \Sigma_{i=1}^n \mathrm{dist}(p_i, u_i)^2 + \lambda\Sigma_{i=1}^{n-1}\mathrm{dist}(u_i, u_{i+1}),$$

where dist is the Riemannian distance on M.

Now we assume that $M = S^2$. First, we need to define a function from $[0, 1]$ to S^2 to get the original image. Afterward, we add a gaussian noise to the image to get the noisy image. Finally, we apply Algorithm 1 to the function F_* defined on M^{100} to get the denoised image; see Figure 1.

Table 1 provides the numerical results for the TV regularization on S^2 using the nonsmooth trust region method and ε-subdifferential method. In this table,

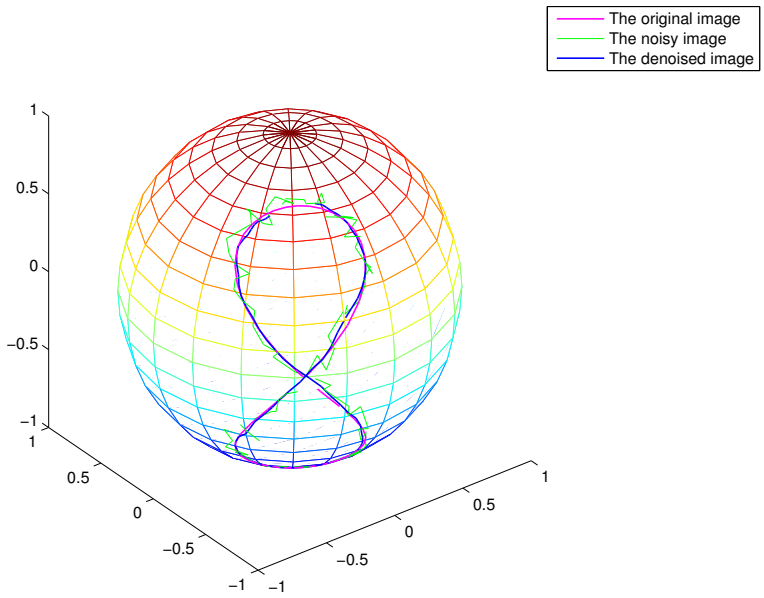

FIGURE 1. TV regularization on S^2

TABLE 1. Numerical results in terms of number of function evaluations and the final obtained value of the function for TV regularization

No.	f in trust region method	nfeval	f in ε-subdifferential method	nfeval
1	3.2266	365	3.2362	934
2	3.2512	327	3.2579	976
3	3.2902	424	3.3145	3316
4	3.2794	509	3.2896	3097
5	3.2366	365	3.2472	934
6	3.2212	327	3.2572	971
7	3.2432	424	3.3066	3216
8	3.2884	509	3.2976	1097
9	3.2187	325	3.2472	1100
10	3.2012	327	3.2972	1906
11	3.2302	424	3.3136	1001
12	3.2394	509	3.2576	1097
13	3.2234	365	3.2365	4934
14	3.2456	327	3.2542	3976
15	3.2912	424	3.3298	3316

"nfval" stands for the number of function evaluations, we also have presented the minimum value of the function. The number of function evaluations is considered as a measure of efficiency for the trust region method.

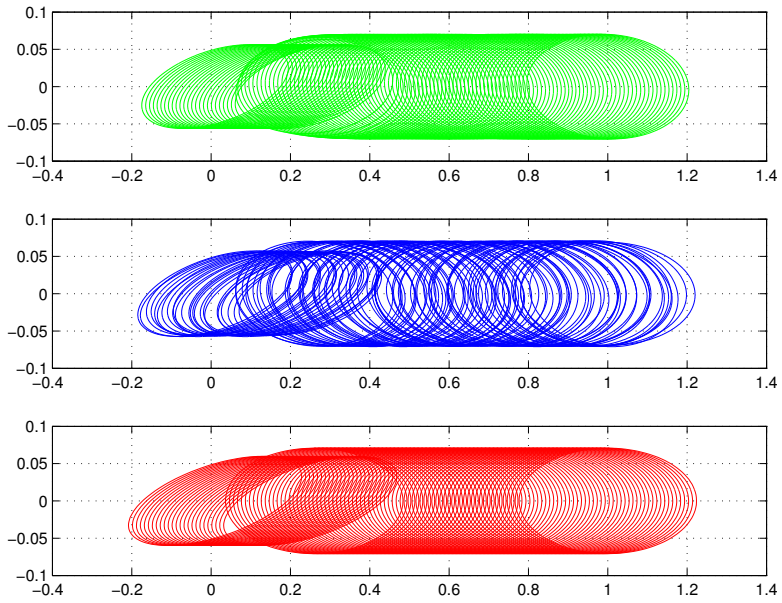

FIGURE 2. TV regularization on $P(2)$. Down-to-up: the original image, the noisy image, the denoised image

5.2. **Denoising on $P(2)$.** Data taking values in a manifold appear naturally in various signal and image processing applications. One example is diffusion tensor imaging where the data live in the Riemannian manifold of positive (definite) matrices; see [9, 40, 54]. The space of all $n \times n$ symmetric, positive definite matrices will be denoted by $P(n)$. The tangent space to $P(n)$ at any of its points P is the space $T_P P(n) = \{P\} \times S(n)$, where $S(n)$ is the space of symmetric $n \times n$ matrices. On each tangent space $T_P P(n)$, the inner product is defined by

$$\langle A, B \rangle_{T_P P(n)} = \mathrm{tr}(P^{-1} A P^{-1} B).$$

The Riemannian distance between $P, Q \in P(n)$ is given by

$$\mathrm{dist}(P, Q) = (\Sigma_{i=1}^{n} \ln^2(\lambda_i))^{(1/2)},$$

where $\lambda_i,\ i = 1, ..., n$ are eigenvalues of $P^{-1}Q$. The exponential map

$$\exp_P : S(n) \to P(n)$$

is defined by

$$\exp_P(v) = P^{1/2} \exp(P^{-1/2} v P^{-1/2}) P^{1/2}.$$

Moreover, if $P \in P(n)$, then

$$\exp_P^{-1} : P(n) \to S(n)$$

is defined by

$$\exp_P^{-1}(Q) = P^{1/2} \log(P^{-1/2} Q P^{-1/2}) P^{1/2},$$

where $\log, \exp$, denote the logarithm and exponential functions on matrix space. For another example, we assume that $M = P(2)$. We add a noise to an original image on $P(2)$. Then we apply our nonsmooth trust region algorithm to F_* on M^{100} to denoise the noisy image. In Figure 2, we present the results regarding

TABLE 2. Numerical results in terms of number of function evaluations and the final obtained value of the function for TV regularization on $P(2)$.

No.	f in trust region method	nfeval	f in ε-subdifferential method	nfeval
1	0.2340	180	0.2354	1344
2	0.1981	111	0.1917	1541
3	0.2102	102	0.2231	1681
4	0.1623	154	0.1734	894
5	0.1689	134	0.1670	800
6	0.1890	159	0.1840	1680
7	0.2301	451	0.2401	970
8	0.1891	167	0.1871	1356
9	0.1924	101	0.1981	1451
10	0.2101	186	0.2191	1189
11	0.1690	192	0.1673	891
12	0.1871	171	0.1780	3911
13	0.1988	161	0.1981	1601
14	0.2121	195	0.2191	761
15	0.2241	108	0.2240	1316

to the minimization of F_* on M^{100}. Table 2 provides the numerical results for the TV regularization on $P(2)$ using the nonsmooth trust region method and ε-subdifferential method.

5.3. **Riemannian geometric median on a sphere.** Our second numerical experiment is concerned with the Riemannian geometric median on S^2. Let M be a Riemannian manifold. Given points $p_1, ..., p_m$ in M and corresponding positive real weights $w_1, ..., w_m$, with $\sum_{i=1}^m w_i = 1$, define the weighted sum of distance functions

$$f(q) = \sum_{i=1}^{m} w_i \text{dist}(p_i, q),$$

where dist is the Riemannian distance function on M. We define the weighted geometric median x as the minimizer of f. When all the weights are equal, $w_i = 1/m$, we call x simply the geometric median. Now, we assume that $M = S^2$.

Table 3 provides the numerical results for finding the geometric median on S^2 with $m = 5000$, using the nonsmooth trust region method and the ε-subdifferential method. The starting point and the points p_i are chosen randomly. As before, "nfval" stands for the number of function evaluations; we also have presented the minimum value of the function. The number of function evaluations is considered as a measure of efficiency for the trust region method.

5.4. **Rayleigh quotients on a sphere.** Now we are going to compare our nonsmooth trust region algorithm with the Riemannian subgradient descent presented in [21]. To this end, we consider the maximum of m Rayleigh quotients on the sphere S^{n-1}, i.e.,

$$f(x) = \max_{i=1,...,m} \frac{1}{2} x' A_i x, \tag{5.3}$$

TABLE 3. Numerical results in terms of number of function evaluations and the final obtained value of the function for the geometric median on S^2

No.	f in trust region method	nfeval	f in ε-subdifferential method	nfeval
1	0.4306	8	0.4365	192
2	0.4302	11	0.4305	67
3	0.4101	2	0.4200	108
4	0.4338	13	0.4390	110
5	0.4206	13	0.4273	93
6	0.4312	5	0.4326	97
7	0.4302	11	0.4321	105
8	0.4394	11	0.4398	109
9	0.4316	11	0.4321	98
10	0.4212	8	0.4237	97
11	0.4302	9	0.4388	128
12	0.4334	10	0.4399	167
13	0.4266	6	0.4280	211
14	0.4292	5	0.4300	145
15	0.4372	4	0.4384	316

where $A_i \in \mathbb{R}^{n\times n}$ is symmetric. Our aim is to find a minimum of f. Table 4 provides the numerical results for finding the minimum of f on S^2 with $m = 20$, using the nonsmooth trust region method and the subgradient descent method.

Moreover, we say that an algorithm solves a problem successfully if the following condition is satisfied:

$$\frac{|f_{opt} - f*|}{|f_{opt} + 1|} \leq \epsilon,$$

where f_{opt} is the minimum value of the function and $f*$ is the minimum value obtained by the algorithm, and $\epsilon = 10^{-4}$ is an accuracy level. Figure 3 shows that the nonsmooth trust region algorithm solve the problem.

5.5. **Sphere packings on Grassmannians.** The sphere packings on Grassmannians have many applications in wireless communication and statistics and seem to be good candidates for use in quantum information theory; see for example [6, 14]. We assume that the Grassmannian $\mathrm{Gr}(n, k)$ is the set of all k-dimensional linear subspaces of $\mathbb{R}^n$. In this section, we consider the problem of the packing of m spherical balls on $\mathrm{Gr}(n, k)$ with respect to the chordal distance. Let $B(P, r)$ denote the ball in $\mathrm{Gr}(n, k)$ with respect to chordal distance. Then we would like to find m points $P_1, ..., P_m$ in $\mathrm{Gr}(n, k)$ such that

$$\max\{r|\ \forall i \neq j :\ B(P_i, r) \cap B(P_j, r) = \emptyset\}, \tag{5.4}$$

is maximized. This problem has been solved in [21] using a subgradient method.

Indeed, $\mathrm{Gr}(n, k)$ can be identified with the set $\{P \in S(n)|\ P^2 = P, \mathrm{tr}(P) = k\}$; see [26]. Moreover, the tangent space of the Grassmannian at the point P, denoted by $T_P\mathrm{Gr}(n, k)$, is the following set

$$T_P\mathrm{Grass}(n, k) = \{P\Omega - \Omega P|\ \Omega \in so(n)\},$$

TABLE 4. Numerical results in terms of number of function evaluations and the final obtained value of the function for Rayleigh quotients on S^2

No.	f in trust region method	nfeval	f in subdifferential method	nfeval
1	0.6529	128	1.7832	501
2	0.3294	116	0.3341	3614
3	0.7321	116	0.7996	5501
4	0.5211	123	0.5458	2123
5	0.5156	117	0.4962	509
6	0.6321	156	0.6578	1504
7	0.4321	167	0.4567	1350
8	0.6421	127	0.6432	680
9	0.7691	117	0.7761	1384
10	0.4021	116	0.4187	890
11	0.6531	170	0.6520	1230
12	0.5090	175	0.5021	2012
13	0.5401	149	0.5721	1890
14	0.3489	110	0.3901	2056
15	0.3209	113	0.3009	1105

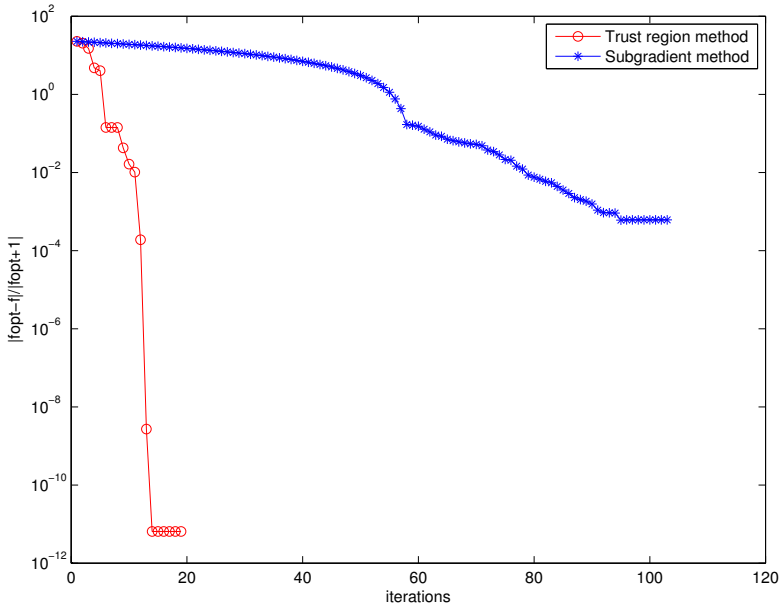

FIGURE 3. Rayleigh qoutients on S^2. We compare the trust region method introduced in this paper with the Riemannian subgradient algorithm from [21].

TABLE 5. Numerical results in terms of number of function evaluations and the final obtained value of the function for sphere packings on Grassmannians

No.	Minimal distance in TR	nfeval	Minimal distance in SB	nfeval
1	1.851	167	1.706	601
2	1.689	104	1.674	231
3	1.795	115	1.779	1109
4	1.862	146	1.796	1123
5	1.801	110	1.892	234
6	1.813	156	1.711	1609
7	1.862	145	1.791	1209
8	1.874	132	1.776	890
9	1.719	112	1.721	1409
10	1.789	123	1.799	980
11	1.708	145	1.769	430
12	1.804	154	1.704	1230
13	1.897	179	1.791	1679
14	1.676	134	1.659	2340
15	1.631	104	1.611	1306

where

$$so(n) = \{\Omega \in \mathbb{R}^{n\times n} |\ \Omega' = -\Omega\}.$$

As $\mathrm{Gr}(n,k)$ is a subset of the Euclidean vector space $S(n)$, the scalar product $\langle P, Q\rangle := \mathrm{tr}(PQ)$ induces a Riemannian metric on it. Therefore the chordal distance on $\mathrm{Gr}(n,k)$, denoted by $\mathrm{dist}(P,Q)$, is defined by

$$\mathrm{dist}(P,Q) = \sqrt{\frac{1}{2}}\|P-Q\|_F,$$

where $\|.\|_F$ denotes the Frobenius norm. On $\mathrm{Gr}(n,k)$ with the induced Riemannian metric, the geodesic γ emanating from P in the direction $\eta \in T_P\mathrm{Grass}(n,k)$ is defined by

$$\gamma(t) = \exp(t(\eta P - P\eta))P\exp(-t(\eta P - P\eta)).$$

The problem (5.4) is equivalent to the minimizing the following nonsmooth function;

$$F(P_1,...,P_m) := \max_{i\neq j} \mathrm{tr}(P_iP_j), \tag{5.5}$$

on $\mathrm{Gr}(n,k)\times ...\times \mathrm{Gr}(n,k)$; see [31].

In Table 5, we illustrate the results of the nonsmooth subgradient (SB) method and nonsmooth trust region method (TR) for the sphere packing in $\mathrm{Gr}(16,2)$ with $m = 10$ and the same arbitrary starting points for both methods.

6. Conclusions

We have presented a practical algorithm in the context of trust region methods for nonsmooth problems on Riemannian manifolds. To the best of our knowledge, this is the first paper on nonsmooth trust region method on Riemannian manifolds. We also introduce a practical local model in our trust region scheme for locally Lipschitz functions. We have seen that the use of exponential map yields trust region

subproblems expressed in Euclidean spaces T_xM. Therefore, all the classical methods for solving the trust region subproblem can be applied. In our implementation, we use the approach based on the Cauchy point and the CG-Steihaug methods; see [3]. The main result is the global convergence property of our trust region method which is stated in Theorem 3.6.

An implementation of our proposed trust region algorithm, along with the subgradient and ε-subgradient methods, is given in Matlab environment and tested on some problems. Numerical results of the considered algorithms show that comparing with the ε-subgradient algorithm, the nonsmooth trust region algorithm has a better performance in terms of the number of function evaluations. Moreover, comparing with the subgradient algorithm, the nonsmooth trust region method gives us a better approximation of the minimum value of the function for some examples.

Acknowledgments

The authors wish to thank R. Yousefpour and C. Lageman for useful discussions and providing us their implementations of subgradient methods.

References

[1] P. A. Absil, C. G. Baker, *Trust-region methods on Riemannian manifolds*, Found. Comput. Math., 7 (2007), pp. 303-330.

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

[3] Z. Akbari, R. Yousefpour, M. Reza Peyghami, *A new nonsmooth trust region algorithm for locally Lipschitz unconstrained optimization problems*, To appear in J. Optim. Theory Appl.

[4] 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), pp. 359-390.

[5] B. Afsari, R. Tron, R. Vidal, *On the convergence of gradient descent for finding the Riemannian center of mass,* SIAM J. Control Optim., 51 (2013), pp. 2230-2260.

[6] A. E. Ashikhmin, A. R. Calderbank, *Grassmannian packings from operator Reed-Muller codes*, IEEE Transactions on Information Theory, 56(11), pp. 5689-5714 (2010)

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

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

[9] P. J. Basser, J. Mattiello, D. L. Bihan, *MR diffusion tensor spectroscopy and imaging*, Biophys. J., 66(1) (1994), pp. 259-267.

[10] 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), pp. 564-572.

[11] 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), pp. 88-107.

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

[13] G. C. Bento, O. P. Ferreira, P. R. Oliveira, *Proximal point method for a special class of nonconvex functions on Hadamard manifolds,* Optimization (2012), pp. 1-31.

[14] B. Ozbek, D. L. Ruyet, *Feedback Strategies for Wireless Communication*, Springer, 2013.

[15] P. B. Borckmans, S. Easter Selvan, N. Boumal, P. A. Absil, *A Riemannian subgradient algorithm for economic dispatch with valve-point effect,* J. Comput. Appl. Math., 255 (2014), pp. 848-866.

[16] R. D. Canary, D. B. A. Epstein, A. Marden, *Fundamentals of Hyperbolic Geometry: Selected Expositions, Cambridge University Press,* Cambridge (2006).

[17] A. R. Conn, N. I. M. Gould, Ph. L. Toint, *Trust Region Methods,* SIAM Series on Optimization, SIAM, Philadelphia, 2000.

[18] J. E. Dennis, S. B. Li, R. A. Tapia, *A unified approach to global convergence of trust region methods for nonsmooth optimization*, Math. Program., 68 (1995), pp. 319-346.
[19] J. X. da Cruz Neto, L. L. Lima, P. R. Oliveira, *Geodesic Algorithm in Riemannian Manifolds,* Balkan J. Geom. Appl., 2 (1998), pp. 89 - 100.
[20] J.X. da Cruz Neto, O. P. Ferreira and L. R. Lucambio Perez, *A proximal regularization of the steepest descent method in Riemannian manifold*, Balkan J. Geom. Appl., 2 (1999), pp. 1-8.
[21] G. Dirr, U. Helmke, C. Lageman, *Nonsmooth Riemannian optimization with applications to sphere packing and grasping*, In Lagrangian and Hamiltonian Methods for Nonlinear Control 2006: Proceedings from the 3rd IFAC Workshop, Nagoya, Japan, 2006, Lecture Notes in Control and Information Sciences, Vol. 366, Springer Verlag, Berlin, 2007.
[22] O. P. Ferreira, P. R. Oliveira, *Proximal point algorithm on Riemannian manifolds,* Optimization. 51 (2002), pp. 257-270.
[23] O. P. Ferreira, P. R. Oliveira, *Subgradient algorithm on Riemannian manifolds,* J. Optim. Theory Appl., 97 (1998), pp. 93-104.
[24] P. Grohs, H. Hardering, O. Sander, *Optimal a priori discretization error bounds for geodesic finite elements,* SAM Report 2013-16, ETH Zürich, (2013), Submitted.
[25] P.Grohs, S. Hosseini, *ε-subgradient algorithms for locally Lipschitz functions on Riemannian manifolds*, Submitted.
[26] U. Helmke, J. Moore, *Optimization and Dynamical System*, Springer-Publ., London, 1994.
[27] S. Hosseini, M. R. Pouryayevali, *Generalized gradients and characterization of epi-Lipschitz sets in Riemannian manifolds*, Nonlinear Anal., 74 (2011), pp. 3884-3895.
[28] S. Hosseini, M. R. Pouryayevali, *Euler characterization of epi-Lipschitz subsets of Riemannian manifolds,* J. Convex. Anal., 20 (2013), No. 1, pp. 67-91.
[29] 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.
[30] W. Klingenberg, *Riemannian Geometry,* Walter de Gruyter Studies in Mathematics, Vol. 1, Walter de Gruyter, Berlin, New York, 1995.
[31] C. Lageman,U. Helmke, *Optimization Methods for the Sphere Packing Problem on Grassmannians*, Taming Heterogeneity and Complexity of Embedded Control, (2013), pp. 393-408.
[32] D. Kressner, M. Steinlechner, B. Vandereycken,*Low-rank tensor completion by Riemannian optimization*, BIT Numer. Math., 54 (2014), pp. 447-468.
[33] S. Lang, *Fundamentals of Differential Geometry,* Graduate Texts in Mathematics, Vol. 191, Springer, New York, 1999.
[34] P. Y. Lee, *Geometric Optimization for Computer Vision,* PhD thesis, Australian National University, 2005.
[35] C. Lemarechal, *Nondifferentiable optimization,* In: Handbook in Operations Research and Management Science (G. L. Nemhauser et al., Eds.), Vol. 1, North Holland, Amsterdam (1989), pp. 529-572.
[36] M. M. Makela, P. Neittaanmaki, *Nonsmooth Optimization*, World Scientific (1992).
[37] C. Li, B. S. Mordukhovich, J. Wang, J. C. Yao, *Weak sharp minima on Riemannian manifolds,* SIAM J. Optim., 21(4) (2011), pp. 1523-1560.
[38] R. E. Mahony, *The constrained Newton method on a Lie group and the symmetric eigenvalue problem,* Linear Algebra. Appl., 248 (1996), pp. 67-89.
[39] N. Mahdavi-Amiri, R. Yousefpour, *An effective nonsmooth optimization algorithm for locally Lipschitz functions*, J. Optim. Theory Appl., 155 (2012), pp. 180-195.
[40] X. Pennec, P. Fillard, N. Ayache, *A Riemannian framework for tensor computing,* Int. J. Comput. Vis., 66(1) (2006), pp. 41-66.
[41] E. A. Papa Quiroz, E. M. Quispe, P. R. Oliveira, *Steepest descent method with a generalized Armijo search for quasiconvex functions on Riemannian manifolds,* J. Math. Anal. Appl., 341 (2008), pp. 467-477.
[42] L. Qi, J. Sun, *A trust region algorithm for minimization of locally Lipschitzian functions*, Math. Program., 66 (1994), pp. 25-43.
[43] C. Qi, K. A. Gallivan, P. A. Absil, *Riemannian BFGS algorithm with applications*, Recent Advances in Optimization and its Applications in Engineering, (2010), pp. 183-192.
[44] W. Ring, B. Wirth, *Optimization methods on Riemannian manifolds and their application to shape space*, SIAM J. Optim., 22(2) (2012), pp. 596-627.

[45] R. C. Riddell, *Minimax problems on Grassmann manifolds. Sums of eigenvalues,* Adv. Math., 54 (1984), pp. 107-199.
[46] L. Rudin, S. Osher, E. Fatemi, *Nonlinear total variation based noise removal algorithms,* Phys. D., 60 (1) (1992), pp. 259-268.
[47] O. Sander, *Geodesic finite elements for Cosserat rods,* Internat. J. Numer. Methods Engrg., 82 (2010), pp. 1645-1670.
[48] M. Sarkis, K. Diepold, *Camera-pose estimation via projective Newton optimization on the manifold* , IEEE Trans. Image Processing., 21 (2012), pp. 1729-1741.
[49] S. T. Smith, *Optimization techniques on Riemannian manifolds*, Fields Institute Communications, 3 (1994), pp. 113-146.
[50] C. Udriste, *Convex Functions and Optimization Methods on Riemannian Manifolds*, Kluwer Academic Publishers, Dordrecht, Netherlands, 1994.
[51] K. Usevich, I. Markovsky,*Optimization on a Grassmann manifold with application to system identification,* Submitted, (http://homepages.vub.ac.be/ imarkovs/t-abstracts.html)
[52] B. Vandereycken, S. Vandewalle, *A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations*, SIAM. J. Matrix Anal. Appl., 31(5) (2010), pp. 2553-2579.
[53] B. Vandereycken, *Low-rank matrix completion by Riemannian optimization,* SIAM J. Optim., 23(2) (2013), pp. 1214-1236.
[54] A. Weinmann, L. Demaret, M. Storath, *Total variation regularization for manifold-valued data,* SIAM J. Imaging Sci., 7(4) (2014), pp. 2226-2257.
[55] L. H. Zhang, *Riemannian Newton method for the multivariate eigenvalue problem,* SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2972-2996.