ETH

Eidgenössische Technische Hochschule Zürich

Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

ε-Subgradient Algorithms for Locally Lipschitz Functions on Riemannian Manifolds

P. Grohs and S. Hosseini

Research Report No. 2013-49
December 2013

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Funding SNF: 140635

ε-SUBGRADIENT ALGORITHMS FOR LOCALLY LIPSCHITZ FUNCTIONS ON RIEMANNIAN MANIFOLDS

P. GROHS*, S. HOSSEINI*

Abstract. This paper presents a descent direction method for finding extrema of locally Lipschitz functions defined on Riemannian manifolds. To this end we define a set-valued mapping $x \to \partial_\varepsilon f(x)$ named ε-subdifferential which is an approximation for the Clarke subdifferential and which generalizes the Goldstein-ε-subdifferential to the Riemannian setting. Using this notion we construct a steepest descent method where the descent directions are computed by a computable approximation of the ε-subdifferential. We establish the convergence of our algorithm to a stationary point. Numerical experiments illustrate our results.

1. Introduction

This paper is concerned with the numerical solution of optimization problems defined on Riemannian manifolds where the objective function may be nonsmooth. Such problems arise in a variety of applications, e.g., in computer vision, signal processing, motion and structure estimation, or numerical linear algebra, see for instance [2, 3, 26, 35].

In the linear case is well known that ordinary gradient descent, when applied to nonsmooth functions, typically fails by converging to a non-optimal point. The fundamental difficulty is that most interesting nonsmooth objective functions assume their extrema at points where the gradient is not defined. This has led to the introduction of the generalized gradient of convex functions defined on a linear space by Rockafellar in 1961 and subsequently for locally Lipschitz functions by Clarke in 1975; [36, 13]. Their use in optimization algorithms began soon after their appearance. Since the Clarke generalized gradient is in general difficult to compute numerically, most of algorithms which are based on it can be efficient only for certain types of functions; see for instance [6, 23, 42].

The paper [19] is among the first works on optimization of Lipschitz functions on Euclidean spaces. In that article a new set valued mapping named $\varepsilon-$subdifferential $\partial_\varepsilon f$ of a function f was introduced, and several properties of this map, which are useful for building optimization algorithms of locally Lipschitz functions on linear spaces, were presented. For the numerical computation of the $\varepsilon-$subdifferential various strategies have been proposed in the literature.

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

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

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

The gradient sampling algorithm (GS), introduced and analyzed by Burke, Lewis, and Overton [11, 12], is a method for minimizing an objective function f that is locally Lipschitz and continuously differentiable in an open dense subset of $\mathbb{R}^n$. At each iteration, the GS algorithm computes the gradient of f at the current iterate and at $m \geq n+1$ randomly generated nearby points. This bundle of gradients is used to find an approximate ε-steepest descent direction as the solution of a quadratic program, where ε denotes the sampling radius. A standard Armijo line search along this direction produces a candidate for the next iterate, which is obtained by perturbing the candidate, if necessary, to stay in the set Ω where f is differentiable; the perturbation is random and small enough to maintain the Armijo sufficient descent property. The sampling radius may be fixed for all iterations or may be reduced dynamically.

The discrete gradient method (DG) approximates $\partial_\varepsilon f(x)$ by a set of discrete gradients. In this algorithm, the descent direction is iteratively computed, and in every iteration the approximation of $\partial_\varepsilon f(x)$ is improved by adding a discrete gradient to the set of discrete gradients, see [6].

In [30], $\partial_\varepsilon f(x)$ is approximated by an iterative algorithm. The algorithm starts with one element of $\partial_\varepsilon f(x)$ in the first iteration, and in every subsequent iteration, a new element of $\partial_\varepsilon f(x)$ is computed and added to the working set to improve the approximation of $\partial_\varepsilon f(x)$. The results of the algorithm presented in [30] as compared to those obtained by the GS is more efficient, and as compared to those by the DG is more accurate, [30].

The extension of the aforementioned optimization techniques 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.

Contributions. Our main contributions are twofold. First, we define a Riemannian generalization of the ε-subdifferential defined in [19]. This is nontrivial since the linear definition of $\partial_\varepsilon f(x)$, $x \in \mathbb{R}^n$ involves subgradients of f at points $y \in \mathbb{R}^n$ different from x. In the linear case this is not an issue since tangent spaces at different points can be identified. In the nonlinear case, with M being a Riemannian manifold and $f : M \to \mathbb{R}$, we move these subgradients at points $y \in M$ to the tangent space in x via the derivative of the logarithm mapping in order to obtain a workable definition of the ε-subdifferential, see Definition 3.1 below. In Section 3.1 we prove several basic properties of the novel Riemannian ε-subdifferential which subsequently enables us to formulate conditions for descent directions in Section 3.2. Using these basic properties of the ε-subdifferential, we are able to generalize (GS) and the algorithm in [30] to the Riemannian setting. In Section 3.3 we present the details for the generalization of [30] which yields the second main contribution of the present paper, namely a proof of convergence of the proposed algorithm. Finally, in Section 4 we present two numerical results in nonlinear signal processing.

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 [1, 3, 14, 29, 34, 39, 40, 43]. The recent monograph 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, [2].

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, [4, 5, 20, 21, 22, 28].

In [7, 10, 16] constrained minimization problems on Hadamard manifolds are solved using a generalization of the proximal point method. Finally we mention work by Ferreira, Bento and Oliveira and their colleagues who generalized subgradient-type methods for convex and quasiconvex functions defined on Riemannian manifolds, see [8, 9, 15, 17, 33].

2. PRELIMINARIES

In this paper, we use the standard notations and known results of Riemannian manifolds, see, e.g. [25]. Throughout this paper, M is an n-dimensional complete manifold endowed with a Riemannian metric $\langle .,. \rangle_x$ on the tangent space T_xM. As usual we denote by $B(x,\delta)$ the open ball centered at x with radius δ, by $\text{int}N(\text{cl}N)$ the interior (closure) of the set N. Also, let S be a nonempty closed subset of a Riemannian manifold M, we define $\text{dist}_S : M \longrightarrow \mathbb{R}$ by

$$\text{dist}_S(x) := \inf\{\text{dist}(x,s) : s \in S \},$$

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

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

By $i_M(x)$ we denote the injectivity radius of M at x, that is the suprimum 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 $B(x, r)$. Note that 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 [24].

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

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

Let us continue with the definition of the Clarke generalized directional derivative for locally Lipschitz functions on Riemannian manifolds; see [20, 22].

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

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

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

Definition 2.3 (Subdifferential)**.** *We define the subdifferential of f, denoted by $\partial f(x)$, as the subset of T_xM whose support function is $f^\circ(x; .)$. It can be proved* [20] *that*

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

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

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

Using the notion of subdifferential we can now define stationary points of a locally Lipschitz mapping f.

Definition 2.4 (Stationary Point, Stationary Set)**.** *A point x is a stationary point of f if $0 \in \partial f(x)$. Z is a stationary set if each $z \in Z$ is a stationary point.*

Proposition 2.5. *A necessary condition that f achieve a local minimum at x is that $0 \in \partial f(x)$.*

Proof. If f has a local minimum at x, then for every $v \in T_xM$, $f^\circ(x, v) \geq 0$ which implies $0 \in \partial f(x)$. □

3. The Riemannian ε-Subdifferential

In smooth optimization, there exist minimization methods, which, instead of using the gradient, use its approximations through finite differences (forward, backward, and central differences). In [27], a very simple convex nondifferentiable function was presented, for which these finite differences may give no information about the subdifferential. It follows that these finite-difference estimates of the gradient cannot be used for the approximation of the subgradient of the nonsmooth functions. In [19] a set valued mapping named ε-subdifferential, to approximate the subdifferential of locally Lipschitz functions defined on $\mathbb{R}^n$ was introduced.

The present section generalizes the concept of the ε-subdifferential of locally Lipschitz functions to functions defined on a Riemannian manifold, generalizing the corresponding Euclidean concept introduced in [19]. The definition is as follows.

Definition 3.1 (ε-subdifferential)**.** *Let $f : M \to \mathbb{R}$ be a locally Lipschitz function on a Riemannian manifold M, and θ_k be any sequence of positive numbers converging downward to zero. For each $\varepsilon > 0$ with $\varepsilon + \theta_k < i_M(x)$ for every k, the $\varepsilon-$subdifferential is defined by*

$$\partial_\varepsilon f(x) := \operatorname{conv} \bigcap_{\mathrm{k=1}}^{\infty} \mathrm{cl}\{\mathrm{d}\exp_{\mathrm{x}}^{-1}(\mathrm{y})(\mathrm{grad}\,\mathrm{f}(\mathrm{y})) :\ \mathrm{y} \in \mathrm{clB}(\mathrm{x}, \varepsilon + \theta_{\mathrm{k}}) \cap \Omega_{\mathrm{f}}\}.$$

Clearly this definition is independent of the choice of the sequence θ_k.

3.1. Basic Properties.

In the present subsection we establish some basic properties of the ε-subdifferential as defined above in Definition 3.1, see [19] for similar results in the linear case. We select ε small enough that f is Lipschitz on $B(x, 2\varepsilon)$ and $\exp_x$ is a diffeomorphism from $B(0_x, 2\varepsilon)$ onto $B(x, 2\varepsilon)$.

Lemma 3.2. *For every $y \in B(x, \varepsilon)$,*

$$d\exp_x^{-1}(y)(\partial f(y)) \subset \partial_\varepsilon f(x).$$

Proof. For every $\xi = \lim_{i\to\infty} \operatorname{grad} f(y_i)$ where $\operatorname{grad} f(y_i)$ exists and $y_i \to y$, we have

$$d\exp_x^{-1}(y)(\xi) = \lim_{i\to\infty} d\exp_x^{-1}(y_i)(\operatorname{grad} f(y_i)),$$

so there exists $N \in \mathbb{N}$, such that for every $i \geq N$, $y_i \in B(x, \varepsilon)$, and

$$\lim_{i\to\infty} d\exp_x^{-1}(y_i)(\operatorname{grad} f(y_i)) \in \bigcap_{k=1}^{\infty} \mathrm{cl}\{d\exp_x^{-1}(y)(\operatorname{grad} f(y)) :\ y \in \mathrm{cl}B(x, \varepsilon+\theta_k)\cap\Omega_f\}$$

which means

$$d\exp_x^{-1}(y)(\{\lim_{i\to\infty} \operatorname{grad} f(y_i) :\ \{y_i\} \subseteq \Omega_f,\ y_i \to y\})$$

is a subset of $\bigcap_{k=1}^{\infty} \mathrm{cl}\{d\exp_x^{-1}(y)(\operatorname{grad} f(y)) :\ y \in \mathrm{cl}B(x, \varepsilon + \theta_k) \cap \Omega_f\}$, which implies

$$d\exp_x^{-1}(y)(\partial f(y)) \subset \partial_\varepsilon f(x).$$

□

Lemma 3.3. *$\partial_\varepsilon f(x)$ is a nonempty compact and convex subset of T_xM.*

Proof. By the Lipschitzness of f and the smoothness of the exponential map,

$$S_k = \mathrm{cl}\{d\exp_x^{-1}(y)(\mathrm{grad}\, f(y)) :\ y \in \mathrm{cl}B(x, \varepsilon + \theta_k) \cap \Omega_f\}$$

is a closed bounded subset of T_xM and $S_{k+1} \subset S_k$. Hence $\bigcap_{k=1}^{\infty} S_k$ is compact and nonempty, and convex hull of a compact set in T_xM is compact. The convexity of $\partial_\varepsilon f(x)$ is deduced by the definition. □

Lemma 3.4.

$$\partial_\varepsilon f(x) = \mathrm{conv}\{\lim_{\mathrm{i}\to\infty} \mathrm{d}\exp_{\mathrm{x}}^{-1}(\mathrm{y_i})(\mathrm{grad}\, \mathrm{f}(\mathrm{y_i})) : \lim_{\mathrm{i}\to\infty} \mathrm{y_i} = \mathrm{y} \in \mathrm{clB}(\mathrm{x}, \varepsilon),\ (\mathrm{y_i}) \in \Omega_{\mathrm{f}}\}.$$

Proof. We start with the inclusion

$$\partial_\varepsilon f(x) \supset \mathrm{conv}\{\lim_{\mathrm{i}\to\infty} \mathrm{d}\exp_{\mathrm{x}}^{-1}(\mathrm{y_i})(\mathrm{grad}\, \mathrm{f}(\mathrm{y_i})) : \lim_{\mathrm{i}\to\infty} \mathrm{y_i} = \mathrm{y} \in \mathrm{clB}(\mathrm{x}, \varepsilon),\ (\mathrm{y_i}) \in \Omega_{\mathrm{f}}\}.$$

Let y_i be a sequence in Ω_f converging to some point $y \in \mathrm{cl}B(x, \varepsilon)$ and $v_i = d\exp_x^{-1}(y_i)(\mathrm{grad}\, f(y_i))$. If y is an interior point of $B(x, \varepsilon)$, then

$$\lim_{i\to\infty} v_i \in \bigcap_{k=1}^{\infty} \mathrm{cl}\{d\exp_x^{-1}(y)(\mathrm{grad}\, f(y)) :\ y \in \mathrm{cl}B(x, \varepsilon + \theta_k) \cap \Omega_f\},$$

which implies

$$\mathrm{conv}\{\lim_{\mathrm{i}\to\infty} \mathrm{d}\exp_{\mathrm{x}}^{-1}(\mathrm{y_i})(\mathrm{grad}\, \mathrm{f}(\mathrm{y_i})) : \lim_{\mathrm{i}\to\infty} \mathrm{y_i} = \mathrm{y} \in \mathrm{clB}(\mathrm{x}, \varepsilon),\ (\mathrm{y_i}) \in \Omega_{\mathrm{f}}\} \subset \partial_\varepsilon \mathrm{f}(\mathrm{x}).$$

Assume that there exists a subsequence y_{i_k} of y_i, such that $\mathrm{dist}(y_{i_k}, x)$ decreases to ε and

$$v_{i_k} \in \mathrm{cl}\{d\exp_x^{-1}(y)(\mathrm{grad}\, f(y)) :\ y \in \mathrm{cl}B(x, \mathrm{dist}(x, y_{i_k})) \cap \Omega_f\} = S_k,$$

since S_k is compact and nested in T_xM we obtain

$$\lim_{i\to\infty} v_i \in \bigcap_{k=1}^{\infty} S_k \subset \partial_\varepsilon f(x),$$

which proves the first inclusion.

For the converse, let

$$w \in \bigcap_{k=1}^{\infty} \mathrm{cl}\{d\exp_x^{-1}(y)(\mathrm{grad}\, f(y)) :\ y \in \mathrm{cl}B(x, \varepsilon + \theta_k) \cap \Omega_f\}.$$

Then for every $k \in \mathbb{N}$ we have

$$w \in \mathrm{cl}\{d\exp_x^{-1}(y)(\mathrm{grad}\, f(y)) :\ y \in \mathrm{cl}B(x, \varepsilon + \theta_k) \cap \Omega_f\},$$

Therefore we can find a sequence $y_i \in \mathrm{cl}B(x, \varepsilon + \theta_i) \cap \Omega_f$ such that

$$\lim_{i\to\infty} \|d\exp_x^{-1}(y_i)(\mathrm{grad}\, f(y_i)) - w\| = 0$$

and (if necessary after passing to a subsequence),

$$\lim_{i\to\infty} y_i = y \in \mathrm{cl}B(x, \varepsilon),$$

as required. □

Using the previous lemma one can prove the following characterization of the Riemannian ε-subdifferential.

Lemma 3.5. *We have*

$$\partial_\varepsilon f(x) = \mathrm{conv}\{\mathrm{d}\exp_{\mathrm{x}}^{-1}(\mathrm{y})(\partial \mathrm{f}(\mathrm{y})) : \mathrm{y} \in \mathrm{clB}(\mathrm{x}, \varepsilon)\}.$$

Proof. Assume that $\eta \in \partial_\varepsilon f(x)$, Lemma 3.4 implies $\eta = \Sigma_{k=1}^n t_k \xi_k$, where

$$\xi_k = \lim_{i\to\infty} d\exp_x^{-1}(y_i)(\operatorname{grad} f(y_i)),$$

$y_i \in \Omega_f$, $\lim_{i\to\infty} y_i = y \in \mathrm{cl}B(x,\varepsilon)$. Hence

$$\xi_k = d\exp_x^{-1}(y)(\lim_{i\to\infty} \operatorname{grad} f(y_i)).$$

Set $\eta_k = \left(d\exp_x^{-1}(y)\right)^{-1}(\xi_k)$ in $\partial f(y)$, then

$$\eta = \Sigma_{k=1}^n t_k d\exp_x^{-1}(y)(\eta_k) \in \mathrm{conv}\{\mathrm{d}\exp_{\mathrm{x}}^{-1}(\mathrm{y})(\partial \mathrm{f}(\mathrm{y})) : \mathrm{y} \in \mathrm{clB}(\mathrm{x},\varepsilon)\}.$$

For the converse, let

$$A = \{d\exp_x^{-1}(y)(\partial f(y)) : y \in \mathrm{cl}B(x,\varepsilon)\}$$

and $\xi \in A$, then $\xi = d\exp_x^{-1}(y)(\eta)$, where $\eta = \lim_{i\to\infty} \operatorname{grad} f(y_i)$, $y_i \in \Omega_f$, $\lim_{i\to\infty} y_i = y$. Hence

$$\xi = \lim_{i\to\infty} d\exp_x^{-1}(y_i)(\operatorname{grad} f(y_i))$$

which implies $A \subset \partial_\varepsilon f(x)$, and the property of convex hull completes the proof. □

Remark 3.6. *Note that for small enough $\varepsilon > 0$, $\partial f(x) \subset \partial_\varepsilon f(x)$. If $\varepsilon_1 > \varepsilon_2$, then $\partial_{\varepsilon_2} f(x) \subset \partial_{\varepsilon_1} f(x)$. Therefore, $\lim_{\varepsilon_k \downarrow 0} \partial_{\varepsilon_k} f(x) = \bigcap_{\varepsilon_k} \partial_{\varepsilon_k} f(x) = \partial f(x)$.*

We recall that a set valued function $F : X \rightrightarrows Y$, where X, Y are topological spaces, is said to be upper semicontinuous at x, if for every open neighborhood U of $F(x)$, there exits an open neighborhood V of x, such that

$$y \in V \Longrightarrow F(y) \subseteq U.$$

Assume that F has compact values, then there is a sequential characterization for the set valued upper semicontinuity as follows: F is upper semicontinuous at x, if and only if for each sequence $\{x_n\} \subset X$ converging to x and each sequence $\{y_n\} \subset F(x_n)$ converging to y; $y \in F(x)$.

The following remark is required in the sequel.

Remark 3.7. Let M be a Riemannian manifold.
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

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

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 $L_{x_nx_0}(\xi_n) \to L_{x_0x_0}(\xi_0) = \xi_0$, for every $(x_0,\xi_0) \in TM$; see [4, Remark 6.11].

Lemma 3.8. *Let U be a compact subset of M and $\varepsilon < i(U)$; then for every open neighborhood Win U, the set valued mapping $\partial_\varepsilon f : W \rightrightarrows TM$ is upper semicontinuous.*

Proof. For every arbitrary fixed $x \in W$, let r be a positive number with $r < \varepsilon$. We define $F : B(x,r) \cap W \rightrightarrows T_xM$ by

$$F(z) = L_{zx}(\{d\exp_z^{-1}(y)(\partial f(y)) :\ y \in \mathrm{cl}B(z,\varepsilon)\}).$$

First, we prove F is upper semicontinuous at x.

Let $\{x_k\} \subset B(x,r) \cap W$ and $\{v_k\} \subset T_xM$ be two sequences converging, respectively, to x and v, where $v_k \in F(x_k)$. So $v_k = L_{x_kx}(d\exp_{x_k}^{-1}(y_k)(\xi_k))$, where $\xi_k \in \partial f(y_k)$, $y_k \in \mathrm{cl}B(x_k,\varepsilon)$.

Note that M is complete, so $\{y_k\}$ has a subsequence convergent to some point y in M. Moreover, f is Lipschitz on $B(x,\varepsilon)$, so by Theorem 2.9 of [20] we deduce that $L_{y_ky}(\xi_k)$ has a subsequence convergent to some vector $\xi \in \partial f(y)$. Therefore, $v = d\exp_x^{-1}(y)(\xi)$, where $\xi \in \partial f(y)$. Since $\mathrm{dist}(x_k,y_k) \leq \varepsilon$ by the continuity of distance function $\mathrm{dist}(x,y) \leq \varepsilon$, which means $v \in F(x)$ and F is upper semicontinuous at x. Note that F has compact values, so the set valued function $\mathrm{convF} : \mathrm{B(x,r) \cap W} \rightrightarrows \mathrm{T_xM}$ defined by

$$\mathrm{convF(z) = L_{zx}(conv\{d\exp_z^{-1}(y)(\partial f(y)) : \ y \in clB(z,\varepsilon)\}).}$$

is upper semicontinuous at x.

Now, we prove the upper semicontinuity of $\partial_\varepsilon f$ at x. Let $\{x_k\} \subset B(x,r) \cap W$ and $\{v_k\} \subset TM$ be two sequences converging, respectively, to x and v, where $v_k \in \partial_\varepsilon f(x_k)$. So $L_{x_kx}(v_k) \in \mathrm{convF(x_k)}$ and by Remark 3.7, $L_{x_kx}(v_k)$ converges to v and by upper semicontinuity of convF at x, $v \in \mathrm{convF(x)} = \partial_\varepsilon \mathrm{f(x)}$.

□

Lemma 3.9. *Let B be a closed ball in a complete Riemannian manifold M, $f : M \to \mathbb{R}$ be locally Lipschitz, Z be the set of all stationary points of f in B and $B_\delta := \{x \in B : dist_Z(x) \geq \delta > 0\}$. Then, there exist $\varepsilon > 0$ and $\sigma > 0$ such that $0 \notin \partial_\varepsilon f(x)$ and $\min\{\|v\| : \ v \in \partial_\varepsilon f(x)\} \geq \sigma$, for all $x \in B_\delta$.*

Proof. Since B is compact, so there exists $\varepsilon > 0$ such that $\partial_\varepsilon f$ is well-defined on B_δ. Assume that $x \in B_\delta$, so $0 \notin \partial f(x)$, we claim that there exists $\varepsilon > 0$ such that $0 \notin \partial_\varepsilon f(x)$. On the contrary suppose that $0 \in \partial_{\frac{1}{i}} f(x)$, for $i = N, N+1, ...$, $1/N < i_M(x)$. Since $\partial_{\frac{1}{i}} f(x)$ is a compact and nested subset of T_xM, so

$$0 \in \cap_{i=N}^{\infty} \partial_{\frac{1}{i}} f(x) = \mathrm{conv} \cap_{\mathrm{i=N}}^{\infty} \{\mathrm{d\exp_x^{-1}(y)(\partial f(y)) : \ dist(y,x) \leq \frac{1}{i}}\}.$$

Hence, $0 = \Sigma_{k=1}^m t_k\xi_k$, where $\xi_k \in \cap_{i=N}^{\infty}\{d\exp_x^{-1}(y)(\partial f(y)) : \ \mathrm{dist}(y,x) \leq \frac{1}{i}\}$. Therefore, there exists $w_{k_i} \in \partial f(y_{k_i})$ such that $\mathrm{dist}(y_{k_i},x) \leq \frac{1}{i+N}$ with $\xi_k = d\exp_x^{-1}(y_{k_i})(w_{k_i})$. Since M is complete, so $\{y_{k_i}\}$ has a subsequence convergent to x in M. By Theorem 2.9 of [20], $L_{y_{k_i}x}(w_{k_i})$ has a subsequence convergent to some vector $\xi \in \partial f(x)$. Since $L_{y_{k_i}x}(w_{k_i}) = L_{y_{k_i}x}((d\exp_x^{-1}(y_{k_i}))^{-1}(\xi_k))$ converges to ξ_k, then $\xi = \xi_k \in \partial f(x)$ and since $\partial f(x)$ is convex, $0 \in \partial f(x)$ which is a contradiction. Note that $\partial_\varepsilon f(x)$ is a compact subset of T_xM, and the norm function is continuous, so there exists $0 \neq w \in \partial_\varepsilon f(x)$, such that $\|w\| = \min\{\|v\| : \ v \in \partial_\varepsilon f(x)\}$. □

3.2. **Descent Directions.** In the present section we treat the problem of finding directions $w_0 \in \partial_\varepsilon f(x)$ such that with suitable step lengths $t > 0$ the objective function f affords a decrease along the geodesic $\exp_x(\frac{-tw_0}{\|w_0\|})$. The next result shows that, whenever one has full knowledge of the ε-subdifferential, a suitable descent direction can be obtained by solving a simple quadratic program. We will use the following theorem, for its proof see [20].

Theorem 3.10. ***(Lebourg's Mean Value Theorem)*** *Let M be a finite dimensional Riemannian manifold, x, $y \in M$ and $\gamma : [0,1] \longrightarrow M$ be a smooth path* *joining x and y. Let f be a Lipschitz function around $\gamma[0,1]$. Then, there exist $0 < t_0 < 1$ and $\xi \in \partial f(\gamma(t_0))$ such that*

$$f(y) - f(x) = \langle \xi, \gamma^{'}(t_0) \rangle.$$

Theorem 3.11. *Assume $\varepsilon > 0$ and δ are given from Lemma 3.9 so that $0 \notin \partial_\varepsilon f(x)$ for all $x \in B_\delta$. Let $x \in B_\delta$ and consider an element of $\partial_\varepsilon f(x)$ with minimum norm,*

$$w_0 := \text{argmin}\{\|v\| : \ v \in \partial_\varepsilon f(x)\},$$

and get $g_0 := -\frac{w_0}{\|w_0\|}$. Then g_0 affords a uniform decrease of f over $B(x,\varepsilon)$, e.g.,

$$f(\exp_x(\varepsilon g_0)) - f(x) \leq -\varepsilon \|w_0\|.$$

Proof. By Lebourg's mean value theorem [20], there exist $0 < t_0 < 1$ and $\xi \in \partial f(\gamma(t_0))$ such that $f(\exp_x(\varepsilon g_0)) - f(x) = \langle \xi, \gamma'(t_0) \rangle$, where $\gamma(t) := \exp_x(t\varepsilon g_0)$ is a geodesic starting at x by initial speed εg_0. So

$$f(\exp_x(\varepsilon g_0)) - f(x) = \langle \xi, d\exp_x(\varepsilon t_0 g_0)(\varepsilon g_0) \rangle = \varepsilon \langle d\exp_x^{-1}(\exp_x(\varepsilon t_0 g_0))(\xi), g_0 \rangle.$$

Since $\text{dist}(\exp_x(\varepsilon t_0 g_0), x) = t_0 \varepsilon \leq \varepsilon$, so $d\exp_x^{-1}(\exp_x(\varepsilon t_0 g_0))(\xi) \in \partial_\varepsilon f(x)$. We claim that $\|w_0\|^2 \leq \langle \phi, w_0 \rangle$, for every $\phi \in \partial_\varepsilon f(x)$ which implies $\langle \phi, g_0 \rangle \leq -\|w_0\|$. Hence, we can deduce that $f(\exp_x(\varepsilon g_0)) - f(x) \leq -\varepsilon \|w_0\|$.

Proof of the claim: assume on the contrary; there exists $\phi \in \partial_\varepsilon f(x)$, such that $\langle \phi, w_0 \rangle < \|w_0\|^2$, and consider $w := w_0 + t(\phi - w_0) \in \partial_\varepsilon f(x)$, then

$$\|w_0\|^2 - \|w\|^2 = -t(2\langle w_0, \phi - w_0 \rangle + t\langle \phi - w_0, \phi - w_0 \rangle),$$

we can assume that t is small enough such that $\|w_0\|^2 > \|w\|^2$, which is a contradiction. □

Definition 3.12 (Descent Direction)**.** *Let $f : M \to \mathbb{R}$ be a locally Lipschitz function on a complete Riemannian manifold M, $w \in T_x M$, $g = -\frac{w}{\|w\|}$ is called a decent direction at x, if there exists $\alpha > 0$ such that*

$$f(\exp_x(tg)) - f(x) \leq -t\|w\|, \ \forall t \in (0, \alpha).$$

In the construction of the previous theorem, g_0 is a descent direction of f at x. So it is clear that we can choose the mentioned descent direction in order to move along a geodesic starting from an initial point toward a neighborhood of a minimum point.

3.3. **Approximation of the ε-subdifferential.** For general nonsmooth optimization problems it may be difficult to give an explicit description of the full subdifferential set. In the present section we generalize ideas of [30] to obtain an iterative procedure to approximate the ε-subdifferential. We start with the gradient of an arbitrary point nearby x and move the gradient to the tangent space in x via the derivative of the logarithm mapping, and in every subsequent iteration, the gradient of a new point nearby x is computed and moved to the tangent space in x to add to the working set to improve the approximation of $\partial_\varepsilon f(x)$. Indeed, we do not want to provide a description of the entire ε-subdifferential set at each iteration, what we do is to approximate $\partial_\varepsilon f(x)$ by the convex hull of its elements. In this way, let $W_k := \{v_1, ..., v_k\} \subseteq \partial_\varepsilon f(x)$, then we define

$$w_k := \underset{v \in \text{conv} \text{W}_\text{k}}{\text{argmin}} \|v\|.$$

Now if we have

$$f(\exp_x(\varepsilon g_k)) - f(x) \le -c\varepsilon\|w_k\|,\ c \in (0,1) \tag{3.1}$$

where $g_k = -\frac{w_k}{\|w_k\|}$, then we can say $\mathrm{convW_k}$ is an acceptable approximation for $\partial_\varepsilon f(x)$. Otherwise, we add a new element of $\partial_\varepsilon f(x) \setminus \mathrm{convW_k}$ to W_k.

Lemma 3.13. *Let $v \in \partial_\varepsilon f(x)$, such that $\langle v, g_k\rangle > -\|w_k\|$, then $v \notin \mathrm{convW_k}$.*

Proof. It can be proved along the same lines as the proof of the claim of Theorem 3.11. □

The following lemma proves that if W_k is not an acceptable approximation for $\partial_\varepsilon f(x)$, then there exists $v_{k+1} \in \partial_\varepsilon f(x)$, such that $\langle v_{k+1}, g_k\rangle \ge -c\|w_k\| > -\|w_k\|$, so by the previous lemma $v_{k+1} \in \partial_\varepsilon f(x) \setminus \mathrm{convW_k}$.

Lemma 3.14. *Let $W_k = \{v_1, ..., v_k\} \subset \partial_\varepsilon f(x)$, $0 \notin \mathrm{convW_k}$ and*

$$w_k = \mathrm{argmin}\{\|v\| :\ v \in \mathrm{convW_k}\}.$$

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

$$\langle d\exp_x^{-1}(\exp_x(\theta_0 g_k))(\bar{v}_{k+1}), g_k\rangle \ge -c\|w_k\|,$$

and $v_{k+1} := d\exp_x^{-1}(\exp_x(\theta_0 g_k))(\bar{v}_{k+1}) \notin \mathrm{convW_k}$.

Proof. We prove this lemma using Lemma 3.1 and Proposition 3.1 in [30]. Define

$$h(t) := f(\exp_x(tg_k)) - f(x) + ct\|w_k\|,\ t \in \mathbb{R},$$

and a new locally Lipschitz function $G : B(0_x, i_M(x)) \subset T_xM \to \mathbb{R}$, by $G(g) = f(\exp_x(g))$, then $h(t) = G(tg_k) - G(0) + ct\|w_k\|$. Assume that $h(\varepsilon) > 0$, then by Proposition 3.1 of [30], there exists $\theta_0 \in [0, \varepsilon]$ such that h is increasing in a neighborhood of θ_0. Therefore, by Lemma 3.1 of [30] for every $\xi \in \partial h(\theta_0)$, one has $\xi \ge 0$. By [20, Proposition 3.1]

$$\partial h(\theta_0) \subseteq \langle \partial f(\exp_x(\theta_0 g_k)), d\exp_x(\theta_0 g_k)(g_k)\rangle + c\|w_k\|,$$

if $\bar{v}_{k+1} \in \partial f(\exp_x(\theta_0 g_k))$ such that

$$\langle \bar{v}_{k+1}, d\exp_x(\theta_0 g_k)(g_k)\rangle + c\|w_k\| \in \partial h(\theta_0),$$

then

$$\langle d\exp_x^{-1}(\exp_x(\theta_0 g_k))(\bar{v}_{k+1}), g_k\rangle + c\|w_k\| \ge 0.$$

Now, Lemma 3.13 implies that

$$v_{k+1} := d\exp_x^{-1}(\exp_x(\theta_0 g_k))(\bar{v}_{k+1}) \notin \mathrm{convW_k}$$

which proves our claim. □

Now we present Algorithm 1 to find a vector $v_{k+1} \in \partial_\varepsilon f(x)$ which can be added to the set W_k 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 [30] that this algorithm terminates after finitely many iterations. Then we give Algorithm 2 for finding a descent direction. Moreover, Theorem 3.15 proves that Algorithm 2 terminates after finitely many iterations.

Algorithm 1 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 $\langle v, d\exp_x(tg)(g)\rangle + c\|w\| \geq 0$

Algorithm 2 A descent direction algorithm; $(g_k, k) = Decent(x, \delta, c, \varepsilon)$.

Input $x \in M; \delta, c, \varepsilon \in (0, 1)$.
Let $g_1 \in T_xM$ such that $\|g_1\| = 1$.
if f is differentiable on $\exp_x(\varepsilon g_1)$, **then**
  $v = d\exp_x^{-1}(\exp_x(\varepsilon g_1))(\operatorname{grad} f(\exp_x(\varepsilon g_1)))$
else select arbitrary $v \in d\exp_x^{-1}(\exp_x(\varepsilon g_1))(\partial f(\exp_x(\varepsilon g_1)))$
  Set $W_1 = \{v\}$ and let $k = 1$
end if
Step 1: (Compute a descent direction)
Solve the following minimization problem and let w_k be its solution:

$$\min_{v \in \mathrm{conv W_k}} \|v\|.$$

if $\|w_k\| \leq \delta$ **then** stop
else let $g_{k+1} = -\frac{w_k}{\|w_k\|}$.
end if
Step 2: (Stopping condition)
if $f(\exp_x(\varepsilon g_{k+1})) - f(x) \leq -c\varepsilon\|w_k\|$, **then** stop.
end if
Step 3: $v = Increasing(x, g_{k+1}, 0, \varepsilon)$.
Set $v_{k+1} = v$, $W_{k+1} = W_k \cup \{v_{k+1}\}$ and $k = k + 1$. Go to step 1.

Theorem 3.15. *Let for the point $x_1 \in M$, the level set $N = \{x : \ f(x) \leq f(x_1)\}$, be bounded, then for each $x \in N$, Algorithm 2 terminates after finitely many iterations.*

Proof. Now we claim that either after a finite number of iterations the stopping condition is satisfied or for some m,

$$\|w_m\| \leq \delta,$$

and the algorithm terminates. If the stopping condition is not satisfied and $\|w_k\| > \delta$, then by Lemma 3.14 we find $v_{k+1} \notin \mathrm{conv W_k}$ such that

$$\langle v_{k+1}, w_k\rangle \leq c\|w_k\|^2.$$

Note that $d\exp_x^{-1}$ on $\mathrm{cl}B(x,\varepsilon)$ is bounded by some $m_1 \geq 0$ and by the Lipschitzness of f of the constant L, Theorem 2.9 of [20] implies that for every $\xi \in \partial_\varepsilon f(x)$, $\|\xi\| \leq m_1 L$. Now, $w_{k+1} \in \mathrm{conv}\{\mathrm{v_{k+1}}\} \cup \mathrm{W_k}$ has the minimum norm, so for all $t \in (0,1)$,

$$\begin{aligned}\|w_{k+1}\|^2 &\leq \|tv_{k+1} + (1-t)w_k\|^2 \\ &\leq \|w_k\|^2 + 2t\langle w_k, (v_{k+1} - w_k)\rangle + t^2\|v_{k+1} - w_k\|^2 \\ &\leq \|w_k\|^2 - 2t(1-c)\|w_k\|^2 + 4t^2L^2m_1^2 \\ &\leq (1 - [(1-c)(2Lm_1)^{-1}\delta]^2)\|w_k\|^2\end{aligned} \tag{3.2}$$

the last inequality is obtained by assuming $t = (1-c)(2Lm_1)^{-2}\|w_k\|^2 \in (0,1)$, $\delta \in (0, Lm_1)$ and $\|w_k\| > \delta$. Now considering $r = 1 - [(1-c)(2Lm_1)^{-1}\delta]^2$, it follows that

$$\|w_{k+1}\|^2 \leq r\|w_k\|^2 \leq ... \leq r^k(Lm_1)^2.$$

Therefore, after a finite number of iterations $\|w_{k+1}\| \leq \delta$. □

A popular inexact line search condition stipulates that ε should first of all give sufficient decrease in the objective function f, as measured by the following inequality named Armijo condition

$$f(\exp_{x_k}(\varepsilon g_{k+1})) - f(x_k) \leq -c_1\varepsilon\|w_k\|, \tag{3.3}$$

for some $c_1 \in (0,1)$.

To rule out unacceptable short steps we introduce a second requirement, called the curvature condition, which requires the existence of $p_k \in \partial f(\exp_{x_k}(\varepsilon g_{k+1}))$,

$$\langle d\exp_{x_k}(\varepsilon g_{k+1})(g_{k+1}), p_k\rangle \geq -c_2\|w_k\|,$$

for some $c_2 \in (c_1, 1)$, where c_1 is the constant in (3.3). We even can consider a more strong curvature condition:

$$|\langle d\exp_{x_k}(\varepsilon g_{k+1})(g_{k+1}), p_k\rangle| \leq c_2\|w_k\|,$$

for some $p_k \in \partial f(\exp(\varepsilon g_{k+1}))$, $c_2 \in (c_1, 1)$.

In the following theorem, it is shown that for every locally Lipschitz function f defined on a complete Riemannian manifold M, there exist step lengths satisfying (3.3), the strong curvature and the curvature conditions.

Theorem 3.16. *Suppose that $f : M \to \mathbb{R}$ is locally Lipschitz. Let g_{k+1} be a descent direction at x_k and assume that f is bounded below along the geodesic $\{\exp_{x_k}(\varepsilon g_{k+1}) : \ \varepsilon > 0\}$, then if $0 < c_1 < c_2 < 1$, there exist step lengths satisfying (3.3), the strong curvature and the curvature conditions.*

Proof. Since $\Phi(\varepsilon) = f(\exp_{x_k}(\varepsilon g_{k+1}))$ is bounded below for all $\varepsilon > 0$ and $0 < c_1 < 1$, the line $l(\varepsilon) = f(x_k) - \varepsilon c_1\|w_k\|$ must intersect the graph ϕ at least once. Let $\varepsilon' > 0$ be the smallest intersecting value of ε, that is,

$$f(\exp_{x_k}(\varepsilon' g_{k+1})) = f(x_k) - \varepsilon' c_1\|w_k\|. \tag{3.4}$$

So for all step lengths less than ε', (3.3) holds.

Now by the mean value theorem, there exist $\varepsilon^* \in (0,1)$ and $\xi \in \partial f(\exp_{x_k}(\varepsilon^*\varepsilon' g_{k+1}))$ such that

$$f(\exp_{x_k}(\varepsilon' g_{k+1})) - f(x_k) = \varepsilon'\langle \xi, d\exp_{x_k}(\varepsilon^*\varepsilon' g_{k+1})(g_{k+1})\rangle. \tag{3.5}$$

By combining (3.4) and (3.5), we obtain

$$\langle \xi, d\exp_{x_k}(\varepsilon^*\varepsilon' g_{k+1})(g_{k+1})\rangle = -c_1\|w_k\| > -c_2\|w_k\|. \tag{3.6}$$

Therefore, $\varepsilon^*\varepsilon'$ satisfies the curvature condition. Moreover, since the term in the left-hand side of (3.6) is negative, the strong curvature condition holds with the same step lengths.

□

Remark 3.17. *Assume that $\varepsilon' < i_M(x_k)$, then ε' satisfies the curvature condition if and only if $\Phi^\circ(\varepsilon', 1) \geq -c_2\|w_k\|$. Since if ε' satisfies the curvature condition then there exists $p_k \in \partial f(\exp_{x_k}(\varepsilon' g_{k+1}))$,*

$$\langle d\exp_{x_k}(\varepsilon' g_{k+1})(g_{k+1}), p_k\rangle \geq -c_2\|w_k\|,$$

for some $c_2 \in (c_1, 1)$, where c_1 is the constant in (3.3). By Proposition 3.4 of [20]

$$\langle d\exp_{x_k}(\varepsilon' g_{k+1})(g_{k+1}), p_k\rangle \in \partial\Phi(\varepsilon').$$

Therefore,

$$\Phi^\circ(\varepsilon', 1) \geq \langle d\exp_{x_k}(\varepsilon g_{k+1})(g_{k+1}), p_k\rangle \geq -c_2\|w_k\|.$$

On the converse, assume that $\Phi^\circ(\varepsilon', 1) \geq -c_2\|w_k\|$, then by [20, Theorem 2.9] *there exists $\xi_k \in \partial\Phi(\varepsilon')$ such that $\Phi^\circ(\varepsilon', 1) = \xi_k$. Hence by Proposition 3.4 of* [20]*, there exists $p_k \in \partial f(\exp_{x_k}(\varepsilon' g_{k+1}))$, such that*

$$\langle d\exp_{x_k}(\varepsilon' g_{k+1})(g_{k+1}), p_k\rangle = \xi_k \geq -c_2\|w_k\|.$$

Given the current iterate x_k and a search direction w_k, as we mentioned before, the task of a line search is to find a step size which decreases the objective function along the geodesics. The mentioned curvature conditions are used in the line search to enforce a sufficient decrease in the objective value, and to exclude unnecessarily small step sizes.

Algorithm 3 is a one dimensional search procedure for the function Φ that is guaranteed to find a step length satisfying (3.3) and the curvature condition. The procedure is a generalization of the algorithm for well-known strong wolf condition for smooth functions, see [32, p. 59-60]. The algorithm has two stages. The first stage begins with a trial estimate a_1 and keeps it increasing until it finds either an acceptable step length or an interval that contains the desired step length. Note that the sequence of trial step lengths $\{a_i\}$ is monotonically increasing, but that the order of the arguments supplied to the next algorithm may vary. The parameter a_{max} is a user-supplied bound on the maximum step length allowed. The last step of Algorithm 3 performs extrapolation to find the next trial value a_{i+1}. To implement this step we can simply set a_{i+1} to some constant multiple of a_i. In the case that Algorithm 3 finds an interval that contains the desired step length, the second stage is invoked by Algorithm 4 called *Zoom* which successively decreases the size of the interval until an acceptable step length is identified. The order of its input arguments is such that each call has the form $Zoom(a, b, c_1, c_2)$ where, the interval bounded by a and b contains a step length that satisfies the curvature condition and (3.3). Moreover, a is, among all step lengths generated so far and satisfying (3.3), the one giving the smallest function value. So the line search algorithm terminates with a^* set to a step length that satisfies the curvature condition.

Algorithm 3 A line search Algorithm; $a^* = Line(x, a_1, c_1, c_2)$

Input $x \in M$, $a_1 > 0$, $c_1, c_2 \in (0, 1)$.
Set $a_0 = 0$, $a_{max} = i_M(x)$, $i = 1$;
Repeat
Evaluate $\Phi(a_i)$
if $\Phi(a_i) > \Phi(0) - c_1 a_i \|w_0\|$ **then**
    a^* must be obtained by $Zoom(a_{i-1}, a_i, c_1, c_2)$ Stop
end if
Evaluate $\Phi^\circ(a_i, 1)$;
if $\Phi^\circ(a_i, 1) \geq -c_2 \|w_0\|$ **then** $a^* = a_i$, Stop.
end if
if $\Phi^\circ(a_i, 1) \geq 0$ **then** a^* must be obtained by $Zoom(a_i, a_{i-1}, c_1, c_2)$, Stop.
end if
Choose $a_{i+1} \in (a_i, a_{max})$
$i = i + 1$.
End(Repeat)

Algorithm 4 $a^* = Zoom(a, b, c_1, c_2)$

Input a, b, c_1, c_2.
$i = 1$;
Repeat
$a_i = \frac{a+b}{2}$
Evaluate $\Phi(a_i)$
if $\Phi(a_i) > \Phi(0) - c_1 a_i \|w_0\|$ **then**
    $b = a_i$
else
    Evaluate $\Phi^\circ(a_i, 1)$;
    if $\Phi^\circ(a_i, 1) \geq -c_2 \|w_0\|$ **then** $a^* = a_i$, Stop.
    end if
    if $\Phi^\circ(a_i, 1)(b - a) \geq 0$ **then** $b = a$, $a = a_i$.
    end if
end if
End(Repeat)

Finally, Algorithm 5 is the minimization algorithm which finds a descent direction in any iteration.

Theorem 3.18. *If $f : M \to \mathbb{R}$ is a locally Lipschitz function on a complete Riemannian manifold M, and*

$$N = \{x : \ f(x) \leq f(x_1)\}$$

is bounded, then either Algorithm 5 terminates after finitely number of iterations with $\|w_k^s\| = 0$, or every accumulation point of the sequence $\{x_k\}$, belongs to the set

$$X = \{x \in M : \ 0 \in \partial f(x)\}.$$

Proof. Note that N is bounded, so $\mathrm{cl}N$ is compact and there exists $\varepsilon < i(\mathrm{cl}N)$ such that $\partial_\varepsilon f$ on N is well-defined. If the algorithm terminates after finite number of

Algorithm 5 A minimization algorithm; $x_k = Min(f, x_1, \theta_\varepsilon, \theta_\delta, \varepsilon_1, \delta_1, c_1, c_2)$.

Input: f (A locally Lipschitz function defined on a complete Riemannian manifold M); $x_1 \in M$ (a starting point); $c_1, c_2, \theta_\varepsilon$, θ_δ, ε_1, $\delta_1 \in (0,1)$; $k = 1$.
Step 1 (Set new parameters) $s = 1$ and $x_k^s = x_k$.
Step 2. (Descent direction) $(g_k^s, n_k^s) = Decent(x_k^s, \delta_k, c_1, \varepsilon_k)$

$$\|w_k^s\| = \min\{\|w\| : \ w \in \mathrm{convW_k^s}\}.$$

if $\|w_k^s\| = 0$ **then** stop
else let $g_s^k = -\frac{w_k^s}{\|w_k^s\|}$ be the descent direction.
end if
if $\|w_k\| \leq \delta_k$ **then** set $\varepsilon_{k+1} = \varepsilon_k \times \theta_\varepsilon$, $\delta_{k+1} = \delta_k \times \theta_\delta$, $k = k+1$, $x_k = x_k^s$. Go to step 1. Otherwise go to Step 3.
end if
Step 3. **A line search**
if

$$f(\exp_{x_k^s}(\varepsilon_k g_k^s)) - f(x_k^s) \leq -c_1 \varepsilon_k \|w_k^s\|,$$

then $\sigma = Line(x_k^s, \varepsilon_k, c_1, c_2)$; and construct the next iterate $x_k^{s+1} = \exp_{x_k^s}(\sigma g_k^s)$. Set $s = s+1$ and go to step 2
end if

iterations then x_k^s is an $\varepsilon-$stationary point of f, which means there exists $y_k^s \in B(x_k^s, \varepsilon_k)$ such that $0 \in \partial f(y_k^s)$. Otherwise, let

$$f(x_k^{s+1}) - f(x_k^s) \leq -c\varepsilon_k \|w_k^s\| < 0,$$

therefore, $f(x_k^{s+1}) < f(x_k^s)$ for $s = 1, 2, ...$ Since f is Lipschitz and N is bounded, f has a minimum in N, so $f(x_k^s)$ is a bounded decreasing sequence in $\mathbb{R}$, so is convergent. Thus $f(x_k^s) - f(x_k^{s+1})$ is convergent to zero and there exists s_k such that

$$f(x_k^s) - f(x_k^{s+1}) \leq c\varepsilon_k \delta_k,$$

for all $s \geq s_k$. Thus

$$\|w_k^s\| \leq \frac{f(x_k^s) - f(x_k^{s+1})}{c\varepsilon_k} \leq \delta_k, \ \ s \geq s_k. \tag{3.7}$$

Hence after finitely many iterations, there exists s_k such that

$$x_{k+1} = x_k^{s_k},$$

and

$$\min\{\|v\| : \ v \in \mathrm{convW}_{\mathrm{n_k^{s_k}+1}}\} \leq \mathrm{\delta_k}.$$

Since M is a complete Riemannian manifold and $\{x_k\} \subset N$ is bounded, there exists a subsequence $\{x_{k_i}\}$ converging to a point $x^* \in M$. Since $\mathrm{convW}_{\mathrm{n_{k_i}^{s_{k_i}}+1}}$ is a subset of $\partial_{\varepsilon_{k_i}} f(x_{k_i}^{s_{k_i}})$ then,

$$\|w_{k_i}\| = \min\{\|v\| : v \in \partial_{\varepsilon_{k_i}} f(x_{k_i}^{s_{k_i}})\} \leq \delta_{k_i}$$

Hence $\lim_{k_i \to \infty} \|w_{k_i}\| = 0$. Note that $w_{k_i} \in \partial_{\varepsilon_{k_i}} f(x_{k_i}^{s_{k_i}})$, so by Lemma 3.8 and Remark 3.6, $0 \in \partial f(x^*)$. $\square$

4. Numerical Experiments

We close this article by giving two numerical experiments. We are going to solve the one dimensional total variation problem for functions which map into a manifold. Therefore, assume that M is a manifold, consider the minimization problem

$$\min_{u\in BV([0,1];M)}\{F(u):=\text{dist}_2(f,u)^2+\lambda\|\nabla u\|_1\} \tag{4.1}$$

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, [37]. Note that for every $w\in[0,1]$, $\nabla u(w):\mathbb{R}\to T_{u(w)}M$, so $\|\nabla u\|_1=\int_{[0,1]}\|\nabla u(w)\|dw$. Now we can formulate a discrete version of the problem (4.1) 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 [38, 18]. We refer to [38] 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,

$$\min_{u\in M^n}\{F_*(u):=\text{dist}_*(\varepsilon(f),u)^2+\lambda\|\nabla(\varepsilon^{-1}(u))\|_1\} \tag{4.2}$$

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

Theorem 4.1. *Let M be a Hadamard manifold, If F_* is defined as in (4.2), then F_* is convex as a function defined on M^n.*

Proof. It is enough to prove that $\|\nabla(\varepsilon^{-1}(u))\|_1$ is convex. So we should prove that $\int_{[0,1]}\|\nabla v_{hu}(w)\|dw$, where v_{hu} is the geodesic finite element function corresponding to u, is convex. To do this, assume that u_1,u_2 are two arbitrary points in M^n, and $\gamma=(\gamma_1,...,\gamma_n)$ is a geodesic connecting them, we first show that for every arbitrary fix $w\in[0,1]$, $f(t)=\|\nabla v_{h\gamma(t)}(w)\|$ as a function of t is convex.

Now define

$$g(t)=1/2\langle\nabla v_{h\gamma(t)}(w),\nabla v_{h\gamma(t)}(w)\rangle.$$

Assume that Γ is a grid on $[0,1]$ and $(s_i,s_{i+1})\in\Gamma$ is such that $w\in(s_i,s_{i+1})$, moreover, σ_{it} is a minimizing geodesic parametrized by arc length connecting $\gamma_i(t)$ and $\gamma_{(i+1)}(t)$.

Since σ_{it} is a geodesic with a constant speed,

$$g(t)=1/2\langle\nabla\sigma_{it}(w),\nabla\sigma_{it}(w)\rangle=\frac{1}{2}\int_0^1\langle\nabla\sigma_{it}(x),\nabla\sigma_{it}(x)\rangle dx.$$

Now we define another smooth function $G:[0,1]\times[0,1]\to M$, by

$$G(t,x)=\sigma_{it}(x).$$

We put $V(t,x):=\frac{\partial G}{\partial t}(t,x)$ and usually write $\nabla\sigma_{it}(x)=\nabla G=\frac{\partial G}{\partial x}dx$. Consider the vector bundle $T([0,1]\times[0,1])^*\otimes G^{-1}TM$ over $[0,1]\times[0,1]$ which admits a natural fiber metric and a standard connection ∇ compatible with the metric. Under the natural identification, we denote $\nabla_x=\nabla_{(0,\frac{\partial}{\partial x})}$ and $\nabla_t=\nabla_{(\frac{\partial}{\partial t},0)}$

$$1/2\frac{\partial^2}{\partial^2 t}\langle\frac{\partial G}{\partial x}dx, \frac{\partial G}{\partial x}dx\rangle = \frac{\partial}{\partial t}\langle\nabla_t\frac{\partial G}{\partial x}dx, \frac{\partial G}{\partial x}dx\rangle$$

since ∇ is metric,

$$= \frac{\partial}{\partial t}\langle\nabla_x\frac{\partial G}{\partial t}dx, \frac{\partial G}{\partial x}dx\rangle$$

$$= \langle\nabla_t\nabla_x\frac{\partial G}{\partial t}dx, \frac{\partial G}{\partial x}dx\rangle + \langle\nabla_x\frac{\partial G}{\partial t}dx, \nabla_x\frac{\partial G}{\partial t}dx\rangle$$

since ∇ is torsion free, and by definition of curvature tensor,

$$= \langle\nabla_x\nabla_t\frac{\partial G}{\partial t}dx, \frac{\partial G}{\partial x}dx\rangle + \langle R(\frac{\partial G}{\partial t}, \frac{\partial G}{\partial x})\ \frac{\partial G}{\partial t}dx, \frac{\partial G}{\partial x}dx\rangle + \langle\nabla_x V dx, \nabla_x V dx\rangle$$

Now since ∇ is metric,

$$0 = \int_0^1 \frac{\partial}{\partial x}\langle\nabla_t\frac{\partial G}{\partial t}, \frac{\partial G}{\partial x}dx\rangle dx =$$

$$\int_0^1 \langle\nabla_x\nabla_t\frac{\partial G}{\partial t}dx, \frac{\partial G}{\partial x}dx\rangle dx.$$

Hence

$$g''(t) = \int_0^1 \|\nabla V\|^2 - trace\langle R(\nabla G, V)V, \nabla G\rangle.$$

Since M is a Hadamard manifold, it is easy to see that $g''(t) \geq 0$ which implies g is convex. By definition of g, it is clear that $g(t) = 1/2f^2(t)$, we assume that $f(t) \neq 0$, then

$$f''(t) = \frac{g''(t)f^2(t) - (g'(t))^2}{f^3(t)}.$$

Hence

$$f''(t) = \frac{1}{f^3(t)}\{\int_0^1 (\|\nabla V\|^2)\langle\nabla v_{h\gamma(t)}(w), \nabla v_{h\gamma(t)}(w)\rangle$$

$$-trace\langle R(\nabla G, V)V, \nabla G\rangle\langle\nabla v_{h\gamma(t)}(w), \nabla v_{h\gamma(t)}(w)\rangle$$

$$-\langle\nabla V, \nabla v_{h\gamma(t)}(w)\rangle^2\} \geq 0,$$

by the Cauchy-Schwarz inequality and the negativity of the sectional curvature.

So we proved that for every $w \in [0,1]$, $f(t) = \|\nabla v_{h\gamma}(w)\|$ is convex, hence

$$\|\nabla v_{h\gamma(t)}(w)\| \leq t\|\nabla v_{hu_1}(w)\| + (1-t)\|\nabla v_{hu_2}(w)\|,$$

which implies

$$\int_{[0,1]} \|\nabla v_{h\gamma(t)}(w)\|dw \leq t\int_{[0,1]} \|\nabla v_{hu_1}(w)\|dw + (1-t)\int_{[0,1]} \|\nabla v_{hu_2}(w)\|dw,$$

which means $\int_{[0,1]} \|\nabla v_{hu}(w)\| = \|\nabla(\varepsilon^{-1}(u))\|_1$ is convex.

Note that if M is a Hadamard manifold, dist^2 is also a convex function on M^n. Hence we can conclude that F_* is convex on M^n. □

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

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

where dist is the Riemannian distance on M. So in order to find the subdifferential of F, we have to find the subdifferential of the distance and the squared distance functions. The distance function is differentiable at $(p, q) \in M \times M$ if and only if there is a unique length minimizing geodesic from p to q. Furthermore, the distance function is smooth in a neighborhood of (p, q) if and only if p and q are not conjugate points along this minimizing geodesic. So the distance function is nondifferentiable at (p, q) if and only if $p = q$ or p and q are the conjugate points. Let the distance function be differentiable at (p, q), then

$$\frac{\partial \text{dist}}{\partial p}(p, q) = \frac{-\exp_p^{-1}(q)}{\text{dist}(p, q)}, \ \frac{\partial \text{dist}^2}{\partial p}(p, q) = -2\exp_p^{-1}(q).$$

In the next lemma we assume that $p = q$, and find a formula for the subdifferential of the distance function.

Lemma 4.2. *Let M be a complete Riemannian manifold. If $dist_p : M \to \mathbb{R}$ is defined by $dist_p(q) = dist(p, q)$, then*

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

where B is the closed unit ball of T_pM.

Before proving the lemma, let us recall a definition of the normal cone and the tangent cone to a closed convex subset of a Riemannian manifold, for more details see [21]. Let S be a closed convex subset of a Riemannian manifold M, the normal cone to S at p denoted by $N_S(p)$ and the tangent cone to S at p denoted by $T_S(p)$ are defined by

$$N_S(p) = \{\xi \in T_pM : \langle \xi, \exp_p^{-1}(q)\rangle \leq 0 \text{ for every } q \in S\}.$$
$$T_S(p) := \{\xi \in T_pM : \langle \xi, v\rangle \leq 0 \ \forall v \in N_S(p)\}.$$

Assume that $S = \{p\}$, then $N_S(p) = T_pM$. Reader can refer to [20, 21, 22] for more details about the normal cone and the tangent cone.

Proof. Let $M \cong \mathbb{R}^n$ and S be a closed convex subset of M, we claim that for every $x \in S$, $\partial \text{dist}_S(x) = N_S(x) \cap B$.

Let $\xi \in N_S(x) \cap B$. For every $y \in M$, there exists $z \in S$ such that $\text{dist}_S(y) = \text{dist}(z, y)$. By the definition of the normal cone

$$\langle \xi, y - x\rangle \leq \langle \xi, z - x\rangle + \langle \xi, y - z\rangle \leq 0 + \|\xi\|\|y - z\| \leq \text{dist}(z, y),$$

which implies $\langle \xi, y - x\rangle \leq \text{dist}_S(y) - \text{dist}_S(x)$, and $\xi \in \partial \text{dist}_S(x)$.

Now assume that $\xi \in \partial \text{dist}_S(x)$, so by [20, Theorem 4.10] $\text{dist}^\circ(x, v) \leq 0$ for every $v \in T_S(x)$. Moreover, by the definition of the support function, $\langle \xi, v\rangle \leq 0$, which means $\xi \in N_S(x)$.

Now we assume that M is a Riemannian manifold, $S = \{p\}$. First, we prove that $\partial \text{dist}_p(p) = \partial \text{dist}_0^*(0)$, where $\text{dist}, \text{dist}^*$ are respectively the Riemannian distance on M and the usual distance on T_pM. By Proposition 2.5 in [20], $\xi \in \partial \text{dist}_p(p)$ if and only if $\xi \in \partial(\text{dist}_p \circ \exp_p)(0)$ if and only if $\langle \xi, v\rangle \leq \text{dist}_p \circ \exp_p(v) - \text{dist}_p \circ \exp_p(0)$, for every $v \in T_pM$, if and only if $\langle \xi, v\rangle \leq \|v\|$ for every $v \in T_pM$, which means $\langle \xi, v\rangle \leq \text{dist}_0^*(v) - \text{dist}_0^*(0)$ for every $v \in T_pM$, and by the definition of the subdifferential $\xi \in \partial \text{dist}_0^*(0)$. Hence, $\partial \text{dist}_p(p) = \partial \text{dist}_0^*(0)$.

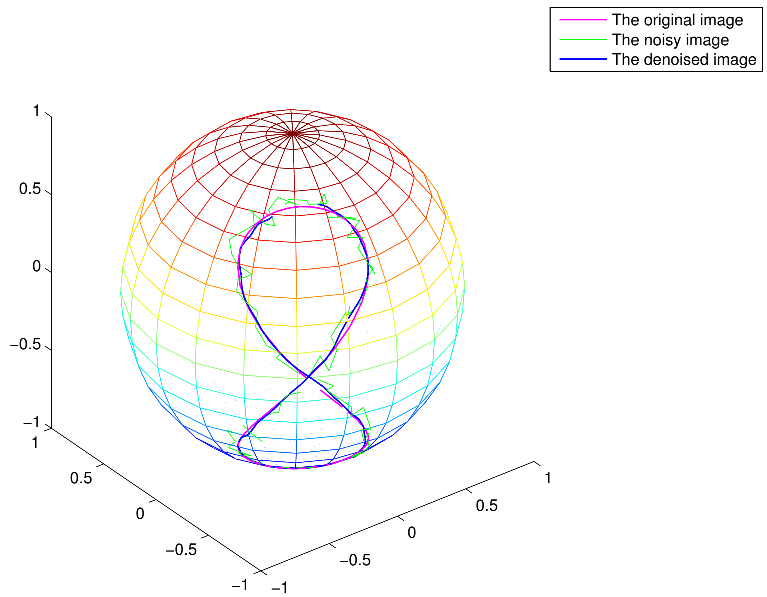

FIGURE 1. TV regularization on S^2.

It is worthwhile to mention that by Proposition 4.3 in [20], $N_p(p) = N_0(0)$. Therefore, by the claim,

$$N_p(p) \cap B = N_0(0) \cap B = \partial \mathrm{dist}_0^*(0) = \partial \mathrm{dist}_p(p).$$

As it was mentioned before $N_p(p) = T_pM$, so $\partial \mathrm{dist}_p(p) = B$. □

In our numerical examples, we consider a two dimensional sphere S^2 and the space of positive-definite matrices which is known as a Hadamard manifold. Therefore F_* is convex on the space of positive definite matrices, F_* is not a convex function on every sphere, see [41].

4.1. **The unit sphere S^2.** 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

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

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]

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

Utilizing the properties of the exponential map on a Riemannian manifold, for fixed point $x \in S^2$, and for each $\varepsilon > 0$, we may find number $\delta_x > 0$ such that:

$$\|d(\exp_x^{-1})(y) - L_{yx}\| \leq \varepsilon, \text{provided that } \mathrm{dist}(x, y) < \delta_x.$$

So we might be able to use parallel transport instead of $d\exp_x^{-1}$.

It is worthwhile to mention that on any sphere antipodal points are conjugate points, but without loss of generality we can assume that u_i and u_{i+1} are not conjugate. In fact we use more than two nodal points for discretization of the function F, so it can be assumed that there is a nodal point between every two antipodal points.

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 5 to the function F_* defined on M^{100} to get the denoised image, see Figure 1.

4.2. **The space of symmetric positive-definite matrices.** The set of symmetric positive definite matrices, as a Riemannian manifold, is the most studied example of manifolds of nonpositive curvature. 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

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

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

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

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

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

is defined by

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

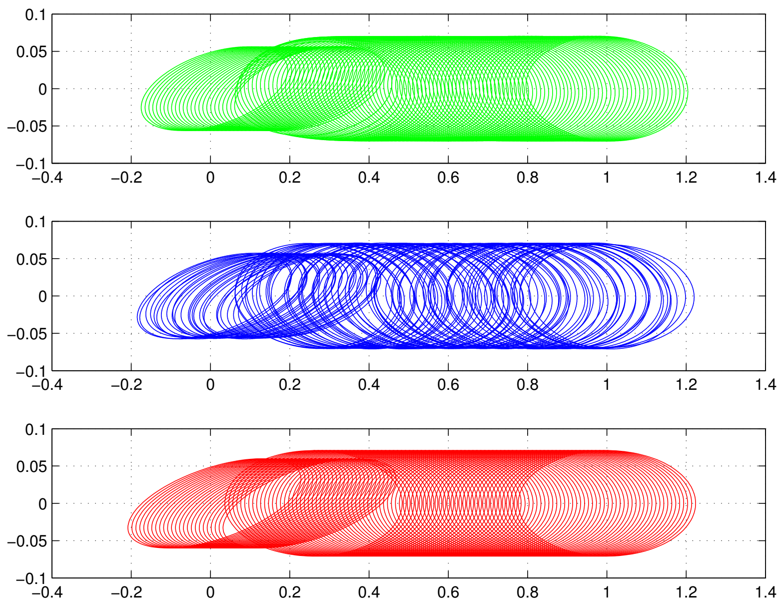

FIGURE 2. TV regularization on $P(2)$. Down-to-up: the original image, the noisy image, the denoised image.

where log, exp, denote the logarithm and the exponential functions on matrix space, for more details see [31].

Now we assume that $M = P(2)$. We add a random noise to an original image on $P(2)$. Then we apply algorithm 5 to F_* on M^{100} to denoise the noisy image. In Figure 2, we present the results regarding to the minimization of F_* on M^{100}.

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] 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.

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

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

[6] A. M. Bagirov, *Continuous subdifferential approximations and their applications*, J. Math. Sci., 115 (2003), pp. 2567-2609.

[7] 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.

[8] 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.

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

[10] 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.
[11] J. V. Burke, A. S. Lewis, M. L. Overton, *Approximating subdifferentials by random sampling of gradients*, Math. Oper. Res., 27 (2002), pp. 567-584.
[12] J. V. Burke, A. S. Lewis, M. L. Overton, *A robust gradient sampling algorithm for nonsmooth, nonconvex optimization,* SIAM J. Optim., 15 (2005), pp. 751-779.
[13] F. H. Clarke, *Necessary Conditions for Nonsmooth Problems in Optimal Control and the Calculus of Variations,* Thesis, University of Washington, Seattle, 1973.
[14] 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.
[15] 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.
[16] O. P. Ferreira, P. R. Oliveira, *Proximal point algorithm on Riemannian manifolds,* Optimization. 51 (2002), pp. 257-270.
[17] O. P. Ferreira, P. R. Oliveira, *Subgradient algorithm on Riemannian manifolds,* J. Optim. Theory Appl., 97 (1998), pp. 93-104.
[18] P. Grohs, H. Hardering, O. Sander, *Optimal a priori discretization error bounds for geodesic finite elements,* SAM Report 2013-16, ETH Zürich, (2013), Submitted.
[19] A. A. Goldstein, *Optimization of Lipschitz continuous functions,* Math. Program., 13 (1977), pp. 14-22.
[20] S. Hosseini, M. R. Pouryayevali, *Generalized gradients and characterization of epi-Lipschitz sets in Riemannian manifolds*, Nonlinear Anal., 74 (2011), pp. 3884-3895.
[21] S. Hosseini, M. R. Pouryayevali, *Euler characterization of epi-Lipschitz subsets of Riemannian manifolds,* J. Convex. Anal., 20 (2013), No. 1, pp. 67-91.
[22] 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.
[23] K. C. Kiwiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics, 1133, Springer-Verlag, Berlin, 1985.
[24] W. Klingenberg, *Riemannian Geometry,* Walter de Gruyter Studies in Mathematics, Vol. 1, Walter de Gruyter, Berlin, New York, 1995.
[25] S. Lang, *Fundamentals of Differential Geometry,* Graduate Texts in Mathematics, Vol. 191, Springer, New York, 1999.
[26] P. Y. Lee, *Geometric Optimization for Computer Vision,* PhD thesis, Australian National University, 2005.
[27] 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.
[28] 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.
[29] R. E. Mahony, *The constrained Newton method on a Lie group and the symmetric eigenvalue problem,* Linear Algebra. Appl., 248 (1996), pp. 67-89.
[30] N. Mahdavi-Amiri, R. Yousefpour, *An effective nonsmooth optimization algorithm for locally Lipschitz functions*, J. Optim. Theory Appl., 155 (2012), pp. 180-195.
[31] M. Moakher, M. Zerai, *The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data*, J. Math. Imaging Vision., 40 (2011), pp. 171-187.
[32] J. Nocedal, S. J. Wright,*Numerical Optimization,* Springer, 1999.
[33] 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.
[34] W. Ring, B. Wirth, *Optimization methods on Riemannian manifolds and their application to shape space*, SIAM J. Optim., 22(2) (2012), pp. 596-627.
[35] R. C. Riddell, *Minimax problems on Grassmann manifolds. Sums of eigenvalues,* Adv. Math., 54 (1984), pp. 107-199.
[36] R. T. Rockafellar, *Convex Functions and Dual Extremum Problems,* Thesis, Harvard, 1963.
[37] L. Rudin, S. Osher, E. Fatemi, *Nonlinear total variation based noise removal algorithms,* Phys. D., 60 (1) (1992), pp. 259-268.

[38] O. Sander, *Geodesic finite elements for Cosserat rods,* Internat. J. Numer. Methods Engrg., 82 (2010), pp. 1645-1670.

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

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

[41] S. T. Yau, *Non-existence of continuous convex functions on certain Riemannian manifolds*, Math. Ann., 207 (1974), pp. 269-270.

[42] L. S. Zhang, X. L. Sun, *An algorithm for minimizing a class of locally Lipschitz functions,* J. Optim. Theory Appl., 90 (1996), pp. 203-212.

[43] L. H. Zhang, *Riemannian Newton method for the multivariate eigenvalue problem,* SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2972-2996.

Recent Research Reports

Nr.	Authors/Title
2013-39	D. Schoetzau and Ch. Schwab and T.P. Wihler hp-dGFEM for Second-Order Mixed Elliptic Problems in Polyhedra
2013-40	S. Mishra and F. Fuchs and A. McMurry and N.H. Risebro Explicit and implicit finite volume schemes for radiation mhd and the effects of radiation on wave propagation in stratified atmospheres.
2013-41	J. Ernest and P. LeFloch and S. Mishra Schemes with Well controlled Dissipation (WCD) I: Non-classical shock waves
2013-42	S. Mishra and N. Risebro and F. Weber Convergence rates of finite difference schemes for the wave equation with rough coeffiicients
2013-43	U. Koley and S. Mishra and N. Risebro and F. Weber Robust finite difference schemes for a nonlinear variational wave equation modeling liquid crystals
2013-44	G. Coclite and S. Mishra and N. Risebro and F. Weber Analysis and Numerical approximation of Brinkman regularization of two-phase flows in porous media
2013-45	M. Hutzenthaler and A. Jentzen Numerical approximations of stochastic differential equations with non-globally Lipschitz continuous coefficients
2013-46	G. Da Prato and A. Jentzen and M. Röckner A mild Itô formula for SPDEs
2013-47	P. Grohs and S. Keiper and G. Kutyniok and M. Schaefer Parabolic Molecules: Curvelets, Shearlets, and Beyond
2013-48	R. Hiptmair and C. Jerez-Hanckes and C. Urzua Optimal Operator Preconditioning for Boundary Elements on Open Curves