Subgradient Method for Convex Feasibility on Riemannian Manifolds

Glaydston C. Bento* Jefferson G. Melo †

March 16, 2011

Abstract

In this paper we present a subgradient type algorithm for solving convex feasibility problem on Riemannian manifold. We prove that the sequence generated by the algorithm converges to a solution of the problem when the sectional curvature of the manifold is non negative. Moreover, assuming a Slater type qualification condition we propose a variant of the first algorithm which ensures finite convergence property, i.e., a feasible point is obtained after a finite number of iteration. We show some examples motivating the application of the algorithm for feasibility problems not necessarily convex (in the usual sense).

Key words: Feasibility problem, Subgradient algorithm, Riemannian manifolds.

1 Introduction

Consider m closed convex subsets $C_1, \ldots, C_m$ of a metric vector space X, such that

$$C_i = \{x \in X : f_i(x) \leq 0\},$$

*IME-Universidade Federal de Goiás, Goiânia-GO (Email: glaydston@mat.ufg.br)

†Email: jefferson@mat.ufg.br

where $f_i : X \to \mathbb{R}$ is a convex function for each $i = 1, \ldots, m$. Note that if some set C_i is given in any other form then it can be represented as above by choosing $f_i = d^2(., C_i)$, the squared distance to the set C_i. The convex feasibility problem is:

$$\text{find a point } x^* \in C = \bigcap_{i=1}^{m} C_i.$$

This problem is fundamental in many areas of application such as image recovery, radiation therapy treatment planning and crystallography. It is studied in many different context, see for instance [1, 2, 3, 4, 5]. Usual approaches for solving convex feasibility problems are projection methods, subgradient type methods and their variants, see e.g., [1, 2] for excellent disscussions on the subject. Subgradient type methods consist of a simple and efficient strategy for solving such a problem. One of its variant is as follows:

Algorithm

Step 1: Take $x^0 \in \mathbb{R}^n$ arbitrary;

Step 2: Guntitled folder 2iven x^k and $i_k \in \{1, 2, \ldots, m\}$ such that $i_k = (k \mod m) + 1$, calculate the next iterate x^{k+1} by

$$x^{k+1} = \begin{cases} x^k - \alpha_k \frac{f_{i_k}(x^k)}{\|s^k\|^2} s^k, & if \quad f_{i_k}(x^k) > 0, \\ x^k, & \text{otherwise}, \end{cases}$$

where $s^k \in \partial f_{i_k}(x^k)$ is a subgradient of f_{i_k} at the point x^k, and the sequence of relaxation parameters, $\{\alpha_k\}$, is such that $\epsilon_1 \leq \alpha_k \leq 2 - \epsilon_2$, for all $k \geq 0$, with $\epsilon_1, \epsilon_2 > 0$ arbitrarily small. See, for example, [2, 6].

Our main interest here is to extend this algorithm for solving convex feasibility problem on Riemannian manifolds. The subgradient method was proposed in the early sixties, see Shor [7] and Polyak [8]. Since then, many extensions and improvements of this method have been studied. It is known that a convexity structure on a Riemannian manifold can be considered, where a set is convex if and only if geodesics connecting any two of its points is entirely inside the set. Convex functions are the ones which composed with geodesics are convex in the usual sense. Therefore it is natural to attempt to extend the subgradient method to the Riemannian context. Such an extension for solving a convex optimization problem on a non compact (connected) complete Riemannian manifold was done by Ferreira and Oliveira [9].

Recently, extension of concepts, techniques and algorithms from Euclidean spaces to Riemannian manifolds with practical and theoretical purposes have been the subject of many research, see [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. The main advantages of such extensions are that non convex problems in the classical sense may become convex through the introduction of an appropriate Riemannian metric (see [19] and Example 4.1 of Section 4), and constrained optimization problems may be seen as unconstrained ones, see [9, 16, 23]. Also, these extensions give rise to interesting theoretical questions.

In this paper we propose and analyze two algorithms to solve the convex feasibility problem on a non compact (connected) complete Riemannian manifold. Assuming that the manifold has a non negative sectional curvature, we prove full convergence of the sequence generated by the first algorithm to a solution of the problem. Moreover, assuming a Slater type qualification condition and considering a variant of the first algorithm, we prove finite convergence of the sequence to a solution of the problem, i.e., a feasible solution is obtained after a finite number of iteration. We present some functions (sets) which are not convex in the usual sense, but a Riemannian metric is considered in such a way that these functions (sets) become convex in the Riemannian context. This is a simple situation where a feasibility problem (not necessarily convex) may be solved by the algorithm proposed here.

This article is organized as follows. In Section 2 we recall some basic results of Riemannian geometry to be used throughout this paper. In Section 3 we recall some results on convex analysis on Riemannian manifolds. In Section 4 we present the convex feasibility problem in the Riemannian context and we propose and analyze an algorithm to solve it. In Section 5 we propose a variant of the first algorithm and we study its convergence property.

2 Preliminaries on Riemannian geometry

In this section, we introduce some properties and notations on Riemannian geometry. These basics facts can be found in any introductory book of Riemannian geometry, for instance [24, 25].

Let M be a n-dimensional connected manifold. We denote by T_xM the n-dimensional *tangent space* of M at x, by $TM = \cup_{x\in M} T_xM$ *tangent bundle* of M and by $\mathcal{X}(M)$ the space of smooth vector fields over M. When M is endowed with a Riemannian metric $\langle \, , \, \rangle$, with corresponding norm denoted by $\|.\|$, then M is a Riemannian manifold. Recall that the metric can be used to define the length of piecewise smooth curves $\gamma : [a, b] \to M$ joining x to y, i.e., such that $\gamma(a) = x$ and $\gamma(b) = y$, by

$$l(\gamma) = \int_a^b \|\gamma'(t)\| dt,$$

and, moreover, by minimizing this length functional over the set of all such curves, we obtain a Riemannian distance $d(x, y)$ which induces the original topology on M. The metric induces a map $f \mapsto grad f \in \mathcal{X}(M)$ which associates to each scalar function smooth over M its gradient via the rule $\langle grad f, X\rangle = df(X),\ X \in \mathcal{X}(M)$. Let ∇ be the Levi-Civita connection associated to $(M, \langle \, , \, \rangle)$. A vector field V along γ is said to be *parallel* if $\nabla_{\gamma'} V = 0$. If γ' itself is parallel we say that γ is a *geodesic*. Because the geodesic equation $\nabla_{\gamma'}\gamma' = 0$ is a second order nonlinear ordinary differential equation, then the geodesic $\gamma = \gamma_v(., x)$ is determined by its position $x \in M$ and velocity v at x. When there is no confusion, the notation γ_v meaning $\gamma_v'(0) = v$ will also be used. It is easy to check that $\|\gamma'\|$ is constant. We say that γ is *normalized* if $\|\gamma'\| = 1$. The restriction of a geodesic to a closed bounded interval is called a *geodesic segment*. A geodesic segment joining x to y in M is said to be *minimal* if its length is equals to $d(x, y)$ and this geodesic is called a *minimizing geodesic*.

A Riemannian manifold is *complete* if and only if the geodesics are defined for any values of $t \in \mathbb{R}$. Hopf-Rinow Theorem ([25, Theorem 1.1, page 84]) asserts that if a Riemannian manifold is complete then any pair of points can be joined by a (not necessarily unique) minimal geodesic segment. Moreover, (M, d) is a complete metric space and bounded and closed subsets are compact.

If R is the *curvature tensor* defined by $R(X, Y) = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$, with

$X, Y, Z \in \mathcal{X}(M)$, where $[X, Y] = YX - XY$, then the *sectional curvature* with respect to X and Y is given by $K(X, Y) = \langle R(X, Y)Y, X\rangle/(||X||^2||Y||^2 - \langle X\, , \, Y\rangle^2)$, where $||X|| = \langle X, X\rangle^{1/2}$.

The following result is an immediate consequence of the known Topogonov Theorem, see [25].

Proposition 2.1. *Let M be a complete Riemannian manifold with sectional curvature $K \geq 0$. If γ_{u_1} and γ_{u_2} are normalized geodesics such that $\gamma_{u_1}(0) = \gamma_{u_2}(0)$, then*

$$d(\gamma_{u_1}(r_1), \gamma_{u_2}(r_2)) \leq ||r_2u_2 - r_1u_1||.$$

From now on, M denotes a non compact (connected) complete Riemannian manifold.

3 Convexity on Riemannian manifolds

In this section we introduce some definitions and notations of convexity on Riemannian manifolds. We also present some properties of the directional derivative and the subdifferential of a convex function, see [26] for more details.

A function $f : M \to \mathbb{R}$ is *convex* if and only if for any geodesic segment $\gamma : [-\delta, \delta] \to M$, $\delta > 0$, the composition $f \circ \gamma : [-\delta, \delta] \to \mathbb{R}$ is convex (in the usual sense). Given $x \in M$, a vector $s \in T_xM$ is said to be a *subgradient* of f at x, if for any geodesic segment $\gamma : [-\delta, \delta] \to \mathbb{R}$, $\delta > 0$, with $\gamma(0) = x$,

$$(f \circ \gamma)(t) \geq f(x) + t\langle s, \gamma'(0)\rangle, \qquad t \in [-\delta, \delta], \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product at T_xM. The set of all subgradients of f at x, $\partial f(x)$, is called the *subdifferential* of f at x.

We mention that a convex function on a Riemannian Manifold is continuous.

In the following, we consider a convex function $f : M \to \mathbb{R}$, a point $x \in M$, a vector $v \in T_xM$ and $\gamma : [-\delta\, , \, \delta] \to M$ the geodesic segment such that $\gamma(0) = x$ and $\gamma'(0) = v$.

Due to the convexity of $f \circ \gamma : [-\delta\, , \, \delta] \to \mathbb{R}$, the function $q_\gamma : (0\, , \, \delta] \to \mathbb{R}$, given by

$$q_\gamma(t) := \frac{f(\gamma(t)) - f(p)}{t}, \tag{2}$$

is non decreasing and bounded near zero. This leads to the following definition:

Definition 3.1. *The directional derivative of f at x in the direction of v is defined by*

$$f'(x,v) := \lim_{t\to 0^+} q_\gamma(t) = \inf_{t>0} q_\gamma(t). \tag{3}$$

Theorem 3.1. *For every $y \in M$, $\partial f(y)$ is non empty, convex and compact. Moreover, the following characterization holds,*

$$\partial f(y) = \left\{ s \in T_yM : f'(y,w) \geq \langle s, w\rangle, \;\; \forall\, w \; \in T_yM \right\}.$$

Proof. See [26, pages 74 and 75]. □

Proposition 3.1. *The functional $f' : TM \to \mathbb{R}$ is upper semicontinuous on TM, i.e., if $\{(x^k, v^k)\}$ converges to $(x,v) \in TM$, then*

$$\limsup_{k\to+\infty} f'(x^k, v^k) \leq f'(x,v).$$

Proof. The proof follows by Proposition 3.1 and Remark 3.3 of [12]. □

Proposition 3.2. *Let $\{x^k\} \subset M$ a bounded sequence. If the sequence $\{v^k\}$ is such that $v^k \in \partial f(x^k)$, for each $k \in \mathbb{N}$, then $\{v^k\}$ is also bounded.*

Proof. Since $\{x^k\}$ is a bounded sequence, it has an accumulation point $\bar{x} \in M$, by Hopf-Rinow Theorem. We may assume, taking a subsequence if necessary, that $\{x^k\}$ converges to $\bar{x}$. Let $U_{\bar{x}}$ be a neighborhood of $\bar{x}$ such that $TU_{\bar{x}} = U_{\bar{x}} \times \mathbb{R}^n$. Since $TU_{\bar{x}} \subset TM$ is an open set and $\{x^k\}$ converges to $\bar{x}$, we may also assume that the whole sequence $\{(x^k, v^k)\}$ is in $TU_{\bar{x}}$. Let us assume, by contradiction, that the sequence $\{v^k\}$ is unbounded. Therefore there exists a subsequence $\{v^{k_j}\}$ such that

$$\lim_{j\to+\infty} \|v^{k_j}\| = +\infty.$$

Considering the sequence $\{w^j\}$ defined by $w^j = v^{k_j}/\|v^{k_j}\|$, for each $j \in \mathbb{N}$, and taking a subsequence if necessary, we may assume that $\{w^j\}$ converges to some $\bar{w} \in \mathbb{R}^n$. Thus, $\{(x^{k_j}, w^j)\}$ converges to $(\bar{x}, \bar{w})$, and by Proposition 3.1 follows that

$$\limsup_{j \to +\infty} f'(x^{k_j}, w^j) \leq f'(\bar{x}, \bar{w}) < +\infty. \tag{4}$$

Since $v^{k_j} \in \partial f(x^{k_j})$, we obtain by Theorem 3.1 that $f'(x^{k_j}, w^j) \geq \langle v^{k_j}, w^j \rangle = \|v^{k_j}\|$. Therefore

$$\limsup_{j \to +\infty} f'(x^{k_j}, w^j) = +\infty,$$

contradicting (4), and the result follows. □

4 Algorithm 1: basic properties and convergence result

In this section we present the convex feasibility problem on Riemannian manifolds. We propose and analyze an algorithm to solve it. We prove some basic results here for the sake of completeness.

The convex feasibility problem CFP in the Riemannian context is:

$$\text{find a point } x^* \in C := \bigcap_{i=1}^{m} C_i,$$

where $C_i = \{x \in M : f_i(x) \leq 0\}$, and $f_i : M \to \mathbb{R}$ is a convex function for each $i = 1, \ldots, m$. We assume that $C \neq \emptyset$.

Next, we present an example illustrating the fact that non convex functions (resp. sets) in the usual sense may become convex functions (resp. sets) in the Riemannian context with the choice of a suitable Riemannian metric. We observe, although, that in general to find such a metric is a difficult task.

Example 4.1. Let $M = \left(\mathbb{R}^2_{++}, g\right)$ be the Riemannian manifold, where $\mathbb{R}^2_{++} = \{(x_1, x_2) \in \mathbb{R}^2 : x_1, x_2 > 0\}$ and $g = (g_{ij})$ is the affine-scaling metric, i.e., $g_{ij}(x) = \frac{\delta_{ij}}{x_i x_j}$. It is well known that M is a complete Riemannian manifold, with sectional curvature $K \equiv 0$ and tangent plane at $x \in M$, denoted by T_xM, equals to $\mathbb{R}^2$. The geodesic curve $\gamma : \mathbb{R} \to M$ satisfying $\gamma(0) = a \in M$ e $\gamma'(0) = s \in T_aM$ is

$$t \mapsto \gamma(t) = \left(a_1\, \mathrm{e}^{(a_1/s_1)t}, a_2\mathrm{e}^{(a_2/s_2)t}\right).$$

Consider the functions $f_1, f_2 : M \to \mathbb{R}$ defined by

$$f_1(x_1, x_2) = \ln x_2 - \ln x_1 \qquad \text{and} \qquad f_2(x_1, x_2) = x_1x_2 - 1,$$

and consider $C_i = \{(x_1, x_2) \in \mathbb{R}^2_{++} : f_i(x_1, x_2) \le 0\}$, $i = 1, 2$. Note that $C = C_1 \cap C_2 \neq \emptyset$, and it is a non convex set in the usual sense, i.e., in the case that $\mathbb{R}^2_{++}$ is endowed with the Euclidean metric. However, since $f_i \circ \gamma$ is convex on $\mathbb{R}$, $i = 1, 2$, it follows that f_1 and f_2 are convex functions on M and consequently C is a convex set on M. It is easy to see that the functions $g_{1,p}(x_1, x_2) = x_1^{1/p} + \ln x_2$, $g_{2,p}(x_1, x_2) = x_1^p\sqrt{x_2}$ and $g_{3,p}(x_1, x_2) = (x_1x_2)^{\frac{1}{p}}$, $p = 1, 2, \ldots$, are convex in M, but they are non convex functions in the usual sense. Consequently, sublevel sets of these functions are convex on M, but are non convex in the usual sense.

Next we present an algorithm to solve CFP.

Algorithm 1

Step 1: Take $x^0 \in M$ arbitrary. Set $k = 0$.

Step 2: Let $i_k \in \{1, 2, \ldots, m\}$ such that $i_k = (k \mod m) + 1$. If $f_{i_k}(x^k) \le 0$ then set $x^{k+1} = x^k$. Otherwise, take $s^k \in \partial f_{i_k}(x^k)$, calculate the geodesic γ_{v^k} satisfying $\gamma_{v^k}(0) = x^k$ and $\gamma'_{v^k}(0) = v^k$, with $v^k = -s^k/\|s^k\|$. Take $t_k = \delta_k \frac{f_{i_k}(x^k)}{\|s^k\|}$ and define

$$x^{k+1} = \gamma_{v^k}(t_k),$$

where $\delta \le \delta_k \le 2 - \delta$ and $0 < \delta < 2$.

Note that if $s^k = 0$ then x^k is a minimizer of f_{i_k} and, therefore, $f_{i_k}(x^k) \le 0$ and $x^{k+1} = x^k$. Hence Algorithm 1 is well defined.

Let M be the Riemannian manifold of Example 4.1 and $h : \mathbb{R}^2_{++} \to \mathbb{R}$ a differentiable function. The Riemannian gradient of h is given by

$$\operatorname{grad} h = g^{-1}\nabla h. \tag{5}$$

So, if f_1 and f_2 are the functions considered in Example 4.1, then

$$\text{grad } f_1(x) = \begin{pmatrix} -x_1 \\ x_2 \end{pmatrix} \qquad \text{and} \qquad \text{grad } f_2(x) = \begin{pmatrix} x_2 x_1^2 \\ x_1 x_2^2 \end{pmatrix}.$$

In this case, if at some iteration k, $f_{i_k}(x^k) > 0$, then the iterate x^{k+1} of Algorithm 1 is given by

$$x^{k+1} = \gamma_{v^k}(t_k) = \left(x_1^k \, \mathrm{e}^{(x_1^k / v_1^k) t_k}, x_2^k \, \mathrm{e}^{(x_2^k / v_2^k) t_k} \right),$$

where $v^k = -\text{grad } f_1(x^k)$ if $i_k = 1$, $v^k = -\text{grad } f_2(x^k)$ if $i_k = 2$, and t_k is as in Step 2 of Algorithm 1.

It is immediate to see that, Algorithm 1 stops at some iteration k if and only if x^k is a feasible point. Therefore, we assume that it generates infinite sequences $\{x^k\}$, $\{s^k\}$, $\{v^k\}$ and $\{t_k\}$. Moreover, from now on we assume that the sectional curvature K of the manifold M is non negative.

Lemma 4.1. *Let $\{x^k\}$, $\{s^k\}$, $\{v^k\}$ and $\{t_k\}$ be the sequences generated by Algorithm 1. For every $y \in M$ and $k \in \{0, 1, 2, \ldots\}$ such that $f_{i_k}(x^k) > 0$, the following inequality holds*

$$d^2(x^{k+1}, y) \leq d^2(x^k, y) + t_k^2 + 2\frac{t_k}{||s^k||}(f_{i_k}(y) - f_{i_k}(x^k)). \tag{6}$$

Proof. Take $y \in M$ and let γ_{u_1} be the normalized minimizing geodesic such that $\gamma_{u_1}(0) = x^k$ and $\gamma_{u_1}(r_1) = y$ where $r_1 = d(x^k, y)$. Moreover, take $u_2 := v^k = -\frac{s^k}{\|s^k\|}$ and let γ_{u_2} be the geodesic such that $\gamma_{u_2}(0) = x^k$ and $\gamma'_{u_2}(0) = u_2$. Therefore $\gamma_{u_2}(t_k) = x^{k+1}$ and we obtain from Proposition 2.1 that

$$d^2(x^{k+1}, y) \leq \| - t_k \frac{s^k}{||s^k||} - r_1 u_1 \|^2 = t_k^2 + d^2(x^k, y) + 2\frac{t_k r_1}{||s^k||}\langle s^k, u_1 \rangle. \tag{7}$$

On the other hand, since $s^k \in \partial f_{i_k}(x^k)$, it follows from the subgradient inequality (1) that

$$r_1 \langle s^k, u_1 \rangle \leq f_{i_k}(y) - f_{i_k}(x^k).$$

Therefore, the result follows from the last inequality combined with (7). □

Remark 4.1. Inequality (6) is well known when $M = \mathbb{R}^n$. It was generalized to Riemannian manifolds of non negative sectional curvature in [23], in the particular case when f_{i_k} is differentiable.

Corollary 4.1. *For every $z \in C$ and $k \in \{0, 1, 2, \ldots\}$ such that $f_{i_k}(x^k) > 0$,*

$$d(x^{k+1}, z) < d(x^k, z).$$

Proof. For each $z \in C$, it follows from Lemma 4.1 that

$$d^2(x^{k+1}, z) \leq d^2(x^k, z) + t_k^2 + 2\frac{t_k}{||s^k||}(-f_{i_k}(x^k)), \qquad k = 0, 1, \ldots.$$

Using the definition of t_k, it follows that

$$d^2(x^{k+1}, z) \leq d^2(x^k, z) + \delta_k(\delta_k - 2)\frac{f_{i_k}^2(x^k)}{||s^k||^2}, \qquad k = 0, 1, \ldots. \tag{8}$$

The result follows from (8) observing that $\delta \leq \delta_k \leq 2 - \delta$ and $0 < \delta < 2$.

□

Definition 4.1. A sequence $\{y_k\}$ in the complete metric space (M, d) is said to be *Fejér convergent* to a set $W \subset M$ iff for every $w \in W$,

$$d(y_{k+1}, w) \leq d(y_k, w) \qquad k = 0, 1, \ldots.$$

Proposition 4.1. *Let $\{y_k\}$ be a sequence in the complete metric space (M, d). If $\{y_k\}$ is Fejér convergent to a non empty set $W \subset M$, then $\{y_k\}$ is bounded. If, furthermore, an accumulation point y of $\{y_k\}$ belongs to W, then $\lim_{k\to\infty} y_k = y$.*

Proof. See for example [16]. □

Next we present our main convergence result.

Theorem 4.1. *The sequence $\{x^k\}$ generated by Algorithm 1 converges to a solution of the feasibility problem.*

Proof. Take $z \in C$. From Corollary 4.1, it follows that $d(x^k, z) \leq d(x^0, z)$ for any $k \geq 1$ and consequently the sequence $\{x^k\}$ is bounded. Thus, from Hopf-Rinow Theorem, there exists a subsequence $\{x^{k_j}\}$ converging to some point $\bar{x}$. Since $i_{k_j} := k_j (\bmod m) + 1 \in \{1, 2, \ldots, m\}$ for all j, there exists at least one index i and an infinite set of indeces $J \subset \mathbb{N}$ such that $i_{k_j} = i$ for all $j \in J$. Without loss of generality we assume that $i = 1$. Let us show that $f_1(\bar{x}) \leq 0$. Observe that if $f_1(x^{k_j}) \leq 0$ infinitely often for $j \in J$, then $f_1(\bar{x}) \leq 0$ because $\{x^{k_j}\}$ converges to $\bar{x}$ and f_1 is continuous. Assuming now that there exists j_0 sufficiently large such that $f_1(x^{k_j}) > 0$ for all $j \geq j_0,\ j \in J$. It follows from (8) that

$$d^2(x^{k_j+1}, z) \leq d^2(x^{k_j}, z) + \delta_k(\delta_k - 2)\frac{f_1^2(x^{k_j})}{||s^{k_j}||^2}, \qquad j \geq j_0,\ j \in J,$$

with $0 < \delta \leq \delta_k \leq 2 - \delta$. Thus,

$$d^2(x^{k_j+1}, z) \leq d^2(x^{k_j}, z) - \delta^2\frac{f_1^2(x^{k_j})}{||s^{k_j}||^2}, \qquad j \geq j_0,\ j \in J.$$

Since $\{x^k\}$ is bounded, it follows by Proposition 3.2 that $\{s^k\}$ is also bounded. Let $L > 0$ such that $\|s^k\| \leq L$, $k = 0, 1, \ldots$. Therefore,

$$d^2(x^{k_j+1}, z) \leq d^2(x^{k_j}, z) - \delta^2\frac{f_1^2(x^{k_j})}{L^2}, \qquad j \geq j_0,\ j \in J. \tag{9}$$

On the other hand, Corollary 4.1 implies that $\{d(x^k, z)\}$ is convergent. Thus, we obtain

$$f_1^2(\bar{x}) = \lim_{j\in J, j\to\infty} f_1^2(x^{k_j}) \leq \frac{L^2}{\delta^2}\lim_{j\in J, j\to\infty}\big(d^2(x^{k_j}, z) - d^2(x^{k_j+1}, z)\big) = 0,$$

proving that the statement is true, that is, $f_1(\bar{x}) \leq 0$. Similarly we prove that $f_2(\bar{x}) \leq 0$. In order to do that we show first that there exists a subset J_1 of J with infinite indeces such that $\{x^{k_j+1}\}_{j\in J_1}$ converges to $\bar{x}$. Indeed, if $f_1(x^{k_j}) \leq 0$ infinitely often for $j \in J$, then $x^{k_j+1} = x^{k_j}$ infinitely often for $j \in J$, by definition of Algorithm 1 and using that $i_{k_j} = 1$. In this case the statement is trivially true, because $\{x^{k_j}\}_{j\in J}$ converges to $\bar{x}$. We may assume, without loss of generality, that $f_1(x^{k_j}) > 0$ for all $j \in J$. Therefore, by Lemma 4.1, we obtain that

$$d^2(x^{k_j+1}, \bar{x}) \leq d^2(x^{k_j}, \bar{x}) + t_{k_j}^2 + 2\frac{t_{k_j}}{||s^{k_j}||}\big(-f_1(x^{k_j})\big), \qquad j \in J,$$

which, combined with definition of t_{k_j}, implies

$$d^2(x^{k_j+1}, \bar{x}) \leq d^2(x^{k_j}, \bar{x}) + \delta_k(\delta_k - 2)\frac{f_1^2(x^{k_j})}{\|s^{k_j}\|^2}, \qquad j \in J.$$

Now, since $0 < \delta_k < 2$, it follows that $d(x^{k_j+1}, \bar{x}) < d(x^{k_j}, \bar{x})$, $j \in J$. Thus, as $\{x^{k_j}\}_{j\in J}$ converges to $\bar{x}$, $\{x^{k_j+1}\}_{j\in J}$ also converges to $\bar{x}$. It is easy to see that

$$i_{(k_j+1)} := (k_j + 1)(\text{mod m}) + 1 = i_{k_j} \text{mod m} + 1.$$

Since we are assuming $i_{k_j} = i = 1$, $j \in J$, it follows that $i_{(k_j+1)} = 2$, $j \in J$. Therefore we may repeat a similar argument as before to prove that $f_2(\bar{x}) \leq 0$. It is easy to see that this argument can be repeated in order to conclude that $f_i(\bar{x}) \leq 0$ for $i = 3, \cdots, m$ and consequently that the point $\bar{x}$ satisfies $f_i(\bar{x}) \leq 0$ for $i = 1, 2, \cdots, m$, that is, $\bar{x}$ belongs to C. Since from Corollary 4.1, $\{x^k\}$ is Fejér convergent to C, we conclude that the full sequence $\{x^k\}$ converges to an element of C, by Proposition 4.1. □

5 Algorithm 2 and convergence result

In this section we propose and analyze a variant of Algorithm 1. We show that, under a Slater type qualification condition, the algorithm has finite convergence property, that is, a feasible point is obtained after a finite number of iterations.

H1. (Slater's qualification) *There exists $\hat{x} \in M$ satisfying $f_i(\hat{x}) < 0$, $i = 1, \ldots, m$.*

In order to state the algorithm we need, for the step size rule, estimates f_i^* such that

$$f_i(\hat{x}) \leq f_i^* < 0, \qquad i = 1, \ldots, m. \tag{10}$$

Remark 5.1. We mention that to obtain estimates f_i^* satisfying (10) is a difficult task. In order to apply this algorithm, a possible choice for these estimates is to consider $f_i^* = -\eta$, where $\eta > 0$ is a small parameter, and to use some measure of feasibility, for example check if $\sum_{i=1}^m \max\{f_i(x^k), 0\}$ is small. If after some iterations of the algorithm, the measure of feasibility is increasing substancially, then a small parameter η may be required. However, we emphasize that we are interested here in theoretical questions.

Algorithm 2

Step 1: Take $x^0 \in M$ arbitrary. Set $k = 0$.

Step 2: Let $i_k \in \{1, 2, \ldots, m\}$ such that $i_k = (k \mod m) + 1$. If $f_{i_k}(x^k) \leq 0$ then set $x^{k+1} = x^k$. Otherwise, take $s^k \in \partial f_{i_k}(x^k)$, calculate the geodesic γ_{v^k} with $\gamma_{v^k}(0) = x^k$, $\gamma'_{v^k}(0) = v^k := -s^k/\|s^k\|$ and let

$$x^{k+1} = \gamma_{v^k}(t_k),$$

where $t_k := \delta_k \frac{f_{i_k}(x^k) - f^*_{i_k}}{\|s^k\|}$ is the step size rule, with $\delta \leq \delta_k \leq 2 - \delta$, $0 < \delta < 2$.

Note that the well definition of Algorithm 2 is similar to the one of Algorithm 1 and x^k is feasible if and only if the algorithm stops.

Lemma 5.1. *Let x^k be a non feasible point generated by Algorithm 2, and f_i^* as in* (10)*, $i = 1, \ldots, m$. Then, for all $y \in S := \{y \in M : f_i(y) \leq f_i^*,\ i = 1, \ldots, m\}$, it holds that*

$$d(x^{k+1}, y) \leq d(x^k, y). \tag{11}$$

Proof. Note that if $f_{i_k}(x^k) \leq 0$ then, by Step 2 of Algorithm 2, $x^{k+1} = x^k$ and (11) trivially holds. Let us assume that $f_{i_k}(x^k) > 0$. Similarly to the proof of Lemma 4.1, for all $z \in M$ we obtain that

$$d^2(x^{k+1}, z) \leq d^2(x^k, z) + t_k^2 + 2\frac{t_k}{||s^k||}(f_{i_k}(z) - f_{i_k}(x^k)). \tag{12}$$

Let $y \in S$. Taking $z := y$ in (12) and using the step size rule of Algorithm 2, we obtain

$$d^2(x^{k+1}, y) \leq d^2(x^k, y) + \delta_k \frac{f_{i_k}(x^k) - f^*_{i_k}}{\|s^k\|^2}\left(\delta_k(f_{i_k}(x^k) - f^*_{i_k}) + 2(f_{i_k}(y) - f_{i_k}(x^k)\right). \tag{13}$$

Using that $f_{i_k}(y) \leq f^*_{i_k}$ and $0 < \delta < \delta_k < 2 - \delta$, (13) yields

$$d^2(x^{k+1}, y) \leq d^2(x^k, y) - \delta^2 \frac{(f_{i_k}(x^k) - f^*_{i_k})^2}{||s^k||^2}, \tag{14}$$

and the result follows. □

Theorem 5.1. *The sequence $\{x^k\}$ generated by Algorithm 2 has finite convergence.*

Proof. Suppose by contradiction that Algorithm 2 generates an infinite sequence $\{x^k\}$. Taking $\hat{x}$ satisfying (10), Lemma 5.1 implies that $\{d(x^k, \hat{x})\}$ is convergent and $\{x^k\}$ is bounded, which in turn implies that $\{s^k\}$ is also bounded, by Proposition 3.2. Let $\{x^{k_j}\}$ be a subsequence converging to some $\bar{x}$, and $L > 0$ such that $\|s^k\| \leq L$, $k = 0, 1, \ldots$. Assuming that $f_{i_{k_j}}(x^{k_j}) > 0$ for sufficiently large j, and taking $y = \hat{x}$ in (14), we obtain

$$\big(f_{i_{k_j}}(x^{k_j}) - f^*_{i_{k_j}})\big)^2 \leq \frac{L^2}{\delta^2}\big(d^2(x^{k_j}, \hat{x}) - d^2(x^{k_j+1}, \hat{x})\big).$$

Since the sequence $\{d(x^k, \hat{x})\}$ is convergent, it follows from the last inequality that

$$\lim_{j\to\infty}(f_{i_{k_j}}(x^{k_j}) - f^*_{i_{k_j}}) = 0. \tag{15}$$

Observing that $i_{k_j} \in \{1, 2, \ldots, m\}$ for all j, there exists at least one index $l \in \{1, \ldots, m\}$ such that $i_{k_j} = l$ for infinite indices j. Considering a subsequence if necessary, we may assume, without loss of generality, that $i_{k_j} = l$, for all j. From (15) follows that $\lim\limits_{j\to\infty} f_l(x^{k_j}) = f^*_l$. Since $\{x^{k_j}\}$ converges to $\bar{x}$ and f_l is continuous, we have $f_l(\bar{x}) = f^*_l < 0$. Now, as $\{x^{k_j}\}$ converges to $\bar{x}$ and $f_l(\bar{x}) < 0$, we obtain that $f_l(x^{k_j}) < 0$ for all $j \geq j_0$ and some j_0. In particular, as $l = i_{k_j}$, for all j, it follows from Step 2 of Algorithm 2 that $x^{k_j+1} = x^{k_j}$ for all $j \geq j_0$. Therefore $\{x^{k_j+1}\}$ also converges to $\bar{x}$. On the other hand if $f_{i_{k_j}}(x^{k_j}) \leq 0$ infinitely often then $x^{k_j+1} = x^{k_j}$ infinitely often, by definition of Algorithm 2. Therefore, in any case, using a subsequence if necessary, we may assume without loss of generality that $\{x^{k_j+1}\}$ converges to $\bar{x}$. Similarly to previous arguments, observing that $i_{k_j+1} = l + 1$ and if l=m then $l + 1 := 1$, we prove that $\lim_{j\to\infty} f_{l+1}(x^{k_j+1}) = f_{l+1}(\bar{x}) = f^*_{l+1} < 0$. Analogously we may prove that $f_i(\bar{x}) = f^*_i < 0$, $i = 1, \ldots, m$, which implies that $\bar{x} \in S$. Combining Lemma 5.1 with Proposition 4.1 we obtain that $\{x^k\}$ converges to $\bar{x}$. Thereby for sufficiently large k, it follows that $f_i(x^k) < 0$ for all $i = 1, 2, \ldots, m$. Thus, Algorithm 2 stops at x^k, which is a contradiction with the assumption on the sequence $\{x^k\}$, concluding the proof. □

References

[1] Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems, SIAM Rev. **38**, 367-426 (1996)

[2] Butnariu, D., Censor, Y., Gurfil, P., Hadar, E.: On the behavior of subgradient projections methods for convex feasibility problems in Euclidean spaces, SIAM J. Optim. **19**, 786-807 (2008)

[3] Butnariu, D., Iusem, A., Burachik, R.: Iterative Methods of Solving Stochastic Convex Feasibility Problems and Applications, Comput. Optim. Appl. **15**, 269-307 (2000)

[4] Combettes, P.L.: The convex feasibility problem in image recovery, Adv. Imag. Elect. Phys. **95**, 155-270 (1996)

[5] O'Hara, J.G., Pillay, P., Xu, H.K.: Iterative approaches to convex feasibility problems in Banach spaces, Nonlinear Anal. **64**, 2022-2042 (2006)

[6] Censor, Y., Lent, A.: Cyclic subgradient projections, Math. Program. **24**, 233-235 (1982)

[7] Shor, N.Z.: Minimization Algorithms for Non-Differentiable Function, Springer Verlag, Berlin, Germany, (1985)

[8] Polyak, B.T.: Minimization of unsmooth functional, USSR Comput. Math. and Math. Phis. **9**, 14-29 (1969)

[9] Ferreira, O.P., Oliveira, P.R.: Subgradient algorithm on Riemannian manifolds, J. Optim. Theory Appl. **97**, 93-104 (1998)

[10] Alvarez, F., Bolte, J., Munier, J.: A Unifying Local Convergence Result for Newton's Method in Riemannian Manifolds, Found. Comput. Math. **8**, 197-226 (2008)

[11] Baker, C.G., Absil, P.-A., Gallivan, K.A.: An Implicit Trust-region Method on Riemannian manifolds, IMA J. Numer. Anal.**28**, 665-689 (2008)

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

[13] Li, C., López, G., Martín-Márquez, V.: Monotone vector fields and the proximal point algorithm on Hadamard manifolds, J. London Math. Soc. **79**, 663-683 (2009)

[14] Ferreira, O.P.: Proximal subgradient and a characterization of Lipschitz function on Riemannian manifolds, J. Math. Anal. Appl. **313**, 587-597 (2006)

[15] Ferreira, O.P.: Dini Derivative and a Characterization for Lipschitz and Convex Functions on Riemannian Manifolds, Nonlinear Anal. **68**, 1517-1528 (2008)

[16] Ferreira, O.P., Oliveira, P.R.: Proximal point algorithm on Riemannian manifold, Optimization. **51**, 257-270 (2002)

[17] Li, S.L., Li, C., Liou, Y.C., Yao, J.C.: Existence of solutions for variational inequalities on Riemannian manifolds, Nonlinear Anal. **71**, 5695-5706 (2009)

[18] Németh, S.Z.: Variational inequalities on Hadamard manifolds, Nonlinear Anal. **52**, 1491-1498 (2003)

[19] da Cruz Neto, J. X., Ferreira, O.P., Lucâmbio Pérez, L.R., Németh, S.Z.: Convex-and Monotone-Transformable Mathematical Programming Problems and a Proximal-Like Point Algorithm, J. Global Optim. **35**, 53-69 (2006)

[20] Ledyaev, Yu. S., Zhu, Q.J.: Nonsmooth analysis on smooth manifolds, Trans. Amer. Math. Soc. **359**, 3687-3732 (2007)

[21] Wang, J.H., Huang, S.C., Li, C.: Extended Newton's Algorithm for mappings on Riemannian manifolds with values in a cone, Taiwanese J. Math. **13**, 633-656 (2009)

[22] Wang, J.H., Lopez, G., Martin-Marquez, V., Li, C.: Monotone and accretive vector fields on Riemannian manifolds. J. Optim. Theory Appl. **146**, 691-708 (2010)

[23] da Cruz Neto, J. X., de Lima, L.L., Oliveira, P.R.: Geodesic algorithms in Riemannian geometry, Balkan J. Geom. Appl. **3**, 89-100 (1998)

[24] do Carmo, M.P.: Riemannian Geometry, Boston, Birkhauser, (1992)

[25] Sakai, T.: Riemannian geometry. Translations of mathematical monographs, **149**, Amer. Math. Soc., Providence, R.I., (1996)

[26] Udriste, C.: Convex functions and optimization Algorithms on Riemannian manifolds,in: Mathematics and its Applications, **297**, Kluwer Academic Publishers, (1994)