Nonlinear Analysis 73 (2010) 564–572

Contents lists available at ScienceDirect

Nonlinear Analysis

journal homepage: www.elsevier.com/locate/na

Local convergence of the proximal point method for a special class of nonconvex functions on Hadamard manifolds

G.C. Bento [a,*], O.P. Ferreira [a], P.R. Oliveira [b]

[a] IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, Brazil
[b] COPPE-Sistemas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21945-970, Brazil

ARTICLE INFO

Article history:
Received 12 December 2008
Accepted 31 March 2010

MSC:
49M30
90C26

Keywords:
Proximal point method
Nonconvex functions
Hadamard manifolds

ABSTRACT

Local convergence analysis of the proximal point method for a special class of nonconvex functions on Hadamard manifold is presented in this paper. The well definedness of the sequence generated by the proximal point method is guaranteed. Moreover, it is proved that each cluster point of this sequence satisfies the necessary optimality conditions and, under additional assumptions, its convergence for a minimizer is obtained.

1. Introduction

The extension of the concepts and techniques of the Mathematical Programming of the Euclidean space R^n to Riemannian manifolds is natural. It has been frequently done in recent years, with a theoretical purpose and also to obtain effective algorithms; see [1–9]. In particular, we observe that, these extensions allow the solving of some nonconvex constrained problems in Euclidean space. More precisely, nonconvex problems in the classic sense may become convex with the introduction of an adequate Riemannian metric on the manifold (see, for example [10]). The proximal point algorithm, introduced by Martinet [11] and Rockafellar [12], has been extended to different contexts; see [4,6] and the references therein. In [4], the authors generalized the proximal point method for solving convex optimization problems of the form

$$(P) \quad \begin{array}{l} \min f(p) \\ \text{s.t. } p \in M, \end{array} \tag{1}$$

where M is a Hadamard manifold and $f : M \to \mathbb{R}$ is a convex function (in the Riemannian sense). The method was described as follows:

$$p^{k+1} := \operatorname*{argmin}_{p \in M} \left\{ f(p) + \frac{\lambda_k}{2} d^2(p, p^k) \right\}, \tag{2}$$

with $p^\circ \in M$ an arbitrary point, d the intrinsic Riemannian distance (to be defined later on) and $\{\lambda_k\}$ a sequence of positive numbers. The authors also showed that this extension is natural. As regards to [6] the authors generalized the proximal point method with Bregman distance to solve quasiconvex and convex optimization problems also on Hadamard manifold.

* Corresponding author. Tel.: +55 62 3521 1418; fax: +55 62 3521 1180.
E-mail addresses: glaydston@mat.ufg.br (G.C. Bento), orizon@mat.ufg.br (O.P. Ferreira), poliveir@cos.ufrj.br (P.R. Oliveira).

0362-546X/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.na.2010.03.057

Spingarn in [13] has, in particular, developed the proximal point method for the minimization of a certain class of non-differentiable nonconvex functions, namely, the lower-C^2 functions defined in Euclidean spaces; see also [14]. Kaplan and Tichatschke in [15] also applied the proximal point method for the minimization of a similar class of the ones of [14,13], namely, functions defined as maximum of a certain collection (finite/infinite) of continuously differentiable functions. In [16] we study, in the Riemannian context, the same class of functions studied in [15]. In that context we applied the proximal point method (2) to solve the problem (1), however we assumed that the collection of functions defining the objective function was finite.

Our goal is to extend the results of [16]. We consider that the objective function is given by the maximum of a collection infinite of continuously differentiable functions. To obtain the results in [16], it was necessary to study the generalized directional derivative in the Riemannian manifolds context. In this paper we go further in the study of properties of the generalized directional derivative in order to analyze the convergence of the proximal point method. Several works have studied such concepts and presented many useful results in the Riemannian optimization context; see for example [17,5,18,19].

The paper is divided as follows. In Section 2 we give the notation and some results on the Riemannian geometry which we will use along the paper. In Section 3 we recall some facts of the convex analysis on Hadamard manifolds. In Section 4 we present definition of generalized directional derivative of a locally Lipschitz function (not necessarily convex) which, in the Euclidean case, coincides with the Clarke generalized directional derivative. Moreover, some properties of that derivative are presented, amongst which the upper semicontinuity of the directional derivative. In Section 5 we study the proximal point method (2) to solve the problem (1), in the case where the objective function is a real-valued function (non-necessarily convex) on a Hadamard manifold M given by the maximum of a certain class of functions. Finally in Section 6 we provide an example where the proximal point method for nonconvex problems is applied.

2. Notation and terminology

In this section we introduce some fundamental properties and notations on Riemannian geometry. These basic facts can be found in any introductory book on Riemannian geometry, such as in [20,21].

Let M be an n-dimensional connected manifold. We denote by T_pM the n-dimensional *tangent space* of M at p, by $TM = \cup_{p\in M} T_pM$ *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 the corresponding norm denoted by $\|\ \|$, then M is now a Riemannian manifold. Recall that the metric can be used to define the length of piecewise smooth curves $\gamma : [a, b] \to M$ joining p to q, i.e., such that $\gamma(a) = p$ and $\gamma(b) = q$, by

$$l(\gamma) = \int_a^b \|\gamma'(t)\|\mathrm{d}t,$$

and, moreover, by minimizing this length functional over the set of all such curves, we obtain a Riemannian distance $d(p, q)$ which induces the original topology on M. The metric induces a map $f \mapsto \operatorname{grad} f \in \mathcal{X}(M)$ which associates to each smooth function on M its gradient via the rule $\langle \operatorname{grad} f, X\rangle = df(X),\ X \in \mathcal{X}(M)$. 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*. Given that geodesic equation $\nabla_{\gamma'}\gamma' = 0$ is a second order nonlinear ordinary differential equation, then geodesic $\gamma = \gamma_v(., p)$ is determined by its position p and velocity v at p. 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 p to q in M is said to be *minimal* if its length equals $d(p, q)$ and this geodesic is called a *minimizing geodesic*. If γ is a curve joining points p and q in M then, for each $t \in [a, b]$, ∇ induces a linear isometry, relative to $\langle\, ,\, \rangle$, $P_{\gamma(a)\gamma(t)} : T_{\gamma(a)}M \to T_{\gamma(t)}M$, the so-called *parallel transport* along γ from $\gamma(a)$ to $\gamma(t)$. The inverse map of $P_{\gamma(a)\gamma(t)}$ is denoted by $P^{-1}_{\gamma(a)\gamma(t)} : T_{\gamma(t)}M \to T_{\gamma(a)}M$. In the particular case of γ is the unique curve joining points p and q in M then parallel transport along γ from p to q is denoted by $P_{pq} : T_pM \to T_qM$.

A Riemannian manifold is *complete* if geodesics are defined for any values of t. Hopf–Rinow's theorem asserts that if this is the case then any pair of points, say p and q, in M 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. Take $p \in M$. The *exponential map* $\exp_p : T_pM \to M$ is defined by $\exp_p v = \gamma_v(1, p)$.

We denote by R *the curvature tensor* defined by $R(X, Y) = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[Y,X]}Z$, 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 $K(X, Y) = \langle R(X, Y)Y, X\rangle/(\|X\|^2\|X\|^2 - \langle X,\ Y\rangle^2)$, where $\|X\| = \langle X, X\rangle^{1/2}$. If $K(X, Y) \leqslant 0$ for all X and Y, then M is called a *Riemannian manifold of nonpositive curvature* and we use the short notation $K \leqslant 0$.

Theorem 2.1. *Let M be a complete, simply connected Riemannian manifold with nonpositive sectional curvature. Then M is diffeomorphic to the Euclidean space $\mathbb{R}^n$, $n = \dim M$. More precisely, at any point $p \in M$, the exponential mapping $\exp_p : T_pM \to M$ is a diffeomorphism.*

Proof. See [20,21]. □

A complete simply connected Riemannian manifold of nonpositive sectional curvature is called a *Hadamard manifold*. The Theorem 2.1 says that if M is Hadamard manifold, then M has the same topology and differential structure of the Euclidean space $\mathbb{R}^n$. Furthermore, are known some similar geometrical properties of the Euclidean space $\mathbb{R}^n$, such as, given two points there exists a unique geodesic that joins them. *In this paper, all manifolds M are assumed to be Hadamard finite dimensional.*

3. Convexity in Hadamard manifold

In this section, we introduce some fundamental properties and notations of convex analysis on Hadamard manifolds which will be used later. We will see that these properties are similar to those obtained in convex analysis on the Euclidean space $\mathbb{R}^n$. References to convex analysis on Euclidean space $\mathbb{R}^n$ are in [22], and on Riemannian manifold are in [23,4,7,21,8,9].

The set $\Omega \subset M$ is said to be *convex* if any geodesic segment with end points in Ω is contained in Ω. Let $\Omega \subset M$ be an open convex set. A function $f : M \to \mathbb{R}$ is said to be *convex* (respectively, *strictly convex*) on Ω if for any geodesic segment $\gamma : [a, b] \to \Omega$ the composition $f \circ \gamma : [a, b] \to \mathbb{R}$ is convex (respectively, strictly convex). Now, a function $f : M \to \mathbb{R}$ is said to be *strongly convex* on Ω with constant $L > 0$ if, for any geodesic segment $\gamma : [a, b] \to \Omega$, the composition $f \circ \gamma : [a, b] \to \mathbb{R}$ is strongly convex with constant $L\|\gamma'(0)\|^2$. Take $p \in M$. A vector $s \in T_pM$ is said to be a *subgradient* of f at p if

$$f(q) \geq f(p) + \langle s, \exp_p^{-1} q\rangle,$$

for any $q \in M$. The set of all subgradients of f at p, denoted by $\partial f(p)$, is called the *subdifferential* of f at p.

Take $p \in M$. Let $\exp_p^{-1} : M \to T_pM$ be the inverse of the exponential map which is also C^∞. Note that $d(q, p) = \|\exp_p^{-1} q\|$, the map $d^2(., p)$: $M \to \mathbb{R}$ is C^∞ and

$$\operatorname{grad} \frac{1}{2} d^2(q, p) = -\exp_q^{-1} p,$$

(remember that M is a Hadamard manifold); see, for example, [21].

Proposition 3.1. *Take $p \in M$. The map $d^2(., p)/2$ is strongly convex.*

Proof. See [23]. □

Definition 3.1. Let $\Omega \subset M$ be an open convex set. A function $f : M \to \mathbb{R}$ is said to be Lipschitz on Ω if there exists a constant $L := L(\Omega) \geq 0$ such that

$$|f(p) - f(q)| \leq Ld(p, q), \quad p, q \in \Omega. \tag{3}$$

Moreover, if it is established that for all $p_0 \in \Omega$ there exists $L(p_0) \geq 0$ and $\delta = \delta(p_0) > 0$ such that the inequality (3) occurs with $L = L(p_0)$ for all $p, q \in B_\delta(p_0) := \{p \in \Omega : d(p, p_0) < \delta\}$, then f is called locally Lipschitz on Ω.

Remark 3.1. As an immediate consequence of the triangular inequality we obtain that $|d(p, p_0) - d(q, p_0)| \leq d(p, q)$ for all p, q and $p_0 \in M$. Then, of Definition 3.1, we get that the Riemannian distance function to a fixed point, $d(\cdot, q)$ is Lipschitzian and therefore Lipschitzian locally. In fact, it is well known that every convex function is locally Lipschitz and consequently continuous. See [24].

Proposition 3.2. *Let $\Omega \subset M$ be an open convex set, $f : M \to \mathbb{R}$ and $p \in M$. If there exists $\lambda > 0$ such that $f + (\lambda/2)\, d^2(., p) : M \to \mathbb{R}$ is convex on Ω, then f is Lipschitz locally on Ω.*

Proof. Because $f + (\lambda/2)\, d^2(., p)$ is convex, it follows from Remark 3.1 that for any $\tilde{p} \in \Omega$ there exist $L_1, \delta_1 > 0$ such that

$$\left|[f(q_1) + (\lambda/2)\, d^2(q_1, p)] - [f(q_2) + (\lambda/2)\, d^2(q_2, p)]\right| \leq L_1 d(q_1, q_2), \quad \forall\, q_1, q_2 \in B(\tilde{p}, \delta_1). \tag{4}$$

Moreover, Proposition 3.1 together with Remark 3.1 imply that there exist $L_2, \delta_2 > 0$ such that

$$|(1/2)d^2(q_1, p) - (1/2)d^2(q_2, p)| \leq L_2 d(q_1, q_2), \quad \forall\, q_1, q_2 \in B(\tilde{p}, \delta_1). \tag{5}$$

Simple algebraic manipulations imply that

$$|f(q_1) - f(q_2)| \leq \left|[f(q_1) + (\lambda/2)\, d^2(q_1, p)] - [f(q_2) + (\lambda/2)\, d^2(q_2, p)]\right| + \left|(\lambda/2)\, d^2(q_2, p) - (\lambda/2)\, d^2(q_1, p)\right|.$$

Therefore, taking $\delta = \min\{\delta_1, \delta_2\}$, using (4) and (5) we conclude from the last inequality that

$$|f(q_1) - f(q_2)| \leq (L_1 + \lambda L_2) d(q_1, q_2), \quad \forall\, q_1, q_2 \in B(\tilde{p}, \delta),$$

and the proof is finished. □

Definition 3.2. Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a continuously differentiable function on Ω. The gradient vector field $\operatorname{grad} f$ is said to be Lipschitz with constant $\Gamma \geq 0$ on Ω always that

$$\|\operatorname{grad} f(q) - P_{pq} \operatorname{grad} f(p)\| \leq \Gamma d(p, q), \quad p, q \in \Omega,$$

where P_{pq} is the parallel transport along the geodesic segment joining p to q.

4. Generalized directional derivatives

In this section we present definitions for the generalized directional derivative and subdifferential of a locally Lipschitz function (not necessarily convex) which, in the Euclidean case, coincide with the Clarke generalized directional derivative and subdifferential, respectively. Moreover, some properties of those concepts are presented, amongst them the upper semicontinuity of the directional derivative and a relationship between the subdifferential of a sum of two Lipschitz locally function (in the particular case that one of them is differentiable) and its subdifferentials.

Definition 4.1. Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a locally Lipschitz function on Ω. The generalized directional derivative $f^\circ : T\Omega \to \mathbb{R}$ of f is defined by

$$f^\circ(p, v) := \limsup_{t \downarrow 0 \; q \to p} \frac{f\left(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v\right) - f(q)}{t} \tag{6}$$

where $(D\exp_p)_{\exp_p^{-1} q}$ denotes the differential of $\exp_p$ at $\exp_p^{-1} q$.

It is worth to pointed out that an equivalent definition has appeared in [17].

Remark 4.1. The generalized directional derivative is well defined. Indeed, let $L_p > 0$ the Lipschitz constant of f in p and $\delta = \delta(p) > 0$ such that

$$|f(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - f(q)| \le L_p\, d(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v,\ q), \quad q \in B_\delta(p),\ t \in [0, \delta).$$

Because $d(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v,\ q) = t\|(D\exp_p)_{\exp_p^{-1} q} v\|$, the above inequality becomes

$$|f(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - f(q)| \le L_p\, t\|(D\exp_p)_{\exp_p^{-1} q} v\|, \quad q \in B_\delta(p),\ t \in [0, \delta).$$

Since $\lim_{q \to p} (D\exp_p)_{\exp_p^{-1} q} v = v$ our statement follows from the last inequality.

Remark 4.2. Note that, if $M = \mathbb{R}^n$ then $\exp_p w = p + w$ and $(D\exp_p)_{\exp_p^{-1} q} v = v$. In this case, (6) becomes

$$f^\circ(p, v) = \limsup_{t \downarrow 0 \; q \to p} \frac{f(q + tv) - f(q)}{t},$$

which is the Clarke generalized directional derivative; see [25]. Therefore, the generalized differential derivative on Hadamard manifold is a natural extension of the Clarke generalized differential derivative.

Now we are going to prove the upper semicontinuity of the generalized directional derivative. The following result will be useful.

Lemma 4.1. *Let c_1 and c_2 be a C^2 curves in M, such that $c_1(0) = c_2(0) = p$, $c_1'(0) = v$ and $c_2'(0) = w$. If $\psi(s) = d(c_1(s), c_2(s))$, then the Taylor's Formula for ψ^2 in some neighborhood of $s = 0$ is given by*

$$\psi^2(s) = \|w - v\|^2 s^2 + \mathcal{O}(s^2), \qquad \lim_{s \to 0^+} \mathcal{O}(s^2)/s^2 = 0.$$

Furthermore, $\lim_{s \to 0^+} d(c_1(s), c_2(s))/s = \|w - v\|$.

Proof. See Lemma 3.2 and Corollary 3.1 of [26]. □

Proposition 4.1. *Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ be a locally Lipschitz function. Then, f° is upper semicontinuous on $T\Omega$, i.e., if $(p, v) \in T\Omega$ and $\{p^k, v^k\}$ is a sequence in $T\Omega$ such that $\lim_{k \to +\infty}(p^k, v^k) = (p, v)$, then*

$$\limsup_{k \to +\infty} f^\circ(p^k, v^k) \le f^\circ(p, v). \tag{7}$$

Proof. Let $(p, v) \in T\Omega$ and $\{(p^k, v^k)\} \subset T\Omega$ such that $\lim_{k \to +\infty}(p^k, v^k) = (p, v)$. For proving the inequality (7) first note that for each k

$$f^\circ(p^k, v^k) \le \limsup_{t \downarrow 0 \; (q,w) \to (p^k, v^k)} \frac{f(\exp_q tw) - f(q)}{t}, \quad (q, w) \in T\Omega.$$

So, by definition of upper limit, there exists $(q^k, w^k) \in T\Omega - \{(p^k, v^k)\}$ and $t_k > 0$ such that

$$f^\circ(p^k, v^k) - \frac{1}{k} < \frac{f(\exp_{q^k} t_k w^k) - f(q^k)}{t_k}, \qquad \tilde{d}((q^k, w^k), (p^k, v^k)) + t_k < \frac{1}{k}, \tag{8}$$

with $\tilde{d}$ being the Riemannian distance in TM. Let $U_p \subset \Omega$ be a neighborhood of p such that $TU_p \approx U_p \times \mathbb{R}^n$ and f is Lipschitz in U_p with constant L_p. From the first inequality in (8), we obtain

$$f^\circ(p^k, v^k) - \frac{1}{k} < \frac{f\left(\exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right) - f(q^k)}{t_k} + \frac{f(\exp_{q^k} t_k w^k) - f\left(\exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right)}{t_k}. \tag{9}$$

On the other hand, as $\lim_{k\to+\infty}(p^k, v^k) = (p, v)$, we conclude from the second inequality in (8) that

$$\exp_{q^k} t_k w^k \in U_p, \qquad \exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v \in U_p, \quad k > k_0,$$

for k_0 sufficiently large. Thus, because f is Lipschitz in U_p, for $k > k_0$ we have

$$\left| f(\exp_{q^k} t_k w^k) - f\left(\exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right)\right| \le L_p\, d\left(\exp_{q^k} t_k w^k,\ \exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right). \tag{10}$$

Since $\lim_{k\to+\infty} p^k = p$, second equation in (8) imply that $\lim_{k\to+\infty} q^k = p$. Thus, we conclude that $\lim_{k\to+\infty}(D\exp_p)_{\exp_p^{-1} q_k} v = v$ which together with Lemma 4.1 and $\lim_{k\to+\infty} t_k = 0$ implies

$$\lim_{k\to+\infty} d\left(\exp_{q^k} t_k w^k,\ \exp_{q^k} t_k (D\exp_p)_{\exp_p^{-1} q^k} v\right) \Big/ t_k = 0.$$

Therefore, combining the last equation, (9), (10), and Definition 4.1 the result follows. □

Lemma 4.2. *Let $f, g : M \to \mathbb{R}$ be locally Lipschitz functions at $p \in M$ and $v \in T_pM$. Then*

$$(f+g)^\circ(p, v) \le f^\circ(p, v) + g^\circ(p, v).$$

Proof. From the definition of the generalized directional derivative we have

$$(f+g)^\circ(p, v) = \limsup_{t\downarrow 0\ q\to p} \frac{(f+g)(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - (f+g)(q)}{t},$$

which immediately implies that

$$(f+g)^\circ(p, v) = \limsup_{t\downarrow 0\ q\to p} \left[\frac{f(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - f(q)}{t} + \frac{g(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - g(q)}{t}\right].$$

Therefore, the result follows by using simple upper limit properties together with the definition of the generalized directional derivative. □

Next we generalize the definition of subdifferential for locally Lipschitz functions defined on Hadamard manifold.

Definition 4.2. Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a locally Lipschitz function on Ω. The generalized subdifferential of f at $p \in \Omega$, denoted by $\partial^\circ f(p)$, is defined by

$$\partial^\circ f(p) := \left\{w \in T_pM : f^\circ(p, v) \ge \langle w, v\rangle, \forall\, v \in T_pM\right\}.$$

Remark 4.3. Let $\Omega \subset M$ be an open convex set. If the function $f : M \to \mathbb{R}$ is convex on Ω, then $f^\circ(p, v) = f'(p, v)$ (respectively, $\partial^\circ f(p) = \partial f(p)$) for all $p \in \Omega$, i.e., the directional derivatives (respectively, subdifferential) for Lipschitz functions is a generalization of the directional derivatives (respectively, subdifferential) for convex functions. See [17] Claim 5.4 in the proof of Theorem 5.3.

Definition 4.3. Let $f : M \to \mathbb{R}$ be locally Lipschitz function. A point $p \in \Omega$ is said to be a stationary point of f always $0 \in \partial^\circ f(p)$.

Proposition 4.2. *Let $f : M \to \mathbb{R}$ be locally Lipschitz function at $p \in M$, $\lambda > 0$ and $\tilde{p} \in M$. Then,*

$$\partial^\circ(f + (\lambda/2)d^2(., \tilde{p}))(p) \subset \partial^\circ f(p) - \lambda \exp_p^{-1} \tilde{p}.$$

Proof. Take $w \in \partial^\circ(f + (\lambda/2)d^2(., \tilde{p}))(p)$. Then,

$$\langle w, v\rangle \le (f + (\lambda/2)d^2(., \tilde{p}))^\circ(p, v), \quad \forall\, v \in T_pM.$$

Because $\lambda > 0$ it follows from Proposition 3.1 that the function $p \ni M \mapsto (\lambda/2)d^2(p, \tilde{p})$ is convex. Thus, Remarks 3.1, 4.3, Lemma 4.2 and the above inequality imply that

$$\langle w, v\rangle \leq f^\circ(p, v) + ((\lambda/2)d^2(., \tilde{p}))'(p, v), \quad \forall\, v \in T_pM.$$

On the other hand, as $p \ni M \mapsto (\lambda/2)d^2(p, \tilde{p})$ is differentiable with $\operatorname{grad}(\lambda/2)d^2(p, \tilde{p}) = -\lambda \exp_p^{-1} \tilde{p}$, the last inequality becomes

$$\langle w, v\rangle \leq f^\circ(p, v) - \langle \lambda \exp_p^{-1} \tilde{p}, v\rangle, \quad \forall\, v \in T_pM,$$

which implies that $w + \lambda \exp_{\tilde{p}}^{-1} p \in \partial^\circ f(p)$. So $w \in \partial^\circ f(p) - \lambda \exp_p^{-1} \tilde{p}$ and the proof is concluded. □

Corollary 4.1. *Let $\Omega \subset M$ be an open convex set, $f : M \to \mathbb{R}$ be locally Lipschitz functions on Ω, $\tilde{p} \in M$ and $\lambda > 0$ such that $f + (\lambda/2)\, d^2(., \tilde{p}) : M \to \mathbb{R}$ is convex on Ω. If $p \in \Omega$ is a minimizer of $f + (\lambda/2)d^2(., \tilde{p})$ then*

$$\lambda \exp_p^{-1} \tilde{p} \in \partial^\circ f(p).$$

Proof. Since p is a minimizer of $f + (\lambda/2)d^2(., \tilde{p})$ we obtain

$$0 \in \partial \left(f + \frac{\lambda}{2} d^2(., \tilde{p})\right)(p).$$

On the other hand, as $f + (\lambda/2)\, d^2(., \tilde{p})$ is convex on Ω, applying Proposition 4.2 we have

$$\partial \left(f + \frac{\lambda}{2} d^2(., \tilde{p})\right)(p) = \partial^\circ \left(f + \frac{\lambda}{2} d^2(., \tilde{p})\right)(p) \subset \partial^\circ f(p) - \lambda \exp_p^{-1} \tilde{p}.$$

Therefore, the result follows by combining two latter inclusions. □

5. Proximal point method for nonconvex problems

In this section we present an application of the proximal point method for minimizing a real-valued function (non-necessarily convex) given by the maximum of a certain class of continuously differentiable functions. Our goal is to prove the following theorem:

Theorem 5.1. *Let $\Omega \subset M$ be an open convex set, $q \in M$ and $T \subset \mathbb{R}$ a compact set. Let $\varphi : M \times T \to \mathbb{R}$ be a continuous function on $\Omega \times T$ such that $\varphi(., \tau) : M \to \mathbb{R}$ is a continosly differentiable function of Ω and continuous on $\bar{\Omega}$ (closure of Ω), for all $\tau \in T$, and $f : M \to \mathbb{R}$ defined by*

$$f(p) := \max_{\tau \in T} \varphi(p, \tau).$$

Assume that $-\infty < \inf_{p\in M} f(p)$, $\operatorname{grad}_p \varphi(., \tau)$ is Lipschitz on Ω with constant L_τ for each $\tau \in T$ such that $\sup_{\tau\in T} L_\tau < +\infty$ and

$$L_f(f(q)) = \{p \in M : f(p) \leq f(q)\} \subset \Omega, \quad \inf_{p\in M} f(p) < f(q).$$

Take $0 < \bar{\lambda}$ and a sequence $\{\lambda_k\}$ satisfying $\sup_{\tau\in T} L_\tau < \lambda_k \leq \bar{\lambda}$ and $\hat{p} \in L_f(f(q))$. Then the proximal point method

$$p^{k+1} := \operatorname*{argmin}_{p\in M} \left\{f(p) + \frac{\lambda_k}{2} d^2(p, p^k)\right\}, \quad k = 0, 1, \ldots, \tag{11}$$

with starting point $p^0 = \hat{p}$ is well defined, the generated sequence $\{p^k\}$ rests in $L_f(f(q))$ and satisfies only one of the following statements

(i) *$\{p^k\}$ is finite, i.e., $p^{k+1} = p^k$ for some k and, in this case, p^k is a stationary point of f,*
(ii) *$\{p^k\}$ is infinite and, in this case, any cluster point of $\{p^k\}$ is a stationary point of f.*

Moreover, assume that the minimizer set of f is non-empty, i. e.,

(h1) *$U^* = \{p : f(p) = \inf_{p\in M} f(p)\} \neq \emptyset$.*

Let $c \in (\inf_{p\in M} f(p),\ f(q))$. If, in addition, the following assumptions hold:

(h2) *$L_f(c)$ is convex and f is convex on $L_f(c)$ and $\varphi(., \tau)$ is continuous on $\bar{\Omega}$ the closure of Ω for $\tau \in T$;*
(h3) *there exist $\tilde{p} \in M$ and $0 < \mu < \bar{\lambda}$ such that $f + (\mu/2)d^2(., \tilde{p})$ is convex and*

$$\|y(p)\| > \delta > 0, \quad \forall\, p \in L_f(f(q)) \setminus L_f(c),\ \forall\, y(p) \in \partial\left(f + (\mu/2)d^2(., \tilde{p})\right)(p) + \mu \exp_p^{-1} \tilde{p},$$

then the sequence $\{p^k\}$ generated by (11) with

$$\max\{\mu, \max_{i\in I} L_i\} < \lambda_k \le \bar{\lambda}, \quad k = 0, 1, \ldots \tag{12}$$

converges to a point $p^* \in U^*$.

Remark 5.1. The continuity of each function $\varphi(., \tau)$ on $\bar{\Omega}$ in (h2) guarantees that the level sets of the function f, in particular the solution set U^*, are closed in the topology of the manifold M.

In the next remark we show that if Ω is bounded and $\varphi(., \tau)$ is convex on Ω and continuous on $\bar{\Omega}$ for all $\tau \in T$ then f satisfies the assumptions (h2) and (h3).

Remark 5.2. If $\varphi(., \tau)$ is a convex function on Ω and continuous in $\bar{\Omega}$ for all $\tau \in T$ then the assumption (h2) is naturally verified and if (h1) hold then (h3) also holds. For details, see [16].

In order to prove the above theorem we need some preliminary results. From now on we assume that every assumptions on Theorem 5.1 hold, with the exception of (h1), (h2) and (h3), which will be considered to hold only when explicitly stated.

Lemma 5.1. *For all $\tilde{p} \in M$ and λ satisfying*

$$\sup_{\tau\in T} L_\tau < \lambda,$$

function $f + (\lambda/2)d^2(., \tilde{p})$ is strongly convex on Ω with constant $\lambda - \sup_{\tau\in T} L_\tau$.

Proof. Since T is compact and φ is continuous the well definition of f follows. To conclude, see Lemma 4.1 in [16]. □

Corollary 5.1. *The proximal point method (11) applied to f with starting point $p^0 = \hat{p}$ is well defined.*

Proof. Since compactness plays no rule, the proof is equal to the proof of Corollary 4.1 in [16]. □

Lemma 5.2. *Let $\{p^k\}$ be the sequence generated by the proximal point method (11). Then*

(i) $0 \in \partial \left(f + \frac{\lambda_k}{2} d^2(., p^k)\right)(p^{k+1}),\ k = 0, 1, \ldots.$
(ii) $\lim_{s\to\infty} d(p^{k+1}, p^k) = 0.$

Moreover, if λ_k satisfies (12) and (h1), (h2) and (h3) hold, then $\{p^k\}$ converges to a point $p^* \in U^*$.

Proof. Since compactness plays no rule, the proof is similar to the proof of Lemma 4.2, 4.3 and 4.4 of [16]. □

Proof of Theorem 5.1. The well definition of the proximal point method follows from the Corollary 5.1. Let $\{p^k\}$ be the sequence generated by proximal point method. Because $p^0 = \hat{p} \in L_f(f(q))$, (11) implies that the whole sequence is in $L_f(f(q))$. From Lemma 5.2 item (i) we have

$$0 \in \partial \left(f + \frac{\lambda_k}{2} d^2(., p^k)\right)(p^{k+1}), \quad k = 0, 1, \ldots.$$

Since $\sup_{\tau\in T} L_\tau < \lambda_k$, Lemma 5.1 implies that $f + (\lambda_k/2)d^2(., p^k)$ is strongly convex on Ω, which together with Proposition 3.2 give us that f is locally Lipschitz in Ω. So, using the definition of p^{k+1} we conclude from Corollary 4.1 with $\lambda = \lambda_k$, $\tilde{p} = p^k$ and $p = p^{k+1}$ that

$$\lambda_k \exp^{-1}_{p^{k+1}} p^k \in \partial^\circ f(p^{k+1}). \tag{13}$$

If $\{p^k\}$ is finite, then $p^{k+1} = p^k$ for some k and the latter inclusion implies that $0 \in \partial^\circ f(p^{k+1})$, i.e., p^k is a stationary point of f. Now assume that $\{p^k\}$ is a infinite sequence. If $\bar{p}$ is a cluster point of $\{p^k\}$, then there exists a subsequence $\{p^{k_s}\}$ such that $\lim_{s\to+\infty} p^{k_s+1} = \bar{p}$ and Lemma 5.2 item (ii) implies

$$\lim_{s\to\infty} \| \exp^{-1}_{p^{k_s+1}} p^{k_s} \| = \lim_{s\to\infty} d(p^{k_s+1}, p^{k_s}) = 0. \tag{14}$$

Now, the relation (13) implies that

$$f^\circ(p^{k_s+1}, v) \ge \lambda_{k_s} \langle \exp^{-1}_{p^{k_s+1}} p^{k_s}, v\rangle, \quad \forall\, v \in T_{p^{k_s+1}}M.$$

Let $\bar{v} \in T_{\bar{p}}M$. Hence, the latter inequality implies that

$$f^\circ(p^{k_s+1}, v^{k_s+1}) \ge \lambda_{k_s} \langle \exp^{-1}_{p^{k_s+1}} p^{k_s}, v^{k_s+1}\rangle, \quad v^{k_s+1} = D(\exp_{\bar{p}})_{\exp^{-1}_{\bar{p}} p^{k_s+1}} \bar{v}.$$

Note that $\lim_{s\to+\infty} p^{k_s+1} = \bar{p}$ implies $\lim_{s\to+\infty} v^{k_s+1} = \bar{v}$. Because $\{\lambda_{k_s}\}$ is bounded, letting s goes to $+\infty$ in the last inequality, Proposition 4.1 together with (14) gives us

$$f^\circ(\bar{p}, \bar{v}) \geq \lim_{s\to+\infty} \sup f^\circ(p^{k_s+1}, v^{k_s+1}) \geq 0,$$

which implies that $0 \in \partial^\circ f(\bar{p})$, i.e., $\bar{p}$ is a stationary point of f and the first part of the theorem is concluded.

The second part follows from the last part of Lemma 5.2 and the proof of the theorem is finished. □

6. Example

Let $(\mathbb{R}_{++}, \langle , \rangle)$ be the Riemannian manifold, where $\mathbb{R}_{++} = \{x \in \mathbb{R} : x > 0\}$ and $\langle , \rangle$ is the Riemannian metric $\langle u, v\rangle = g(x)uv$ with $g : \mathbb{R}_{++} \to (0, +\infty)$. So, the Christoffel symbol and the geodesic equation are given by

$$\Gamma(x) = \frac{1}{2} g^{-1}(x) \frac{\mathrm{d}g(x)}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x} \ln\sqrt{g(x)}, \qquad \frac{\mathrm{d}^2 x}{\mathrm{d}t^2} + \Gamma(x)\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right)^2 = 0,$$

respectively. Besides, in relation to the twice differentiable function $h : \mathbb{R}_{++} \to \mathbb{R}$, the Gradient and the Hessian of h are given by

$$\operatorname{grad} h = g^{-1}h', \qquad \operatorname{hess} h = h'' - \Gamma h',$$

respectively, where h' and h'' denote the first and second derivatives of h in the Euclidean sense. For more details see [9]. So, in the particular case of $g(x) = x^{-2}$,

$$\Gamma(x) = -x^{-1}, \qquad \operatorname{grad} h(x) = x^2 h'(x), \qquad \operatorname{hess} h(x) = h''(x) + x^{-1}h'(x). \tag{15}$$

Moreover, the map $\psi : \mathbb{R} \to \mathbb{R}_{++}$ defined by $\psi(x) = e^x$ is an isometry between the Euclidean space $\mathbb{R}$ and the manifold $(\mathbb{R}_{++}, \langle , \rangle)$ and the Riemannian distance $d : \mathbb{R}_{++} \times \mathbb{R}_{++} \to \mathbb{R}_+$ is given by

$$d(x, y) = |\psi^{-1}(x) - \psi^{-1}(y)| = |\ln(x/y)|, \tag{16}$$

(see, for example [10]). Therefore, $(\mathbb{R}_{++}, \langle , \rangle)$ is a Hadamard manifold and the unique geodesic $x : \mathbb{R} \to \mathbb{R}_{++}$ with initial conditions $x(0) = x_0$ and $x'(0) = v$ is given by

$$x(t) = x_0 e^{(v/x_0)t}.$$

Now let $f_1, f_2, f : \mathbb{R}_{++} \to \mathbb{R}$ and $\varphi : \mathbb{R}_{++} \times [0, 1] \to \mathbb{R}$ be real-valued functions such that

$$\varphi(x, \tau) = f_1(x) + t(f_2(x) - f_1(x)), \qquad f(x) = \max_{\tau\in[0,1]} \varphi(x, \tau), \tag{17}$$

and consider the problem

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } x \in \mathbb{R}_{++}. \end{aligned}$$

Take a sequence $\{\lambda_k\}$ satisfying $0 < \lambda_k$. From (16), the proximal point method (11) becomes

$$x^{k+1} := \operatorname*{argmin}_{x\in\mathbb{R}_{++}} \left\{ f(x) + \frac{\lambda_k}{2} \ln^2\left(\frac{x}{x^k}\right) \right\}, \quad k = 0, 1, \ldots.$$

If f_1 and f_2 are given, respectively, by $f_1(x) = \ln(x)$ and $f_2(x) = -\ln(x) + e^{-2x} - e^{-2}$ then φ is continuous and $\varphi(., \tau)$ is continuously differentiable for each $\tau \in [0, 1]$. The last expression in (15) implies that

$$\operatorname{hess} f_1(x) = 0, \qquad \operatorname{hess} f_2(x) = (4 - 2/x)e^{-2x}, \quad x \in \mathbb{R}_{++}, \tag{18}$$

as a consequence, first equation in (17) gives us

$$\operatorname{hess}_x \varphi(x, \tau) = \tau \operatorname{hess} f_2(x), \quad \forall x \in \mathbb{R}_{++}\ \forall \tau \in [0, 1].$$

Note that, for $0 < \epsilon < 1/4$ and $\Omega = (\epsilon, +\infty)$, $\operatorname{hess} f_2$ is bounded on Ω and therefore $\operatorname{grad} f_2$ is Lipschitz on Ω. We denote by L the constant of Lipschitz of $\operatorname{grad} f_2$. From the last equality $\operatorname{hess}_x \varphi(., \tau)$ is also bounded on Ω and $\operatorname{grad}_x \varphi(., \tau)$ is Lipschitz on Ω with constant $L_\tau = \tau L$ for all $\tau \in [0, 1]$. Besides, $\sup_{\tau\in[0,1]} L_\tau = L < +\infty$.

We claim that $f(x) = \max_{j=1,2} f_j(x)$. Indeed, note that $f_2(x) - f_1(x) > 0$ for $x \in (0, 1)$, $f_2(x) - f_1(x) < 0$ for $x \in (1, +\infty)$ and $f_1(1) = f_2(1)$. Thus the affine function $[0, 1] \ni \tau \mapsto \varphi(x, \tau)$ satisfies

$$\max_{\tau\in[0,1]} \varphi(x, \tau) = \begin{cases} f_1(x), & x \in (0, 1), \\ f_2(x), & x \in (1, +\infty) \end{cases}$$

and the claim follows. With that characterization for f all assumptions of Theorem 5.1 are verified, with $q = 5/16$, $c = f(3/4)$ and $\delta = 2/5$; see Example in [16]. Hence, letting $x^0 \in \mathbb{R}_{++}$ and $\bar{\lambda} > 0$ such that $x^0 \in L_f(f(q))$ and $L < \mu < \lambda_k \leq \bar{\lambda}$, the proximal point method, characterized in Theorem 5.1, can be applied for solving the above nonconvex problem.

Remark 6.1. Function $f(x) = \max_\tau \varphi(x, \tau)$, in the above example, is nonconvex (in the Euclidean sense) when restricted to any open neighborhood containing its minimizer $x^* = 1$. Therefore, the local classical proximal point method (see [15]) cannot be applied to minimize that function. Also, as f is nonconvex in the Riemannian sense, the Riemannian proximal point method (see [4]) cannot be applied to minimize that function; see Example in [16] for more details.

7. Final remarks

We have extended the range of application of the proximal point method to solve nonconvex optimization problems on Hadamard manifold in the case that the objective function is given by the maximum of a certain infinite collection of continuously differentiable functions. Convexity of the auxiliary problems is guaranteed with the choice of appropriate regularization parameters in relation to the constants of Lipschitz of the field gradients of the functions which they compose the class in question. As regards Theorem 5.1, in the particular case where $\varphi(., \tau)$ is convex for $\tau \in T$, convexity of the auxiliary problems is guaranteed without the need of restrictive assumptions on the regularization parameters. Besides, as observed in Remark 5.2, the additional assumptions (h2) and (h3) are satisfied whenever Ω is bounded.

Acknowledgements

The second author was supported in part by CNPq Grant 302618/2005-8, PRONEX–Optimization (FAPERJ/CNPq) and FUNAPE/UFG. The third author was supported in part by CNPq.

References

[1] P.-A. Absil, C.G. Baker, K.A. Gallivan, Trust-region methods on Riemannian manifolds, Found. Comput. Math. 7 (3) (2007) 303–330.
[2] H. Attouch, J. Bolte, P. Redont, M. Teboulle, Singular Riemannian barrier methods and gradient-projection dynamical systems for constrained optimization, Optimization 53 (5–6) (2004) 435–454.
[3] O.P. Ferreira, P.R. Oliveira, Subgradient algorithm on Riemannian manifolds, J. Optim. Theory Appl. Appl. 97 (1) (1998) 93–104.
[4] O.P. Ferreira, P.R. Oliveira, Proximal point algorithm on Riemannian manifolds, Optimization 51 (2) (2000) 257–270.
[5] Yu.S. Ledyaev, Qiji J. Zhu, Nonsmooth analysis on smooth manifolds, Trans. Amer. Math. Soc. 359 (8) (2007) 3687–3732 (electronic).
[6] E.A. Papa Quiroz, P.R. Oliveira, Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds, J. Convex Anal. 16 (1) (2009) 49–69.
[7] T. Rapcsák, Smooth Nonlinear Optimization in R^n, Kluwer Academic Publishers, Dordrecht, 1997.
[8] S.T. Smith, Optimization techniques on Riemannian Manifolds, in: Fields Institute Communications, vol. 3, Amer. Math. Soc., Providence, RI, 1994, pp. 113–146.
[9] C. Udriste, Convex Functions and Optimization Methods on Riemannian Manifolds, in: Mathematics and its Applications, vol. 297, Kluwer Academic Publishers, 1994.
[10] J.X. da Cruz Neto, O.P. Ferreira, L.R. Lucâmbio Pérez, S.Z. Németh, Convex-and monotone-transformable Mathematical Programming problems and a proximal-like point method, J. Global Optim. 35 (2006) 53–69.
[11] B. Martinet, Régularisation, d'inéquations variationelles par approximations successives, Rev. Française Informat. Recherche Opérationnelle Ser. R-3 4 (1970) 154–158 (in French).
[12] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control. Optim. 14 (1976) 877–898.
[13] Jonathan E. Spingarn, Submonotone mappings and the proximal point algorithm, Numer. Funct. Anal. Optim. 4 (2) (1981/82) 123–150.
[14] W. Hare, C. Sagastizábal, Computing proximal points of nonconvex functions, Math. Program. Ser. B (116) (2009) 221–258.
[15] A. Kaplan, R. Tichatschke, Proximal point methods and nonconvex optimization, J. Global Optim. 13 (4) (1998) 389–406.
[16] G.C. Bento, O.P. Ferreira, P.R. Oliveira, Proximal point methods for a special class of nonconvex functions on hadamard manifolds, arXiv:0809.2594v6 [math.OC], 2010, 37 pages.
[17] D. Azagra, J. Ferrera, M. López-Mesas, Nonsmooth analysis and Hamilton–Jacobi equations on Riemannian manifolds, J. Funct. Anal. 220 (2005) 304–361.
[18] D. Motreanu, N.H. Pavel, Quasitangent vectors in flow-invariance and optimization problems on Banach manifolds, J. Math. Anal. Appl. 88 (1) (1982) 116–132.
[19] W. Thämelt, Directional derivatives and generalized gradients on manifolds, Optimization 25 (2–3) (1992) 97–115.
[20] M.P. do Carmo, Riemannian Geometry, Birkhauser, Boston, 1992.
[21] T. Sakai, Riemannian Geometry, in: Translations of Mathematical Monographs, vol. 149, Amer. Math. Soc., Providence, RI, 1996.
[22] J.-B Hiriart-Urruty, C. Lemaréchal, Convex Analysis and Minimization Algorithm I and II, Springer-Verlag, 1993.
[23] J.X. da Cruz Neto, O.P. Ferreira, L.R. Lucâmbio Pérez, Contribution to the study of monotone vector fields, Acta Mathematica Hungarica 94 (4) (2002) 307–320.
[24] R.E. Greene, H. Wu, On the subharmonicity and plurisubharmonicity of geodesically convex functions, Indiana Univ. Math. J. 22 (1973) 641–654.
[25] F.H. Clarke, Optimization and Nonsmooth Analysis, in: Classics in Applied Mathematics, vol. 5, SIAM, New York, 1983.
[26] J.X. da Cruz Neto, O.P. Ferreira, L.R. Lucâmbio Pérez, Monotone point-to-set vector fields. Dedicated to Professor Constantin Udriste, Balkan J. Geom. Appl. 5 (1) (2000) 69–79.