Noname manuscript No.
(will be inserted by the editor)

A proximal point algorithm for DC functions on Hadamard manifolds

J.C.O. Souza · P.R. Oliveira

Received: date / Accepted: date

Abstract An extension of a proximal point algorithm for difference of two convex functions is presented in the context of Riemannian manifolds of nonposite sectional curvature. If the sequence generated by our algorithm is bounded it is proved that every cluster point is a critical point of the function (not necessarily convex) under consideration, even if minimizations are performed inexactly at each iteration. Application in maximization problems with constraints, within the framework of Hadamard manifolds is presented.

Keywords Nonconvex optimization · proximal point algorithm · DC functions · Hadamard manifolds

Mathematics Subject Classification (2000) 49M30 · 90C26 · 90C48

1 Introduction

It is well known that the class of Proximal Point Algorithm (PPA) is one of the most studied methods for finding zeros of maximal monotone operators and, in particular it's used to solve convex optimization problems. The classical PPA was introduced into optimization literature by Martinet [1]. It is based on the notion of proximal mapping J_λ^f,

$$J_\lambda^f(x) = \arg\min_{z\in\mathbb{R}^n}\{f(z) + \frac{1}{2\lambda}||z-x||^2\}, \tag{1}$$

This research was partially supported by CNPq, Brazil.

J.C.O. Souza
COPPE-Sistemas, Universidade Federal do Rio de Janeiro, Caixa Postal 68511, CEP 21945-970, Rio de Janeiro, RJ, Brazil and CEAD, Universidade Federal do Piauí, Teresina, PI, Brazil
E-mail: joaocos.mat@ufpi.edu.br

P.R. Oliveira
COPPE-Sistemas, Universidade Federal do Rio de Janeiro, Caixa Postal 68511, CEP 21945-970, Rio de Janeiro, RJ, Brazil
E-mail: poliveir@cos.ufrj.br

introduced earlier by Moreau [2]. The PPA was popularized by Rockafellar [3], who showed the algorithm converges even if the auxiliary minimizations in (1) are performed inexactly, which is an important consideration in practice. The algorithm is useful, however, only for convex problems, because the idea underlying the results is based on the monotonicity of subdifferential operators of convex functions. Therefore, PPA for nonconvex functions has been investigated by many authors (cf. [4],[5] and references therein). In Rockafellar [3], the algorithm starting with any $x^0 \in \mathbb{R}^n$, iteratively updates x^{k+1} conforming to the following recursion

$$0 \in c_k T(x^{k+1}) + x^{k+1} - x^k, \tag{2}$$

where $0 < c \leq c_k$, is a sequence of scalars and T is a multivalued maximal monotone operator from $\mathbb{R}^n$ to itself.

On the other hand, extension to Riemannian manifolds of the concepts and techniques that fit in Euclidean spaces is natural and nontrivial. Actually, in recent years, some algorithms defined to solve minimization problems have been extended from Hilbert space framework to the more general setting of Riemannian manifolds (see, for example [6]-[20]). The main advantages of these extensions are that nonconvex problems in the classic sense may become convex and constrained optimization problems may be seen as unconstrained ones through the introduction of an appropriate Riemannian metric (see [6]-[10]). Numerical solution of optimization problems defined on Riemannian manifolds arise in a variety of applications, e.g., in computer vision, signal processing, motion and structure estimation, or numerical linear algebra (see for instance [21]-[24]). Also, these extensions give rise to interesting theoretical questions. To extend (1) and (2) to the context of Riemannian manifolds was the subject of [6] and [11], respectively.

We will consider a special class of nonconvex optimization problem of the form

$$\min_{x \in M} f(x) = g(x) - h(x), \tag{3}$$

where $g, h : M \to \mathbb{R}$ are proper, convex and lower semi-continuous (lsc) functions and M is a complete Riemannian manifold. The function f is called a DC function (i.e. difference of two convex functions). The interest in the theory of DC functions has much increased in the last years (see for instance [25]-[29] and references therein), but only a few have proposed some specific algorithms or numerical experiments (for example [30]-[31]). Some mathematical reasons used to explain interest in DC functions can be found in [26], for instance, the class of DC functions defined on a compact convex set $X \subset \mathbb{R}^n$ is dense in the set of continuous function over X, endowed with the topology of uniform convergence over X.

Sun et al [30] proposed a proximal point algorithm for minimization of DC functions which use convex properties of the two convex functions separately. The purpose of this paper is to extend the PPA presented in [30] to Riemannian manifolds framework. Also, two different inexact methods of our algorithm are considered. Moreover, an application to the constrained optimization problems on Hadamard manifolds is given. To the best of our knowledge a proximal point algorithm to solve DC optimization problems in the context of Riemannian manifolds has not been established yet.

The paper is organized as follows. In Sect. 2, some fundamental definitions, properties and notations of Riemannian manifolds are presented. In Sect. 3, some definitions, notations and properties of convex analysis on Riemannian manifolds are presented. Convergence analysis of the exact version and inexact versions of the algorithm are provided in Sect. 4 and 5, respectively. In Sect. 6, an application to constrained optimization problems on Hadamard manifolds is given.

2 Basic Concepts

In this section, we introduce some fundamental properties and notations of Riemannian manifold. These basics facts can be found in any introductory book of Riemannian geometry, for example [32], [33].

Let M be a connected m-dimensional C^∞ manifold and let

$$TM = \{(x, v) : x \in M, v \in T_xM\}$$

be its *tangent bundle*, where T_xM is the *tangent space* of M at x. T_xM is a linear space and has same dimension of M, moreover, because we restrict ourselves to real manifolds, it is isomorphic to R^m. If M is endowed with a Riemannian metric g, then M is a *Riemannian manifold* and we denoted it by (M, g). The inner product of two vectors u and v in T_xM is written $\langle u, v\rangle := g_x(u, v)$ where g_x is the metric at the point x. The norm of a vector $u \in T_xM$ is defined by $\|u\| := \langle u, u\rangle_x^{1/2}$. Recall that the metric can be used to define the length of piecewise smooth curve $c : [a, b] \to M$ joining x' to x, i.e., such that $c(a) = x'$ and $c(b) = x$, by $L(c) = \int_a^b \|c'(t)\|dt$. Minimizing this length functional over the set of all such curves we obtain a Riemannian distance $d(x, x')$ which induces the original topology on M. Let ∇ be the Levi-Civita connection associated to (M, g). A vector field V along c is said to be *parallel* if $\nabla_{c'}V = 0$. If c' itself is parallel we say that c is a *geodesic*. The geodesic equation $\nabla_{\gamma'}\gamma' = 0$ is a second order nonlinear ordinary differential equation, then $\gamma = \gamma_v(\cdot, x)$ is determined by its position x and velocity v at x. 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 x in M is said to be *minimal* if its length equals $d(x, x')$ and this geodesic is called a *minimizing geodesic*.

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 x' and x, 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. In this paper, all manifolds are assumed to be complete. Take $x \in M$, the *exponential map* $\exp_x : T_xM \to M$ is defined by $\exp_x(v) = \gamma_v(1, x)$.

We denote by R the *curvature tensor* defined by $R(X, Y) = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[Y,X]}Z$, where X, Y and Z are vector fields of M and $[X, Y] = YX - XY$. 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 $\|X\|^2 = \langle X, X\rangle$. If $K(X, Y) \leq 0$ for all X and Y, then M is called a Riemannian manifold of nonpositive curvature and we use the short notation $K \leq 0$. A complete simply connected Riemannian manifold of nonpositive sectional curvature is called a *Hadamard manifold*. The following result is well known (see, for example [33], Theorem 4.1, p.221).

Theorem 1 *Let M be a Hadamard manifold and let $p \in M$. Then $\exp_p : T_pM \to M$ is a diffeomorphism, and for any two points $p, q \in M$ there exists a unique normalized geodesic joining p to q, which is, in fact, a minimal geodesic.*

This theorem shows that M is diffeomorphic to the Euclidean space $\mathbb{R}^n$. Thus, we see that M has the same topology and differential structure as R^n. Moreover, Hadamard manifolds and Euclidean spaces have some similar geometrical properties. One of the most important properties is described in the following theorem, which is taken from ([33], Proposition 4.5, p.223) and will be useful in our study. Recall that a *geodesic triangle* $\triangle(p_1, p_2, p_3)$ of a Riemannian manifold is the set consisting of three distinct points p_1, p_2 and p_3 called the *vertices* and three minimizing geodesic segments γ_{i+1} joining p_{i+1} to p_{i+2} called the *sides*, where $i = 1, 2, 3(\text{mod}3)$.

Theorem 2 (Comparison theorem for triangles) *Let M be a Hadamard manifold and $\triangle(x_1, x_2, x_3)$ a geodesic triangle. Denote by $\gamma_{i+1} : [0, l_{i+1}] \to M$ geodesic segments joining x_{i+1} to x_{i+2} and set $l_{i+1} := L(\gamma_{i+1})$, $\theta_{i+1} = \angle(\gamma'_{i+1}(0), -\gamma'_i(l_i))$, where $i = 1, 2, 3(mod3)$. Then*

$$\theta_1 + \theta_2 + \theta_3 \leq \pi \tag{4}$$

$$l_{i+1}^2 + l_{i+2}^2 - 2l_{i+1}l_{i+2}cos\theta_{i+2} \leq l_i^2. \tag{5}$$

Let $\gamma : [a, b] \to M$ be a normalized geodesic segment. A differentiable variation of γ is by definition a differentiable mapping $\alpha : [a, b] \times (-\epsilon, \epsilon) \to M$ satisfying $\alpha(t, 0) = \gamma(t)$. The vector field along γ defined by $V(t) = (\partial\alpha/\partial s)(t, 0)$ is called the variational vector field of α. The first variational formula of arc length on α is given as follows:

$$L'(\gamma) := \frac{d}{ds} L(c_s) \mid_{s=0} = \langle V, \gamma \rangle \mid_a^b, \tag{6}$$

where $c_s(t) = \alpha(t, s)$ with $s \in (-\epsilon, \epsilon)$.

The Riemannian distance plays a fundamental role in the next sections. We proceed now stating a result which we will go to use. Let M be a Hadamard manifold. For any $x \in M$ we can define the exponential inverse map

$$\exp_x^{-1} : M \to T_x M$$

which is C^∞. Since $d(x, x') = \| \exp_{x'}^{-1}(x) \|$, then the map $\rho_{x'} : M \to R$ defined by $\rho_{x'}(x) = \frac{1}{2} d^2(x, x')$ is C^∞ and its gradient at x is $grad\rho_{x'}(x) = -\exp_x^{-1}(x')$ (see, [33]).

Using the properties of the parallel transport and the exponential map, we obtain the following proposition that will be used in the next sections.

Proposition 1 *Let M be a Hadamard manifold. Let $x^0 \in M$ and $\{x^k\} \subset M$ be such that $x^k \to x^0$. Then the following assertions hold.*

1. *For any $y \in M$, we have*
$$\exp_{x^k}^{-1} y \to \exp_{x^0}^{-1} y \text{ and } \exp_y^{-1} x^k \to \exp_y^{-1} x^0.$$
2. *If $v^k \in T_{x^k} M$ and $v^k \to v^0$, then $v^0 \in T_{x^0} M$.*
3. *Given $u^k, v^k \in T_{x^k} M$ and $u^0, v^0 \in T_{x^0} M$, if $u^k \to u^0$ and $v^k \to v^0$, then*
$$\langle u^k, v^k \rangle \to \langle u^0, v^0 \rangle.$$
4. *For any $u \in T_{x^0} M$, the function $F : M \to TM$ defined by $F(x) = P_{x,x^0} u$ for each $x \in M$ is continuous on M.*

Proof See [11], Lemma 2.4, p. 666. □

3 Convexity on Riemannian Manifolds

In this section, we introduce some definitions and notation of convexity on Riemannian manifolds. We also present some properties of the subdifferential of a convex function; see [34] for more details.

A subset $C \subset M$ is said to be *convex* if, for any points p and q in C, the geodesic joining p to q is contained in C, that is, if $\gamma : [a, b] \to M$ is a geodesic such that $\gamma(a) = p$ and $\gamma(b) = q$, then $\gamma((1-t)a + tb) \in C$ for all $t \in [0, 1]$. Let $f : M \to \mathbb{R}$ be a proper extended real-valued function. The domain of the function f is denoted by $\mathrm{dom}(f)$ and defined by $\mathrm{dom}(f) = \{x \in M : \ f(x) \neq +\infty\}$. The function f is said to be convex (respectively, strictly convex) if, for any geodesic segment $\gamma : [a, b] \to M$, the composition $f \circ \gamma : [a, b] \to \mathbb{R}$ is convex (respectively, strictly convex), that is,

$$(f \circ \gamma)(ta + (1-t)b) \leq t(f \circ \gamma)(a) + (1-t)(f \circ \gamma)(b),$$

for any $a, b \in \mathbb{R}$ and $0 \leq t \leq 1$. The *subdifferential* of f at x is defined by

$$\partial f(x) = \{u \in T_x M \ ; \ \langle u, \exp_x^{-1} y\rangle \leq f(y) - f(x), \ \forall y \in M\}. \tag{7}$$

Then $\partial f(x)$ is a closed convex (possible empty) set. The proofs of the above assertions and the following propositions can be found in [6] and [34].

Proposition 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.*

Proposition 3 *Let M be a Hadamard manifold and let $f : M \to \mathbb{R}$ be convex function. Then, for any $x \in M$, there is $s \in T_x M$ such that*

$$f(y) \geq f(x) + \langle s, \exp_x^{-1} y\rangle, \ \forall y \in M.$$

In other words, the subdifferential $\partial f(x)$ of f at $x \in M$ is nonempty.

Proposition 4 *If a function $f : M \to \mathbb{R}$ is convex, then for any $x \in M$ and $\lambda > 0$, there exists a unique point, denoted by $p_\lambda(x)$, such that*

$$f(p_\lambda(x)) + \frac{\lambda}{2} d^2(p_\lambda(x), x) = f_\lambda(x)$$

characterized by $\lambda(\exp_{p_\lambda(x)}^{-1} x) \in \partial f(p_\lambda(x))$, where $f_\lambda(x) = \inf_{y \in M}\{f(y) + \lambda d^2(x, y)\}$.

4 Proximal Point Algorithm

Let M be a Hadamard manifold and let $f : M \to \mathbb{R}$ be a DC function, i.e., $f(x) = g(x) - h(x)$, where $g, h : M \to \mathbb{R}$ are proper, convex and lsc functions satisfying $\mathrm{dom}(g) \cap \mathrm{dom}(h) \neq \emptyset$. A necessary condition for x to be a local minimum of f is that $0 \in \partial f(x) \subset \partial g(x) - \partial h(x)$. In other words, the subdifferentials $\partial g(x)$ and $\partial h(x)$ must overlap:

$$\partial g(x) \cap \partial h(x) \neq \emptyset.$$

A similar condition holds true when x is a local maximum of f. So, we will focus our attention on finding critical points of f. The set of *critical points* of f is defined by

$S = \{x \in M; \partial h(x) \cap \partial g(x) \neq \emptyset\}$. Observe that a necessary and sufficient condition for a point x to be a critical point of a DC function f is that

$$\frac{1}{c} \exp_x^{-1} y \in \partial g(x),$$

where $y = \exp_x(cw)$, for any $w \in \partial h(x)$ and $c > 0$ a real number.

Throughout the remainder of this paper, we always assume that M is a Hadamard manifold, $f : M \to \mathbb{R}$ is a bounded from below DC function, such that $f(x) = g(x) - h(x)$, and $S \neq \emptyset$. For finding critical points of a DC functions on Hadamard manifolds, which satisfies necessary optimality conditions, we consider the following algorithm:

Algorithm (DCPPA)

Step 1: Given an initial point $x^0 \in M$ and a bounded sequence of positive numbers $\{c_k\} \subset [b, c]$.

Step 2: Compute

$$w^k \in \partial h(x^k) \text{ and set } y^k := \exp_{x^k}(c_k w^k). \tag{8}$$

Step 3: Compute

$$x^{k+1} := \arg\min_{x \in M} \{g(x) + \frac{1}{2c_k} d^2(x, y^k)\}. \tag{9}$$

If $x^{k+1} = x^k$, stop. Otherwise, $k := k + 1$ and return to Step 2.

The well definition of the sequences $\{x^k\}$ and $\{y^k\}$ follows immediately from Proposition 3 and 4.

Note that when $h(x) = 0$, algorithm DCPPA becomes exactly the algorithm proposed by [6]. If $M = \mathbb{R}^n$ algorithm DCPPA reduces to the algorithm proposed in [30]. Therefore, the algorithm DCPPA on Hadamard manifolds is a natural generalization of the proximal point algorithm for DC functions on $\mathbb{R}^n$ defined by Sun at al [30] and more general than proximal point algorithm proposed by Ferreira and Oliveira [6].

Now we shall establish the convergence of the algorithm. We begin by showing that algorithm DCPPA is a decent algorithm.

Theorem 3 *The sequence $\{x^k\}$ generated by algorithm DCPPA satisfies:*

1. *either the algorithm stops at a critical point;*
2. *or f decreases strictly, i.e., $f(x^{k+1}) < f(x^k)$, $\forall k \geq 0$.*

Proof It follows from (8) and (9) that

$$w^k = \frac{1}{c_k} \exp_{x^k}^{-1} y^k \in \partial h(x^k) \tag{10}$$

and

$$\frac{1}{c_k} \exp_{x^{k+1}}^{-1} y^k \in \partial g(x^{k+1}). \tag{11}$$

If $x^{k+1} = x^k$ the algorithm stops and, this clearly implies that $\frac{1}{c_k} \exp_{x^k}^{-1} y^k \in \partial h(x^k) \cap \partial g(x^k)$, which means, $x^k \in S$. Now, suppose $x^{k+1} \neq x^k$. Using (10) and (11) in (7), we obtain that

$$h(x) \geq h(x^k) + \frac{1}{c_k}\langle \exp^{-1}_{x^k} y^k, \exp^{-1}_{x^k} x\rangle, \ \forall x \in M$$

and

$$g(x) \geq g(x^{k+1}) + \frac{1}{c_k}\langle \exp^{-1}_{x^{k+1}} y^k, \exp^{-1}_{x^{k+1}} x\rangle, \ \forall x \in M.$$

Adding the last inequalities with $x = x^{k+1}$ in the first one and $x = x^k$ in the second one, we have

$$f(x^k) \geq f(x^{k+1}) + \frac{1}{c_k}\left[\langle \exp^{-1}_{x^k} y^k, \exp^{-1}_{x^k} x^{k+1}\rangle + \langle \exp^{-1}_{x^{k+1}} y^k, \exp^{-1}_{x^{k+1}} x^k\rangle\right]. \quad (12)$$

Now, consider the geodesic triangle $\triangle(y^k, x^k, x^{k+1})$ and set $\theta = \angle(\exp^{-1}_{x^k} y^k, \exp^{-1}_{x^k} x^{k+1})$. By Theorem 2, we have

$$d^2(y^k, x^k) + d^2(x^k, x^{k+1}) - 2d(y^k, x^k)d(x^k, x^{k+1})\cos\theta \leq d^2(y^k, x^{k+1}).$$

Since $\langle \exp^{-1}_{x^k} y^k, \exp^{-1}_{x^k} x^{k+1}\rangle = 2d(y^k, x^k)d(x^k, x^{k+1})\cos\theta$, so that

$$\langle \exp^{-1}_{x^k} y^k, \exp^{-1}_{x^k} x^{k+1}\rangle \geq \frac{1}{2}d^2(y^k, x^k) + \frac{1}{2}d^2(x^k, x^{k+1}) - \frac{1}{2}d^2(y^k, x^{k+1}).$$

Similarly, considering the geodesic triangle $\triangle(y^k, x^{k+1}, x^k)$ and setting $\theta = \angle(\exp^{-1}_{x^{k+1}} y^k, \exp^{-1}_{x^{k+1}} x^k)$, we have

$$\langle \exp^{-1}_{x^{k+1}} y^k, \exp^{-1}_{x^{k+1}} x^k\rangle \geq \frac{1}{2}d^2(y^k, x^{k+1}) + \frac{1}{2}d^2(x^k, x^{k+1}) - \frac{1}{2}d^2(y^k, x^k).$$

Adding the last two inequalities, we obtain that

$$\langle \exp^{-1}_{x^k} y^k, \exp^{-1}_{x^k} x^{k+1}\rangle + \langle \exp^{-1}_{x^{k+1}} y^k, \exp^{-1}_{x^{k+1}} x^k\rangle \geq d^2(x^k, x^{k+1}).$$

Combining the above inequality with (12), we have that

$$f(x^k) \geq f(x^{k+1}) + \frac{1}{c_k}d^2(x^k, x^{k+1}), \quad (13)$$

this means that $f(x^{k+1}) < f(x^k)$. □

Corollary 1 *Consider $\{x^k\}$ generated by algorithm DCPPA, then the sequence $\{f(x^k)\}$ is convergent.*

Proof Being f bounded from below due to the last theorem that the sequence $\{f(x^k)\}$ is bounded and thus has at least one cluster point. Indeed, let $\{f(x^k)\}$ admit two different cluster points $f_1 < f_2$. Let $f(x^{k_j})$ and $f(x^{k_l})$ be two subsequences converging to f_1 and f_2, respectively. Set $\epsilon = \frac{f_2 - f_1}{2}$, then there exist $k_{j_0}, k_{l_0} \in \mathbb{N}$ such that

$$f(x^{k_j}) < f_1 + \epsilon$$
$$f_2 - \epsilon < f(x^{k_l}),$$

for all $k_j, k_l \geq k_0 = \max\{k_{j_0}, k_{l_0}\}$. By virtue of item 2 of last theorem, we have

$$f(x^{k_j}) \leq f(x^{k_0}) < f_1 + \epsilon = f_2 - \epsilon.$$

This is a contradiction, and hence $\{f(x^k)\}$ has at the most one cluster point. □

Corollary 2 *If f is a continuous function and $\{x^k\}$ is bounded, then* $\lim_{k\to\infty} f(x^k) = f(\overline{x})$, *for some cluster point $\overline{x}$ of $\{x^k\}$.*

Proof Let $\{x^{k_j}\}$ be any convergent subsequence with limit $\overline{x} \in M$. Since f is continuous, then $f(x^{k_j}) \to f(\overline{x})$. Thus, for a given $\epsilon > 0$, there exists $j_0 \in \mathbb{N}$ such that $\forall j \geq j_0$, we have

$$f(x^{k_j}) - f(\overline{x}) < \epsilon.$$

Since f is a decent function, we obtain that

$$f(x^k) - f(\overline{x}) = f(x^k) - f(x^{k_{j_0}}) + f(x^{k_{j_0}}) - f(\overline{x}) \leq f(x^{k_{j_0}}) - f(\overline{x}) < \epsilon,$$

$\forall k \geq k_{j_0}$, for an arbitrary $\epsilon > 0$ and the proof is concluded. □

Proposition 5 *Consider $\{x^k\}$ generated by algorithm DCPPA, then* $\sum_{k=0}^{\infty} d^2(x^k, x^{k+1}) < \infty$. *In particular* $\lim_{k\to+\infty} d(x^k, x^{k+1}) = 0$.

Proof From (13), we have that

$$\frac{1}{c_k} d^2(x^k, x^{k+1}) \leq f(x^k) - f(x^{k+1})$$

and, therefore

$$\sum_{k=0}^{n-1} \frac{1}{c_k} d^2(x^k, x^{k+1}) \leq f(x^0) - f(x^n).$$

Since f is bounded from below and $\{c_k\}$ is bounded, we obtain $\sum_{k=0}^{\infty} d^2(x^k, x^{k+1}) < \infty$, and it follows that $\lim_{k\to+\infty} d(x^k, x^{k+1}) = 0$. □

According to Proposition 2, in Algorithm DCPPA, if $\{x^k\}$ is bounded, then $\{w^k\}$ is also bounded. Since the exponential mapping is a diffeomorphism, we also have that $\{y^k\}$ is bounded.

Theorem 4 *Suppose that $\{x^k\}$ is bounded. Then every cluster-point of $\{x^k\}$ is a critical point of the function f.*

Proof Let x and y be cluster points of $\{x^k\}$ and $\{y^k\}$, respectively. So, consider two subsequences x^{k_j} and y^{k_l} converging respectively to x and y, i.e., $x^{k_j} \to x$ and $y^{k_l} \to y$. From definition of the sequences $\{x^k\}$, $\{y^k\}$ and $\{c_k\} \subset [b, c]$, we have

$$h(z) \geq h(x^{k_j}) + \frac{1}{c}\langle \exp^{-1}_{x^{k_j}} y^{k_l}, \exp^{-1}_{x^{k_j}} z\rangle, \ \forall z \in M$$

and

$$g(z) \geq g(x^{k_j+1}) + \frac{1}{c}\langle \exp^{-1}_{x^{k_j+1}} y^{k_l}, \exp^{-1}_{x^{k_j+1}} z\rangle \ \forall z \in M.$$

Taking $k_j, k_l \to \infty$ and making use of the Proposition 5, we have that $\frac{1}{c}\exp^{-1}_x y \in \partial h(x)$ and $\frac{1}{c}\exp^{-1}_x y \in \partial g(x)$. In other words that x is a critical point of f. □

Corollary 3 *Suppose that $\{x^k\}$ is bounded and S is a singleton, i.e., $S = \{x^*\}$, then the entire sequence $\{x^k\}$ converges to x^*.*

Proof Suppose, by contradiction, that there exist an $\epsilon > 0$, such that

$$d(x^k, x^*) \geq \epsilon, \tag{14}$$

$\forall k \geq k_0$. Note that there exist a subsequence $\{x^{k_j}\} \subset \{x^k\}$ such that $x^{k_j} \to \overline{x}$. By the theorem we have $\overline{x} \in S$. But $S = \{x^*\}$, and thus $\overline{x} = x^*$. Therefore, for all $\epsilon > 0$, there exist $k_0 \in \mathbb{N}$ such that $d(x^{k_j}, x^*) = d(x^{k_j}, \overline{x}) < \epsilon$, $\forall k \geq k_0$, violating (14). This completes the proof. □

Remark 1 If the level set of f and the subdifferential of h are compact and bounded, respectively, then the sequences $\{x^k\}$ and $\{y^k\}$ are bounded. If f is strictly convex and coercive or strongly convex, then S is a singleton.

Remark 2 It is worthwhile to point out that under the assumptions of Corollary 2, if f satisfies the sharp minima condition (see Polyak [35]), then the whole sequence $\{x^k\}$ converges to some point $x^* \in S$. Regarding weak sharp minima introduced by Ferris [36], it was recently considered on the context of Riemannian manifolds by Li et. al. [16] and finite termination of Proximal Point Algorithm on Hadamard manifolds by Bento and Cruz Neto [37]. We hope that this paper may stimulate further research involving Algorithm DCPPA and these concepts.

5 Inexact Version

Here we consider the approximate version obtained by replacing the exact subdifferential by the approximate one, since the function h and g are assumed to be convex, proper and lower semicontinuous. We define $\partial_0 h(x) = \partial h(x)$ and $\partial_0 g(x) = \partial g(x)$, for any $x \in M$. Furthermore, directly from the definition it follows that $0 \leq \epsilon_1 \leq \epsilon_2$, then

$$\partial_{\epsilon_1} h(x) \subseteq \partial_{\epsilon_2} h(x)$$

and

$$\partial_{\epsilon_1} g(x) \subseteq \partial_{\epsilon_2} g(x).$$

We recall that a vector $w \in T_x M$ is called an ϵ-subgradient (with $\epsilon \geq 0$) of f at $x \in \mathrm{dom}(f)$, denoted by $w \in \partial_\epsilon f(x)$, if

$$f(y) \geq f(x) + \langle w, \exp_x^{-1} y\rangle - \epsilon, \ \forall y \in M.$$

Thus $\partial_\epsilon h(x)$ and $\partial_\epsilon g(x)$ are an enlargement of $\partial h(x)$ and $\partial g(x)$, respectively. The use of elements in $\partial_\epsilon h(x)$ and $\partial_\epsilon g(x)$ instead of $\partial h(x)$ and $\partial g(x)$ allows an extra degree of freedom which is very useful in various applications. Setting $\epsilon = 0$ one retrieves the exact subdifferential. To this reason we consider the following inexact version of Algorithm DCPPA:

Algorithm (IDCPPA-1)

Step 1: Given an initial point $x^0 \in M$, a bounded sequence of positive numbers

$\{c_k\} \subset [b, c]$ and $\epsilon_k \geq 0$.
Step 2: Compute

$$w^k \in \partial_{\epsilon_k} h(x^k) \text{ and set } y^k := \exp_{x^k}(c_k w^k). \tag{15}$$

Step 3: Compute

$$x^{k+1} :\approx \arg\min_{x \in M} \{g(x) + \frac{1}{2c_k} d^2(x, y^k)\} \Leftrightarrow \frac{1}{c_k} \exp^{-1}_{x^{k+1}} y^k \in \partial_{\epsilon_k} g(x^{k+1}) \tag{16}$$

If $x^{k+1} = x^k$, stop. Otherwise, $k := k + 1$ and return to Step 2.

Theorem 5 *Let $\{x^k\}$ be a sequence generated by Algorithm IDCPPA-1. Suppose that $\{x^k\}$ is bounded and $\sum_{k=0}^{+\infty} \epsilon_k < \infty$. Then the sequence $\{f(x^k)\}$ is convergent and every cluster-point of $\{x^k\}$ is critical point of the function f.*

Proof Similar to Theorem 3, we have that

$$f(x^k) \geq f(x^{k+1}) + \frac{1}{c_k} d^2(x^k, x^{k+1}) - 2\epsilon_k.$$

Then,

$$\frac{1}{c} \sum_{k=0}^{n-1} d^2(x^k, x^{k+1}) \leq f(x^0) - f(x^n) + 2 \sum_{k=0}^{n-1} \epsilon_k.$$

Since f is bounded from below, the inequality above clearly implies that $\sum_{k=0}^{\infty} d^2(x^k, x^{k+1}) < \infty$, thanks to the summable assumption of $\{\epsilon_k\}$. Thus, $\lim_{k \to +\infty} d(x^k, x^{k+1}) = 0$.

Now, let x and y be cluster points of $\{x^k\}$ and $\{y^k\}$, respectively. So, consider two subsequences x^{k_j} and y^{k_l} converging respectively to x and y, i.e., $x^{k_j} \to x$ and $y^{k_l} \to y$. From definition of Algorithm IDCPPA-1, we have

$$h(z) \geq h(x^{k_j}) + \frac{1}{c} \langle \exp^{-1}_{x^{k_j}} y^{k_l}, \exp^{-1}_{x^{k_j}} z \rangle - \epsilon_{k_j}, \ \forall z \in M$$

and

$$g(z) \geq g(x^{k_j+1}) + \frac{1}{c} \langle \exp^{-1}_{x^{k_j+1}} y^{k_l}, \exp^{-1}_{x^{k_j+1}} z \rangle - \epsilon_{k_l}, \ \forall z \in M.$$

Since, $\lim_{k \to \infty} \epsilon_k = 0$ and $\lim_{k \to +\infty} d(x^k, x^{k+1}) = 0$, taking $k_j, k_l \to \infty$, we have that $\frac{1}{c} \exp^{-1}_x y \in \partial h(x)$ and $\frac{1}{c} \exp^{-1}_x y \in \partial g(x)$. The proof is complete. □

As remarked in Rockafellar [3], for a proximal point method to be practical, it is also important that it should work with approximate solutions of the subproblems. To the best of our knowledge, approximate solutions of proximal points has not been found to be explored in DC functions setting. For this reason, we provide an inexact proximal point algorithm for DC functions with a relative error tolerance.

Algorithm (IDCPPA-2)
Step 1: Given an initial point $x^0 \in M$ and a bounded sequence of positive numbers

$\{c_k\} \subset [b, c]$.
Step 2: Compute

$$w^k \in \partial h(x^k) \text{ and set } y^k := \exp_{x^k}(c_k w^k). \tag{17}$$

Step 3: Compute

$$e^{k+1} \in \partial g(x^{k+1}) - \frac{1}{c_k} \exp^{-1}_{x^{k+1}} y^k \tag{18}$$

where

$$||e^{k+1}|| \leq \eta d(x^{k+1}, x^k), \ \eta c_k \in [0, 1). \tag{19}$$

If $x^{k+1} = x^k$, stop. Otherwise, $k := k + 1$ and return to Step 2.

When $x^{k+1} = x^k$ or $\eta = 0$, (19) obviously implies that $e^{k+1} = 0$ and the Algorithm IDCPPA-2 reduces to the Algorithm DCPPA.

Theorem 6 *Let $\{x^k\}$ be a sequence generated by Algorithm IDCPPA-2. Suppose that $\{x^k\}$ is bounded, then the sequence $\{f(x^k)\}$ is convergent and every cluster-point of $\{x^k\}$ is a critical point of the function f.*

Proof Similar to Theorem 3, we have

$$f(x^k) \geq f(x^{k+1}) + \frac{1}{c_k} d^2(x^k, x^{k+1}) + \langle e^{k+1}, \exp^{-1}_{x^{k+1}} x^k \rangle,$$

which by (19) imply that

$$f(x^k) \geq f(x^{k+1}) + \frac{(1 - \eta c_k)}{c_k} d^2(x^k, x^{k+1}) > f(x^{k+1}),$$

if $x^{k+1} \neq x^k$. Otherwise, the algorithm stops. Since f is bounded from below, we have that the sequence $\{f(x^k)\}$ is convergent. Furthermore,

$$\frac{(1 - \eta c)}{c} \sum_{k=0}^{n-1} d^2(x^k, x^{k+1}) \leq f(x^0) - f(x^n).$$

The inequality above obviously implies that

$$\sum_{k=0}^{\infty} d^2(x^k, x^{k+1}) < \infty.$$

Thus $\lim_{k \to +\infty} d(x^k, x^{k+1}) = 0$.

Now, let x and y be cluster points of $\{x^k\}$ and $\{y^k\}$, respectively. So, consider two subsequences x^{k_j} and y^{k_j} converging respectively to x and y (here we will use the same notation for the index even if it needs extracting other subsequences), i.e., $x^{k_j} \to x$ and $y^{k_j} \to y$. From definition of the Algorithm IDCPPA-2, we have

$$h(z) \geq h(x^{k_j}) + \frac{1}{c_{k_j}} \langle \exp^{-1}_{x^{k_j}} y^{k_j}, \exp^{-1}_{x^{k_j}} z \rangle, \ \forall z \in M$$

and

$$g(z) \geq g(x^{k_j+1}) + \frac{1}{c_{k_j}} \langle \exp^{-1}_{x^{k_j+1}} y^{k_j}, \exp^{-1}_{x^{k_j+1}} z \rangle + \langle e^{k_j+1}, \exp^{-1}_{x^{k_j+1}} x^{k_j} \rangle, \ \forall z \in M.$$

By passing to the limit in the above relations, since $\lim_{k\to\infty} e^k = 0$, $\lim_{k\to+\infty} d(x^k, x^{k+1}) = 0$ and taking into account the fact that the functions g, h are lsc and $\{c_k\}$ is bounded, we have that $\partial h(x) \cap \partial g(x) \neq \emptyset$, in other words that x is a critical point of f. □

6 Example and application

In this section we present an example of a nonconvex minimization problem where the objective function is defined on the Poincaré half plane (a Hadamard manifold with curvature identically to -1). In the example, the proximal point algorithm proposed by Ferreira and Oliveira [6] does not apply. However, the method proposed in this article applies. Also, an application to constrained maximization problems on Hadamard manifolds is given.

6.1 Example

Consider the Poincaré upper half-plane $\mathbb{H} = \{(u,v) \in \mathbb{R}^2 \,:\, v > 0\}$ endowed with the Riemannian metric defined for every $(u,v) \in \mathbb{H}$ by

$$g_{ij}(u,v) = \frac{1}{v^2}\delta_{ij}, \text{ for } i,j = 1,2.$$

The pair $(\mathbb{H}, g)$ is a Hadamard manifold with constant sectional curvature -1 and the geodesics in $\mathbb{H}$ are the semi-lines and the semicircles orthogonal to the line $v = 0$ (see [34] page 20) with the following natural parameterizations

$$\gamma_a : u = a,\ v = e^s,\ s \in (-\infty, +\infty);$$
$$\gamma_{b,r} : u = b - r\tanh s,\ v = \frac{r}{\cosh s},\ s \in (-\infty, +\infty).$$

The geodesic passing at moment $s = s_0$ through the point $p = (x,y)$ tangent to the vector $w = (u,v) \in T_p\mathbb{H}$ is

$$\gamma(s) = \begin{cases} (x, ye^{s-s_0}), \text{ for } u = 0, v = y; \\ \left(x + \frac{v}{u}||w|| + \frac{||w||^2}{u}\tanh(s), -y\frac{||w||}{u}\frac{1}{\cosh(s)}\right), \text{ for } u \neq 0, \end{cases}$$

where $s \in [s_0, \infty)$ and $||w|| = u^2 + v^2$.

Consider the geodesic passing at moment $t = 0$ through the point $p = (x,y)$ tangent to the vector $w = (u,v) \in T_p\mathbb{H}$. Hence, the exponential map is defined by

$$\exp_p w = \gamma(1) = \begin{cases} (x, ye), \text{for } u = 0, v = y; \\ \left(x + \frac{v}{u}||w|| + \frac{||w||^2}{u}\tanh(1+s_0), -y\frac{||w||}{u}\frac{1}{\cosh(1+s_0)}\right), \text{ for } u \neq 0. \end{cases}$$

The Riemannian distance between two points $(u_1,v_1),(u_2,v_2) \in \mathbb{H}$ is given by

$$d((u_1,v_1),(u_2,v_2)) = \operatorname{arccosh}\left(1 + \frac{(u_2-u_1)^2 + (v_2-v_1)^2}{2v_1v_2}\right).$$

Let $f : \mathbb{H} \to \mathbb{R}$ be a function given by $f(x,y) = x^4 + y^4 - 2x^2 - 2y^2 + 3$. Note that f is bounded from below and f is not a convex function on $\mathbb{H}$ while $g(x,y) = x^4 + y^4$ and $h(x,y) = 2x^2 + 2y^2 + 3$ are convex functions on $\mathbb{H}$. Clearly, the set of critical points of f is nonempty. Therefore, f agree with the assumptions made. Thus, the Algorithm DCPPA can be applied.

6.2 Application to constrained maximization problems

We consider the problem of maximizing a convex lower semi-continuous function h on a closed convex set $C \subset M$, namely

$$\max_{x \in C} h(x). \tag{20}$$

This problem can be rewritten as a DC problem, and (20) is equivalent to the following problem:

$$-\min_{x \in M} \{\delta_C(x) - h(x)\}, \tag{21}$$

where $\delta_C(x)$ is the indicate function defined by $\delta_C(x) = 0$ if $x \in C$ and $\delta_C(x) = +\infty$ otherwise. Let $N_C(x)$ denote the normal cone of the set C at a point $x \in C$:

$$N_C(x) := \{u \in T_xM;\ \langle u, \exp_x^{-1} y\rangle \leq 0\ \forall y \in C\}.$$

Then

$$\partial \delta_C(x) = N_C(x),\ \forall x \in C.$$

In this context Algorithm DCPPA takes the following form:

Compute $w^k \in \partial h(x^k)$ and set $y^k = exp_{x^k}(c_k w^k)$. Define $x^{k+1} \in M$ as the solution of the following variational inequality problem:

$$\langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} y\rangle \leq 0,\ \forall y \in C.$$

Existence and uniqueness theorems for variational inequalities on Hadamard manifolds can be found, for instance in [11], [13].

Acknowledgements The authors wish to express their gratitude to the anonymous referee for his helpful comments.

References

1. Martinet, B., Regularisation d'inéquations variationelles par approximations succesives, Rev. Franaise d'Inform. Recherche Oper., **4**, 154–159 (1970)
2. Moreau, J.J., Proximité et dualité dans un espace Hilbertien, Bull. Soc. Math. France, **93**, 273–299 (1965)
3. Rockafellar, R.T., Monotone operators and the proximal point algorithm, SIAM J. control. optim., **14**, 877–898 (1976)
4. Bento, G.C., Ferreira, O.P., Oliveira, P.R., Local convergence of the proximal point method for a special class of nonconvex functions on Hadamard manifolds. Nonlinear Anal. **73**, 564–572 (2010)
5. Papa Quiroz, E.A.., Oliveira, P.R., Proximal point method for minimization quasiconvex locally Lipschitz functions on Hadamard manifolds. Nonlinear Anal. **75**, 5924–5932 (2012)
6. Ferreira, O.P., Oliveira, P.R., Proximal point algorithm on Riemannian manifolds, Optimization, **51** 257–270 (2002)
7. da Cruz Neto, J.X., Ferreira, O.P., Lucambio Pérez, L.R., Németh, S.Z.: Convex-and monotone- transformable mathematical programming problems and a proximal-like point algorithm. J. Glob. Optim. **35**, 53–69 (2006)
8. Ferreira, O.P., Oliveira, P.R.: Subgradient algorithm on Riemannian manifolds. J. Optim. Theory Appl. **97**, 93–104 (1998)
9. da Cruz Neto, J.X., de Lima, L.L., Oliveira, P.R.: Geodesic algorithms in Riemannian geometry. Balk. J. Geom. Appl. **3**, 89–100 (1998)

10. Kristály, A.: Nash-type equilibria on Riemannian manifolds: a variational approach. J. Math. Pures Appl. **101**, 660–688 (2014)
11. Li, C., López, G., Martín-Márquez, V., Monotone vector fields and the proximal point algorithm on Hadamard manifolds. J. Lond. Math. Soc. **79**, 663–683 (2009)
12. 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)
13. Németh, S.Z., Variational inequalities on Hadamard manifolds. Nonlinear Anal. **52**, 1491–1498 (2003)
14. Wang, J.H., López, G., Martín-Márquez, V., Li, C., Monotone and accretive vector fields on Riemannian manifolds. J. Optim. Theory Appl. **146**, 691–708 (2010)
15. Li, C., Wang, J.H., Newton's method for sections on Riemannian manifolds: generalized covariant α-theory. J. Complex. **24**, 423–451 (2008)
16. Li, C., Mordukhovich, B.S., Wang, J.H., Yao, J.C., Weak sharp minima on Riemannian manifolds. SIAM J. Optim. **21(4)**, 1523–1560 (2011)
17. Bento, G.C., Ferreira, O.P., Oliveira, P.R., Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds, J. Optim. Theory Appl., **154**, 88–107 (2012)
18. Li, C., Yao, J.C., Variational inequalities for set-valued vector fields on Riemannian manifolds: convexity of the solution set and the proximal point algorithm. SIAM J. Control Optim. **50(4)**, 2486–2514 (2012)
19. Li, C., López, G., Wang, J.H., Yao, J.C., Convergence analysis of inexact proximal point algorithms on Hadamard manifolds, J Glob Optim, DOI 10.1007/s10898-014-0182-2 (2014)
20. Huang, N., Tang, G., An inexact proximal point algorithm for maximal monotone vector fields on Hadamard manifolds, Operations Research Letters, **41**, 586–591 (2013)
21. Absil, P.A., Baker, C.G., Trust-region methods on Riemannian manifolds, Found. Comput. Math., **7**, 303–330 (2007)
22. Adler, R.L., Dedieu, J.P., Margulies, J.Y., Martens, M., Shub, M., Newton's method on Riemannian manifolds and a geometric model for the human spine, IMA Journal of Numerical Analysis, **22**, 359–390 (2002)
23. Lee, P.Y., Geometric Optimization for Computer Vision, PhD thesis, Australian National University (2005)
24. Riddell, R.C., Minimax problems on Grassmann manifolds. Sums of eigenvalues, Advances in Mathematics, 54, 107–199 (1984)
25. Hiriart-Urruty, J.B., From convex optimization to nonconvex optimization: necessary and sufficient conditions for global optimization. Nonsmooth Optimization and Related Topics, Springer US, 219–239 (1989)
26. Hiriart-Urruty, J.B., Generalized differentiabity, duality and optimization for problems dealing with difference of convex functions. Convexity and Duality in Optimization, Springer Berlin Heidelberg, 37–70 (1985)
27. Hiriart-Urruty, J.B., Tuy, H., Essays on nonconvex optimization, Mathematical Programming, **41**, North-Holland (1988)
28. Elhilali Alaoui, A. Caractrisation des fonctions D.C. (Characterization of D. C. functions), Ann. Sci. Math. Qu. **20**, No.1, 1–13 (1996)
29. Toland, J.F., Duality in nonconvex optimization, Journal of Mathematical Analysis and Applications, **66**, 399–415 (1978)
30. Sun, W., Sampaio, R.J.B., Candido, M.A.B., Proximal point algorithm for minimization of DC Functions, Journal of Computational Mathematics, **21**, 451–462 (2003)
31. Moudafi, A., Maing, P-E., On the convergence of an approximate proximal method for d.c. functions, Journal of Computational Mathematics, **24**, 475–480 (2006)
32. do Carmo, M.P.: Riemannian Geometry. Birkhauser, Boston (1992)
33. Sakai, T.: Riemannian Geometry. Translations of Mathematical Monographs, **149**, American Mathematical Soc., Providence (1996)
34. Udriste, C., Convex Functions and Optimization Algorithms on Riemannian Manifolds. Mathematics and Its Applications, **297**, Kluwer Academic, Dordrecht (1994)
35. Polyak, B. T., Sharp Minima., Institute of Control Sciences Lecture Notes, Moscow, USSR, 1979; Presented at the IIASA Workshop on Generalized Lagrangians and Their Applications, IIASA, Laxenburg, Austria (1979)
36. Ferris, M. C., Weak Sharp Minima and Penalty Functions in Mathematical Programming, Ph.D. Thesis, University of Cambridge, UK (1988)
37. Bento G.C., Cruz Neto, J.X., Finite Termination of the Proximal Point Method for Convex Functions on Hadamard Manifolds, Optimization, **63**, 1281–1288 (2014)