An inertial proximal point method for difference of maximal monotone vector fields in Hadamard manifolds

João S. Andrade, Jurandir de O. Lopes, João Carlos de O. Souza

An inertial proximal point method for difference of maximal monotone vector fields in Hadamard manifolds

João S. Andrade[1,2] · Jurandir de O. Lopes[2] · João Carlos de O. Souza[2,3]

Abstract
We propose an inertial proximal point method for variational inclusion involving difference of two maximal monotone vector fields in Hadamard manifolds. We prove that if the sequence generated by the method is bounded, then every cluster point is a solution of the non-monotone variational inclusion. Some sufficient conditions for boundedness and full convergence of the sequence are presented. The efficiency of the method is verified by numerical experiments comparing its performance with classical versions of the method for monotone and non-monotone problems.

Keywords Variational inclusion · Proximal point method · Monotone vector fields · DC functions · Hadamard manifolds

Mathematics Subject Classification 65K05 · 90C26 · 90C48

1 Introduction

The problem of finding a zero of maximal monotone operators T, i.e.,

$$0 \in T(x) \tag{1}$$

includes, as special cases, optimization and min-max problems, complementarity problems, and variational inequalities. It finds many important applications in scientific fields such as image processing, computer vision, machine learning and signal processing. For this reason, in recent years much attention has been given to develop efficient and implementable

✉ João Carlos de O. Souza
joaocos.mat@ufpi.edu.br

João S. Andrade
joaosa.mat@ufpi.edu.br

Jurandir de O. Lopes
jurandir@ufpi.edu.br

1 CSHNB, Federal University of Piauí, Picos, PI 64607-670, Brazil

2 Department of Mathematics, CCN, Federal University of Piauí, Teresina, PI 64049-550, Brazil

3 Aix Marseille University, CNRS, AMSE, Marseille, France

numerical methods for solving this problem in different contexts; see for instance [1–3] and references therein.

One of the fundamental approaches for solving (1) is the proximal point method which its origin can be traced back to Martinet [4] in the context of convex minimization and Rockafellar [38] in the general setting of maximal monotone operators in Hilbert space. In the Riemannian setting this method was studied by Li et al. [5] based on first extension of the proximal point method to the Riemannian context proposed by Ferreira and Oliveira [6] for convex minimization problems.

The proximal point method computes at each iteration the well known resolvent operator introduced by Moreau [7]. Unfortunately, in many interesting cases, the evaluation of the resolvent operator is as difficult as solving the original problem. On the other hand, in many problems, the operator T can be written as the sum of two maximal monotone operators, namely $T = A + B$, such that the resolvent operator of each component is much easier to compute than the original operator T. The so-called forward–backward methods overcome this drawback combining the resolvents of each component to find a solution of the original problem; for instance the Douglas–Rachford algorithm [8] among others.

It is well known that the sum of two monotone operators is a monotone operator, whereas the difference of two monotone operators is not necessarily a monotone operator. Therefore, the problem of finding a zero of the difference of two monotone operators can be very difficult. It generalizes the problem of finding the critical points of the difference of two convex functions (DC functions) and was not studied extensively yet (even in the linear setting); see for instance [9–13]. In this direction, Souza and Oliveira [14] proposed the first extension of the proximal point method for DC functions in Hadamard manifolds.

On the other hand, Polyak [15] introduced the so-called heavy ball method for minimizing a smooth convex function. The difference compared to the gradient method is that each iteration, an extrapolation point (which combines the current and the previous iterates) is used instead of the current iterate. This minor change improves the performance of the gradient method. Alvarez and Attouch [16] adapted this idea (called inertial method) to the proximal point method for maximal monotone operators. Inertial methods were considered in the particular case of DC functions in the linear context; see for instance [17, 18]. However, to the best of our knowledge, there does not exist any study of inertial proximal method for variational inclusion involving the difference two monotone vector fields in Hadamard manifolds and, in particular, in the linear setting. The present work is a contribution towards this goal.

The aim of this paper is to proposed an inertial version of the proximal point method for the variational inclusion problem

$$0 \in A(x) - B(x), \tag{2}$$

where A and B are maximal monotone vector fields on finite dimensional Hadamard manifold M. It is equivalent to the problem of finding $x \in M$ such that

$$A(x) \cap B(x) \neq \emptyset. \tag{3}$$

We illustrate with some preliminaries numerical experiments that the inertial version of the proximal point method has a better performance compared to its classical version.

Although we propose an inertial version of the proximal point method for variational inclusion involving the difference of two maximal monotone vector fields in Hadamard manifolds, some results related to the proposed method are new for DC functions in the linear setting as well as in some particular instances in Hadamard manifolds, namely, if $B(x) = 0$

in (2) we have an inertial version of the proximal point method considered by Li et al. [5] for finding a singularity of a maximal monotone vector field; if $A(x) = \partial g(x)$ and $B(x) = \partial h(x)$ in (3), where g, h are convex functions, then our method becomes an inertial version of the proximal point method for DC functions ($f(x) = g(x) - h(x)$) proposed by Souza and Oliveira [14]; if $A(x) = \partial f(x)$ and $B(x) = 0$ in (2), where f is a convex function, then we have an inertial version of the proximal point method proposed by Ferreira and Oliveira [6] for convex minimization problems; and finally, if the parameter which appears in the direction of the inertial method is zero, we have the (non-boosted) proximal point method for difference of maximal monotone operators extending the algorithm proposed by Souza and Oliveira [14].

In our convergence analysis, we prove that every cluster point of the sequence generated by the inertial method, if any, is a solution of the variational inclusion problem. Furthermore, we present some sufficient conditions for boundedness and full convergence of the method which are new even for DC functions in Hadamard manifolds. To show the efficiency of the method, we provide some numerical experiments solving minimization problems in a genuine (with non-constant curvature different from zero) Hadamard manifold involving convex and non-convex functions as well as variational inclusion involving operators which are not the subdifferential of convex functions in Euclidean space.

The study of algorithms for solving non-monotone variational inclusion is a very difficult problem and interesting by itself. The variational inclusion problem involving the difference of maximal monotone vector fields generalizes the very important problem of minimizing DC functions which has a lot of applications. Another interesting application of difference of monotone operators in linear spaces is given by Attouch and Théra [19]. They mention that most of the equations arising in physics, economics, among other, can be written as follows

$$0 \in C(x) = A(x) + B(x), \tag{4}$$

where A, B are possibly multivalued operators. The splitting of the operator C into the sum of two operators A and B has usually a deep physical or economical meaning since A and B may have very distinct properties. Instead of directly study (4), the authors in [19] proposed an equivalent duality transformation and they consider the following difference of multivalued operators

$$0 \in A^{-1}(y) - B^{-1}(-y)$$

offering a new duality approach to some central questions in the theory of variational inequalities and maximal monotone operators.

The remainder of this paper is organized as follows. In Sect. 2, we recall some notations, definitions and preliminary results in Riemannian manifolds, convexity and vector fields in Hadamard manifolds which will be used for further analysis. In Sect. 3, we present our algorithm as well as its well definition. In Sect. 4, we establish convergence analysis of the proposed algorithm under some mild conditions for the non-monotone case. In Sect. 5, we present the convergence analysis of the method for the monotone case. In Sect. 6 some numerical experiments are reported to support the theoretical results obtained and illustrate the feasibility and efficiency of the proposed algorithm comparing its performance with the classical version of the proximal point method in both Euclidean space and Hadamard manifolds. Some concluding remarks are given in the last section.

2 Notation and basic concepts

The standard notations, results and preliminary concepts of Riemannian geometry used throughout the paper can be found, for instance, in Sakai [20], Udriste [21] and do Carmo [22]. We follow the notation, terminology and results of [14] and [23].

Throughout this paper, we will assume that M a finite dimensional Hadamard manifold. We denote by T_xM the tangent space of M at x. Recall that the parallel transport along the geodesic γ from $\gamma(a)$ to $\gamma(b)$ is denoted by $P_{\gamma,\gamma(b),\gamma(a)} : T_{\gamma(a)}M \to T_{\gamma(b)}M$. For any a, b, the parallel transport $P_{\gamma,\gamma(b),\gamma(a)}$ is an isometry from $T_{\gamma(a)}M$ to $T_{\gamma(b)}M$. Note that, for any a, b, a_1, b_1, we have

$$P_{\gamma,\gamma(b_2),\gamma(b_1)} \circ P_{\gamma,\gamma(b_1),\gamma(a)} = P_{\gamma,\gamma(b_2),\gamma(a)} \text{ and } P^{-1}_{\gamma,\gamma(b),\gamma(a)} = P_{\gamma,\gamma(a),\gamma(b)}.$$

If $\gamma(a) = p$ and $\gamma(b) = q$, we will write $P_{q,p}$ instead of $P_{\gamma,\gamma(b),\gamma(a)}$ in the case when γ a minimal geodesic joining p to q and no confusion arises. The restriction of a geodesic to a closed bounded interval is called a geodesic segment. Given points $x, y \in M$, we denote the geodesic segment from x to y by $[x, y]$. We usually do not distinguish between a geodesic and its geodesic segment, as no confusion can arise. We denote by $\exp_x : T_xM \to M$ the exponential map. For any $x \in M$ we can define the exponential inverse mapping $\exp_x^{-1} : M \to T_xM$ which is C^∞. Since $d(x, x') = \| \exp^{-1}_{x'}(x)\|$, then the map $\rho_{x'} : M \to \mathbb{R}$ defined by $\rho_{x'}(x) = \frac{1}{2}d^2(x, x')$ is C^∞ and its gradient at x, denoted by $\mathrm{grad}\rho_{x'}(x)$, is given by $\mathrm{grad}\rho_{x'}(x) = -\exp_x^{-1}(x')$; see for instance [6, Proposition 3.3].

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 $x \in M$ and $\{x^k\} \subset M$ be such that $x^k \to x$. Then the following assertions hold.*

1. *For any $y \in M$, we have*

$$\exp^{-1}_{x^k} y \to \exp^{-1}_x y \text{ and } \exp^{-1}_y x^k \to \exp^{-1}_y x.$$

2. *If $v^k \in T_{x^k}M$ and $v^k \to v$, then $v \in T_xM$.*
3. *Given $u^k, v^k \in T_{x^k}M$ and $u, v \in T_xM$, if $u^k \to u$ and $v^k \to v$, then*

$$\langle u^k, v^k\rangle \to \langle u, v\rangle.$$

4. *For any $u \in T_xM$, the function $F : M \to TM$ defined by $F(z) = P_{z,x}u$ for each $z \in M$ is continuous on M.*

Proof See [5, Lemma 2.4]. □

Let $\mathscr{X}(M)$ denote the set of all multivalued vector fields $A : M \longrightarrow 2^{TM}$ such that $A(x) \subset T_xM$, for each $x \in M$, $TM = \cup_{x\in M} T_xM$ and the domain of A is closed and convex, where the domain $D(A)$ of A is defined by

$$D(A) = \{x \in M;\ A(x) \neq \varnothing\}.$$

Definition 1 Let $A \in \mathscr{X}(M)$. Then A is said to be:

(i) Monotone if the following condition holds for any $x, y \in D(A)$:

$$\langle u, \exp_x^{-1} y\rangle \le \langle v, -\exp_y^{-1} x\rangle, \quad \forall u \in A(x),\ v \in A(y); \tag{5}$$

(ii) Strictly monotone if (5) holds with strict inequality for any $x, y \in D(A)$ with $x \neq y$, that is,

$$\langle u, \exp_x^{-1} y\rangle < \langle v, -\exp_y^{-1} x\rangle, \quad \forall u \in A(x),\ v \in A(y);$$

(iii) Strongly monotone if there exists $\rho > 0$ such that, for any $x, y \in D(A)$, we have

$$\langle u, \exp_x^{-1} y\rangle - \langle v, -\exp_y^{-1} x\rangle \leq -\rho d^2(x, y), \quad \forall u \in A(x),\ v \in A(y);$$

(iv) Maximal monotone if it is monotone and the following implication holds for any $x \in D(A)$ and $u \in T_x M$:

$$\langle u, \exp_x^{-1} y\rangle \leq \langle v, -\exp_y^{-1} x\rangle, \quad \forall y \in D(A),\ v \in A(y) \Longrightarrow u \in A(x).$$

Definition 2 Let $A \in \mathscr{X}(M)$ and $p \in D(A)$. Then A is said to be

(i) Upper semicontinuous at p if, for any open set V satisfying $A(p) \subset V \subset T_p M$, there exists an open neighbourhood $U(p)$ of p such that $P_{p,x} A(x) \subset V$ for any $x \in U(p)$;
(ii) Upper Kuratowski semicontinuous at p if, for any sequences $\{x^k\} \subseteq D(A)$ and $\{u^k\} \subset TM$ with each $u^k \in A(x^k)$, the relations $\lim_{k\to\infty} x^k = p$ and $\lim_{k\to\infty} u^k = u$ imply that $u \in A(p)$.
(iii) Locally bounded at p if there exists an open neighbourhood $U(p)$ of p such that the set $\cup_{x\in U(p)} A(x)$ is bounded;
(iv) Upper semicontinuous (resp. upper Kuratowski semicontinuous, locally bounded) on M if it is upper semicontinuous (resp. upper Kuratowski semicontinuous, locally bounded) at each point $p \in D(A)$.

Proposition 2 *Suppose that $A \in \mathscr{X}(M)$ is maximal monotone and $D(A) = M$. Then A is locally bounded on M.*

Proof See [5, Lemma 3.6]. □

Remark 1 Suppose that $A \in \mathscr{X}(M)$ is maximal monotone and $D(A) = M$. If $\{x^k\}$ is bounded and $v^k \in A(x^k)$ for all $k \in \mathbb{N}$, then $\{v^k\}$ is bounded. Indeed by boundedness of $\{x^k\}$ there exists $C > 0$ such that $d(x^0, x^k) \leq C$, for all $k \in \mathbb{N}$. Define the open neighbourhood of x^0 given by $V = \{x \in M; d(x^0, x) < C + 1\}$. It follows from Proposition 2 and Definition 2 (iii) that $\cup_{x\in V} A(x)$ is bounded. Since $x^k \in V$ and $v^k \in A(x^k)$, we conclude that $\{v^k\}$ is bounded and the claim is proved.

Definition 3 Given $\lambda > 0$ and $A \in \mathscr{X}(M)$, the resolvent of order λ is the set-valued mapping $J_\lambda^A : M \longrightarrow 2^M$ defined by

$$J_\lambda^A(x) = \{z \in M;\ x \in \exp_z \lambda A(z)\}, \quad \forall x \in M.$$

Remark 2 As mentioned in [24, Remark 3] by the definition of the resolvent of a vector field, the domain of the resolvent J_λ^A is the range of the vector field defined by $x \mapsto \exp_x \lambda A(x)$. We will denote this range as $\mathscr{R}(\exp_{(\cdot)} \lambda A(\cdot))$. Then we have that

$$D(J_\lambda^A) = \mathscr{R}(\exp_{(\cdot)}\ \lambda A(\cdot)).$$

Definition 4 Let $K \subseteq M$ be a non-empty, closed and convex set. Given a mapping $T : K \subseteq M \to M$, we say that T is firmly non-expansive if for any $x, y \in K$, the function $\Phi : [0, 1] \to \mathbb{R}_+$ defined by

$$\Phi(t) = d(\exp_x t \exp_x^{-1} Tx, \exp_y t \exp_y^{-1} Ty), \quad \forall t \in [0, 1],$$

is non-increasing.

Next, we state a well known result on firmly non-expansive mappings.

Proposition 3 *Let $K \subseteq M$ be a non-empty, closed and convex set and $T : K \subseteq M \to M$. The following assertions are equivalent.*

1. *T is firmly non-expansive;*
2. *For any $x, y \in K$ and $t \in [0, 1]$, one has*

$$d(T(x), T(y)) \leq d(\exp_x t \exp_x^{-1} Tx, \exp_y t \exp_y^{-1} Ty).$$

Proof See [24, Proposition 5]. □

The next result establishes the relation between the firm non-expansivity of the resolvent and the monotonicity of the corresponding vector field.

Proposition 4 *Let $A \in \mathscr{X}(M)$. The following assertions hold for any $\lambda > 0$.*

(i) The vector field A is monotone if and only if J_λ^A is single-valued and firmly non-expansive.
(ii) if $D(A) = M$, the vector field A is maximal monotone if and only if J_λ^A is single-valued, firmly non-expansive and the domain $D(J_\lambda^A) = M$.

Proof See [24, Theorem 4]. □

From Proposition 4 and Remark 2, we have the following result which constitutes a counterpart in the setting of Hadamard manifolds of the well known Minty's theorem.

Corollary 1 *Let $A \in \mathscr{X}(M)$ be monotone such that $D(A) = M$, and let $\lambda > 0$. Then A is maximal monotone if and only if $\mathscr{R}(exp_{(\cdot)} \lambda A(\cdot)) = M$.*

3 An inertial proximal point method

Throughout this paper, we assume that M is a finite dimensional Hadamard manifold and $A, B \in \mathscr{X}(M)$ are maximal monotone vector fields with $D(A) = D(B) = M$.

We are interested in solving the following problem:

$$\text{find } x^* \in M \text{ such that } A(x^*) \cap B(x^*) \neq \varnothing. \tag{6}$$

A point $x \in M$ satisfying (6) is said to be the critical point of the difference $A - B$. The set of critical points of $A - B$ is defined by

$$S = \{x \in M; \quad A(x) \cap B(x) \neq \varnothing\}.$$

It is well known that the sum of two monotone vector fields remains monotone. However, the difference of two monotone vector fields is possibly non-monotone even in the Euclidean setting as we can see in the following simple example.

Example 1 Let $g, h : \mathbb{R} \to \mathbb{R}$ be convex functions given by $g(x) = \frac{1}{4}x^4$ and $h(x) = \frac{1}{2}x^2$. Recall that the subdifferential of a convex function is a monotone operator, i.e., $A(x) = \nabla g(x)$ and $B(x) = \nabla h(x)$ are monotone operators. However, we can easily see that the function f given by $f(x) = g(x) - h(x)$ is not convex and its subdifferential is not monotone. On the other hand, the set of critical points of f satisfies (6).

In this paper, we suppose that $S \neq \varnothing$ and A or B is ρ-strongly monotone.

Remark 3 Although in our proof we need the strongly monotonicity of only one of the vector fields A and B, it is worth to mention that the strongly monotonicity of both A and B is not a restrictive assumption. In fact, if it does not hold for A and B, then we can obtain another decomposition satisfying this condition as follows

$$T(x) = A(x) - B(x) = [A(x) - \rho \exp_x^{-1} y] - [B(x) - \rho \exp_x^{-1} y],$$

for $\rho > 0$ arbitrary and $y \in M$ fixed. One has that $\tilde{A}(x) = A(x) - \rho \exp_x^{-1} y$ and $\tilde{B}(x) = B(x) - \rho \exp_x^{-1} y$ are ρ-strongly monotone; see [5, Remark 4.4].

Inertial proximal point method (IPPM)

Step 1: Given an initial point $x^0 \in M$, $\gamma_k \in [0, \frac{\rho}{2})$ and a bounded sequence of positive numbers $\{\mu_k\}$ (to be specified latter). Define $x^{-1} = x^0$.

Step 2: Given $x^k \in M$, define $d^k = \gamma_k \exp_{x^k}^{-1} x^{k-1}$. Find

$$w^k \in B(x^k) \text{ and set } y^k = \exp_{x^k} \mu_k (w^k + d^k). \tag{7}$$

Step 3: Compute $x^{k+1} \in M$ such that

$$0 \in A(x^{k+1}) - \frac{1}{\mu_k} \exp_{x^{k+1}}^{-1} y^k. \tag{8}$$

If $x^{k+1} = x^k$ and $d^k = 0$, stop. Otherwise, set $k = k + 1$ and return to Step 2.

Remark 4 The well definition of the sequences $\{y^k\}$ and $\{x^k\}$ directly follows from the fact the the exponential map is a global diffeomorphism in Hadamard manifolds and the vector field defined by $x \mapsto A(x) - \frac{1}{\mu_k} \exp_x^{-1} y^k$ is strongly monotone, and hence, it has a unique singularity, respectively; see [5, Theorem 4.3 and Remark 4.4]. Moreover, from (7) and (8), we have

$$\frac{1}{\mu_k} \exp_{x^k}^{-1} y^k - d^k \in B(x^k) \quad \text{and} \quad \frac{1}{\mu_k} \exp_{x^{k+1}}^{-1} y^k \in A(x^{k+1}).$$

Thus, if $x^{k+1} = x^k$ and $d^k = 0$, we obtain

$$\frac{1}{\mu_k} \exp_{x^k}^{-1} y^k \in B(x^k) \quad \text{and} \quad \frac{1}{\mu_k} \exp_{x^k}^{-1} y^k \in A(x^k).$$

Therefore, $\frac{1}{\mu_k} \exp_{x^k}^{-1} y^k \in A(x^k) \cap B(x^k)$, i.e., $A(x^k) \cap B(x^k) \neq \varnothing$ and hence x^k is a solution of (6).

Remark 5 In the linear setting, Alimohammady and Ramazannejad [11] proposed an inertial proximal point algorithm for the difference of maximal monotone operators based on the method considered by Moudafi [12]. These methods are based on the resolvent operator and the Yosida approximate. It is worth to mention that in the linear setting our method is different from the algorithms proposed in [9, 11, 12]. Note that if $\gamma_k = 0$, for all $k \in \mathbb{N}$ in IPPM, then $d^k = 0$ and (7) becomes

$$w^k \in B(x^k) \text{ and } y^k = \exp_{x^k} \mu_k w^k.$$

Since, $x^{k+1} \in M$ is defined as (8), if $B(x) = \partial h(x)$ and $A(x) = \partial g(x)$, where $g, h : M \to \mathbb{R}$ are convex functions, then (8) is equivalent to

$$x^{k+1} = \arg\min_{x \in M} \{g(x) + \frac{1}{\mu_k} d^2(x, y^k)\}$$

which is the proximal point method for DC functions proposed by [14]. Therefore, IPPM with $\gamma_k = 0$, for all $k \in \mathbb{N}$, ca be viewed as an extension to difference of maximal monotone vector fields of the method considered in [14] for difference of convex functions. In this case, our method is different from the regularized methods for the difference of maximal monotone operators proposed by Moudafi [9, 12] in the linear setting.

It is worth to mention that IPPM is new, in the Hadamard setting, even for solving monotone vector fields. Therefore, we consider a monotone version of IPPM by doing $B(x) = 0$. Then, IPPM becomes the following:

IPPM: monotone version (mIPPM)

Step 1: Given an initial point $x^0 \in M$, $\gamma_k \in [0, \frac{\rho}{2})$ and a bounded sequence of positive numbers $\{\mu_k\}$ (to be specified latter). Define $x^{-1} = x^0$.

Step 2: Given $x^k \in M$, define $d^k = \gamma_k \exp^{-1}_{x^k} x^{k-1}$. Find

$$y^k = \exp_{x^k} \mu_k d^k. \tag{9}$$

Step 3: Compute $x^{k+1} \in M$ such that

$$0 \in A(x^{k+1}) - \frac{1}{\mu_k} \exp^{-1}_{x^{k+1}} y^k. \tag{10}$$

If $x^{k+1} = x^k$ and $d^k = 0$, stop. Otherwise, set $k = k + 1$ and return to Step 2.

In the sequel, we present a convergence analysis for both algorithms IPPM for the difference of maximal monotone vector fields (which is a possibly non-monotone vector field) and mIPPM for finding a zero of a monotone vector field. To this end, we will consider their convergence analysis separately.

4 Convergence analysis: possibly non-monotone case

From now on, we consider $\{x^k\}$ the sequence generated by IPPM and we assume that $x^{k+1} \neq x^k$, for all $k \in \mathbb{N}$, otherwise the algorithm returns a solution of the problem. Now we shall establish its convergence properties.

Proposition 5 *Suppose that A or B is ρ-strongly monotone. Then,*

$$\left(\rho + \frac{1}{b}\right) d(x^{k+1}, x^k) \leq \| P_{x^k, x^{k+1}} u^{k+1} - v^k \| + \frac{\rho}{2} d(x^k, x^{k-1}), \tag{11}$$

for any $v^k \in A(x^k)$ and $u^{k+1} \in B(x^{k+1})$.

Proof We will suppose that A is monotone and B is ρ-strongly monotone. The other case is analogous. By (7) e (8), we have

$$\frac{1}{\mu_k} \exp^{-1}_{x^k} y^k - d^k \in B(x^k) \quad \text{and} \quad \frac{1}{\mu_k} \exp^{-1}_{x^{k+1}} y^k \in A(x^{k+1}).$$

Given $x \in M$, since B is ρ-strongly monotone, then

$$\langle \frac{1}{\mu_k} \exp^{-1}_{x^k} y^k - d^k, \exp^{-1}_{x^k} x \rangle - \langle u, -\exp^{-1}_x x^k \rangle \leq -\rho d^2(x, x^k), \quad \forall\, u \in B(x). \tag{12}$$

It follows from monotonicity of A that

$$\langle \frac{1}{\mu_k} \exp^{-1}_{x^{k+1}} y^k, \exp^{-1}_{x^{k+1}} x \rangle \leq \langle v, -\exp^{-1}_x x^{k+1} \rangle, \quad \forall\, v \in A(x). \tag{13}$$

Taking $x = x^{k+1}$ in (12), $x = x^k$ in (13) and summing up these inequalities, we have

$$\frac{1}{\mu_k}[\langle \exp_{x^k}^{-1} y^k - d^k, \exp_{x^k}^{-1} x^{k+1}\rangle + \langle exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k\rangle] \\ \leq \langle u^{k+1}, -\exp_{x^{k+1}}^{-1} x^k\rangle + \langle v^k, -\exp_{x^k}^{-1} x^{k+1}\rangle - \rho d^2(x^{k+1}, x^k) + \langle d^k, \exp_{x^k}^{-1} x^{k+1}\rangle, \tag{14}$$

for any $v^k \in A(x^k)$ and $u^{k+1} \in B(x^{k+1})$. Let $\triangle(y^k, x^k, x^{k+1})$ be the geodesic triangle with $\theta = \angle(\exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1})$. Then, by the Comparison Theorem for Triangles [20, Proposition 4.5], we have

$$d^2(y^k, x^k) + d^2(x^k, x^{k+1}) - 2\langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1}\rangle \leq d^2(y^k, x^{k+1}). \tag{15}$$

Similarly, to the geodesic triangle $\triangle(y^k, x^{k+1}, x^k)$ with $\theta = \angle(\exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k)$, we have

$$d^2(y^k, x^{k+1}) + d^2(x^{k+1}, x^k) - 2\langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k\rangle \leq d^2(y^k, x^k). \tag{16}$$

Adding (15) and (16), we obtain

$$d^2(x^{k+1}, x^k) \leq \langle \exp_{x^k}^{-1} y^k, \exp_{x^k}^{-1} x^{k+1}\rangle + \langle \exp_{x^{k+1}}^{-1} y^k, \exp_{x^{k+1}}^{-1} x^k\rangle. \tag{17}$$

Since $\mu_k > 0$, for all $k \in \mathbb{N}$, and using (17) in (14), we have

$$\frac{1}{\mu_k} d^2(x^{k+1}, x^k) \\ \leq \langle u^{k+1}, -\exp_{x^{k+1}}^{-1} x^k\rangle + \langle v^k, -\exp_{x^k}^{-1} x^{k+1}\rangle - \rho d^2(x^{k+1}, x^k) + \langle d^k, \exp_{x^k}^{-1} x^{k+1}\rangle. \tag{18}$$

Thus, using the fact that $0 < \mu_k \leq b$, we obtain

$$\left(\rho + \frac{1}{b}\right) d^2(x^{k+1}, x^k) \leq \langle u^{k+1}, -\exp_{x^{k+1}}^{-1} x^k\rangle + \langle v^k, -\exp_{x^k}^{-1} x^{k+1}\rangle + \langle d^k, \exp_{x^k}^{-1} x^{k+1}\rangle, \tag{19}$$

for any $v^k \in A(x^k)$ and $u^{k+1} \in B(x^{k+1})$.

On the other hand, using the parallel transport properties, we have that $\exp_{x^k}^{-1} x^{k+1} = -P_{x^k, x^{k+1}} \exp_{x^{k+1}}^{-1} x^k$, and hence

$$\langle u^{k+1}, -\exp_{x^{k+1}}^{-1} x^k\rangle + \langle v^k, -\exp_{x^k}^{-1} x^{k+1}\rangle = \langle P_{x^k, x^{k+1}} u^{k+1} - v^k, \exp_{x^k}^{-1} x^{k+1}\rangle$$

and hence, using this fact in (19), one has

$$(\rho + \frac{1}{b}) d^2(x^{k+1}, x^k) \leq \langle P_{x^k, x^{k+1}} u^{k+1} - v^k + d^k, \exp_{x^k}^{-1} x^{k+1}\rangle, \tag{20}$$

for any $v^k \in A(x^k)$ and $u^{k+1} \in B(x^{k+1})$.

Therefore, the desired result follows applying the Cauchy-Schwarz inequality and triangular inequality in (20) and using the fact that, in Hadamard manifolds, $\| \exp_{x^k}^{-1} x^{k+1}\| = d(x^{k+1}, x^k)$ and $\|d^k\| = \gamma_k d(x^k, x^{k-1})$. ◻

Proposition 6 *Suppose that $\{x^k\}$ is bounded, then there exist constants $L, M \geq 0$ such that $||P_{x^k,x^{k+1}}u^{k+1} - v^k|| \leq L$ and $\limsup_{k\to\infty} d(x^{k+1}, x^k) = M$, for any $v^k \in A(x^k)$, $u^{k+1} \in B(x^{k+1})$ and k large enough. Moreover, if A or B is ρ-strongly monotone and*

$$\left(\frac{\rho}{2} + \frac{1}{b}\right)M > L, \tag{21}$$

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

Proof Since $\{x^k\}$ is bounded it follows from the maximal monotonicity of A and B (see Proposition 2) that there exist constants $K_1, K_2 > 0$ such that $||v^k|| < K_1$ and $||u^k|| < K_2$ for k large enough. Thus, from triangular inequality, we have

$$||P_{x^k,x^{k+1}}u^{k+1} - v^k|| \leq ||P_{x^k,x^{k+1}}u^{k+1}\| + \|v^k||.$$

Using the fact that the parallel transport mapping is an isometry, we have

$$||P_{x^k,x^{k+1}}u^{k+1} - v^k|| < L = K_1 + K_2, \tag{22}$$

for k large enough. On the other hand, since $\{x^k\}$ is bounded, we have that $\{d(x^{k+1}, x^k)\}$ is also bounded, and hence, there exists $M = \limsup_{k\to\infty} d(x^{k+1}, x^k)$ and the first inequality is proved. Now, if $M = 0$, then the second assertion directly follows. Otherwise, suppose that $M > 0$ and, by assumption, we have that A or B is ρ-strongly monotone with $\rho > 0$ satisfying (21). Thus, combining (11) with (22), we have

$$\left(\rho + \frac{1}{b}\right) d(x^{k+1}, x^k) \leq \|P_{x^k,x^{k+1}}u^{k+1} - v^k\| + \frac{\rho}{2}d(x^k, x^{k-1}) \leq L + \frac{\rho}{2}d(x^k, x^{k-1}).$$

Taking the $\limsup_{k\to\infty}$ in the above inequality, this implies that $\left(\frac{\rho}{2} + \frac{1}{b}\right) M \leq L$ which is a contradiction. Therefore, the assertion is proved. □

Remark 6 If $A(\cdot) = \partial g(\cdot)$ and $B(\cdot) = \partial h(\cdot)$ with g and h convex functions (and hence, maximal monotone vector fields) and $f(x) = g(x) - h(x)$, then (21) can be replaced by the lower boundedness of f. However, the key of the proof still follows from an inequality like (11), more precisely, $\frac{1}{b}d^2(x^{k+1}, x^k) \leq f(x^0) - f^*$, where $f^* = \inf_{x\in M} f(x)$; see [14, Proposition 5]. Dealing with general maximal monotone vector fields, it is worth to mention that condition (21) is not restrictive due to the fact that the parameter ρ can be taken large enough so that (21) holds; see Remark 3.

Next, we present partial and full convergence results for IPPM for the more general case of difference of monotone vector fields in Hadamard manifolds. Furthermore, we introduce an assumption on the proximal parameter $\{\mu_k\}$ and a Lipschitz continuity of B in order to obtain full convergence of the sequence to a critical point of $A - B$.

4.1 Partial convergence analysis

In this subsection, we suppose that the assumptions in Proposition 6 hold, i.e., we assume that $\{x^k\}$ is bounded, A or B is ρ-strongly monotone and (21) holds. Furthermore, we consider $a, b > 0$ and $\{\mu_k\}$ such that

$$a \leq \mu_k \leq b. \tag{23}$$

Theorem 1 *Every cluster point of $\{x^k\}$ is a critical point of $A - B$.*

Proof Let μ, x and y be cluster points of $\{\mu_k\}$, $\{x^k\}$ and $\{y^k\}$, respectively. Without loss of generality we can take subsequences $\{\mu_{k_j}\}$, $\{x^{k_j}\}$ and $\{y^{k_j}\}$ converging respectively to μ, x and y (we can extract another subsequence if necessary). From Proposition 6, we have that $x^{k_j+1} \to x$. It follows from (7) that

$$\frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j}} y^{k_j} - d^{k_j} \in B(x^{k_j}).$$

By the monotonicity of B, we have

$$\langle \frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j}} y^{k_j} - d^{k_j}, \exp^{-1}_{x^{k_j}} z \rangle \leq \langle u, -\exp^{-1}_{z} x^{k_j} \rangle, \quad \forall u \in B(z),\ z \in M.$$

Letting $j \to +\infty$ in last inequality and using Proposition 1, we obtain

$$\langle \frac{1}{\mu} \exp^{-1}_{x} y, \exp^{-1}_{x} z \rangle \leq \langle u, -\exp^{-1}_{z} x \rangle, \quad \forall u \in B(z),\ z \in M,$$

and hence, by the maximality of B, we have

$$\frac{1}{\mu} \exp^{-1}_{x} y \in B(x). \tag{24}$$

Now, from (8), we have

$$\frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j+1}} y^{k_j} \in A(x^{k_j+1}).$$

Similarly, by the maximal monotonicity of A, we have

$$\langle \frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j+1}} y^{k_j}, \exp^{-1}_{x^{k_j+1}} z \rangle \leq \langle u, -\exp^{-1}_{z} x^{k_j+1} \rangle, \quad \forall u \in A(z),\ \forall z \in M,$$

and letting $j \to +\infty$, we show that

$$\frac{1}{\mu} \exp^{-1}_{x} y \in A(x). \tag{25}$$

Therefore, from (24) and (25), we have that $\frac{1}{\mu} \exp^{-1}_{x} y \in A(x) \cap B(x)$ which means that (6) holds, i.e., x is critical point of $A - B$. □

4.2 Sufficient conditions for boundedness

Now, for the sake of completeness, we state the following technical tool which will be used in the sequel.

Lemma 1 *Let $\{\alpha_k\}$, $\{\beta_k\}$ and $\{\Gamma_k\}$ three sequences of non-negative numbers satisfying $\alpha_{k+1} \leq (1 + \beta_k)\alpha_k + \Gamma_k$. If $\sum_{k=0}^{\infty} \beta_k < +\infty$ and $\sum_{k=0}^{\infty} \Gamma_k < +\infty$, then $\{\alpha_k\}$ is convergent.*

Proof See Polyak [25]. □

Recently, some works have proved the convergence of the whole sequence for several methods applied to DC functions supposing the Kurdyka–Łojasiewicz property of the objective function; see [26–29]. However, even in this case, it is supposed that the sequence generated by the method is bounded. The aim of this subsection is to introduce sufficient conditions in order to guarantee the boundedness of the sequence generated by our method. It is worth to mention that this result is new even for DC problems in the Euclidean setting.

In the remain of this section, we consider Algorithm IPPM with the following condition instead of (23):

$$0 < \mu_k \leq b, \quad k \in \mathbb{N}. \tag{26}$$

Furthermore, we consider the following assumptions in the vector field B, the proximal parameters μ_k and to sequence of non-negative numbers $\{\gamma_k d(x^k, x^{k-1})\}$.
$(A1)$ Given $x, y \in M$, there exists a constant $L > 0$ such that $\|u - P_{x,y}v\| \leq Ld(x, y)$, for any $u \in B(x)$ and $v \in B(y)$.
$(A2)$ The sequence of positive numbers $\{\mu_k\}$ satisfies $\sum_{k=0}^{\infty} \mu_k < \infty$.
$(A3)$ The sequence $\{\gamma_k d(x^k, x^{k-1})\}$ is bounded.

Remark 7 Note that $(A1)$ is Lipschitz type assumption for vector fields. It is worth to mention that assumption $(A1)$ is a natural extension of the Lipschitz continuity of the gradient function from DC function to difference of maximal monotone vector fields. In the literature of algorithm for DC functions some papers have considered the non-smooth case where $f(x) = g(x) - h(x)$ with g, h convex function, g possibly non-smooth and h a function $C^{1,1}$, i.e., h is differentiable and its gradient is Lipschitz continuous; see for instance [26–29]. In the context of vector fields, condition (A1) implies that the vector field B is single-valued. On the other hand, $(A2)$ is a classical assumption on proximal point methods; see for instance [6]. Dealing with DC functions it has been natural to suppose boundedness assumption of the sequence for different methods; see for instance [14, 17, 18, 23, 26–30]. Some works have replaced this assumption by the concept of Kurdyka–Łojasiewicz inequality; see for instance [26–29]. Here, we consider the boundedness assumption $(A3)$ in order to obtain that $\{x^k\}$ is bounded. It is quite natural because $\gamma_k \in [0, \frac{\rho}{2})$, for all $k \in \mathbb{N}$, and the proximal point method and its variants usually have the property that $\{d(x^k, x^{k-1})\}$ converges to zero.

Theorem 2 *Suppose that assumptions $(A1)$, $(A2)$ and $(A3)$ hold. Then, $\{d(x^k, x^*)\}$ is convergent for every $x^* \in S$. In particular, $\{x^k\}$ is bounded.*

Proof By (8), we have that $\exp^{-1}_{x^{k+1}} y^k \in \mu_k A(x^{k+1})$ and hence

$$y^k \in \exp_{x^{k+1}} \mu_k A(x^{k+1}).$$

This means that $x^{k+1} = J^A_{\mu_k}(y^k)$ and from (7), we have

$$x^{k+1} = J^A_{\mu_k}(\exp_{x^k} \mu_k(w^k + d^k)),$$

where $w^k = B(x^k)$. Let $x^* \in S$ be fixed and take $w^* \in A(x^*) \cap B(x^*)$ such that

$$x^* = J^A_{\mu_k}(\exp_{x^*} \mu_k w^*).$$

Since A is maximal monotone from Proposition 4 its resolvent is firmly nonexpansive. Thus, from Proposition 3, we have

$$\begin{aligned}
d(x^{k+1}, x^*) &= d(J^A_{\mu_k}(\exp_{x^k} \mu_k(w^k + d^k)), J^A_{\mu_k}(\exp_{x^*} \mu_k w^*)) \\
&\leq d(\exp_{x^k} \mu_k(w^k + d^k), \exp_{x^*} \mu_k w^*) \\
&\leq d(\exp_{x^k} \mu_k(w^k + d^k), x^*) + d(x^*, \exp_{x^*} \mu_k w^*) \\
&\leq d(x^k, x^*) + d(\exp_{x^k} \mu_k(w^k + d^k), x^k) + \|\mu_k w^*\|
\end{aligned}$$

$$
\begin{aligned}
&= d(x^k, x^*) + \mu_k \|w^k + d^k\| + \mu_k \|w^*\| \\
&\le d(x^k, x^*) + \mu_k (\|w^k - P_{x^k,x^*} w^*\| + \|P_{x^k,x^*} w^*\| + \|d^k\|) + \mu_k \|w^*\| \\
&\le d(x^k, x^*) + \mu_k (L d(x^k, x^*) + \|w^*\| + \|d^k\|) + \mu_k \|w^*\| \\
&= (1 + \mu_k L) d(x^k, x^*) + \mu_k \gamma_k d(x^k, x^{k-1}) + 2\mu_k \|w^*\|,
\end{aligned}
$$

where the triangular inequality was successively applied and in last inequality we used $(A1)$.

Therefore, applying Lemma 1 with $\alpha_k = d(x^k, x^*)$, $\beta_k = L\mu_k$ and $\Gamma_k = \mu_k \gamma_k d(x^k, x^{k-1}) + 2\mu_k \|w^*\|$ taking into account that assumptions $(A2)$ and $(A3)$ hold, then we conclude that $\{d(x^k, x^*)\}$ is convergent and hence $\{x^k\}$ is bounded. □

In Proposition 6, we proved that $\lim_{k\to\infty} d(x^{k+1}, x^k) = 0$ under the assumption (21). In the absence of this assumption, we prove in the sequel that this assertion still holds under the conditions (A1), (A2) and (A3) for the choice of μ_k as in (26).

Corollary 2 *If assumptions* $(A1)$, $(A2)$ *and* $(A3)$ *hold, then* $\lim\limits_{k\to+\infty} d(x^{k+1}, x^k) = 0$.

Proof From (18), we have

$$
\frac{1}{\mu_k} d^2(x^{k+1}, x^k) \le \langle u^{k+1}, -\exp^{-1}_{x^{k+1}} x^k \rangle + \langle v^k, -\exp^{-1}_{x^k} x^{k+1} \rangle + \langle d^k, \exp^{-1}_{x^k} x^{k+1} \rangle,
$$

$u^{k+1} = B(x^{k+1})$ and $v^k \in A(x^k)$.

Using the parallel transport, we have that $\exp^{-1}_{x^k} x^{k+1} = -P_{x^k,x^{k+1}} \exp^{-1}_{x^{k+1}} x^k$ and then,

$$
\frac{1}{\mu_k} d^2(x^{k+1}, x^k) \le \|P_{x^k,x^{k+1}} u^{k+1} - v^k + d^k\| \, \|\exp^{-1}_{x^k} x^{k+1}\|.
$$

This implies that

$$
0 \le d(x^{k+1}, x^k) \le \mu_k (\|u^{k+1}\| + \|v^k\| + \gamma_k d(x^k, x^{k-1})) \tag{27}
$$

taking into account that $\|\exp^{-1}_{x^k} x^{k+1}\| = d(x^{k+1}, x^k) > 0$. Since, from Theorem 2, $\{x^k\}$ is bounded, we have from Proposition 2 and $(A3)$ that the sequences $\{u^k\}$, $\{v^k\}$ and $\{\gamma_k d(x^k, x^{k-1})\}$ are bounded. Therefore, letting $k \to +\infty$ in (27) taking into account that from $(A2)$ we have that $\lim_{k\to+\infty} \mu_k = 0$, we obtain that $\lim\limits_{k\to+\infty} d(x^{k+1}, x^k) = 0$ and the assertion is proved. □

4.3 Sufficient condition for full convergence

In this section, we provide a sufficient condition to obtain the convergence of the whole sequence generated by our method. Under the assumption $(A1)$, $(A2)$ and $(A3)$ we proved that $\lim\limits_{k\to+\infty} d(x^{k+1}, x^k) = 0$ which is a classical behaviour of the proximal point method. It follows from $(A2)$ that $\lim\limits_{k\to+\infty} \mu_k = 0$. Since this auxiliary sequence of parameters is freely chosen satisfying some suitable conditions we will take it in such a way that it does not go to zero faster than $d(x^{k+1}, x^k)$. This leads us to the following condition:

$(A3^*)$ Suppose that $\lim\limits_{k\to\infty} \dfrac{d(x^{k+1}, x^k)}{\mu_k} = 0$.

Remark 8 Note that under assumption $(A2)$, condition $(A3^*)$ implies $(A3)$. Indeed, $(A3^*)$ implies that $\lim_{k\to\infty} d(x^{k+1}, x^k) = 0$ due to the fact that $\lim_{k\to\infty} \mu_k = 0$ from $(A2)$.

Since $0 \leq \gamma_k < \frac{\rho}{2}$, we have that $\lim_{k\to\infty} \gamma_k d(x^k, x^{k-1}) = 0$ and hence, it is bounded. Therefore, replacing assumption $(A3)$ by $(A3^*)$ the results of Sect. 4.2 remain true and we additionally obtain the convergence of the method as stated in the next result. Recall that we are considering algorithm IPPM with (26).

Theorem 3 *If assumptions* $(A1)$*,* $(A2)$ *and* $(A3^*)$ *hold, then* $\{x^k\}$ *converges to a critical point of* $A - B$*.*

Proof In view of Theorem 2 it is enough to show that every cluster point of $\{x^k\}$ belongs to S. Indeed, if x^* is a cluster point of $\{x^k\}$ which belongs to S, then there exists a subsequence $\{x^{k_j}\}$ such that $\{d(x^{k_j}, x^*)\}$ converges to zero. Since from Theorem 2, we have that $\{d(x^k, x^*)\}$ is convergent, thus it converges to zero, and hence, $\{x^k\}$ converges to $x^* \in S$.

Now, we shall prove that an arbitrary cluster point $\overline{x}$ of $\{x^k\}$ belongs to S. Let $\{x^{k_j}\}$ be a subsequence of $\{x^k\}$ such that $x^{k_j} \to \overline{x}$. From Corollary 2, we have that $\lim_{j\to\infty} x^{k_j+1} = \overline{x}$. Moreover, combining (7) with Theorem 2 and Proposition 2, we have that $\{w^k\}$ is bounded. Therefore, without loss of generality, we may assume that there exists a subsequence $\{w^{k_j}\}$ converging to $\overline{w}$. Since $w^{k_j} = B(x^{k_j})$ from the monotonicity of B, we have

$$\langle w^{k_j}, \exp^{-1}_{x^{k_j}} z\rangle \leq \langle u, -\exp^{-1}_z x^{k_j}\rangle, \quad \forall u = B(z),\ z \in M.$$

Letting $j \to \infty$, we obtain that $\overline{w} = B(\overline{x})$. On the other hand, from (8), we have

$$\frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j+1}} y^{k_j} \in A(x^{k_j+1}),$$

where $y^{k_j} = \exp_{x^{k_j}} \mu_{k_j} w^{k_j}$. We claim that $\lim_{j\to\infty} \frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j+1}} y^{k_j} = \overline{w}$, and thus, $\overline{w} \in A(\overline{x})$. Indeed,

$$\begin{aligned}
&\|\frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j+1}} y^{k_j} - P_{x^{k_j+1},\overline{x}}\overline{w}\| \\
&\quad\leq \|\frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j+1}} y^{k_j} - P_{x^{k_j+1},x^{k_j}}(w^{k_j} + d^{k_j})\| + \|P_{x^{k_j+1},x^{k_j}}(w^{k_j} + d^{k_j}) - P_{x^{k_j+1},\overline{x}}\overline{w}\|.
\end{aligned} \tag{28}$$

Let us prove that the right-hand side of the above inequality vanishes as $j \to \infty$. First, note that

$$\begin{aligned}
&\|\frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j+1}} y^{k_j} - P_{x^{k_j+1},x^{k_j}}(w^{k_j} + d^{k_j})\|^2 \\
&\quad= \frac{1}{\mu^2_{k_j}} \|\exp^{-1}_{x^{k_j+1}} y^{k_j}\|^2 + \frac{1}{\mu^2_{k_j}} \|\exp^{-1}_{x^{k_j}} y^{k_j}\|^2 - 2\frac{1}{\mu_{k_j}}\Big\langle \exp^{-1}_{x^{k_j+1}} y^{k_j}, P_{x^{k_j+1},x^{k_j}}(w^{k_j} + d^{k_j})\\
&\quad= \frac{1}{\mu^2_{k_j}}\Big(d^2(x^{k_j+1}, y^{k_j}) + d^2(y^{k_j}, x^{k_j})\Big) - 2\frac{1}{\mu_{k_j}}\Big\langle \exp^{-1}_{x^{k_j+1}} y^{k_j}, P_{x^{k_j+1},x^{k_j}}(w^{k_j} + d^{k_j})\\
&\quad= \frac{1}{\mu^2_{k_j}}\Big(d^2(x^{k_j+1}, y^{k_j}) + d^2(y^{k_j}, x^{k_j})\Big)\\
&\qquad -\frac{2}{\mu_{k_j}}\Big\langle P_{y^{k_j},x^{k_j+1}}(\exp^{-1}_{x^{k_j+1}} y^{k_j}), P_{y^{k_j},x^{k_j+1}}(P_{x^{k_j+1},x^{k_j}}(w^{k_j} + d^{k_j}))\\
&\quad= \frac{1}{\mu^2_{k_j}}\Big(d^2(x^{k_j+1}, y^{k_j}) + d^2(y^{k_j}, x^{k_j})\Big) - \frac{2}{\mu_{k_j}}\Big\langle -\exp^{-1}_{y^{k_j}} x^{k_j+1}, P_{y^{k_j},x^{k_j}}(\frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j}} y^{k_j})
\end{aligned}$$

$$= \frac{1}{\mu_{k_j}^2}\Big(d^2(x^{k_j+1}, y^{k_j}) + d^2(y^{k_j}, x^{k_j})\Big) - 2\frac{1}{\mu_{k_j}}\Big\langle -\exp^{-1}_{y^{k_j}} x^{k_j+1}, -\frac{1}{\mu_{k_j}}\exp^{-1}_{y^{k_j}} x^{k_j}$$

$$= \frac{1}{\mu_{k_j}^2}\Big(d^2(x^{k_j+1}, y^{k_j}) + d^2(y^{k_j}, x^{k_j})\Big) - 2\frac{1}{\mu_{k_j}^2}\Big\langle \exp^{-1}_{y^{k_j}} x^{k_j+1}, \exp^{-1}_{y^{k_j}} x^{k_j} \ . \tag{29}$$

From the geodetic triangle $\triangle(x^{k_j+1}, y^{k_j}, x^{k_j})$, we have

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

From the geodetic triangle $\triangle(x^{k_j+1}, y^{k_j}, x^{k_j})$, we have

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

Using last inequality in (29), we have

$$\|\frac{1}{\mu_{k_j}}\exp^{-1}_{x^{k_j+1}} y^{k_j} - P_{x^{k_j+1}, x^{k_j}}(w^{k_j} + d^{k_j})\| \leq \frac{1}{\mu_{k_j}} d(x^{k_j+1}, x^{k_j})$$

which from $(A4)$ implies

$$\lim_{j\to\infty} \|\frac{1}{\mu_{k_j}}\exp^{-1}_{x^{k_j+1}} y^{k_j} - P_{x^{k_j+1}, x^{k_j}}(w^{k_j} + d^{k_j})\| = 0. \tag{30}$$

Now, by $(A1)$ and triangular inequality, we have that

$$\|P_{x^{k_j+1}, x^{k_j}}(w^{k_j} + d^{k_j}) P_{x^{k_j+1}, \bar{x}}\overline{w}\| = \|P_{x^{k_j+1}, x^{k_j}}(w^{k_j} + d^{k_j}) - P_{x^{k_j+1}, x^{k_j}}(P_{x^{k_j}, \bar{x}}\overline{w})\|$$
$$= \|w^{k_j} + d^{k_j} - P_{x^{k_j}, \bar{x}}\overline{w}\| \leq L d(x^{k_j}, \bar{x}) + \gamma_{k_j} d(x^{k_j}, x^{k_j-1}).$$

From Corollary 2 and $x^{k_j} \to \overline{x}$, letting $j \to \infty$ in the above equality we obtain

$$\lim_{j\to\infty} \|P_{x^{k_j+1}, x^{k_j}}(w^{k_j} + d^{k_j}) - P_{x^{k_j+1}, \bar{x}}\overline{w}\| = 0. \tag{31}$$

Therefore, using (30) and (31) in (28), we have

$$\lim_{j\to\infty} \frac{1}{\mu_{k_j}}\exp^{-1}_{x^{k_j+1}} y^{k_j} = \overline{w}.$$

Using now the monotonicity of A, we have

$$\langle \frac{1}{\mu_{k_j}}\exp^{-1}_{x^{k_j+1}} y^{k_j}, \exp^{-1}_{x^{k_j+1}} z\rangle \leq \langle u, -\exp^{-1}_{z} x^{k_j+1}\rangle, \quad \forall u \in A(z),\ z \in M.$$

Letting $j \to \infty$ and using the maximality of A, we have that $\overline{w} \in A(\overline{x})$ and thus $\overline{w} \in A(\overline{x}) \cap B(\overline{x})$ implying that $\overline{x} \in S$. This completes the proof. $\square$

5 Convergence analysis: monotone case

As we mention before, IPPM is new even for finding a zero of a maximal monotone vector field. In this section, we consider the variational problem (2) with $B(x) = 0$ and the algorithm mIPPM. Here, we will not need to suppose assumption $(A1)$.

In order to prove the convergence of mIPPM let us consider $A^{-1}(0) \subset M$ the set of singularities of A and assume that $A^{-1}(0) \neq \varnothing$, i.e., $A^{-1}(0) = \{x \in M;\ 0 \in A(x)\}$.

In this section, we will consider two possible choices for $\{\mu_k\}$ (26) and (23) as in the non-monotone case, i.e., $0 < \mu_k \leq b$ and $0 < a \leq \mu_k \leq b$. In the first case, we assume that assumptions $(A2)$ and $(A3^*)$ hold. In the second one, we replace these assumptions by

$$\sum_{k=0}^{\infty} \gamma_k d(x^k, x^{k-1}) < \infty. \tag{32}$$

Remark 9 In the linear setting, condition (32) is considered for instance by Alimohammady and Ramazannejad [11]. In Alvarez and Attouch [16], a similar condition is proposed in Hilbert spaces. More precisely, it is assumed that $\sum_{k=0}^{\infty} \gamma_k \|x^k - x^{k-1}\|^2 < \infty$; see [16, Theorem 2.1]. Furthermore, such a condition holds for some special cases which can be verified *a priori*; see [16, Proposition 2.1].

Next results ensure that the sequence $\{x^k\}$ generated by mIPPM converges to a singularity of A. We will consider both possible choices of μ_k as mentioned before.

Theorem 4 *The sequence $\{d(x^k, x^*)\}$ is convergent for every $x^* \in A^{-1}(0)$ and, in particular, $\{x^k\}$ is bounded.*

Proof By (10), we have that $\exp^{-1}_{x^{k+1}} y^k \in \mu_k A(x^{k+1})$ and hence

$$y^k \in \exp_{x^{k+1}} \mu_k A(x^{k+1}).$$

This means that $x^{k+1} = J^A_{\mu_k}(y^k)$ and from (9), we have

$$x^{k+1} = J^A_{\mu_k}(\exp_{x^k} \mu_k(d^k)).$$

Let $x^* \in A^{-1}(0)$ be fixed, then $0 \in A(x^*)$ such that

$$x^* = J^A_{\mu_k}(\exp_{x^*} \mu_k 0) = J^A_{\mu_k}(x^*).$$

Since A is maximal monotone from Proposition 4 its resolvent is firmly nonexpansive. Thus, from Proposition 3, we have

$$\begin{aligned}
d(x^{k+1}, x^*) &= d(J^A_{\mu_k}(\exp_{x^k} \mu_k d^k), J^A_{\mu_k}(x^*)) \\
&\leq d(\exp_{x^k} \mu_k d^k, x^*) \\
&\leq d(x^k, x^*) + d(\exp_{x^k} \mu_k d^k, x^k) \\
&= d(x^k, x^*) + \mu_k \|d^k\| \\
&= d(x^k, x^*) + \mu_k \gamma_k d(x^k, x^{k-1}),
\end{aligned}$$

where the triangular inequality was applied. Therefore, the desired result follows from Lemma 1 with $\alpha_k = d(x^k, x^*)$, $\beta_k = 0$ and $\Gamma_k = \mu_k \gamma_k d(x^k, x^{k-1})$ taking into account that $\sum_{k=0}^{\infty} \Gamma_k < \infty$ for both possible choices of μ_k. Indeed, if $0 < \mu_k \leq b$, then assumptions $(A2)$ and $(A3^*)$ hold, and hence, $\{\gamma_k d(x^k, x^{k-1})\}$ is bounded and $\sum_{k=0}^{\infty} \mu_k < \infty$ which implies that $\sum_{k=0}^{\infty} \Gamma_k < \infty$. If $0 < a \leq \mu_k \leq b$, then $\{\mu_k\}$ is bounded and from (32) we obtain that ${}_{k=0}^{\infty} \Gamma_k < \infty$. This completes the proof. □

As mentioned in Remark 8, under assumption $(A2)$ we have that condition $(A3^*)$ implies $\lim_{k\to\infty} d(x^{k+1}, x^k) = 0$. This is the case if $\{\mu_k\}$ is given by (26), namely, $0 < \mu_k \leq b$. On the other hand, if $\{\mu_k\}$ is given by (23), i.e., $0 < a \leq \mu_k \leq b$, we prove next that this condition still holds under assumption (32).

Corollary 3 *Let $\{x^k\}$ be a sequence generated by mIPPM with $\{\mu_k\}$ given by (23). Then,* $\lim_{k\to+\infty} d(x^{k+1}, x^k) = 0$.

Proof From (10), we have that $\exp^{-1}_{x^{k+1}} y^k \in \mu_k A(x^{k+1})$. Take $x^* \in A^{-1}(0)$. By monotonicity of A, we have

$$\langle \frac{1}{\mu_k} \exp^{-1}_{x^{k+1}} y^k, \exp^{-1}_{x^{k+1}} x^* \rangle \leq \langle 0, -\exp^{-1}_{x^*} x^{k+1} \rangle = 0. \tag{33}$$

From the geodetic triangle $\triangle(y^k, x^{k+1}, x^*)$ with $\theta = \angle(\exp^{-1}_{x^{k+1}} y^k, \exp^{-1}_{x^{k+1}} x^*)$, we obtain

$$d^2(y^k, x^{k+1}) + d^2(x^{k+1}, x^*) - 2\langle \exp^{-1}_{x^{k+1}} y^k, \exp^{-1}_{x^{k+1}} x^* \rangle \leq d^2(y^k, x^*). \tag{34}$$

Combining (33) with (34), we have

$$d^2(y^k, x^{k+1}) + d^2(x^{k+1}, x^*) \leq d^2(y^k, x^*)$$

which applying the triangular inequality leads to

$$d^2(y^k, x^{k+1}) \leq d^2(y^k, x^k) + 2d(y^k, x^k)d(x^k, x^*) + [d^2(x^k, x^*) - d^2(x^{k+1}, x^*)]. \tag{35}$$

Since $y^k = \exp_{x^k} \mu_k d^k$, one has $d(y^k, x^k) = \mu_k \gamma_k d(x^k, x^{k-1})$. Thus, we obtain that

$$\lim_{k\to\infty} d(y^k, x^k) = 0$$

taking into account that $\{\mu_k\}$ is bounded and $\lim_{k\to\infty} \gamma_k d(x^k, x^{k-1}) = 0$ from (32). Therefore, letting $k \to \infty$ in (35) and using the fact that $\{d(x^k, x^*)\}$ is convergent, we have that $\lim\limits_{k\to\infty} d(y^k, x^{k+1}) = 0$. Thus, from triangular inequality, we have

$$0 \leq \lim_{k\to\infty} d(x^{k+1}, x^k) \leq \lim_{k\to\infty} [d(x^{k+1}, y^k) + d(y^k, x^k)] = 0,$$

and hence, $\lim\limits_{k\to\infty} d(x^{k+1}, x^k) = 0$. □

Theorem 5 *The sequence $\{x^k\}$ converges to a singularity of A.*

Proof In view of Theorem 4 it is enough to show that every cluster point of $\{x^k\}$ belongs to S. Indeed, if x^* is a cluster point of $\{x^k\}$ which belongs to $A^{-1}(0)$, then there exists a subsequence $\{x^{k_j}\}$ such that $\{d(x^{k_j}, x^*)\}$ converges to zero. Since from Theorem 4, we have that $\{d(x^k, x^*)\}$ is convergent, thus it converges to zero, and hence, $\{x^k\}$ converges to $x^* \in A^{-1}(0)$.

Now, we shall prove that an arbitrary cluster point $\bar{x}$ of $\{x^k\}$ belongs to $A^{-1}(0)$. Let $\{x^{k_j}\}$ be a subsequence of $\{x^k\}$ such that $x^{k_j} \to \bar{x}$. From assumption $(A3^*)$ and Corollary 3, for both possible choices of μ_k, we have that $\lim_{k\to\infty} d(x^{k+1}, x^k) = 0$ and hence $\lim_{j\to\infty} x^{k_j+1} = \bar{x}$. From the definition of mIPPM, we have that $d^{k_j} = \gamma_{k_j} \exp^{-1}_{x^{k_j}} x^{k_j-1}$ and $y^{k_j} = \exp_{x^{k_j}} \mu_{k_j} d^{k_j}$. Thus,

$$\begin{aligned}\left\| \frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j+1}} y^{k_j} \right\| &= \frac{1}{\mu_{k_j}} d(x^{k_j+1}, y^{k_j}) \leq \frac{1}{\mu_{k_j}} d(x^{k_j+1}, x^{k_j}) + \frac{1}{\mu_{k_j}} d(x^{k_j}, y^{k_j}) \\ &\leq \frac{1}{\mu_{k_j}} d(x^{k_j+1}, x^{k_j}) + ||d^{k_j}||,\end{aligned}$$

for both choices of μ_k. The right-hand side of the above inequality goes to zero as $j \to \infty$ from $(A3^*)$, for $0 < \mu_k \leq b$, and from $\frac{1}{\mu_{k_j}} d(x^{k_j+1}, x^{k_j}) \leq \frac{1}{a} d(x^{k_j+1}, x^{k_j})$, for $0 < a \leq$

$\mu_k \leq b$, taking into account that $\lim_{k\to\infty} d(x^{k+1}, x^k) = 0$ in both cases. Now using the monotonicity of A for $\frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j+1}} y^{k_j} \in A(x^{k_j+1})$, we have

$$\langle \frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j+1}} y^{k_j}, \exp^{-1}_{x^{k_j+1}} z\rangle \leq \langle u, -\exp^{-1}_{z} x^{k_j+1}\rangle, \quad \forall u \in A(z),\ z \in M.$$

Letting $j \to +\infty$ in last inequality and using the fact that $\frac{1}{\mu_{k_j}} \exp^{-1}_{x^{k_j+1}} y^{k_j} \to 0 \in T_{\overline{x}}M$ together with Proposition 1 and the maximality of A, we obtain that $0 \in A(\overline{x})$ and the proof is completed. □

6 Numerical experiments

In this section, we solve unconstrained minimization DC problems in a genuine Hadamard manifold considered by Almeida et al. [23]. We run IPPM and mIPPM comparing its performance with proximal point methods [6] (PPM in the convex case) and [14] (DCPPM in the non-convex DC case). We also present some experiments for monotone and non-monotone operators including an operator which its components are not the subdifferential (or gradient) of a convex function and an example of a point-to-set operator in the Euclidean setting.

Let $\mathbb{P}^n$ be the set of the symmetric matrices, $\mathbb{P}^n_+$ be the cone of the symmetric positive semi-definite matrices and $\mathbb{P}^n_{++}$ be the cone of the symmetric positive definite matrices both $n \times n$. Let $M = (\mathbb{P}^n_{++}, \langle\cdot,\cdot\rangle)$ be the Riemannian manifold endowed with the Riemannian metric given by

$$\langle U, V\rangle = \operatorname{tr}(VX^{-1}UX^{-1}),\ X \in M,\ U, V \in T_X M, \tag{36}$$

where $\operatorname{tr}(X)$ denotes the trace of $X \in \mathbb{P}^n$ and $T_X M \approx \mathbb{P}^n$, with the corresponding norm denoted by $\|\ .\ \|$; see Rothaus [31]. In this case, for any $X, Y \in M$ the unique geodesic joining those two points is given by:

$$\gamma(t) = X^{1/2}\left(X^{-1/2}YX^{-1/2}\right)^t X^{1/2}, \quad \gamma'(0) = X^{1/2}\ln(X^{-1/2}YX^{-1/2})X^{1/2}, \quad t \in [0, 1],$$

see, for instance, Nesterov and Todd [32] and Bento et al. [30]. More precisely, M is a Hadamard manifold; see, for instance, [33, Theorem 1.2., page 325]. One can compute the curvature of M and verify that it has non-constant curvature; see Lenglet et al. [34, page 428]. We consider the exponential map and its inverse as follows

$$\exp_X V = X^{1/2}e^{X^{-1/2}VX^{-1/2}}X^{1/2}, \quad \exp^{-1}_X V = X^{1/2}\ln(X^{-1/2}VX^{-1/2})X^{1/2};$$

see Bhatia [35, Chapter 6].

The numerical experiments are coded in MATLAB R2020b on a machine with a Intel(R) Core(TM) i7, 2.3 GHz CPU and 8 GB memory. We consider two kind of problems. In Problems 1 and 2, we consider minimization problems in the Riemannian manifolds M described above with $n \times n$ matrices for $n = 5$, $n = 25$, $n = 50$ and $n = 100$. We perform the methods performed using "MANOPT" which is a toolbox for optimization on manifolds; see Boumal et al. [36]. We generate random matrices using "manifold.rand" to start the method. The subproblems are solved using the routine "steepestdescent" with the inner loop stopping if the norm of the gradient drops below $\epsilon = 10^{-5}$ as well as the stopping criterion used to the outer loop was $d(x^{k+1}, x^k) < \epsilon$. In Problem 3 and 4, we consider examples in $\mathbb{R}^2$ with operators which are not the subdifferential of a convex function or multi-valued operators. In this context, we use the same stopping rule where the Riemannian distance is the Euclidean norm.

For each problem below, we perform the algorithms 100 times using different random starting points. In all problems, we compare the performance of the methods using random initial starting points but the same at each running for all the methods. We compare the performance of the methods for different choices of the parameters μ_k and γ_k.

Problem 1 In this problem, we consider the convex minimization problem

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

where $f(X) = \frac{1}{2} \ln(\det(X))^2$. Clearly, this problem can be seen as $0 \in A(X) - B(X)$ for $A(X) = \operatorname{grad} f(X)$ and $B(X) = 0$, where A is a maximal monotone vector field. The solution of the equation $0 = \operatorname{grad} f(X^*)$ is a matrix such that $\det(X^*) = 1$, and hence, $f(X^*) = 0$. Note that the Euclidean gradient and hessian of f, denoted by $f'(X)$ and $f''(X)$, are given by $f'(X) = \ln(\det(X))X^{-1}$ and $f''(X)V = \operatorname{tr}(X^{-1}V)X^{-1} - \ln(\det(X))X^{-1}VX^{-1}$. Thus, one can verify that f is not convex in the Euclidean sense while it is in the Riemannian setting using the metric given in (36) and having in mind that the Riemannian gradient and hessian of f, denoted by $\operatorname{grad} f(X)$ and $\operatorname{hess} f(X)$, are given by $\operatorname{grad} f(X) = \ln(\det(X))X$ and $\operatorname{hess} f(X)V = \operatorname{tr}(X^{-1}V)X$.

Problem 2 In this problem, we consider the non-convex DC problem

$$\min_{X \in M} f(X) = g(X) - h(X),$$

where $f(X) = \ln(\det(X))^4 - 2\ln(\det(X))^2 + 1$ with $g(X) = \ln(\det(X))^4 + 1$ and $h(X) = 2\ln(\det(X))^2$. Clearly, this problem can be seen as $0 \in A(X) - B(X)$ for $A(X) = \operatorname{grad} g(X)$ and $B(X) = \operatorname{grad} h(X)$, where A, B are maximal monotone vector fields. The solution of the equation $0 = \operatorname{grad} f(X^*)$ is a matrix such that $\det(X^*) = 1$, $\det(X^*) = e$ or $\det(X^*) = e^{-1}$ with $\min_{x \in M} f(X) = 0$. It is worth to mention that g and h are not convex functions in the Euclidean sense and become convex in the Riemannian sense using the metric (36). Despite the convexity of g and h in M, we have that f is a non-convex function in both Euclidean and Riemannian setting.

Next, we consider some examples where the vector fields are defined in the Euclidean space. In Problem 3, we have an operator which its components are not the subdifferential (or gradient) of a convex function. In Problem 4, we have an example of a point-to-set operator.

Problem 3 Let $A, B : \mathbb{R}^2 \to \mathbb{R}^2$ given by $A(x) = \left(\frac{x_1}{2} - x_2, x_1 + \frac{x_2}{2}\right)$ and $B(x) = (x_2, -x_1)$ with $x = (x_1, x_2)$. One has that A is strongly maximal monotone and B is maximal monotone but neither of them is the subdifferential of a convex function because they are not cyclically monotone operators; see [38, Theorem 3]. Let us consider the problem of finding

$$0 \in C(x) = A(x) - B(x)$$

which is equivalent to show that $A(x) \cap B(x)$ is non-empty, for some $x \in \mathbb{R}^2$. One can easily show that taking $x^* = (0, 0)$, we have that $0 \in A(x^*) \cap B(x^*)$. Thus, $x^* = (0, 0)$ is the solution of the above (non-monotone) variational inclusion.

Problem 4 In this problem, we consider the non-monotone variational problem

$$0 \in A(x) - B(x),$$

where $A(x) = \partial g(x)$ and $B(x) = \partial h(x)$ with $x = (x_1, x_2)$ and

$$g(x) = ||x||^2 + ||x||, \quad h(x) = 0.5||x||^2 + \max\{-x_1, 0\} + \max\{-x_2, 0\}.$$

Clearly, A, B are maximal monotone (point-to-set) operators and this problem can be seen as the non-differentiable and non-convex DC problem

$$\min_{x \in \mathbb{R}^2} f(x) = g(x) - h(x) = 0.5||x||^2 + ||x|| - \max\{-x_1, 0\} - \max\{-x_2, 0\}.$$

The results are presented in Tables 1, 2, 3 and 4 where the column n denotes the size of the $n \times n$ matrices, the columns μ_k and γ_k present the values of these parameters used in the algorithms, the columns min. iter. (k) (resp. min. time), max. iter. (k) (resp. max. time) and med. iter. (k) (resp. med. time) stand to the minimal, maximal and median of iterates (resp. CPU time in seconds) of the methods in 100 runs. The columns $||\text{grad } f(X^k)||$ and $|f(X^k)|$ (or $||C(x^k)||$) show the median of these values in 100 runs. In Table 4, this last column is considered as $\max ||C(x^k)||$ which means the maximal value for $||C(x^k)||$ because in this example the methods converge to different limit points depending of the initial point, so in this case does not make sense to consider the median.

We also present the behavior of the methods in terms of $||d^k||$, $\frac{d(x^{k+1}, x^k)}{\mu_k}$ and $|f(X^k)|$ (or $||C(x^k)||$). It is shown in Figures 1, 2, 3, 4, 5 and 6.

In Tables 1, the results show that for Problem 1, the method mIPMM outperforms the non-boosted version of the method (the case where $\gamma_k = 0$) in both number of iterates and CPU time for different choices of μ_k and γ_k in all the dimensions considered. In this example, our method with $\gamma_k = 0$ coincides with the classical proximal point method (PPM) considered in [6].

In Tables 2, the results show that for Problem 2, the method IPMM outperforms the non-boosted version of the method (the case where $\gamma_k = 0$) in CPU time for different choices of μ_k and γ_k in all the dimensions considered but underperform it in number of iterates. In this example, our method with $\gamma_k = 0$ coincides with the classical proximal point method for DC functions (PPMDC) considered in [14].

In Tables 3, the results show that for Problem 3, the method IPMM outperforms the non-boosted version of the method (the case where $\gamma_k = 0$) in both number of iterates and CPU time for different choices of μ_k and γ_k in all the dimensions considered.

In Tables 4, the results show that for Problem 4, the method IPMM outperforms the non-boosted version of the method (the case where $\gamma_k = 0$) in both number of iterates and CPU time for different choices of μ_k and γ_k in all the dimensions considered. In this example, our method with $\gamma_k = 0$ coincides with the classical proximal point method for DC functions (PPMDC) considered in [14].

7 Conclusions

We have considered an inertial version of the proximal point method for finding a zero of (non-monotone) difference of two maximal monotone vector fields in Hadamard manifolds. We have proved that every cluster point of the sequence, if any, is a solution of the problem. Furthermore, we have presented some sufficient conditions for boundedness and full convergence of the proximal point method for difference of monotone vector fields which are new

Table 1 Running 100 times PPM [6] and $mIPPM$ for Problem 1

n	μ_k	γ_k	Min. iter.(k)	Max. iter.(k)	Med. iter.(k)	Min. time	Max. time	Med. time	$\|\|\text{grad } f(X^k)\|\|$	$\|f(X^k)\|$
5	0.1	0	25	28	26.67	0.31162	1.3126	0.36525	8.4047e − 05	7.1832e − 10
5	5	0.01	6	7	6.6	0.047077	0.16567	0.058094	4.9033e − 06	3.174e − 12
5	1	0.1	9	10	9.91	0.075131	0.15946	0.095228	3.7638e − 06	2.0527e − 12
5	$\frac{10}{(k+1)^2}$	0.1	10	12	11	0.077351	0.12221	0.095089	5.5684e − 05	3.1871e − 10
25	0.1	0	11	11	11	0.25215	0.30694	0.27212	5.0102e − 05	5.0643e − 11
25	5	0.01	5	6	5.98	0.090076	0.13349	0.10827	4.8911e − 06	6.3716e − 13
25	1	0.1	7	7	7	0.11813	0.16889	0.13914	4.5943e − 06	6.0743e − 13
25	$\frac{10}{(k+1)^2}$	0.1	7	7	7	0.098635	0.14243	0.12109	3.5828e − 06	4.4739e − 13
50	0.1	0	8	8	8	0.57692	1.029	0.67213	8.1799e − 05	6.7226e − 11
50	5	0.01	5	5	5	0.24895	0.43507	0.34478	4.6578e − 06	3.136e − 12
50	1	0.1	7	7	7	0.35853	0.51106	0.43423	4.7164e − 06	2.9162e − 13
50	$\frac{10}{(k+1)^2}$	0.1	6	7	6.99	0.2952	0.42262	0.36601	3.8439e − 06	2.2537e − 13
100	0.1	0	7	7	7	1.2405	1.828	1.4523	1.9767e − 05	1.9599e − 12
100	5	0.01	5	5	5	0.59282	0.95723	0.78223	4.2715e − 06	1.333e − 12
100	1	0.1	6	6	6	0.83413	1.2492	0.98612	4.8153e − 06	1.563e − 13
100	$\frac{10}{(k+1)^2}$	0.1	6	6	6	0.62338	1.0542	0.81505	5.3529e − 06	1.8962e − 13

Table 2 Running 100 times PPMDC [14] and $IPPM$ for Problem 2

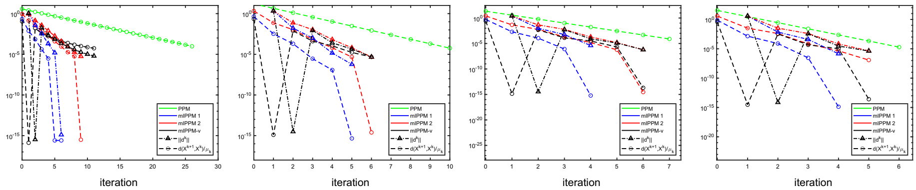
n	μ_k	γ_k	Min. iter.(k)	Max. iter.(k)	Med. iter.(k)	Min. time	Max. time	Med. time	$\|\mathrm{grad}\, f(X^k)\|$	$\lvert f(X^k)\rvert$
5	1	0	8	11	10.69	0.16139	0.3325	0.25947	1.2042e − 04	1.9652e − 10
5	0.1	0.01	11	14	13.76	0.1568	0.24653	0.22278	8.6164e − 05	9.6829e − 11
5	0.5	0.1	9	13	11.86	0.14281	0.24853	0.20879	5.4389e − 05	3.9383e − 11
5	$\frac{2}{(k+1)^2}$	0.1	10	17	15.76	0.14788	0.26679	0.22738	8.4186e − 04	9.1446e − 09
25	1	0	11	11	11	0.94859	1.4397	1.1847	5.4563e − 04	7.4563e − 10
25	0.1	0.01	12	12	12	0.55303	0.71977	0.63063	3.4307e − 04	2.9478e − 10
25	0.5	0.1	12	12	12	0.69976	0.96456	0.81632	2.0127e − 04	1.015e − 10
25	$\frac{2}{(k+1)^2}$	0.1	13	13	13	0.53209	0.91111	0.65692	4.4119e − 04	4.8757e − 10
50	1	0	11	11	11	3.0508	4.3938	3.3717	9.7601e − 05	1.1912e − 09
50	0.1	0.01	12	12	12	1.784	2.3953	1.9991	4.4926e − 04	2.5242e − 10
50	0.5	0.1	12	12	12	2.1716	3.2668	2.4341	3.429e − 04	1.4707e − 10
50	$\frac{2}{(k+1)^2}$	0.1	12	12	12	1.5111	2.4	1.9059	8.12e − 04	8.2456e − 10
100	1	0	11	11	11	7.4373	11.1317	8.3472	1.6744e − 03	1.7525e − 09
100	0.1	0.01	12	12	12	4.1062	6.053	4.8483	6.5633e − 04	2.693e − 10
100	0.5	0.1	12	12	12	5.2686	8.4117	6.3045	5.7401e − 04	2.0599e − 10
100	$\frac{2}{(k+1)^2}$	0.1	12	12	12	4.2017	6.2018	4.988	9.1262e − 04	5.2065e − 10

Table 3 Running 100 times $IPPM$ for Problem 3

μ_k	γ_k	Min. iter.(k)	Max. iter.(k)	Med. iter.(k)	Min. time(s)	Max. time(s)	Med. time(s)	$\|\|C(x^k)\|\|$
0.1	0	187	259	244.15	0.0003684	0.001284	0.00070324	7.1621e-06
0.5	0.1	52	68	64.88	0.0001189	0.000402	0.00022221	1.2119e-06
1	0.5	22	29	27.7	$5.19e-05$	0.0001852	$9.7068e-05$	5.4863e-07
$\frac{10^3}{(k+1)^2}$	0.5	41	44	42.86	$8.99e-05$	0.0003203	0.00014966	4.948e-08

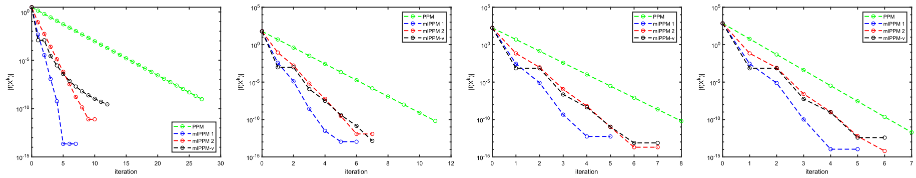
Table 4 Running 100 times PPMDC functions [14] and $IPPM$ for Problem 4

μ_k	γ_k	Min. iter.(k)	Max. iter.(k)	Med. iter.(k)	Min. time(s)	Max. time(s)	Med. time(s)	max $\|\|C(x^k)\|\|$
1	0	4	56	26.99	0.0025276	0.18057	0.017934	2.2505e-07
5	0.1	4	45	22.44	0.0021067	0.026851	0.01377	9.8168e-08
1	0.01	5	42	26.05	0.0026901	0.036193	0.015676	2.0754e-07



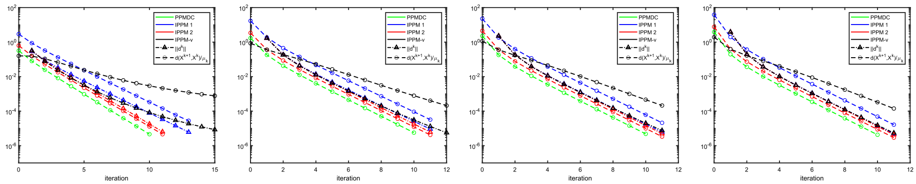
(a) $n = 5$ **(b)** $n = 25$ **(c)** $n = 50$ **(d)** $n = 100$

Fig. 1 Computing $||d^k||$ and $d(X^{k+1}, X^k)/\mu_k$ with PPM [6] and mIPPM for Problem 1



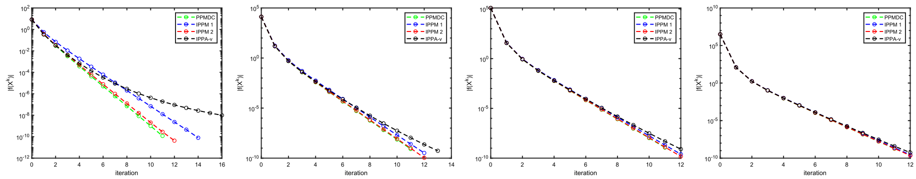
(a) $n = 5$ **(b)** $n = 25$ **(c)** $n = 50$ **(d)** $n = 100$

Fig. 2 Computing $|f(X^k)|$ with PPM [6] and mIPPM for Problem 1



(a) $n = 5$ **(b)** $n = 25$ **(c)** $n = 50$ **(d)** $n = 100$

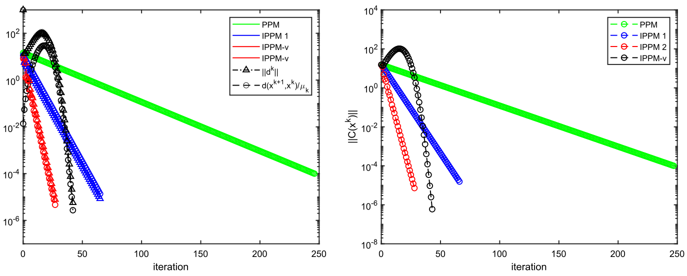
Fig. 3 Computing $||d^k||$ and $d(X^{k+1}, X^k)/\mu_k$ with PPMDC [14] and IPPM for Problem 2



(a) $n = 5$ **(b)** $n = 25$ **(c)** $n = 50$ **(d)** $n = 100$

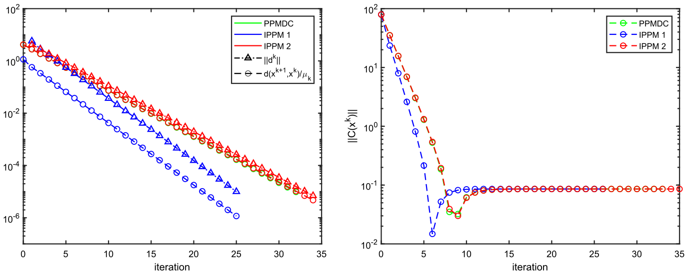
Fig. 4 Computing $|f(X^k)|$ with PPMDC [14] and IPPM for Problem 2

even for DC functions. We have illustrated the method with some numerical experiments in a genuine (with non-constant curvature different from zero) Hadamard manifold and Euclidean space.



(a) $||d^k||$ and $||x^{k+1} - x^k||/\mu_k$ **(b)** $||C(x^k)||$

Fig. 5 Running IPPM for Problem 3



(a) $||d^k||$ and $||x^{k+1} - x^k||/\mu_k$ **(b)** $||C(x^k)||$

Fig. 6 Running PPMDC [14] and IPPM for Problem 4

Acknowledgements We would like to thank the referees for their constructive remarks which allow us to improve our work. J.C.O. Souza was supported in part by CNPq Grants 424169/2018-5 and 313901/2020-1. The project leading to this publication has received funding from the French government under the "France 2030" investment plan managed by the French National Research Agency (reference: ANR-17-EURE-0020) and from Excellence Initiative of Aix-Marseille University - A*MIDEX.

References

1. Ansari, Q.H., Babu, F.: Existence and boundedness of solutions to inclusion problems for maximal monotone vector fields in Hadamard manifolds. Optim. Lett. **14**(3), 711–727 (2020)
2. Ansari, Q.H., Babu, F., Yao, J.-C.: Inexact proximal point algorithms for inclusion problems on Hadamard manifolds. J. Nonlinear Convex Anal. **21**(10), 2417–2432 (2020)
3. Ansari, Q.H., Babu, F.: Proximal point algorithm for inclusion problems in Hadamard manifolds with applications. Optim. Lett. **15**(3), 901–921 (2021)
4. Martinet, B.: Regularisation d'inéquations variationelles par approximations succesives. Rev. Française d'Informatique. et de Rech. Oper. **4**, 154–159 (1970)
5. 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)
6. Ferreira, O.P., Oliveira, P.R.: Proximal point algorithm on Riemannian manifolds. Optimization **51**, 257–270 (2002)
7. Moreau, J.J.: Proximité et dualité dans un espace Hilbertien. Bull. Soc. Math. Fr. **93**, 273–299 (1965)
8. Douglas, J., Rachford, H.H.: On the numerical solution of heat conduction problems in two and three space variables. Trans. Am. Math. Soc. **82**, 421–439 (1956)
9. Moudafi, A.: On the difference of two maximal monotone operators: regularization and algorithmic approach. Appl. Math. Comput. **202**, 446–452 (2008)

10. Alimohammady, M., Ramazannejad, M., Roohi, M.: Notes on the difference of two monotone operators. Optim. Lett. **8**(1), 81–84 (2014)
11. Alimohammady, M., Ramazannejad, M.: Inertial proximal algorithm for difference of two maximal monotone operators. Indian J. Pure Appl. Math. **47**(1), 1–8 (2016)
12. Moudafi, A.: On critical points of the difference of two maximal monotone operators. Afr. Mat. **26**(3), 457–463 (2015)
13. Noor, M.A., Noor, K.I., Hamdi, A., El-Shemas, E.H.: On difference of two monotone operators. Optim. Lett. **3**, 329–335 (2009)
14. Souza, J.C.O., Oliveira, P.R.: A proximal point method for DC functions on Hadamard manifolds. J. Glob. Optim. **63**, 797–810 (2015)
15. Polyak, B.T.: Some methods of speeding up the convergence of iteration methods. U.S.S.R. Comput. Math. Math. Phys. **4**(5), 1–17 (1964)
16. Alvarez, F., Attouch, H.: An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping. Set-Valued Anal. **9**(1–2), 3–11 (2001)
17. Maingé, P.E., Moudafi, A.: Convergence of new inertial proximal methods for DC programming. SIAM J. Optim. **19**, 397–413 (2008)
18. Oliveira, W., Tcheou, M.: Level an inertial algorithm for DC programming. Set-Valued Var. Anal. **27**, 895–919 (2019)
19. Attouch, H., Théra, M.: A general duality principle for the sum of two operators. J. Convex Anal. **3**, 1–24 (1996)
20. Sakai, T.: Riemannian geometry. Translations of mathematical monographs. American Mathematical Society, Providence (1996)
21. Udriste, C.: Convex functions and optimization algorithms on Riemannian manifolds. Mathematics and its applications, Kluwer Academic, Dordrecht (1994)
22. do Carmo, M.P.: Riemannian geometry. Birkhauser, Boston (1992)
23. Almeida, Y.T., Cruz Neto, J.X., Oliveira, P.R., Souza, J.C.O.: A modified proximal point method for DC functions on Hadamard manifolds. Comput. Optim. Appl. **76**, 649–673 (2020)
24. Li, C., López, G., Martín-Márquez, V., Wang, J.H.: Resolvents of set valued monotone vector fields in Hadamard manifolds. Set-Valued Anal. **19**, 361–383 (2011)
25. Polyak, B.T.: Introduction to optimization. Optimization Software Inc., New York (1987)
26. Aragon Artacho, F.J., Vuong, P.T.: The boosted difference of convex functions algorithm for non-smooth functions. SIAM J. Optim. **30**, 980–1006 (2020)
27. Cruz Neto, J.X., Oliveira, P.R., Soubeyran, A., Souza, J.C.O.: A generalized proximal linearized algorithm for DC functions with application to the optimal size of the firm problem. Ann. Oper. Res. **289**, 313–339 (2020)
28. Ferreira, O.P., Santos, E.M., Souza, J.C.O.: Boosted scaled subgradient method for DC programming. arXiv:2103.10757 (2021)
29. Le Thi, H.A., Huynh, V.N., Dinh, T.P.: Convergence analysis of difference-of-convex algorithm with sub-analytic data. J. Optim. Theory Appl. **179**, 103–126 (2018)
30. Bento, G.C., Ferreira, O.P., Oliveira, P.R.: Proximal point method for a special class of non-convex functions on Hadamard manifolds. Optimization **64**(2), 289–319 (2015)
31. Rothaus, O.S.: Domains of positivity. Abh. Math. Sem. Univ. Hambg. **24**, 189–235 (1960)
32. Nesterov, Y.E., Todd, M.J.: On the Riemannian geometry defined by self-concordant barriers and interior-point methods. Found. Comput. Math. **2**, 333–361 (2002)
33. Lang, S.: Fundamentals of differential geometry. Volume 191 of graduate texts in mathematics. Springer, New York, (1999)
34. Lenglet, C., Rousson, M., Deriche, R., Faugeras, O.: Statistics on the manifold of multivariate normal distributions: theory and application to diffusion tensor MRI processing. J. Math. Imaging Vis. **25**, 423–444 (2006)
35. Bhatia, R.: Positive definite matrices, vol. 24. Princeton University Press, Princeton (2009)
36. Boumal, N., Mishira, B., Absil, P.-A., Sepulchre, R.: Manopt, a Matlab toolbox for optimization on manifolds. J. Mach. Learn. Res. **15**, 1455–1459 (2014) http://www.manopt.org
37. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
38. Rockafellar, R.: Characterization of the subdifferentials of convex functions. Pac. J. Math. **17**(3), 497–510 (1966)