Inexact Proximal Point Methods for Quasiconvex Minimization on Hadamard Manifolds

N. Baygorrea, E.A. Papa Quiroz and N. Maculan

Federal University of Rio de Janeiro
COPPE-PESC-UFRJ
PO Box 68511, Rio de Janeiro, CEP 21941-972, Brazil.
{nbaygorrea,erik,maculan}@cos.ufrj.br

Abstract

In this paper we present two inexact proximal point algorithms to solve minimization problems for quasiconvex objective functions on Hadamard manifolds. We prove that under natural assumptions the sequence generated by the algorithms are well defined and converge to critical points of the problem. We also present an application of the method to demand theory in economy.

Keywords: Proximal Point Method, Quasiconvex Function, Hadamard Manifolds, Nonsmooth Optimization, Abstract Subdifferential.

1 Introduction

Many of the algorithms for solving optimization problems have been generalized from linear spaces (Euclidean, Hilbert, Banach) to differentiable manifolds, in particular in the context of Riemannian manifolds, see for example Udriste [46], Rapcsák[40], Németh [32], Da Cruz Neto et al.[18], Chong Li [14], Bento [11], Ferreira and Oliveira [25].

In this paper we are interested in solving optimization problems on Hadamard manifolds, that is, we consider the following problem

$$\min_{x \in M} f(x), \tag{1.1}$$

where M is a Hadamard manifold, which is a complete simply connected riemannian manifold with non positive sectional curvature, and $f : M \to I\!R \cup \{+\infty\}$ is a lower semicontinuous proper function. This class of manifolds is a natural motivation to study more general spaces of nonpositive curvature such as, for instance, CAT(0) (geodesic metric spaces) introduced by Alexandrov, see [3] and [9]. For a further motivation in continuous optimization see the introduction of Papa Quiroz and Oliveira, [36].

One of the most popular and studied methods by the optimization community is the proximal point algorithm. This method has been introduced, in the setting of Riemannian geometry, by Ferreira and Oliveira[25], when f is a convex function on a Hadamard manifold, which is described by:

$$x^k := \text{argmin}_{x\in M} \left\{ f(x) + \frac{\lambda_k}{2} d^2(x, x^{k-1}) \right\}, \qquad k = 1, 2, \ldots \tag{1.2}$$

where x^0 is an arbitrary start point, λ_k is a positive parameter and d is a riemannian distance defined on M. The authors proved that, for each $k \in I\!N$, the function $f(\cdot)+\frac{\lambda_k}{2}d^2(\cdot, x^{k-1}) : M \to I\!R$ is 1-coercive and the sequence $\{x^k\}$ is well defined, with x^k being uniquely determined. Furthermore, assuming $\sum_{k=0}^{+\infty} \frac{1}{\lambda_k} = +\infty$ and that f has a minimizer, they proved that the sequence $\{f(x^k)\}$ converges to the minimum value and the sequence $\{x^k\}$ converges to a minimizer point.

In order to extend the range of applications of the problem (1.1), Papa Quiroz and Oliveira[35] considered the iteration (1.2) when f is quasiconvex on Hadamard manifolds, and they proved convergence to an optimal solution of the problem when the parameter λ_k converges to zero. Furthermore, they generalized the proximal point method using Bregman distances on Hadamard manifolds.

Other contribution in this framework was done by Bento et al.[13], these researchers presented the proximal point method for a special class of nonconvex function. They proved, under some additional assumptions, the convergence to a minimum of the problem (1.1) whenever each accumulation point of the generated sequence satisfies certain conditions. Local convergence analysis of the proximal point method for this special class of nonconvex function is presented in Bento et al.[12]. Moreover, when in (1.1), f satisfies a Kurdyka-Lojasiewicz inequality, motivated from the work of Attouch[4, 5], the proximal point algorithm has been extending by Papa Quiroz and Oliveira[33] and Da Cruz Neto et al[19] in the setting of riemannian geometry. This inequality has been introduced by Kurdyka[28] for a o-minimal structure in $I\!R^n$.

The proximal point method has also been extended on Hadamard manifolds to obtain finite convergence, find singularities and solve variational inequality problems. Indeed, Bento and Cruz Neto[11], proved that the sequence generated by the proximal point method associated to the problem (1.1) has finite termination when the objective function is convex and the solution set of this problem is a set of weak sharp minimizers. For finding a singularity of a multivalued vector field and solving variational inequality problems for monotone operators see the works of [2, 20, 29, 43] and references therein. Under weaker assumptions than monotonicity, Tang et al.[42] showed that the sequence generated by the proximal point algorithm for pseudomonotone vector fields is well defined and proved that this sequence converges to a solution of variational inequality, whenever it exists.

In recent years, also applications of that method to solve optimization problems on Riemannian manifolds have been studied. For instance, da Cruz Neto et al.[19] showed that if the set of constraints of a given problem is a Riemannian manifold of non positive curvature and the objective function is lower semicontinuous and satisfies the Kurdyka-Lojasiewicz property, then the alternating proximal algorithm is naturally extended to solve that class of problems. The authors proved that the sequence generated by the proximal point algorithm is well defined and converges to an inertial Nash equilibrium under mild assumptions about the objective function. Worthwhile to mention that Nash equilibrium problems and its generalizations have become increasingly popular among academics and practitioners. Classical applications of Nash equilibrium include computer science, telecommunications, energy markets, and many others, see [24] for a recent survey.

Another topic for applications is in economy where demand theory is one of the core theories of microeconomy. It aims to answer basic questions about preferences of economic agents (companies, governments, states, etc.) and how demand is impacted by income levels and satisfaction (utility). Based on the perceived utility of goods and services by agents, markets adjust the supply available and the prices charged. Indeed, mathematical models in economic theory of decision consist of optimal choice of economic agents into all possible alternatives to choose, see Section 6 of this paper.

From the computational point of view, note that when f is a nonconvex function, iteration (1.2) could be a as hard subproblem as original problem to solving. Based on this idea in the quasiconvex framework, Papa Quiroz and Oliveira[36], instead of considering iteration (1.2), proposed the following iteration

$$0 \in \partial(f(\cdot) + (\lambda_k/2)d^2(\cdot, x^{k-1}))(x^k), \tag{1.3}$$

where $\partial = \partial^F$ is the regular subdifferential on Hadamard manifolds. They proved full convergence of the sequence to a generalized critial point of f. Besides, Papa Quiroz and Oliveira[34] also considered iteration (1.3), when $\partial = \partial^C$ is the Clarke subdifferential on M, for solving quasiconvex locally Lipschitz optimization problem, and they proved that the sequence $\{x^k\}$ given by (1.3) there exists and converges to a critical point of f.

With the objective to construct implementable proximal point algorithms for quasiconvex functions we are interested in this paper for introducing two inexact algorithms. We emphasize three features of the proposed algorithms:

- It is generally difficult to find an exact value for x^k given by (1.3). Thus, (1.3) should be solved approximately. A way to obtain this goal is to replace (1.3) by

  $$\epsilon_k \in \lambda_k \partial f(x^k) - \exp^{-1}_{x^k} x^{k-1}; \tag{1.4}$$

  where ∂ is an abstract subdifferential, defined in Section 3, which can recover a large class of subdifferencial already studied.

- To solve the minimization problem (1.1), it is desirable that the sequence $\{f(x^k)\}$ is decreasing. To obtain this property we impose the following criterion:

  $$d(\exp_{x^k} \epsilon^k, x^{k-1}) \le \max\{\|\epsilon^k\|, d(x^k, x^{k-1})\}. \tag{1.5}$$

  This condition may be considered as a generalization of the proximal point algorithm for quasiconvex function in euclidean spaces studied by Papa Quiroz et al.[37].

- The first inexact algorithm is based on the conditions:

$$\sum_{k=1}^{+\infty} \|\epsilon^k\| < +\infty \text{ and } \sum_{k=1}^{+\infty} |\langle \epsilon^k, \exp_z^{-1} x^k \rangle| < +\infty,$$

$\forall\, z \in U$, where U is certain nonempty set which contains the optimal solutions set. These criterion are motivate from the work of Eckstein [23] to solve singularity of monotone operators and Papa Quiroz and Oliveira [37] to solve quasiconvex minimization problems. With the above conditions we obtain that the sequence $\{x^k\}$ is Quasi-Fejér convergent to U.

The Second inexact algorithm consider the conditions:

$$\|\epsilon^k\| \leq \eta_k d(x^k, x^{k-1}) \text{ and } \sum_{k=1}^{+\infty} \eta_k^2 < +\infty,$$

which have been recently studied by Tang and Huang [43] for maximal monotone vector fields on Hadamard manifolds. The authors proved the convergence, see [43], and the superlinear rate of convergence,[44], of the iterations to a singularity of a monotone operator. In this work we extend the above convergence result for the quasiconvex case.

The paper is organized as follows: Section 2, we recall some definitions and results about Riemannian geometry and quasiconvex analysis. In Section 3, we will introduce the definition of abstract subdifferential on Riemannian manifolds, studed by Aussel [7] in Banch spaces. We also show that this subdifferential definition to allow to recover a large class of subdifferential, which it will be studied in this section. In Section 4, we define the general inexact algorithm for solving quasiconvex minimization problems, we will show its well-definedness and develop of the preliminar convergence analysis. In section 5, we introduz the two inexact algorithms an we prove the convergence of the method. In Section 6, we give an application to demand theory on Hadamard manifolds and in Section 7 we discuss the globally and locally strongly convexity of the objective function of the subproblems.

2 Some Basic Facts on Metric Spaces and Riemannian Manifolds

In this section we recall some fundamental properties and notation on Riemannian manifolds. Those basic facts can be seen, for example, in do Carmo [22], Sakai [41], Udriste [46] and Rapcsák [40].

Let M be a differential manifold with finite dimension n. We denote by T_xM the tangent space of M at x and $TM = \bigcup_{x\in M} T_xM$. T_xM is a linear space and has the same dimension of M. Because we restrict ourselves to real manifolds, T_xM is isomorphic to $\mathbb{R}^n$. If M is endowed with a Riemannian metric g, then M is a Riemannian manifold and we denote it by (M, G)

or only by M when no confusion can arise, where G denotes the matrix representation of the metric g. The inner product of two vectors $u, v \in T_xM$ is written as $\langle u, v\rangle_x := g_x(u, v)$, where g_x is the metrics at point x. The norm of a vector $v \in T_xM$ is set by $||v||_x := \langle v, v\rangle_x^{1/2}$. If there is no confusion we denote $\langle , \rangle = \langle , \rangle_x$ and $||.|| = ||.||_x$. The metrics can be used to define the length of a piecewise smooth curve $\alpha : [t_0, t_1] \to M$ joining $\alpha(t_0) = p'$ to $\alpha(t_1) = p$ through $L(\alpha) = \int_{t_0}^{t_1} \|\alpha'(t)\|_{\alpha(t)} dt$. Minimizing this length functional over the set of all curves we obtain a Riemannian distance $d(p', p)$ which induces the original topology on M.

Given two vector fields V and W in M, the covariant derivative of W in the direction V is denoted by $\nabla_V W$. In this paper ∇ is the Levi-Civita connection associated to (M, G). This connection defines an unique covariant derivative D/dt, where, for each vector field V, along a smooth curve $\alpha : [t_0, t_1] \to M$, another vector field is obtained, denoted by DV/dt. The parallel transport along α from $\alpha(t_0)$ to $\alpha(t_1)$, denoted by P_{α,t_0,t_1}, is an application $P_{\alpha,t_0,t_1} : T_{\alpha(t_0)}M \to T_{\alpha(t_1)}M$ defined by $P_{\alpha,t_0,t_1}(v) = V(t_1)$ where V is the unique vector field along α so that $DV/dt = 0$ and $V(t_0) = v$. Since ∇ is a Riemannian connection, P_{α,t_0,t_1} is a linear isometry, furthermore $P^{-1}_{\alpha,t_0,t_1} = P_{\alpha,t_1,t_0}$ and $P_{\alpha,t_0,t_1} = P_{\alpha,t,t_1} P_{\alpha,t_0,t}$, for all $t \in [t_0, t_1]$. A curve $\gamma : I \to M$ is called a geodesic if $D\gamma'/dt = 0$.

A Riemannian manifold is complete if its geodesics are defined for any value of $t \in I\!R$. Let $x \in M$, the exponential map $\exp_x : T_xM \to M$ is defined $\exp_x(v) = \gamma(1, x, v)$, for each $x \in M$. If M is complete, then $\exp_x$ is defined for all $v \in T_xM$. Besides, there is a minimal geodesic (its length is equal to the distance between the extremes).

Given the vector fields X, Y, Z on M, we denote by R the curvature tensor defined by $R(X, Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z + \nabla_{[X,Y]} Z$, where $[X, Y] := XY - YX$ is the Lie bracket. Now, the sectional curvature as regards X and Y is defined by

$$K(X, Y) = \frac{\langle R(X, Y)Y, X\rangle}{\|X\|^2\|Y\|^2 - \langle X, Y\rangle^2}.$$

Given an extended real valued function $f : M \to I\!R \cup \{+\infty\}$ we denote its domain by $\texttt{domf} := \{x \in M : f(x) < +\infty\}$ and its epigraph by $\texttt{epi}f := \{(x, \beta) \in M \times I\!R : f(x) \le \beta\}$. f is said to be proper if $\texttt{domf} \neq \emptyset$ and $\forall x \in \texttt{domf}$ we have $f(x) > -\infty$. f is a lower semicontinuous function if $\texttt{epi}f$ is a closed subset of $M \times I\!R$.

A function $f : M \to I\!R \cup \{+\infty\}$ is convex on M iff for any geodesic $\gamma : I\!R \to M$, the composition $f \circ \gamma : I\!R \to I\!R \cup \{+\infty\}$ is convex. The gradient of a differentiable function $f : M \to I\!R$, $\texttt{grad}f$, is a vector field on M defined by $df(X) = \langle \texttt{grad}f, X\rangle = X(f)$, where X is also a vector field on M. Let $f : M \to I\!R \cup \{+\infty\}$ be a convex function, $x \in M$ and a vector $v \in T_xM$. The directional derivative of f at $x \in M$ in the direction of a vector v is defined by

$$f'(x, v) = \lim_{t \to 0^+} \frac{f(\gamma(t)) - f(x)}{t}, \tag{2.6}$$

where $\gamma : I\!R \to M$ is a geodesic such that $\gamma(0) = x$ and $\gamma'(0) = v$.

Given $x \in M$, a vector $s \in T_xM$ is said to be a subgradient of a convex function f at x iff for any geodesic $\gamma : I\!R \to M$ with $\gamma(0) = x$,

$$(f \circ \gamma)(t) \ge f(x) + t\langle s, \gamma'(0)\rangle.$$

The set of all subgradients of f at x, $\partial^{FM} f(x)$, is called the Fenchel-Moreau subdifferential of f at x.

Let $f : M \to I\!R \cup \{+\infty\}$ to be a convex proper function, which is continuous at a point x where f is finite. It holds that

$$f'(x, v) = \max_{s \in \partial f(x)} \langle s, v \rangle, \tag{2.7}$$

for all $v \in T_x M$. This proof can be found in Proposition 3.2 of da Cruz Neto et al. [20].

Let f be a proper function f, it is called quasiconvex if for all $x, y \in M$, $t \in [0, 1]$, it holds that

$$f(\gamma(t)) \leq \max\{f(x), f(y)\},$$

for the geodesic $\gamma : [0, 1] \to I\!R$, so that $\gamma(0) = x$ and $\gamma(1) = y$.

Complete simply-connected Riemannian manifolds with nonpositive curvature are called *Hadamard manifolds*. Some examples of Hadamard manifolds may be find in Section 4 of Papa Quiroz and Oliveira [36].

Theorem 2.1 *Let M be a Hadamard manifold. Then M is diffeomorphic to the Euclidian space $I\!R^n$, $n = \dim M$. More precisely, at any point $x \in M$, the exponential mapping $\exp_x : T_x M \to M$ is a global diffeomorphism.*

Proof. See Sakai, [41], Theorem 4.1, page 221. ∎

A consequence of the preceding theorem is that Hadamard manifolds have the property of uniqueness of geodesic between any two points. Another useful property is the following: let $[x, y, z]$ be a geodesic triangle, which consists of *vertices* and the geodesics joining them. We have:

Theorem 2.2 *Given a geodesic triangle $[x, y, z]$ in a Hadamard manifold, it holds that:*

$$d^2(x, z) + d^2(z, y) - 2\langle \exp_z^{-1} x, \exp_z^{-1} y \rangle \leq d^2(x, y),$$

where $\exp_z^{-1}$ denotes the inverse of $\exp_z$.

Proof. See Sakai, [41], Proposition 4.5, page 223. ∎

Theorem 2.3 *Let M be a Hadamard manifold and let y be a fixed point. Then, the function $g(x) = d^2(x, y)$ is strictly convex and $\operatorname{grad} g(x) = -2 \exp_x^{-1} y$.*

Proof. See Ferreira and Oliveira, [25], Proposition II.8.3. ∎

Proposition 2.1 *Take $x \in M$. The map $d^2(\cdot, x)/2$ is strongly convex.*

Proof. See Da Cruz Neto et al. [21].

Let (M, d) be a complete metric space with metric d. We recall some fundamental results on Quasi-Fejér convergence, which it will be an useful tool to prove convergence result of the first proposed algorithm.

Definition 2.1 *A sequence $\{y^k\}$ in (M, d) is said to be Quasi-Fejér convergent to a set $W \subset M$ if, for every $w \in W$, there exists a non-negative, summable sequence $\{\delta_k\} \subset I\!R$ such that*

$$d^2(y^{k+1}, w) \leq d^2(y^k, w) + \delta_k$$

The next proposition is well-known and it shows the main properties of Quasi-Fejér convergence theory on Hadamard manifolds.

Proposition 2.2 *Let $\{y^k\}$ be a sequence in (M, d). If $\{y^k\}$ is Quasi-Fejér convergent to a non-empty set $W \subset M$, then $\{y^k\}$ is bounded. If furthermore, an accumulation point y of $\{y^k\}$ belongs to W, then $\lim_{k\to\infty} y^k = y$.*

3 An Abstract Subdifferential on Hadamard Manifolds

In order to solve nonsmooth and non convex optimization poblems in vectorial spaces, it has been generalized several concepts of subgradient and subdifferential in different way, as a consequence, there exists various subdifferential for non convex functions such as Fréchet, Clarke, Clarke-Rockafellar, Hadamard, Dinni, abstract subdifferential, between others, see for example Ioffe [27], Correa et. al [17], Thibault and Zagrodny [45], Clarke [15, 16], Aussel et al.[7], Rockafellar [39] and references therein.

In this section, we extend the definition of abstract subdifferential from Banach spaces, introduced by Aussel et al.[7], to Hadamard manifolds which will be used along the paper. We will note that the extension to Riemannian manifold is natural due that does not require any assumption about the sectional curvature.

Definition 3.1 *We call abstract subdifferential, denoted by ∂, any operator which associates a subset $\partial f(x)$ of T_xM to any lower semicontinuous function $f : M \to I\!R \cup \{+\infty\}$ and any $x \in M$, satisfying the following properties:*

a. *If f is convex, then $\partial f(x) = \{g \in T_xM \mid \langle g, \exp_x^{-1} z\rangle + f(x) \leq f(z), \quad \forall\, z \in M\}$;*

b. *$0 \in \partial f(x)$, if $x \in M$ is a local minimum of f;*

c. *$\partial(f + g)(x) \subset \partial f(x) + \partial g(x)$, whenever $g : M \to I\!R \cup \{+\infty\}$ is a convex continuous function which is ∂-differentiable at $x \in M$.*

Here, g is ∂-differentiable at x means that both $\partial g(x)$ and $\partial(-g)(x)$ are nonempty. We say that a function f is ∂-subdifferentiable at x when $\partial f(x)$ is nonempty.

Due to Clarke-Rockafellar and upper Dinni subdifferential are the biggest subset among the classical subdifferentials, see Aussel et al.[7], in this paper we restrict our attention to the following subdifferentials:

i). The Clarke-Rockafellar derivative at $x \in M$ in the direction $v \in T_xM$ is given by

$$f^{\uparrow}(x,v) = \sup_{\varepsilon>0} \limsup_{\substack{u \xrightarrow{f} x \\ t \searrow 0}} \inf_{d \in B_\epsilon(v)} \frac{f(\exp_u t(D\exp_x)_{\exp_x^{-1}u} d) - f(u)}{t},$$

where $(D\exp_x)_{\exp_x^{-1}u}$ is the differential of the exponential function, $\exp_x$, at $\exp_x^{-1}u$, $B_\epsilon(v) = \{d \in T_xM : \|d - v\| < \varepsilon\}$, $t \searrow 0$ indicates the fact that $t > 0$ and $t \to 0$, and $u \xrightarrow{f} x$ means that both $u \to x$ and $f(u) \to f(x)$.

ii). The upper Dinni derivative at $x \in M$ in the direction $v \in T_xM$ is defined by

$$f^{D^+}(x,v) = \limsup_{t \searrow 0} \frac{f(\exp_x(tv)) - f(x)}{t}.$$

Associated with the Clarke-Rockafellar and upper Dinni derivative are the following set-valued mappings given by

$$\begin{aligned} \partial^{CR} f(x) &= \Big\{ s \in T_xM \mid \langle s, v \rangle \le f^{\uparrow}(x,v), \ \forall\, v \in T_xM \Big\} \\ \partial^{D^+} f(x) &= \Big\{ s \in T_xM \mid \langle s, v \rangle \le f^{D^+}(x,v), \ \forall\, v \in T_xM \Big\}; \end{aligned}$$

called the Clarke-Rockafellar and upper Dinni subdifferential mapping, respectively.

Besides ∂^{CR} and ∂^{D^+}, one can define others subdifferential such as the Clarke and Fréchet subdifferential. We recall that the Clarke subdifferential at the point $a \in M$ is the set

$$\partial^C f(x) = \{s \in T_xM \mid \langle s, v \rangle \le f^\circ(x,v), \ \forall\, v \in T_xM\},$$

where

$$f^\circ(x,v) = \limsup_{\substack{u \to x \\ t \searrow 0}} \frac{f(\exp_u t(D\exp_x)_{\exp_x^{-1}u} v) - f(u)}{t}.$$

Another useful subdifferential is the Fréchet ones, which will be denoted by ∂^F. We say that $s \in \partial^F f(x)$ if for all $y \in M$ we have

$$f(y) \ge f(x) + \langle s, \exp_x^{-1} y \rangle + o(d(x,y)),$$

where

$$\lim_{\substack{y \to x \\ y \ne x}} \frac{o(d(x,y))}{d(x,y)} = 0.$$

In this paper we also use the following (limiting) subdifferential concept of f at $x \in M$, which is defined by

$$\partial^{Lim} f(x) := \{s \in T_xM \mid \exists\, x^k \to x, \ f(x^k) \to f(x), \exists\, s^k \in \partial f(x^k) : P_{\gamma_k,0,1} s^k \to s\},$$

4 Definition of the problem and the Algorithm.

Let M be a Hadamard manifold. We are interested in solving the problem:

$$\min_{x \in M} f(x) \tag{4.8}$$

where $f : M \to I\!R \cup \{+\infty\}$ is an extended real-valued function which satisfies the following assumption:

(H1) f is a proper, bounded from below and lower semicontinuous quasiconvex function.

4.1 Algorithm

The proximal point algorithm to solve the problem (4.8), given an arbitrary initial point $x^0 \in M$, generates a sequence $\{x^k\}$, given by

$$x^{k+1} \in \arg\min_{x \in M} \left\{ f(x) + \frac{1}{2\lambda_k} d^2(x, x^k) \right\}, \tag{4.9}$$

where d the intrinsic Riemannian distance and $\{\lambda_k\}$ is a sequence of positive numbers. The convergence of the algorithm under inexact computation of the iterates, that is, when (4.9) is replaced by

$$\epsilon_k \in \lambda_k \partial f(x^k) - \exp^{-1}_{x^k} x^{k-1}; \tag{4.10}$$

comes up for at least two reasons: first of all, it is generally difficult to find an exact value for x^k given by (4.9); second, it is clearly inefficient to spend too much effort in the computation of a given iterate x^k when only the limit of the sequence has the desired properties.

On the other hand, the proximal algorithm, for the convex case, as many minimizing algorithms to solve (4.8) has the property to generate a minimizing sequence, that is, a sequence $\{x^k\}$ such that $\{f(x^k)\}$ is decreasing. One way to ensure such a property for the quasiconvex case, is to focus on a result for quasiconvex functions, see Lemma 4.2,

$$\text{If there exists } g^k \in \partial f(x^k) \, : \, \langle g^k, \exp^{-1}_{x^k} x^{k-1} \rangle > 0 \implies f(x^k) \leq f(x^{k-1}).$$

Indeed, from (4.10) we have $\langle \epsilon^k + \exp^{-1}_{x^k} x^{k-1}, \exp^{-1}_{x^k} x^{k-1} \rangle > 0$. It implies that

$$\langle \epsilon^k, \exp^{-1}_{x^k} x^{k-1} \rangle > -d(x^{k-1}, x^k).$$

Therefore, a sufficient condition to obtain $f(x^k) \leq f(x^{k-1})$ is given by

$$\langle \epsilon^k, \exp^{-1}_{x^k} x^{k-1} \rangle \geq 0;$$

that is, $0 \leq \|\epsilon^k\| d(x^k, x^{k-1}) \cos\alpha$, where α is the angle between vectors e^k and $\exp^{-1}_{x^k} x^{k-1}$. Then x^k is surely better than x^{k-1} whenever α is an acute angle.

If we consider the triangle $[x^{k-1}, x^k, \exp_{x^k} \epsilon^k]$, shown in the figure 1, it is easy to see that α will be acute if the side opposite to it is not the largest one, or equivalently, it is suffice have the following criterion:

$$d(\exp_{x^k} \epsilon^k, x^{k-1}) \leq \max\{\|\epsilon^k\|, d(x^k, x^{k-1})\}. \tag{4.11}$$

Considering (4.10) and (4.11) we will present an inexact proximal point algorithm on Hadamard manifolds to solve the problem (4.8), which we will denote by HMIP algorithm.

HMIP Algorithm.

Initialization: Take $x^0 \in M$. Set $k = 0$.

Iterative step: Given $x^{k-1} \in M$, find $x^k \in M$ and $\epsilon^k \in T_{x^k}M$ such that

$$\epsilon^k \in \lambda_k \partial f(x^k) - \exp^{-1}_{x^k} x^{k-1} \tag{4.12}$$

where

$$d(\exp_{x^k} \epsilon^k, x^{k-1}) \leq \max\left\{||\epsilon^k||, d(x^k, x^{k-1})\right\}. \tag{4.13}$$

Stopping rule: If $x^{k-1} = x^k$ or $0 \in \partial f(x^k)$. Otherwise, $k + 1 \leftarrow k$ and go to *Iterative step*.

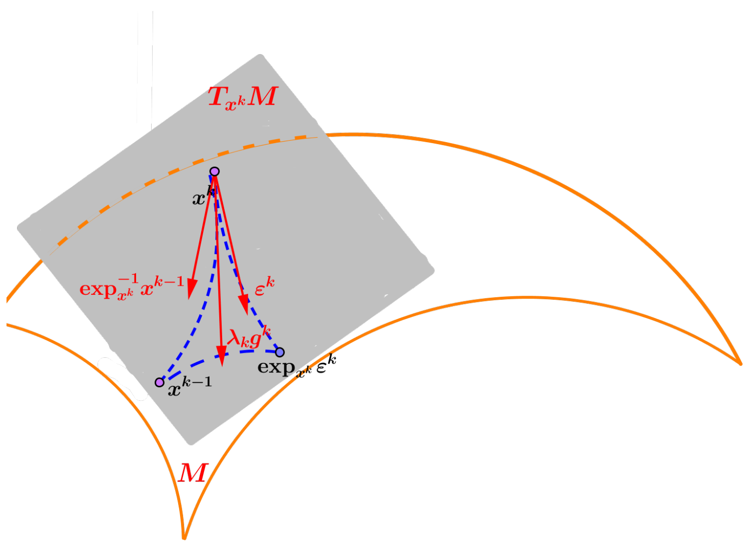

Figure 1: geodesic triangle $[x^{k-1}, x^k, \exp^k_x \epsilon^k]$. Iterates of the HMIP algorithm on the Hadamard manifolds.

Remark 4.1 *From the HMIP Algorithm we have*

1. *If* $x^k = x^{k+1}$, *it follows that* $\epsilon^k \in \lambda_k \partial f(x^k)$, *which implies that* x^k *is an approximate solution of a critical point whenever* $\frac{\varepsilon_k}{\lambda_k}$ *is small enough.*

2. *If $M = I\!R^n$, then* (4.12) *and* (4.13) *reduce, respectively, to*

$$\epsilon^k \in \lambda_k \partial f(x^k) + x^k - x^{k-1},$$

and

$$\|x^{k-1} - x^k - \epsilon^k\| \le \max\{\|\epsilon^k\|, \|x^k - x^{k-1}\|\}$$

Then, the HMIP Algorithm extends the inexact proximal point algorithm proposed by Papa Quiroz et al. [37] from $I\!R^n$ to Hadamard manifolds.

3. *If $\epsilon^k = 0$ in* (4.12) *for every $k \ge 0$, we get*

$$\frac{1}{\lambda_k}\exp^{-1}_{x^k} x^{k-1} \in \partial f(x^k).$$

Thus, if f is convex and $\varepsilon_k = 0$ then the HMIP Algorithm reduces to the proximal point algorithm on Hadamard manifolds studied by Ferreira and Oliveira [25].

Theorem 4.1 *Let M be a Hadamard manifold. If $f : M \to \{+\infty\}$ is a proper, lower bounded and lower semicontinuous function. Then, the sequence $\{x^k\}$, generated by the proximal method is well defined.*

Proof. As $d^2(\cdot, x^{k-1})$ is coercive then there exists a minimum $\tilde{z} \in M$ of $f(\cdot) + \frac{1}{2\lambda_k} d^2(\cdot, x^{k-1})$. From Definition 3.1, items a and b, we obtain

$$\begin{aligned} 0 &\in \partial\left(f(\cdot) + \frac{1}{2\lambda_k} d^2(\cdot, x^{k-1})\right)(\tilde{z}). \\ &\subset \lambda_k \partial f(\tilde{z}) - \exp^{-1}_{\tilde{z}} x^{k-1},. \end{aligned}$$

where the last expression is obtained by the fact that $d^2(\cdot, x^{k-1})$ is a convex continuous function which is also ∂-differentiable at $\tilde{z} \in M$. Thus, considering $e^k = 0$, such a minimizer $\tilde{z}$ satisfies (4.12) and (4.13) and can be taken as x^k. Thus, we conclude the existence of iterates x^k and e^k satisfying the expressions (4.12) and (4.13). ■

The following two results are used to show that HMIP Algorithm is a descent method.

Lemma 4.1 *Let M be a Hadamard manifold. Suposse equation* (4.13) *holds. Then*

$$\langle \epsilon^k, \exp^{-1}_{x^k} x^{k-1} \rangle \ge 0.$$

Proof. Taking $z = x^k$, $x = \exp_{x^k} \epsilon^k$ and $y = x^{k-1}$ in Theorem 2.2 that

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

Therefore, the result follows from (4.13). ■

Next, using the same arguments as in Theorem 2.1 by Aussel[8], we show a property of l.s.c. quasiconvex function on Hadamard manifolds. This result will play a central role in the sequel.

Lemma 4.2 *Let M be a Hadamard manifold. Suposse that $f : M \to I\!R \cup \{+\infty\}$ satisfies the assumption* (H1) *and there exists $g \in \partial f(x)$ such that $\langle g, \exp_x^{-1} z\rangle > 0$. Then it holds that*

i. *If $\partial f \subset \partial^{CR} f$ and assuming f to be continuous, then $f(x) \le f(z)$.*

ii. *If $\partial f \subset \partial^{D^+} f$, then $f(x) < f(z)$.*

Proof. i. Case $\partial f \subset \partial^{CR} f$. Let $x, z \in M$ and $g \in \partial f(x)$ such that $0 < \langle g, \exp_x^{-1} z\rangle \le f^{\uparrow}(x, \exp_x^{-1} z)$. Then, there exists $\varepsilon' > 0$ such that

$$0 \quad < \quad \inf_{\eta>0} \sup_{d(u,x)+t+|f(u)-f(x)|<\eta} \inf_{d\in B_{\epsilon'}(\exp_x^{-1} z)} \frac{f(\exp_u t(D\exp_x)_{\exp_x^{-1} u} d) - f(u)}{t}.$$

Hence, for all $\eta > 0$ there exists $u(\eta) \in M$ and $t(\eta) \in I\!R_{++}$ such that $d(u(\eta), x) + t(\eta) + |f(u(\eta)) - f(x)| < \eta$ for all $d \in B_{\varepsilon'}(\exp_x^{-1} z)$ we obtain

$$0 < \frac{f(\exp_{u(\eta)} t(\eta)(D\exp_x)_{\exp_x^{-1} u(\eta)} d) - f(u(\eta))}{t(\eta)}.$$

Taking $\eta = 1/k$, from the infimum property, there exists sequences $\{u^k\} \in M$ and $\{t^k\} \in I\!R_{++}$ so that $d(u^k, x) + t^k + |f(u^k) - f(x)| < 1/k$ and

$$f(\exp_{u^k} t^k (D\exp_x)_{\exp_x^{-1} u^k}(\exp_x^{-1} z)) > f(u^k). \tag{4.14}$$

Denote $r^k = (D\exp_x)_{\exp_x^{-1} u^k}(\exp_x^{-1} z)$, which yields

$$\lim_{k\to+\infty} r^k = (D\exp_x)_0(\exp_x^{-1} z) = (\exp_x^{-1} z) \tag{4.15}$$

Let $\alpha : [0, 1] \to M$ be the minimizing geodesic such that $\alpha(0) = u^k$ and $\alpha'(0) = r^k$. Thus, expression (4.14) can be written as

$$f(u^k) < f(\alpha(u^k, 1, t^k r^k)) = f(\alpha(u^k, t^k, r^k)), \tag{4.16}$$

where the last equality is obtained by the well known homogeneity property for geodesics.

On the other hand, since $t^k < 1$, it follows that $0 < t^k < 1/k - d(u^k, x) < 1$, and hence, using the quasiconvexity of f,

$$\begin{aligned} \max\left\{ f(\alpha(u^k, 0, r^k), f(\alpha(u^k, 1, r^k)) \right\} &\ge f(\alpha(u^k, t^k, r^k)) \\ &> f(u^k), \end{aligned}$$

where the last inequality is obtained by (4.16). Since $f(\alpha(u^k, 0, r^k) = f(u^k)$, we have that $f(\alpha(u^k, 1, r^k) > f(u^k)$. Assuming continuity, the claim follows taking limit for $k \to +\infty$ and considering expression (4.15). This completes the proof of i..

ii. Case $\partial f \subset \partial^{D^+}$. The proof is very simple. Indeed, if $x, z \in M$ and $g \in T_xM$ such that $0 < \langle g, \exp_x^{-1} z\rangle \leq f^{D+}(x, \exp_x^{-1} z)$. Hence, there exists $\bar{t} \in]0,1[$ such that

$$f(x) < f(\exp_x(\bar{t}\exp_x^{-1} z)). \tag{4.17}$$

Let

$$\gamma(t) = \exp_x(t\exp_x^{-1} z) = \gamma(1, x, t\exp_x^{-1} z) = \gamma(t, x, \exp_x^{-1} z),$$

where the last equality is due to the homogeneous property of the geodesic γ, see [22, Lemma 2.6]. Thus $\gamma(0) = x$ and $\gamma(1) = z$, then from the quasiconvexity of f and as $\bar{t} \in]0,1[$ we have

$$f(\exp_x(\bar{t}\exp_x^{-1} z)) = f(\gamma(\bar{t})) \leq \max\{f(x), f(z)\}$$

From (5.29) we conclude that, $f(x) < f(z)$. ■

Using the same ideas of Aussel[8] we consider the following assumption:

(H2) $\quad\big(\partial \subset \partial^{D+}\big)$ or $\big(\ \partial \subset \partial^{CR}$ and f is continuous in $M.\big)$

Next we can state the assertion that HMIP Algorithm is a descent method.

Proposition 4.1 *Suppose that assumptions* (H1) *and* (H2) *are satisfied and let* $\{x^k\}$ *be the sequence given by the HMIP algorithm* (4.12)-(4.13)*, then*

a. If $\partial \subset \partial^{CR}$ *then* $f(x^k) \leq f(x^{k-1})$

b. If $\partial \subset \partial^{D^+}$ *then* $f(x^k) < f(x^{k-1})$

thus, we have that $\{f(x^k)\}$ *is nonincreasing (decreasing in the case b) and converges.*

Proof. From (4.12) and Theorem 2.3, there exists $g^k \in \partial f(x^k)$ such that

$$\begin{aligned} \epsilon^k &= \lambda_k g^k - \exp_{x^k}^{-1} x^{k-1} && (4.18)\\ &= \lambda_k g^k + \operatorname{grad}\frac{1}{2}d^2(\cdot, x^{k-1})(x^k). && (4.19)\end{aligned}$$

On the other hand, due to Proposition 2.1, $d^2(\cdot, x^{k-1})$ is strongly convex and consequently $\operatorname{grad}\frac{d^2}{2}(\cdot, x^{k-1})$ is strongly monotone and because $x^k \neq x^{k-1}$ we have

$$\langle \mathcal{P}_{x^k,x^{k-1}}\operatorname{grad}\frac{d^2}{2}(\cdot, x^{k-1})(x^k) - \operatorname{grad}\frac{d^2}{2}(., x^{k-1})(x^{k-1}), \exp_{x^{k-1}}^{-1} x^k\rangle > 0, \tag{4.20}$$

where $\mathcal{P}_{x,y}$ is the parallel transport of x to y along the (minimal) geodesic.
As $\operatorname{grad}\frac{d^2}{2}(., y)(y) = 0$, $\mathcal{P}_{x,y}\exp_x^{-1} y = -\exp_y^{-1} x$ and due to the isometric property of the parallel transport $\mathcal{P}_{x,y}$, (4.20) yields

$$\langle \operatorname{grad}\frac{d^2}{2}(\cdot, x^{k-1})(x^k), \exp_{x^k}^{-1} x^{k-1}\rangle < 0. \tag{4.21}$$

Combining (4.21) and (4.19), we obtain

$$\langle \epsilon^k - \lambda_k g^k, \exp_{x^k}^{-1} x^{k-1} \rangle < 0,$$

and consequently,

$$\langle g^k, \exp_{x^k}^{-1} x^{k-1} \rangle > \frac{1}{\lambda_k} \langle \epsilon^k, \exp_{x^k}^{-1} x^{k-1} \rangle.$$

Applying now Lemma 4.1 gives

$$\langle g^k, \exp_{x^k}^{-1} x^{k-1} \rangle > 0.$$

Thus, the result follows from Lemma 4.2. ■

Now, define

$$U := \left\{ x \in M \;\mid\; f(x) \leq \inf_k f(x^k) \right\}.$$

Note that, if $U = \emptyset$ then for all $x \in M$ it follows that $f(x) > \inf_k f(x^k) \geq \inf_{x \in M} f(x)$, and consequently, it is clear that $\lim_{k\to\infty} f(x^k) = \inf_{x\in I\!N} f(x^k) = \inf_{x\in M} f(x)$.

From now on we suppose that U is nonempty.

5 Some Variants of the PPA Algorithm

In this section, we present and analyze two variants of the HMIP algorithm based on two error criteria, which ensures that, under certain assumptions about error tolerance, convergence results for both algorithms are guaranteed.

5.1 HMIP1 Algorithm and Convergence Results

We now state an algorithm for solving (4.8), denoted by HMIP1, motivated from the work of Eckstein [23]. Among the features of this algorithm, Quasi-Fejér convergence theory plays a crucial role in the proof of its convergence results.

HMIP1 Algorithm.

Considering the *Initialization* of the HMIP Algorithm. Let $\{x^k\}$ and $\{\epsilon^k\}$ be sequences generated by (4.12)-(4.13) such that

(B1) $\sum_{k=1}^{+\infty} \|\epsilon^k\| < +\infty$.

(B2) $\sum_{k=1}^{+\infty} |\langle \epsilon^k, \exp_z^{-1} x^k \rangle| < +\infty, \quad \forall\, z \in U$.

Remark 5.1 *According to assumption* (B1)*, it is important to observe that there is a certain leeway in choosing* ϵ^k*. In fact, this assumptions was also considered natural in the classical inexact proximal point, see Rockafellar* [38]*.* (B2) *is a theoretical condition which induces us to think that iterates is closer to the set* U *whenever ones progress.*

The following result is used to show that sequences $\{x^k\}$, generated by MHIP1 Algorithm, is bounded.

Proposition 5.1 *If all the assumptions of the problem:* (H1) *and* (H2) *are satisfied, then*

$$d^2(x, x^k) \le d^2(x, x^{k-1}) - d^2(x^{k-1}, x^k) - 2\langle \epsilon^k, \exp^{-1}_{x^k} x\rangle, \quad \forall\ x \in U.$$

Proof. We consider two cases:

i. If $\partial f \subset \partial^{CR} f$. Let $x \in U$. Since that $U \subset U_k := \{x \in M : f(x) \le f(x^k)\}$ and $int(U_k) = \{x \in M : f(x) < f(x^k)\}$ is nonempty (because $0 \notin \partial f(x^k)$, then x^k is not a minimum of f which implies that $inf_{x\in M} f(x) < f(x^k)$) there exists $\{x^l\} \subset M$ with $f(x^l) < f(x^k)$ such that $x^l \to x$. It follows from Lemma 4.2 and (4.18) that

$$\langle \epsilon^k, \exp^{-1}_{x^k} x^l\rangle \le -\langle \exp^{-1}_{x^k} x^{k-1}, \exp^{-1}_{x^k} x^l\rangle \tag{5.22}$$

Taking $z = x^k, x = x^{k-1}$ and $y = x^l$ in Theorem 2.2, we obtain

$$-2\langle \exp^{-1}_{x^k} x^{k-1}, \exp^{-1}_{x^k} x^l\rangle \le d^2(x^{k-1}, x^l) - d^2(x^{k-1}, x^k) - d^2(x^k, x^l). \tag{5.23}$$

Combining (5.22) and (5.23), we get

$$2\langle \epsilon^k, \exp^{-1}_{x^k} x^l\rangle \le d^2(x^{k-1}, x^l) - d^2(x^{k-1}, x^k) - d^2(x^k, x^l).$$

Therefore, the result follows taking the limit as $l \to +\infty$

ii. If $\partial f \subset \partial^{D^+} f$. Let $x \in U$, then $f(x) \le f(x^k)$, It follows from Lemma 4.2 and (4.18) that

$$\langle \epsilon^k, \exp^{-1}_{x^k} x\rangle \le -\langle \exp^{-1}_{x^k} x^{k-1}, \exp^{-1}_{x^k} x\rangle \tag{5.24}$$

Taking $z = x^k, x = x^{k-1}$ and $y = x$ in Theorem 2.2, we obtain

$$-2\langle \exp^{-1}_{x^k} x^{k-1}, \exp^{-1}_{x^k} x\rangle \le d^2(x^{k-1}, x) - d^2(x^{k-1}, x^k) - d^2(x^k, x). \tag{5.25}$$

Combining (5.24) and (5.25), we get

$$2\langle \epsilon^k, \exp_{x^k} x\rangle \le d^2(x^{k-1}, x) - d^2(x^{k-1}, x^k) - d^2(x^k, x).$$

■

The convergence analysis of HMIP1 algorithm will be greatly simplified by the following results.

Lemma 5.1 *Let* $\{x^k\}$ *and* $\{\epsilon^k\}$ *be sequences generated by* HMIP1 Algorithm . *Assume that assumptions* $(H1), (H2)$ *and* $(B2)$ *are satisfied. Then it holds that*

(i) *The sequence* $\{x^k\}$ *is bounded;*

(ii) *For all* $x \in U$*, the sequence* $\{d(x^k, x)\}$ *is convergent;*

(iii) $\lim_{k\to\infty} d(x^k, x^{k-1}) = 0$*;*

(iv) *If* $x^{k_j} \to \bar{x}$ *then* $\lim_{k\to\infty} x^{k_j-1} \to \bar{x}$.

Proof. In order to prove the boundedness of $\{x^k\}$, we will show that the sequence $\{x^k\}$ is Quasi-Fejér convergent to U, in that case the result follow from Proposition 2.2. Indeed, Proposition 5.1 implies that

$$d^2(x, x^k) \le d^2(x, x^{k-1}) + \delta_k, \quad \forall\, x \in U$$

where $\delta_k = 2|\langle \epsilon^k, \exp^{-1}_{x^k} x\rangle|$. Since assumption (B2) is satisfied, it follows that $\{x^k\}$ is a Quasi-Fejér convergent sequence and thus, as pointed out before, $\{x^k\}$ is bounded.
(ii): By Proposition 5.1, $d(x, x^k) \le d(x, x^{k-1})$. Then, $\{d(x, x^k\})$ is a bounded below non-decreasing sequence and hence convergent.
(iii): Once again, using Proposition 5.1, we have $0 \le d^2(x^{k-1}, x^k) \le d^2(x, x^{k-1}) - d^2(x, x^k)$. Upon taking the limit as $k \to +\infty$ which, together with (ii), we conclude that $\lim_{k\to+\infty} d(x^{k-1}, x^k) = 0$.
(iv): From the "triangle inequality property", we have that

$$d(x^{k_j-1}, \bar{x}) \le d(x^{k_j-1}, x^{k_j}) + d(x^{k_j}, \bar{x}).$$

As the right-hand side converges by (iii) and from the hypothesis, we obtain the claim. ∎

We are now ready to prove the two main results of this section.

Theorem 5.1 *Let* $\{x^k\}$ *and* $\{\epsilon^k\}$ *be sequences generated by* HMIP1 Algorithm . *Considering the assumptions* (H1), (H2) *and* (B1), (B2)*. Then, the sequence generated by this ones,* $\{x^k\}$*, converges to a point of* U*.*

Proof. This proof is rather similar to the proof of Theorem 4.3 by Papa Quiroz and Oliveira [36]. ∎

Theorem 5.2 *Let* $\{x^k\}$ *and* $\{\epsilon^k\}$ *be sequences generated by* HMIP1 Algorithm . *Considering the assumptions* (H1), (H2) *and* (B1)*,* (B2) *with* $\tilde{\lambda}$ *such that* $\tilde{\lambda} < \lambda_k$ *and* f *be a continuous function on* $dom(f)$*, then* $\{x^k\}$ *converges to some* $\bar{x}$ *with* $0 \in \partial^{Lim} f(\bar{x})$*.*

Proof. As direct consequence of Theorem 5.1, there exists a point $\bar{x} \in U$ such that $x^k \to \bar{x}$. Since that f is continuous, we have $f(x^k) \to f(\bar{x})$. From (4.18), Let $g^k \in \partial f(x^k)$ such that

$$g^k = \frac{1}{\lambda_k}(\epsilon^k + \exp^{-1}_{x^k} x^{k-1}).$$

It remains to show that $\mathcal{P}_{\psi_k,0,1} g^k$ converges to 0, for the geodesic ψ_k joining x^k to $\bar{x}$. Indeed, by assumption (B1), the continuity of $\exp^{-1}_{x^k}(\cdot)$ and since the operator $\mathcal{P}_{\psi_k,0,1}$ is linear, consequently continuous, we have

$$\lim_{k\to\infty} \mathcal{P}_{\psi_k,0,1}(\epsilon^k + \exp^{-1}_{x^k} x^{k-1}) = \lim_{k\to\infty} \mathcal{P}_{\psi_k,0,1}\epsilon^k + \lim_{k\to\infty} \mathcal{P}_{\psi_k,0,1}\exp^{-1}_{x^k} x^{k-1} = 0 + 0,$$

The claim therefore follows from the boundedness of $1/\lambda_k$. ■

Theorem 5.3 *Let* $\{x^k\}$ *and* $\{\epsilon^k\}$ *be sequences generated by* HMIP1 Algorithm *. If the assumptions* (H1)-(H2), (B1) *and* (B2) *are satisfied with* $\tilde{\lambda}$ *such that* $\tilde{\lambda} < \lambda_k$ *and* f *is locally Lipschitz, then* $\{x^k\}$ *converges to some* $\bar{x}$ *with* $0 \in \partial^o f(\bar{x})$.

Proof. From (4.18), $(1/\lambda_k)(\epsilon_k + \exp_{x^k} x^{k-1}) \in \partial^\circ f(x^k)$, so

$$f^\circ(x^k, v) \geq (1/\lambda_k)\langle \epsilon^k + \exp_{x^k} x^{k-1}, v\rangle, \ \forall v \in T_{x^k}M.$$

Let $\bar{x}$ be the limit of $\{x^k\}$, and $\bar{v} \in T_{\bar{x}}M$, then

$$f^\circ(x^k, v^k) \geq (1/\lambda_k)\langle \epsilon^k + \exp_{x^k} x^{k-1}, v^k\rangle,$$

where $v^k = D(\exp_{\bar{x}})_{\exp_{\bar{x}}^{-1} x^k}\bar{v}$. Taking lim sup in the above inequality and using Proposition 4.1 of [13] we obtain that

$$f^\circ(\bar{x}, \bar{v}) \geq \limsup_{k\to+\infty} f^\circ(x^k, v^k) \geq 0$$

It follows that $0 \in \partial^\circ f(\bar{x})$. ■

5.2 Algorithm HMIP2 and Convergence Results

We now state a class of HMIP for solving (4.8). It is worth mentioning that the convergence results of the below algorithm do not really need Quasi- Fejér convegence theory, as in the previous HMIP1 Algorithm.

Algorithm HMIP2

Considering the *Initialization* of the HMIP Algorithm. Let $\{x^k\}$ and $\{\epsilon^k\}$ be sequences generated by (4.12)-(4.13) such that

(C1) $\|\epsilon^k\| \leq \eta_k d(x^k, x^{k-1})$.

(C2) $\sum_{k=1}^{+\infty} \eta_k^2 < +\infty$.

Remark 5.2 *At first sight, this algorithm seems computational more practical and many implementable when compared to the* HMIP1 *Algorithm, since knowledge of vectors of* U *is not required.*

Lemma 5.2 *If all the assumptions of the problem:* (H1)-(H2), (C1) *and* (C2) *are satisfied, then there exists an integer* $k_0 \geq 0$ *such that for all* $k \geq k_0$ *we have*

$$d^2(x^k, x) \leq \left(1 + \frac{2\eta_k^2}{1 - 2\eta_k^2}\right) d^2(x^{k-1}, x) - \frac{1}{2}d^2(x^k, x^{k-1}), \quad \forall \in x \in U \tag{5.26}$$

Furthermore, $\{x^k\}$ *is a bounded sequence and* $\lim_{k\to+\infty} d(x^k, x^{k-1}) = 0$.

Proof. The proof is obtained using an argument quite similar to the one used in Lemma 3.2 by Tang and Huang [43]. Indeed, for $\eta_k > 0$ and since η_k converges to 0, there exists $k_0 \geq 0$ such that for all $k \geq k_0$, $1 - 2\eta_k^2 > 0$. Hence expression (5.26) can be easily obtained.

To prove that $\{x^k\}$ is a bounded sequence, we can use the recurrence inequality given by (5.26). Thus,

$$d^2(x^k, x) \leq \prod_{i=k_0}^{k} \left(1 + \frac{2\eta_i^2}{1 - 2\eta_i^2}\right) d^2(x^{k_0}, x). \tag{5.27}$$

Since $\{\eta_k^2\}$ is a convergent sequence, for all $\varepsilon > 0$ there exists $\tilde{k}_0$ so that $\eta_k^2 < \varepsilon$, for $k \geq \tilde{k}_0$. Namely, for all $\tilde{k}_0 > k_0$ we get

$$0 < 1 - 2\varepsilon < 1 - 2\eta_k^2 < 1.$$

Thus, for all $\varepsilon < 1/2$ we have

$$\frac{2\eta_k^2}{1 - 2\eta_k^2} < \frac{2\eta_k^2}{1 - 2\varepsilon}, \qquad \forall\ k > k_0. \tag{5.28}$$

Summing up inequality (5.28) and considering (C2),

$$\sum_{k=1}^{+\infty} \frac{2\eta_k^2}{1 - 2\eta_k^2} < +\infty. \tag{5.29}$$

On the other hand, from inequality of arithmetic and geometric means, we obtain that

$$\begin{aligned} \prod_{k=k_0}^{n} \left(1 + \frac{2\eta_k^2}{1 - 2\eta_k^2}\right) &\leq \left(\frac{1}{n - k_0 + 1} \sum_{k=k_0}^{n} \left(1 + \frac{2\eta_k^2}{1 - 2\eta_k^2}\right)\right)^{n-k_0+1} \\ &= \left(1 + \frac{1}{n - k_0 + 1} \sum_{k=k_0}^{n} \frac{2\eta_k^2}{1 - 2\eta_k^2}\right)^{n-k_0+1}. \end{aligned}$$

Taking limit, as $n \to +\infty$, we have

$$\lim_{n \to +\infty} \prod_{k=k_0}^{n} \left(1 + \frac{2\eta_k^2}{1 - 2\eta_k^2}\right) \leq e^{\sum_{k=k_0}^{n} \frac{2\eta_k^2}{1-2\eta_k^2}},$$

which, together with (5.29), gives $\prod_{k=k_0}^{+\infty} \left(1 + \frac{2\eta_k^2}{1-2\eta_k^2}\right) < +\infty$. Therefore, By (5.27), we conclude that $\{x^k\}$ is bounded.

Finally, it is easy to check the last result. Summing up (5.26), we have for all $k \geq k_0$

$$\begin{aligned}\frac{1}{2}\sum_{k=k_0}^{n} d^2(x^{k+1}, x^k) &\leq \sum_{k=k_0}^{n}(d^2(x^k, x) - d^2(x^{k+1}, x)) + \sum_{k=k_0}^{n}\frac{2\eta_k^2}{1-2\eta_k^2}d^2(x^k, x) \\ &\leq d^2(x^{k_0}, x) - d^2(x^n, x) + \max_{n\geq k\geq k_0}\{d^2(x^k, x)\}\sum_{k=k_0}^{n}\frac{2\eta_k^2}{1-2\eta_k^2}.\end{aligned}$$

Hence note that since $\{x^k\}$ is a bounded sequence, we have that $d(x^k, x)$ is one for all $k \geq 0$. Therefore, taking limit as $n \to +\infty$ and considering expression (5.29), we conclude that $d(x^{k+1}, x^k)$ converges to 0. ■

Theorem 5.4 *Let* $\{x^k\}$ *and* $\{\epsilon^k\}$ *be sequences generated by* HMIP2 Algorithm . *If all the assumptions of the problem:* (H1)-(H2), (C1) *and* (C2) *are satisfied with* $\tilde{\lambda}$ *such that* $\tilde{\lambda} < \lambda_k$ *and* f *be a continuous function, then* $\{x^k\}$ *converges to some* $\bar{x}$ *with* $0 \in \partial^{Lim} f(\bar{x})$.

Proof. The proof is rather similar to the one of Theorem 5.2 combined with Theorem 3.1 by Tang and Huang [43] ■

Theorem 5.5 *Let* $\{x^k\}$ *and* $\{\epsilon^k\}$ *be sequences generated by the* HMIP2 Algorithm . *If the assumptions* (H1)-(H2), (C1) *and* (C2) *are satisfied with* $\tilde{\lambda}$ *such that* $\tilde{\lambda} < \lambda_k$ *and* f *is a locally Lipschitz function, then* $\{x^k\}$ *converges to some* $\bar{x}$ *with* $0 \in \partial^o f(\bar{x})$.

Proof. Simmilar to Theorem 5.3. ■

6 Application to demand theory

In demand theory, there are three elements which characterize the decision problem of the agent:

i. Set of choice, denoted by $X \subset M$ which represents the universe of alternatives.

ii. Valuation criteria, which studies how agents evaluate different alternatives that are offered. This criterion is defined by a reflexive and transitive binary relation, $\precsim$, economically interpreted as “it is at least as preferable”, which reflect their preferences over the set of choice X. In economics this is called preference relation.

iii. Set of constraints on X, denoted by C. It defines the set of opportunities on which the agent can effectively choose.

Given the choice set X, we say that $\mu : X \to I\!R$ is a utility function representing $\precsim$ if for all $x, y \in X$ we have

$$y \precsim x, \iff \mu(y) \leq \mu(x).$$

Note that the utility function is not always available. Indeed, define on the set $X = M = I\!R^2$ the lexicographical relation:

$$\text{For } x = (x_1, x_2), y = (y_1, y_2) \in I\!R^2, x \succeq y \quad \Longleftrightarrow \quad x_1 > y_1 \text{ or } (x_1 = y_1 \text{ and } x_2 \geq y_2)\,.$$

Fortunately a large class of preferences $\precsim$, under certain conditions on X, can be represented by a utility function μ. See Mas Colell et al.[31, Proposition 3.c.1] for the Euclidean case. If a preference relation $\precsim$ is represented by a utility function μ, then the problem of choosing the best alternative is equivalent to solving the maximization problem:

$$(P)\max\{\mu\,(x) : x \in C \subset X\}.$$

A particular type of utility function is the quasi-concave one defined on the convex set X, ie, μ satisfies:

$$\mu(\gamma(t)) \geq \min\{\mu(x), \mu(y)\}, \text{ for all } x, y \in X \text{ and } \lambda \in [0, 1],$$

and for all geodesic $\gamma : [0, 1] \longrightarrow M$ such that $\gamma(0) = x$ and $\gamma(1) = y$.

It can be proven that this function is intimately related to the convexity hypothesis of the preference, ie, $\precsim$ is convex if given $x, y \in X$ with $x \precsim y$ and $x \precsim z$ in X and $0 \leq \alpha \leq 1$ then

$$x \precsim \gamma(t).$$

for all geodesic curve $\gamma : [0, 1] \longrightarrow M$, such that $\gamma(0) = z$ and $\gamma(1) = y$.

In economic terms, this definition is interpreted as a natural tendency of the agent to diversify its choice among all best, that is, if z and y are two alternatives at least as preferable than x, then it is natural to think that the convex combination of y and z, $\gamma(\alpha)$, for all $\alpha \in [0, 1]$, remains at least as preferable that x. Thus, if the hypothesis of diversification of the choice is taken and X and C are (geodesically) convex, then the optimization problem (P) has a quasi-concave objective function. Finally, defining $f(x) = -\mu(x)$, we obtain the following quasiconvex minimization problem:

$$(P')\min\{f(x) : x \in C \subset X\}.$$

On the other hand, Attouch and Soubeyran [6] stablished a procedure for finding, using an algorithm which is a mixture between local search and proximal algorithms, minimal points of f in a dynamical decision model. In the Riemannian framework, the proximal regularization term $d^2(x, x^k)$ caracterizes the "cost to change", where the agent (taken in a broad sense) has to explore and learn in order to find a satisfactory action.

For simplicity, assume that $C = X = M$ and consider that, at each step, the agent explores a set $E(x, \epsilon(x)) \subset M$ around x, where $\epsilon(x)$ is a control parameter which measures the intensity of the exploration effort. Given $x^{k-1} \in M$, $\epsilon(x^{k-1}) = \epsilon^k > 0$, $k \in I\!N$, a set $E(x^{k-1}, \epsilon^k)$ around x^{k-1} (it may preferably be a ball in which the function $f(\cdot) + \frac{1}{2\lambda_k} d^2(\cdot, x^{k-1})$ is strongly convex) such that we can find a point x^k where a transition from x^{k-1} to x^k on M is acceptable if the estimated advantages $f(x^{k-1})-f(x^k)$ from x^{k-1} to x^k are higher than some parameter positive $\frac{1}{\lambda_k}$ of the estimated "cost to move" $d^2(x^{k-1},x^k)$. This defines the following worthwhile-to-move relation

$$f(x^{k-1}) - f(x^k) \geq \frac{1}{2\lambda_k} d^2(x^k, x^{k-1}). \tag{6.30}$$

In this context, the agent considers maximizing his net gain $-\frac{\lambda_k}{2}d^2(x,x^{k-1}) - f(x)$(the gain which is attached to x minus the cost to move from x^{k-1} to x) over the exploration $E(x^{k-1},\epsilon^k)$. It is described as follows:

$$x^k \in \text{argmax}_{x\in E(x^{k-1},\epsilon^k)} \left\{ -f(x) - \frac{\lambda_k}{2} d^2(x, x^{k-1}) \right\}.$$

Observe that equation (6.30) can be obtained from the exact version of proximal point algorithm. Indeed, if we assume that the function $x \mapsto f(x) + \frac{\lambda_k}{2} d^2(x, x^{k-1})$ is strongly convex on some ball $B(x^{k-1}, \epsilon^k) \in M$ around x^{k-1} , then

$$x^k \in \text{argmin}_{x\in B(x^{k-1},\epsilon^k)} \left\{ f(x) + \frac{\lambda_k}{2} d^2(x, x^{k-1}) \right\}.$$

It means that, for all $x \in B(x^{k-1}, \epsilon^k)$, we have

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

By taking $x = x^{k-1}$ it holds (6.30). Therefore, the proximal point algorithm can be interpreted as a especial case of a "whorthwhile-to-move problem on dynamical decision model".

7 On the Strongly Convexity of the Subproblems

The computational effort in applying the proximal point method is the solution, in each iteration, of the subproblems

$$\min \left\{ f(x) + \frac{1}{2\lambda_k} d^2(x, x^{k-1}) : x \in M \right\}. \tag{7.31}$$

When f is convex, the regularized function $f_k := f(.)+\frac{1}{2\lambda_k}d^2(., x^{k-1})$ is strongly convex and thus (7.31) can be solvable efficiently. On the other hand, when f is not convex generally the function f_k is not convex and we may obtain that (7.31) may be as hard to solve globally as the original one due to the existence of multiple isolated local minimizer. Considering this problem Papa Quiroz and Oliveira, [36], have been introduced the following iteration $0 \in \partial f_k(.)$ which is a local condition to local minimizer, local maximizer or saddle point. Also, in Section 5 of this paper we present two inexact proximal point method based on the iteration $\epsilon^k \in \lambda_k \partial f(x^k) - \exp^{-1}_{x^k} x^{k-1}$.

A natural question is to known when f_k is strongly or locally strongly convex. In this section we give a sufficient condition on the function f which is related to the concept of locally weakly monotone vector field.

Definition 7.1 *Let $f : M \to I\!R \cup \{+\infty\}$ be a proper lower semicontinuous function. The operator ∂f is called locally weakly monotone if for each $x \in \operatorname{dom}(\partial f)$ there exists ϵ_x and $\rho_x > 0$ such that for all $z, y \in B(x, \epsilon_x)$ we have*

$$\left\langle P_\gamma u - v, \exp_y^{-1} z \right\rangle \geq -\rho_x d(z, y)^2. \tag{7.32}$$

for all $u \in \partial f(z)$, and for all $v \in \partial f(y)$, where P_γ is the parallel transport from z to y. The operator ∂f is called weakly monotone if $\epsilon_x = +\infty$.

Lemma 7.1 *Let $f : M \to I\!R \cup \{+\infty\}$ be a proper lower semicontinuous function,* $\operatorname{dom}(\partial^\circ f) \neq \emptyset$*. Given an arbitrary point $y \in M$, if $\partial^\circ f$ is a (locally) weakly monotone operator with constant $\rho > 0$, and β_k is a positive number satisfying*

$$\beta_k \geq \beta > \rho,$$

then $F(.) := \partial^\circ f(.) + \beta_k \mathrm{grad} \frac{1}{2} d^2(\cdot, y)$ is locally strongly monotone, that is, $f(.) + \beta_k \frac{1}{2} d^2(., y)$ is (locally) strongly convex with constant $\beta - \rho$.

proof. It is immediate. ∎

Lemma 7.2 *Let $f : I\!R^n \to I\!R \cup \{+\infty\}$ be a proper locally Lipschitz function, if $\partial^\circ f$ is (locally) Lipschitz then it is (locally) weakly monotone.*

Proof. It is immediate. ∎

Remark 7.1 *The above result guarantees that if the gradient of f is a Lipschitz function then $f_k := f(.) + \frac{1}{2\lambda_k} d^2(., y)$ is strongly convex for each $y \in M$.*

8 Conclusion and Future Works

In this paper we introduce two inexact proximal point methods to solve the problem (1.1) when the objective function is quasiconvex using an abstract general subdiffefrential approach which includes Clarke-Rockafellar and Frechet Subdifferentials. It is also important to mention that the second inexact version of our algorithm, denoted by HMIP2, was studed by Tang and Huang [43] for monotone vector fields in Hadamard manifolds. Consequently, that work is extended for quasiconvex optimization problems, particularly in euclidean spaces.

We establish that the sequence generated by the algorithm is well-defined and it converges to a critical point of the function when the parameter λ_k satisfies $\bar{\lambda} < \lambda_k$, for some $\bar{\lambda} > 0$.

We apply the results of this paper to introduce a method to solve a worthwhile to change exploration-exploitation model of Attouch and Soubeyran[6] for the quasiconvex case.

It is analyzed some conditions on the function to obtain the strong convexity of the subproblems.

On the other hand, results and properties of convegence rate of the inexact proximal point method for quasiconvex functions on Hadamard manifolds is being done by the authors in [10].

A future work may be, based on the paper of Chong Li and Yao Jen-Ching[14], the extension of the proximal method to the quasiconvex case on arbitrary Riemannian manifolds.

References

[1] Absil, P. A., Mahony, R., and Andrews, B., Convergence of the Iterates of Descent Methods for Analytic Cost Function, SIAM J. Optim., 16, 531-547 (2005)

[2] Ahmadi P. and Khatibsadeh H., On the convergence of inexact proximal point algorithm on Hadamard manifolds, Taiwanese Journal of Mathematics, vol 18(2), 419-433 (2014)

[3] Alexandrov A. D., Ruled surfaces in metric spaces, Vestnik Leningrad Univ 12, 5-26 (1957)

[4] Attouch H.,Bolte J.,Redont P. and Soubeyran A., Proximal Alternating Minimization and Projection Methods for Nonconvex Problems: An Approach Based on the Kurdyka-Lojasiewicz Inequality. Math. Oper. Res. 35(2) 438457 (2010)

[5] Attouch, H. and Bolte, J., On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. Math. Programming. 116, 5-16 (2009)

[6] Attouch, H. and Soubeyran, A., Local search proximal algorithms as decision dynamics with costs to move. Set-Valued and Variational Analysis 19 (1): 157-177. (2011)

[7] Aussel, D., Corvellec, J.N. and Lassonde, M., Mean-value Property and Subdifferential Criteria for Lower Semicontinuous Functions, Trans. of the American Math. Society, 347, 4147-4161 (1995)

[8] Aussel, D., Subdifferential Properties of Quasiconvex and Pseudoconvex Functions: Unified Approach, Journal of Optim. Theory and App., 97(1), 29-45 (1998)

[9] Ballmann W., Gromov M. and Schroeder V., Manifolds of nonpositive curvature, Progress in Mathematics, vol. 61, Birkhäuser Boston Inc., Boston, MA (1985)

[10] Baygorrea N., Papa Quiroz E.A. and Maculan N., On the convergence rate of inexact proximal point algorithm for quasiconvex functions on Hadamard manifolds, Federal University Rio of Janeiro, working paper. (2015)

[11] Bento G.C., Cruz Neto J.X., Finite termination of the proximal point method for convex functions on Hadamard manifolds , Mathematical Programming and Operations Research, 63(9), 1281-1288 (2014)

[12] Bento G.C., Ferreira O.P. and Oliveira P.R., Local convergence of the proximal point method for a special class of nonconvex functions on Hadamard manifolds, Nonlinear Anal. Theory Methods and App., 73(2), 564-572 (2010)

[13] Bento G.C., Ferreira O.P. and Oliveira P.R., Proximal point method for a special class of nonconvex functions on Hadamard manifolds, Optimization, 64(2), 289-319 (2015).

[14] Chong Li and Jen-Chih Yao, Variational Inequalities for Set-Valued and Vector Fields on Riemannian Manifolds: Convexity of the Solution Set and Proximal Point Algorithm, SIAM J. Control Optim, v. 50, n4, 2486 -2514 (2012)

[15] Clarke F.H., Generalized gradients and applications. Transaction of the American Mathematical Society, 205: 247-262 (1975)

[16] Clarke F.H., Optimization and nonsmooth analysis. New York: Wiley (1990)

[17] Correa R., Jofré A. and Thibault, Subdifferential monotonicity as characterization of convex functions, Numer. Funct. Anal. Optim. 15, 531-535 (1994).

[18] J.X da Cruz Neto, L.L. de Lima and P.R. Oliveira, Geodesic algorithms in Riemannian geometry, Balkan Journal of Geometry and its Applications, 3(2) 89 -100, (1998).

[19] Da Cruz Neto, J.X., Oliveira, P.R., Soares P.A. and Soubeyran A., Learning how to Play Nash, Potential Games and Alternating Minimization Method for Structured Nonconvex Problems on Riemannian Manifolds, J. Convex Analysis 20(2), 395–438, 2013.

[20] Da Cruz Neto, J.X., Ferreira, O.P. and Lucâmbio Pérez L.R., Monotone point-to-set vector fields, Balkan J. Geom. Appl. 5, 6979 (2000)

[21] Da Cruz Neto, J.X., Ferreira, O.P. and Lucâmbio Pérez L.R., Contribution to the study of monotone vector fields, Acta Mathematica Hungarica, 94(4), 307-320 (2002)

[22] Do Carmo, M. P., Riemannian Geometry, Bikhausen, Boston (1992)

[23] Eckstein J., Approximate Iterations in Bregman-function-based Proximal Algorithms, Mathematical Programming 83, 113-123 (1998).

[24] Facchinei F. and Kanzow C., Generalized Nash equilibrium problems. 4OR, 5 (3), 173-210, 2007

[25] Ferreira, O. P., and Oliveira, P. R., Proximal Point Algorithm on Riemannian Manifolds, Optimization, 51(2) 257-270 (2002)

[26] Ioffe A. D., Approximate subdifferentials and applications I: The finite dimensional theory, Trans. Amer. Math. Soc. 281, 389-416, (1984).

[27] Kurdyka K., On gradients of functions definable in o-minimal structures. Ann. Inst. Fourier 48(3)769783 (1998)

[28] Li, C., Lopez, G., Martin-Marquez, V., Monotone vector fields and the proximal point algorithm on Hadamard manifolds. J. Lond. Math. Soc. 79(2), 663683 (2009)

[29] Martinet, B., Regularisation d'inéquations variationelles par approximations successives, Revue Franąise Informatique Recherche Opériaionnelle, 4, 154-158 (1972)

[30] Mas-Colell A, Whinston MD, Green JR (1995) Microeconomic theory, Oxford University Press: New York, NY, USA (1995)

[31] Németh S.Z., Monotone vector fields and the proximal point algorithm on Hadamard manifolds. J. Lond. Math. Soc. 79(2), 663683 (2009)

[32] Papa Quiroz, E.A. and Oliveira, P.R.,Proximal Point Methods for Functions Involving Lojasiewicz, Quasiconvex and Convex Properties on Hadamard Manifolds, http://www.optimization-online.org/DB_FILE/2008/06/1996.pdf. (2008)

[33] Papa Quiroz, E.A. and Oliveira, P.R., Proximal Point Methods for Minimizing Quasiconvex Locally Lipschitz Functions on hadamard manifolds, Nonlinnear Analysis. 75, 5924-5932, (2012)

[34] Papa Quiroz, E. A., O. P. and Oliveira, P. R., Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds. J. Convex Anal. 16(1) 49-69 (2009)

[35] Papa Quiroz, E.A. and Oliveira, P.R., Full Convergence of the Proximal Point Method for Quasiconvex Function on Hadamard manifolds, ESAIM: Control, Optimisation and Calculus of Variations, 18, 483-500 (2012)

[36] Papa Quiroz, E.A., Mallma Ramirez L. and Oliveira, P.R., An Inexact Proximal Method for Quasiconvex Minimizations, http://www.optimization-online.org/DB_FILE/2013/08/3982.pdf, acepted for publication in EJOR (2015) .

[37] Rockafellar, R.T., Monotone Operators and the Proximal point Algorithm, SIAM Journal of control and Opt., 877-898 (1976)

[38] Rockafellar R.T., Generalized directional derivatives and subgradients of nonconvex functions, Canad. J. Math., 32, 257-280, (1980)

[39] Rapcsák, T., Smooth Nonlinear Optimization, Kluwer Academic Publishers (1997)

[40] Sakai, T. Riemannian Geometry, American Mathematical Society, Providence, RI. (1996)

[41] Tang G.J., Zhou L.W. and Huang N.J., The proximal point algorithm for pseudomonotone variational inequalities on Hadamard manifolds. Operations Research Letters 7(4),779-790 (2012)

[42] Tang G.J. and Huang N.J., An inexact proximal point algorithm for maximal monotone vector fields on Hadamard manifolds. Operations Research Letters, (41)6, 586-591 (2013)

[43] Tang G.J. and Huang N.J., An inexact proximal point algorithm for maximal monotone vector fields on Hadamard manifolds. Operations Research Letters, (42), 383-387 (2014)

[44] Thibault L., and Zagrodny D., Integration of subdifferential of lower semicontinuous functions on Banach spaces, J. Math. Anal. Appl. 189, 33-58 (1995)

[45] Udriste, C., Convex Function and Optimization Methods on Riemannian Manifolds, Kluwer Academic Publishers, 1994.

[46] Wolkowicz, H., Saigal, R. and Vanderberge, L., (Eds.), Handbook of Semidefinite Programming Theory, Algorithms and Applications, First ed., Internat. Ser. Oper. Management Sci., Springer (2005)