Proximal Point Method for Minimizing Quasiconvex Locally Lipschitz Functions on Hadamard Manifolds

E. A. Papa Quiroz*and P. Roberto Oliveira

Universidad Nacional del Callao
Universidad Nacional Mayor de San Marcos
Lima, Peru
erikpapa@gmail.com

PESC-COPPE
Federal University of Rio de Janeiro
Rio de Janeiro, Brazil
poliveir@cos.ufrj.br

Abstract

In this paper we propose an extension of the proximal point method to solve minimization problems with quasiconvex locally Lipschitz objective functions on Hadamard manifolds. To reach this goal, we use the concept of Clarke subdifferential on Hadamard manifolds and assuming that the function is bounded from below, we prove the global convergence of the sequence generated by the method to a critical point of the function.

Keywords: Proximal point method, quasiconvex functions, locally Lipschitz functions, Hadamard manifolds, global convergence.

*Corresponding author: E.A. Papa Quiroz, telephone: 51(1)5685537//Mobil Phone: 51(1)992332681

1 Introduction

In this paper we introduce an extension of the proximal point method for solving minimization problems with quasiconvex locally Lipschitz objective functions on Hadamard manifolds, that is,

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

where $f : M \to I\!R \cup \{+\infty\}$ is a proper lower semicontinuous quasiconvex locally Lipschitz function and M is a Hadamard manifold (recalling that a Hadamard manifold is a simply connected finite dimensional Riemannian manifold with nonpositive sectional curvature).

The proximal point method in Riemannian manifolds generates a sequence $\{x^k\}$ given by $x^0 \in M$, and

$$x^k \in \texttt{arg}\min\{f(x) + (\lambda_k/2)d^2(x, x^{k-1}) : x \in M\}, \tag{1.2}$$

where λ_k is a certain positive parameter and d is the Riemannian distance in M. Observe that especially when $M = I\!R^n$ we obtain the classical proximal method introduced by Martinet, [23], and further developed by Rockafellar, [33] (in a general framework):

$$x^k \in \texttt{arg}\min\{f(x) + (\lambda_k/2)||x - x^{k-1}||^2 : x \in I\!R^n\},$$

where $||.||$ is the Euclidian norm, i.e, $||x|| = \sqrt{\langle x, x\rangle}$.

It is well known, see Ferreira and Oliveira, [17], that if M is a Hadamard manifold, f is convex in (1.2) and $\{\lambda_k\}$ satisfies

$$\sum_{k=1}^{+\infty}(1/\lambda_k) = +\infty, \tag{1.3}$$

then $\lim_{k\to\infty} f(x^k) = \inf\{f(x) : x \in M\}$. Furthermore, if the optimal set is nonempty, we obtain that $\{x^k\}$ converges to an optimal solution of the problem.

On the other hand, several applications in diverse Science and Engineering areas are sufficient motivation to work with nonconvex objective functions and proximal point methods, see for example [3, 35]. In particular, the class of quasiconvex minimization problems has been receiving special attention from many researchers due to the broad range of applications, for example, in location theory [18] and specially in economic theory [36]. Furthermore, we point out that an important class of nonconvex problems is given by nonconvex quadratic problems (with SDP relaxations as a possible issue), as can be seen in [39], Chapter 13.

Proximal point methods to solve the problem (1.1) for nonconvex objective functions in Euclidean spaces, i.e. $M = I\!R^n$, was studied by some researchers. Tseng, [37], proved a weak convergence result, that is, assuming that f is a lower semicontinous and bounded from below function and $\lambda_k = \lambda > 0$ then every cluster point z is a stationary point of f, i.e.

$$f'(z, d) := \liminf_{\lambda\downarrow 0}\left(\frac{f(z + \lambda d) - f(z)}{\lambda}\right) \geq 0.$$

Kaplan and Tichatschke, [21], studied the method for a class of nonconvex functions when the auxiliary function $f(.) + (\lambda_k/2)||. - x^{k-1}||^2$ becomes strongly convex on certain convex set under a suitable choice of λ_k. The authors proved that the sequence stops in a finite number of iterations at a stationary point or any accumulation point of $\{x^k\}$ is a stationary point of f. This does not mean that the whole sequences converge, a property known as full or global convergence or single limit-point in the literature. Indeed, it is known, see [1], that the sequences of iterates of basic descent methods may fail to converge even when the trajectory is bounded and f is smooth. So a natural question arose: under what minimal condition may the full convergence (convergence of all the sequence) of the proximal point method be proved?

An advance in this direction was given by Attouch and Bolte, [3], where, under the assumptions that f satisfies a Lojasiewicz property and $\{x^k\}$ is bounded, they have proved the convergence of the method to some generalized critical point of the problem. For smooth quasiconvex minimization on the nonnegative orthant, there are some recent works in the literature. Attouch and Teboulle, [4], with a regularized Lotka-Volterra dynamical system, have proved the convergence of the continuous method to a point which belongs to a certain set which contains the set of optimal points; see also Alvarez et al., [2], that treats a general class of dynamical systems that includes the one of Attouch and Teboulle [4], and includes also the case of quasiconvex objective functions in connection with continuous in time models of generalized proximal point algorithms. Cunha et al. [13] and Chen and Pan [10], with a particular $\phi-$divergence distance, have proved the full convergence of the proximal method to the KKT-point of the problem when parameter λ_k is bounded and convergence to an optimal solution when $\lambda_k \to 0$. Pan and Chen [26], with the second-order homogeneous distance, and Souza et al. [35] with a class of separated Bregman distances, have proved the same convergence result of [10] and [13].

The iteration (1.2) on Hadamard manifolds has been previously considered by some researches. Ferreira and Oliveira have proved that if f is convex in (1.2) and $\{\lambda_k\}$ satisfies (1.3), then $\lim_{k\to\infty} f(x^k) = \inf\{f(x) : x \in M\}$. Furthermore, if the optimal set is nonempty, $\{x^k\}$ converges to an optimal solution of the problem. Bento et al. [7, 8] have proved that if f is a special class of nonconvex functions each cluster point of the sequence satisfies the necessary optimality condition. On the other hand, Papa Quiroz and Oliveira in [29] have proved that if f is continuous and quasiconvex in (1.2), $\{\lambda_k\}$ satisfies $\lambda_k \to 0$ and the optimal solution is nonempty then, $\{x^k\}$ converges to an optimal solution of the problem. Observe that in [29, 7, 8] was assumed that (1.2) is solved exactly, which is a strong assumption.

In [30] was observed that the main difficulty in extending the proximal method for nonconvex functions is that due to the nonconvexity of f the subproblems of (1.2) may not be convex and thus, from a practical point of view, we may obtain that minimization subproblems may be as hard to solve globally as the original one due to the existence of multiple isolated local minimizers. To solve this disadvantage the authors have been proposed the following iteration:

$$0 \in \hat{\partial}\left(f(.) + (\lambda_k/2)d^2(., x^{k-1})\right)(x^k) \tag{1.4}$$

where $\hat{\partial}$ is the regular subdifferential on Hadamard manifolds. Assuming that the function is bounded from below, continuous, quasiconvex and the sequence of proximal parameters is bounded, the authors proved the convergence of the sequence to a generalized critical point, that is, a point $\bar{x}$ such that $0 \in \partial f(\bar{x})$, where ∂f is the limmiting subdifferential of f at $\bar{x}$, see Theorem 4.4 of [30].

Unfortunately, although the extended proximal point method was proved to share the favorable convergence results, the limmiting subdifferential may be a set so large that it can contain points that have nothing to do with local, global or sadle points. This is the main motivation for the present work.

On the other hand, the motivation to study the proximal point method on Hadamard manifolds comes from two fields. One of them is that the relative interior of some important constraints in optimization can be seen as Hadamard manifolds, for example:

i) the hypercube $(0,1)^n$ with the metric $X^{-2}(I-X)^{-2} = diag(x_1^{-2}(1-x_1)^{-2}, ..., x_n^{-2}(1-x_n)^{-2})$, see theorems 3.1 and 3.2 of [28];

ii) the positive orthant $I\!R^n_{++}$ with the Dikin metric $X^{-2} = diag(1/x_1^2, ..., 1/x_n^2)$, see Section 4.1, Example 1, of [27];

iii) the set of symmetric positive definite matrices $\mathcal{S}^n_{++}$ with the metric given by the Hessian of the barrier $-\log\det X$; see Corollary 3.1 of [25] and Corollary 5.10 of [32].

iv) the cone $K := \{z = (\tau, x) \in I\!R^{1+n} : \tau > ||x||\}$ endowed with the Hessian of the barrier $-\ln(\tau^2 - ||x||^2)$, see Corollary 3.1 of [25] and Corollary 5.10 of [32].

Thus, we can solve more general optimization problems with nonconvex objective functions, that is, if we can transform those problems into convex or quasiconvex ones on the manifold, we can then use the proposed algorithm. Another one is that the class of Hadamard manifolds is the natural motivation to study more general spaces of nonpositive curvature such as, for example, Hadamard (also called CAT(0)) and Alexandrov spaces. Observe that spaces of nonpositive curvature play a significant role in many areas: Lie group theory, combinatorial and geometric group theory, dynamical system, harmonic maps and vanishing theorems, geometric topology, Kleinian group theory and Theichmüller theory, see the books [6, 9, 16, 20] for details.

In this paper we propose the iteration

$$0 \in \partial^\circ \left(f(.) + (\lambda_k/2) d^2(., x^{k-1}) \right) (x^k) \tag{1.5}$$

where ∂° is the Clarke subdiferential on M, see Section 4. Under the assumption that the function f is proper, lower semicontinuous, quasiconvex, locally Lipschitz and bounded from below we prove that the sequence $\{x^k\}$ given by (1.5) there exists and converges to a critical point of the problem. Of course, in the actual quasiconvex case the best that one can expect from (1.5) is convergence towards a critical point (not necessary a global minimum), this is so because the local nature of (1.4). But when there is no critical point of f in M, the paper ensures the convergence of $\{f(x^k)\}$ to be (possibly not realized) infimum value of f on M (see Remark 4.4). This is interesting in view of the applications, where M is identified with the relative interior of the feasible set. Indeed, take for instance the minimization of a nonconvex function f on the non negative orthant $I\!R^n_+$. The approach consists in taking the interior of the positive orthant as a manifold, namely $M = I\!R^n_{++}$ and endow it with a nonpositive sectional curvature metric such that f becomes quasiconvex in M. But the most interesting case is when the minimum is realized at the boundary ∂M of M in $I\!R^n$, while the constraint $x \geq 0$ is irrelevant for those critical points in $I\!R^n_{++}$. Moreover, in many interesting situations it is not possible to extend the metric to the boundary of M due to some singularities (as the metric induced by the Hessian of the log barrier), thus it is not clear how to prove actual convergence of $\{x^k\}$ in that case. In this context, the convergence in value result is interesting.

The paper is divided as follows: Section 2 gives some results on metric spaces and Riemannian manifolds. In section 3, some results of Clarke subdifferential on Hadamard manifolds are presented, providing some characterization and calculus rules. In Section 4, we extend and analyze the proximal point method with Riemannian distances to solve minimization problems on Hadamard manifolds for proper lower semicontinuos quasiconvex and locally Lipschitz functions. Then, we prove the full convergence of the sequence generated by this method for a critical point of the problem. Finally, in Section 5 some conclusions are provided.

2 Some Basic Facts on Metric Spaces and Riemannian Manifolds

In this section we recall some fundamental properties and notation on Fejér convergence in metric spaces and convex analysis on Riemannian manifolds. Those basic facts can be seen, for example, in do Carmo [15], Sakai [34], Udriste [38] and Rapcsák [31].

Definition 2.1 *Let (X, ρ) be a complete metric space with metric ρ. A sequence $\{z^k\}$ of X is* Fejér convergent *to a set $U \subset X$, if for every $u \in U$ we have*

$$\rho(z^{k+1}, u) \leq \rho(z^k, u).$$

Theorem 2.1 *In a complete metric space (X, ρ), if $\{z^k\}$ is Fejér convergent to a nonempty set $U \subseteq X$, then $\{z^k\}$ is bounded. If, furthermore, a cluster point $\bar{z}$ of $\{z^k\}$ belongs to U, then $\{z^k\}$ converges and $\lim_{k\to+\infty} z^k = \bar{z}$.*

Proof. See, for example, Lemma 6.1 by Ferreira and Oliveira [17]. ■

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 $I\!R^n$. 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 as $\exp_x(v) = \gamma(1)$, where γ is the geodesic such that $\gamma(0) = x$ and $\gamma'(0) = v$. 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 $\mathtt{domf} := \{x \in M : f(x) < +\infty\}$ and its epigraph by $\mathtt{epi}f := \{(x, \beta) \in M \times I\!R : f(x) \le \beta\}$. f is said to be proper if $\mathtt{domf} \neq \emptyset$ and $\forall x \in \mathtt{domf}$ we have $f(x) > -\infty$. f is a lower semicontinuous function if $\mathtt{epi}f$ is a closed subset of $M \times I\!R$.

The gradient of a differentiable function $f : M \to I\!R$, $\mathtt{grad}f$, is a vector field on M defined through $df(X) = \langle \mathtt{grad}f, X\rangle = X(f)$, where X is also a vector field on M.

The complete simply-connected Riemannian manifolds with nonpositive curvature are called *Hadamard manifolds.*

Theorem 2.2 *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 $\exp_x : T_xM \to M$ is a global diffeomorphism.*

Proof. See Sakai, [34], 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 2.3 *Given a geodesic triangle $[x, y, z]$ in a Hadamard manifold, it holds that:*

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

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

Proof. See Sakai, [34], Proposition 4.5, page 223. ∎

Definition 2.2 *Let M be a Hadamard manifold. A subset A is said to be convex in M if given $x, y \in A$, the geodesic curve $\gamma : [0, 1] \to M$ such that $\gamma(0) = x$ and $\gamma(1) = y$ verifies $\gamma(t) \in A$, for all $t \in [0, 1]$.*

Definition 2.3 *Let M be a Hadamard manifold and $f : M \to I\!R \cup \{+\infty\}$ be a proper function. f is called convex if for all $x, y \in M$ and $t \in [0, 1]$, it holds that*

$$f(\gamma(t)) \le tf(y) + (1 - t)f(x),$$

for the geodesic curve $\gamma : [0, 1] \to M$ so that $\gamma(0) = x$ and $\gamma(1) = y$. When the preceding inequality is strict, for $x \neq y$ and $t \in (0, 1)$, the function f is called strictly convex.

Definition 2.4 *Let M be a Hadamard manifold and $f : M \to I\!R \cup \{+\infty\}$ be a proper function. f is called quasiconvex if for all $x, y \in M$, $t \in [0, 1]$, it holds that*

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

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

Theorem 2.4 *Let M be a Hadamard manifold and $f : M \to I\!R \cup \{+\infty\}$ be a proper function. f is quasiconvex if and only if the set $\{x \in M : f(x) \le c\}$ is convex for each $c \in I\!R$.*

Proof. See Udriste, [38], page 98, Theorem 10.2. ∎

Theorem 2.5 *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 $\mathtt{grad}g(x) = -2\exp_x^{-1} y$.*

Proof. See Ferreira and Oliveira, [17], Proposition II.8.3. ∎

3 Clarke Subdifferential on Hadamard Manifolds

In this section we present some results on generalized directional derivatives on Hadamard manifolds which will be used later. For the interested reader in the literature we refer to [7, 24, 19, 5].

Along the paper M will be a Hadamard manifold.

Definition 3.1 *Let $f : M \to I\!R \cup \{+\infty\}$ be proper function. We said that f is locally Lipschitz on M if for each $x \in \mathtt{dom} f$ there exists $\epsilon_x > 0$ such that*

$$|f(z) - f(y)| \leq L_x d(z, y), \forall z, y \in B(x, \epsilon_x),$$

where L_x is some positive number (called the Lipschitz constant of f in a neighborhood of x) and $B(x, \epsilon_x) := \{y \in M : d(x, y) < \epsilon_x\}$.

Remark 3.1 *From the above definition we obtain that if $f : M \to I\!R \cup \{+\infty\}$ is a proper locally Lipschitz function, then $\mathtt{dom} f$ is open in M.*

Definition 3.2 *Let $f : M \to I\!R \cup \{+\infty\}$ be proper locally Lipschitz function. Given $p \in \mathtt{dom} f$ the generalized directional derivative of f at p in the direction $v \in T_pM$, denoted by $f^\circ(x, v)$, is defined as*

$$f^\circ(p, v) := \limsup_{t \downarrow 0 \;\; q \to p} \frac{f\left(\exp_q t (D \exp_p)_{\exp_p^{-1} q} v\right) - f(q)}{t}. \tag{3.6}$$

where we use the (accepted) identification $T_{\exp_p^{-1} q}(T_pM) \approx T_pM$.

It is worth to point out that another definitions has appeared in [5, 24].

Remark 3.2 *It easy to prove that the generalized directional derivative is well defined, see Remark 4.1 in [7].*

Remark 3.3 *Note that, if $M = I\!R^n$ then $\exp_p w = p + w$ and $(D \exp_p)_{\exp_p^{-1} q} v = v$. In this case,* (3.6) *becomes*

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

which is the Clarke's generalized directional derivative, see [11, 12]. Therefore, the generalized differential derivative on Hadamard manifold is a natural extension of the Clarke's generalized differential derivative.

Proposition 3.1 *Let $f : M \to I\!R \cup \{+\infty\}$ be a proper locally Lipschitz function. Then, f° is upper semicontinuous, i.e., if $\{(p^k, v^k)\}$ is a sequence in TM such that $\lim_{k \to +\infty}(p^k, v^k) = (p, v)$, then*

$$\limsup_{k \to +\infty} f^\circ(p^k, v^k) \leq f^\circ(p, v). \tag{3.7}$$

Proof. See Bento et al. [8], Proposition 4.1 ∎

Definition 3.3 *Let $f : M \to I\!R \cup \{+\infty\}$ be a proper locally Lipschitz function. The generalized subdifferential of f at $p \in \mathtt{dom} f$, denoted by $\partial^\circ f(p)$, is defined by*

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

Remark 3.4 *If the proper locally Lipschitz function $f : M \to I\!R \cup \{+\infty\}$ is convex, then $f^\circ(p, v) = f'(p, v)$ (respectively, $\partial^\circ f(p) = \partial_F f(p)$, where $\partial_F f$ is the Fenchel Subdifferential of f) for all $p \in \mathtt{dom}f$*

Definition 3.4 *Let $f : M \to I\!R \cup \{+\infty\}$ be a proper locally Lipschitz functions. A point $p \in M$ is said to be a stationary point of f if $0 \in \partial^\circ f(p)$.*

Theorem 3.1 *If a proper locally Lipschitz function $g : M \to I\!R \cup \{+\infty\}$, has a local minimum at $\bar{x}$ then $0 \in \partial^\circ g(\bar{x})$.*

Proof. Immediate. ■

Proposition 3.2 *Let $f : M \to I\!R \cup \{+\infty\}$ be a proper locally Lipschitz functions at $p \in M$, $\lambda > 0$ and $\tilde{p} \in M$. Then,*

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

Proof. See [7], Proposition 4.2. ■

4 Proximal Point Method on Hadamard Manifolds

Along this section we are interested in solving the problem:

$$(p) \;\; \min\{f(x) : x \in M\}$$

where $f : M \to I\!R \cup \{+\infty\}$ is a proper locally lipschitzian function on a Hadamard manifold M.

The proposed algorithm is as follows:

PPM Algorithm

Initialization:
Let $\{\lambda_k\}$ be a sequence of positive parameters and an initial point

$$x^0 \in M. \tag{4.8}$$

Main Steps:
For $k = 1, 2, 3, ...$
If $0 \in \partial^\circ f(x^{k-1})$ then, stop.
Otherwise, find $x^k \in M$ such that

$$0 \in \partial^\circ \left(f(.) + (\lambda_k/2)d^2(., x^{k-1})\right)(x^k) \tag{4.9}$$

Take $k = k + 1$.

Remark 4.1 *Observe that the proposed method is a natural extension (for nonconvex functions) of the proximal point method in Hadamard manifolds studied by Ferreira and Oliveira, [17]. In fact, as M is a Hadamard manifold, then $d^2(., x^{k-1})$ is strictly convex and by the convexity of f on M and Remark 3.4 then (4.9) becomes*

$$0 \in \partial_F \left(f(.) + (\lambda_k/2)d^2(., x^{k-1})\right)(x^k),$$

which is equivalent to

$$x^k = \mathtt{arg}\min\{f(x) + (\lambda_k/2)d^2(x, x^{k-1}) : x \in M\}.$$

Remark 4.2 *As we are interested in solving the problem (p) when f is nonconvex, it is important to observe that the method (4.8)-(4.9) only needs, in each iteration, to find an stationary point of the regularized function $f(.) + (\lambda_k/2)d^2(., x^{k-1})$. So we believe that the proposed algorithm may be more practical than a previous work with quasiconvex functions.*

Theorem 4.1 *Let M be a Hadamard manifold. If $f : M \to I\!R \cup \{+\infty\}$ is a proper lower semicontinuous and locally Lipschitz function and lower bounded, then the sequence $\{x^k\}$, given by (4.8)-(4.9) exists.*

Proof. We proceed by induction. It holds for $k = 0$, due to (4.8). Assume x^k exists. As f is lower semicontinuous and bounded below then, the function $f(.) + (\lambda_{k+1}/2)d^2(., x^k)$ has compact level sets, and thus this function has a global minimum x^{k+1} and from Theorem 3.1 we have $0 \in \partial^\circ \left(f(.) + (\lambda_{k+1}/2)d^2(., x^k)\right)(x^{k+1})$. ∎

Remark 4.3 *Under the assumptions of the preceding theorem, from (4.9), and Proposition 3.2, we obtain that there exists $g^k \in \partial^\circ f(x^k)$ such that*

$$g^k = \lambda_k \exp^{-1}_{x^k} x^{k-1}.$$

We impose the following assumptions:

Assumption *A***.** M is a Hadamard manifold and $f : M \to I\!R \cup \{+\infty\}$ is a proper locally Lipschitz function bounded from below.

Assumption *B***.** f is lower semicontinuous and quasiconvex.

As we are interested in the asymptotic convergence of the method we also assume that in each iteration $0 \notin \partial^\circ f(x^k)$ which implies that $x^k \neq x^{k-1}$, for all k.

Lemma 4.1 *Let M be a Hadamard manifold and $f : M \to I\!R \cup \{+\infty\}$ be a proper locally Lipschitz and quasiconvex function. If $g \in \partial^\circ f(x)$ and $\langle g, \exp^{-1}_x z\rangle > 0$, then $f(x) < f(z)$.*

Proof. Let $g \in \partial^\circ f(x)$, then from Definition 3.3

$$f^\circ(x, \exp^{-1}_x z) > \alpha > 0,$$

for some positive α. Then, exists $\{y_l\} \subset M$ and $\{t_l\} \subset I\!R_{++}$ such that $d(x, y_l) + t_l < \frac{1}{l}$ satisfying

$$f(y_l) + \alpha < f\left(\exp_{y_l}\left[t_l (D\exp_x)_{\exp^{-1}_x y_l}(\exp^{-1}_x z)\right]\right).$$

From the quasiconvexity of f we have

$$f(y_l) + \alpha < f\left(\exp_{y_l}\left[(D\exp_x)_{\exp^{-1}_x y_l}(\exp^{-1}_x z)\right]\right).$$

Finally, taking $l \to +\infty$ and using the continuity of f we obtain the desired result. ∎

Proposition 4.1 *Under assumptions A and B we have that $\{f(x^k)\}$ is decreasing and converges.*

Proof. As $x^k \neq x^{k-1}$, and from Remark 4.3 we have $\langle g^k, \exp_{x^k}^{-1} x^{k-1}\rangle > 0$. Using the quasiconvexity of f and Lemma 4.1, this implies that $f(x^k) < f(x^{k-1})$. The convergence of $\{f(x^k)\}$ is immediate from the lower boundedness of f. ■

Now, we define the following set

$$U := \{x \in M : f(x) \leq \inf_{j\geq 0} f(x^j)\}$$

Observe that this set depends on the choice of the initial iterates x^0 and sequence $\{\lambda_k\}$. Furthermore, if $U = \emptyset$ then it can be easily proven that $\lim_{k\to+\infty} f(x^k) = \inf_{x\in M} f(x)$, and $\{x^k\}$ is unbounded.
From now on we assume that $U \neq \emptyset$, so from assumptions A and B, it is a nonempty closed and convex set (see Theorem 2.4 for the convexity property).

Theorem 4.2 *Under assumptions A and B and $U \neq \emptyset$, the sequence $\{x^k\}$, generated by the Proximal algorithm, is Fejér convergent to U.*

Proof. Let $x \in U$, then $f(x) \leq f(x^k)$. As $g^k = \lambda_k(\exp_{x^k}^{-1} x^{k-1}) \in \partial^\circ f(x^k)$ (see Remark 4.3) and f is quasiconvex, using Lemma 4.1 we have

$$\langle \exp_{x^k}^{-1} x, \exp_{x^k}^{-1} x^{k-1}\rangle \leq 0. \tag{4.10}$$

On the other hand, for all $x \in U$ from Theorem 2.3, taking $y = x^{k-1}$ and $z = x^k$, we have

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

Now, the last inequality and (4.10) imply, in particular,

$$0 \leq d^2(x^k, x^{k-1}) \leq d^2(x, x^{k-1}) - d^2(x, x^k) \tag{4.11}$$

for every $x \in U$. Thus

$$d(x, x^k) \leq d(x, x^{k-1}). \tag{4.12}$$

This means that $\{x^k\}$ is Fejér convergent to U. ■

Proposition 4.2 *Under the assumptions of the preceding theorem, the following facts are true*

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

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

Proof.
a. From (4.12), $\{d(x, x^k)\}$ is a bounded below nonincreasing sequence and hence convergent.
b. Taking $k \to +\infty$ in (4.11) and using **a**, we obtain the result. ■

Theorem 4.3 *Suppose that assumptions A and B and $U \neq \emptyset$ are satisfied, then the sequence $\{x^k\}$ converges to a point of U.*

Proof. From the previous theorem, $\{x^k\}$ is Fejér convergent to U, and therefore $\{x^k\}$ is bounded (see Theorem 2.1). Then, there exist $\bar{x}$ and a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ converging to $\bar{x}$. From the lower semicontinuity of f we obtain $\liminf_{j\to+\infty} f(x^{k_j}) \geq f(\bar{x})$. As $\{f(x^k)\}$ is a decreasing sequence and converges then $f(\bar{x}) \leq \lim_{k\to+\infty} f(x^k) \leq f(x^k)$, $\forall k \in \mathbb{N}$. This implies that $\bar{x} \in U$. Now, from Theorem 2.1 we conclude that $\{x^k\}$ converges to $\bar{x}$. ■

To obtain strong results for the proximal method we impose some conditions to $\{\lambda^k\}$.

Theorem 4.4 *Suppose that assumptions A and B and $U \neq \emptyset$ are satisfied. If $0 < \lambda_k < \bar{\lambda}$, where $\bar{\lambda}$ is a positive real number, then sequence $\{x^k\}$ converges to critical point of f.*

Proof. The convergence was proved in Theorem 4.3. From Remark 4.3, $\lambda_k \exp_{x^k} x^{k-1} \in \partial^\circ f(x^k)$, so

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

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

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

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 3.1 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 $0 \in \partial f(\bar{x})$. ∎

Remark 4.4 *Observe that along this section we did not assume in M the existence of the optimal solution or critical points of the problem (p). This fact, in particular, is important in the applications of the Riemannian geometry to solve constrained optimization problems. For example, consider the following problem*

$$\min\{f(x) : x \in X\},$$

where $X \subset I\!R^n$ (probably implicit constraints), $f : X \to I\!R$ is a nonconvex function. Assuming that the Hadamard manifold M is modelling the interior of X, i.e $M = int(X)$, f is locally Lipschitz and quasiconvex on the manifold and there are no critical points in M, then we have that $U = \emptyset$ and so

$$\lim_{k \to +\infty} f(x^k) = \inf_{x \in M} f(x) = \inf_{x \in int(X)} f(x).$$

Furthermore, if X is closed and convex in $I\!R^n$, we obtain that $\inf_{x \in int(X)} f(x) = \inf_{x \in X} f(x)$ (see Kiwiel [22], Lemma 1), and thus

$$\lim_{k \to +\infty} f(x^k) = \inf_{x \in X} f(x).$$

That is, if there are no critical points in M, the algorithm gives us enough information to obtain an approach of the infimum value of the constrained problem. In particular, if the problem only has optimal solutions in the boundary ∂X and f is convex as regards the manifold M, then using Remark 3.4 we conclude that $\{f(x^k)\}$ converges to the minimum of the original problem.

5 Conclusion

We have obtained new results on the convergence of the proximal point method for solving minimization problems with proper lower semicontinuous, locally Lipschitz quasiconvex objective functions. For a computational implementation of the proposed method it is needed to solve the iteration (4.9) using a local algorithm, which only provides an approximate solution. Therefore, we consider that in a future work it is important to analyze the convergence of the proposed algorithm considering now an inexact iteration.

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] Alvarez, F., Bolte, J., and Brahic, O., Hessian Riemannian Gradient Flows in Convex Programming, SIAM J. Optim., 43, 2, 477-501, 2004.

[3] Attouch, H., and Bolte, J., On the Convergence of the Proximal Algorithm for Nonsmooth Functions Involving Analytic Features, Mathematical Programming, 116, 1-2, B, 5-16, 2009.

[4] Attouch, H., and Teboulle, M., Regularized Lotka-Volterra Dynamical System as Continuous Proximal-Like Method in Optimization, Journal Optimization Theory and Applications, 121, 541-580, 2004.

[5] D. Azagra, J. Ferrera, F. Lop-Mesas, Nonsmooth analysis and Hamilton-Jacobi equations on Riemannian manifolds, Journal of Functional Analysis 220, 2005, 304-361.

[6] Ballmann, W., Lectures on Spaces of Nonpositive Curvature, Birkhäuser, Basel, 1995.

[7] 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 Analysis, 73, 564-572, 2010.

[8] Bento G.C., Ferreira O.P., and Oliveira P.R., Proximal Point Method for a Special Class of Nonconvex Functions on Hadamard Manifolds, Preprint PESC-COPPE.

[9] Bridson, M.R., and Haefliger, A., Metric Spaces of Non-Positive Curvature, Springer-Verlag, Berlin, 1999.

[10] Chen, J. S., and S. Pan, S. H., Proximal-Like Algorithm for a Class of Nonconvex Programming, Pacific Journal of Optimization, 4, 319-333, 2008.

[11] Clarke, F.H., *Generalized Gradients and Applications.* Transaction of the American Mathematical Society, Vol.205, pp. 247-262, 1975.

[12] Clarke, F.H., *Optimization and Nonsmooth Analysis*, New York:Wiley, 1990.

[13] Cunha, G. F. M., da Cruz Neto, J. X., and Oliveira, P. R., A Proximal Point Algorithm with ϕ-Divergence to Quasiconvex Programming, Optimization, 59, 777-792, 2010.

[14] da Cruz Neto, J. X., Ferreira, O. P., Lucambio Perez, L., and Németh, S. Z., Convex-and Monotone-Transformable Mathematical Programming and a Proximal-Like Point Method, Journal of Global Optimization, 35, 1 53-69, 2006.

[15] do Carmo, M. P., Riemannian Geometry, Bikhausen, Boston, 1992.

[16] Eberlein, P.B., Geometry of Nonpositively Curved Manifolds, University of Chicago Press, Chicago, 1996.

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

[18] Gromicho, J., Quasiconvex Optimization and Location Theory, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1998.

[19] S. Hosseini, M.R. and M.R. Pouryayevali, Generalized gradients and characterization of epi-Lipschitz sets in Riemannian manifolds, Non Linear Analysis 74, 2001, 3884-3895.

[20] Jost, J., Non Positive Curvature: Geometric and Analytic Aspects, Lectures in Mathematics, Base; Boston, Birkhauser, 1997.

[21] Kaplan, A., and Tichatschke, R., Proximal Point Methods and Nonconvex Optimization, Journal of global Optimization, 13, 389-406, 1998.

[22] Kiwiel, K. C., Convergence and Efficiency of Subgradient Methods for Quasiconvex Minimization, Mathematical Programming, A, 90,1-25, 2001.

[23] Martinet, B., Régularisaton, D'inéquations Variationelles par Approximations Successives, Revue Française D'Informatique et de Recherche Operationelle, 154-159, 1970.

[24] D. Motreanu, N.H. Pavel, Quasitangent vectors in flow-invariance and optimization problems on Banach manifolds, J. Math. Anal. Appl. 88 (1982),116-132.

[25] Nesterov, Y. E., and Todd, M. J., On the Riemannian Geometry Defined by Self-Concordant Barrier and Interior-Point Methods, Foundations of Computational Mathematics, 2, 333-361, 2002.

[26] Pan, S. H., and Chen, J. S., Entropy-Like Proximal Algorithms Based on a Second-Order Homogeneus Distances Function for Quasiconvex Programming, Journal of Global Optimization, 39, 555-575, 2007.

[27] Papa Quiroz, E. A., and Oliveira, P. R., New Results on Linear Optimization Through Diagonal Metrics and Riemannian Geometry Tools, Technical Report, ES-645/04, PESC COPPE, Federal University of Rio de Janeiro, 2004.

[28] Papa Quiroz, E. A., and Oliveira, P. R., A New Self-concordant Barrier for the Hypercube, Journal of Optimization Theory and Applications, 135, 475-490, 2007.

[29] Papa Quiroz, E. A., and Oliveira, P. R., Proximal Point Methods for Quasiconvex and Convex Functions with Bregman Distances on Hadamard manifolds, Journal of Convex Analysis, 16, 1, 46-69, 2009.

[30] Papa Quiroz, E. A., and Oliveira, P. R., Full Convergence of the Proximal Point Method for Quasiconvex Functions on Hadamard manifolds, ESAIM: Control, Optimisation and Calculus of Variations, to appear.

[31] Rapcsák, T. Smooth Nonlinear Optimization, Kluwer Academic Publishers, 1997.

[32] Rothaus, O.S., Domains of Positivity, Abh. Math. Sem. Univ. Hamburg, 24, 189-235, 1960.

[33] Rockafellar, R.T., Monotone Operations and the Proximal Point Method, SIAM J. Control Optimization, 14, 877-898, 1976.

[34] Sakai, T. Riemannian Geometry, American Mathematical Society, Providence, RI, 1996.

[35] Souza, S. S., Oliveira, P. R., da Cruz Neto, J. X., and Soubeyran, A., A Proximal Method with Separable Bregman Distance for Quasiconvex Minimization on the Nonnegative Orthant, European Journal of Operational Research, 201, 365-376, 2010.

[36] Takayama A., Mathematical Economics, 2*nd* Edition, Cambrigde Univ. Press, 1995.

[37] Tseng, P., Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization, Journal of Optimization Theory and Applications, 109, 3, 475-494, 2001.

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

[39] 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.