

## FULL CONVERGENCE OF THE PROXIMAL POINT METHOD FOR QUASICONVEX FUNCTIONS ON HADAMARD MANIFOLDS

ERIK A. PAPA QUIROZ<sup>1</sup> AND P. ROBERTO OLIVEIRA<sup>2</sup>

**Abstract.** In this paper we propose an extension of the proximal point method to solve minimization problems with quasiconvex objective functions on Hadamard manifolds. To reach this goal, we initially extend the concepts of regular and generalized subgradient from Euclidean spaces to Hadamard manifolds and prove that, in the convex case, these concepts coincide with the classical one. For the minimization problem, assuming that the function is bounded from below, in the quasiconvex and lower semicontinuous case, we prove the convergence of the iterations given by the method. Furthermore, under the assumptions that the sequence of proximal parameters is bounded and the function is continuous, we obtain the convergence to a generalized critical point. In particular, our work extends the applications of the proximal point methods for solving constrained minimization problems with non-convex objective functions in Euclidean spaces when the objective function is convex or quasiconvex on the manifold.

**Mathematics Subject Classification.** 90C26.

Received May 7, 2009. Revised January 29, 2010.  
Published online June 22, 2011.

### 1. INTRODUCTION

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

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

where  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper quasiconvex 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

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

---

*Keywords and phrases.* Proximal point method, quasiconvex function, Hadamard manifolds, full convergence.

<sup>1</sup> Universidad Nacional del Callao, Universidad Nacional Mayor de San Marcos, Lima, Peru. [erikpapa@gmail.com](mailto:erikpapa@gmail.com)

<sup>2</sup> PESC-COPPE Federal University of Rio de Janeiro, Rio de Janeiro, Brazil. [poliveir@cos.ufrj.br](mailto:poliveir@cos.ufrj.br)

where  $\lambda_k$  is a certain positive parameter and  $d$  is the Riemannian distance in  $M$ . Observe that especially when  $M = \mathbb{R}^n$  we obtain the classical proximal method introduced by Martinet [21], and further developed by Rockafellar [29] (in a general framework):

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

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

It is well known, see Ferreira and Oliveira [16], 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  $\lim_{k \rightarrow \infty} f(x^k) = \inf\{f(x) : x \in M\}$ . 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 [3,4,32]. 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 [17], control theory [7] and specially in economic theory [33]. For the interested reader in the literature on convex optimization problems, we refer to [8,9]. 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 [36], Chapter 13.

Proximal point methods to solve the problem (1.1) for nonconvex objective functions in Euclidean spaces, *i.e.*  $M = \mathbb{R}^n$ , was studied by some researchers. Tseng [34] proved a weak convergence result that is, assuming that  $f$  is a lower semicontinuous 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 [19] studied the method for a class of nonconvex functions when the auxiliary function  $f(\cdot) + (\lambda_k/2) \|\cdot - x^{k-1}\|^2$  becomes strongly convex on certain convex set under a suitable choice of  $\lambda_k$ . 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 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 [5], 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 [5], and includes also the case of quasiconvex objective functions in connection with continuous in time models of generalized proximal point algorithms. Cunha *et al.* [12] and Chen and Pan [11], with a particular  $\phi$ -divergence distance, have proved the full convergence of the proximal method to the KKT-point of the problem when parameter  $\lambda_k$  is bounded and convergence to an optimal solution when  $\lambda_k \rightarrow 0$ . Pan and Chen [23], with the second-order homogeneous distance, and Souza *et al.* [32] with a class of separated Bregman distances, have proved the same convergence result of [11,12].

The iteration (1.2) has been previously considered by Ferreira and Oliveira in [16] and Papa Quiroz and Oliveira in [26] for convex and quasiconvex functions, respectively. Ferreira and Oliveira have proved that if  $f$  is convex in (1.2) and  $\{\lambda_k\}$  satisfies (1.3), then  $\lim_{k \rightarrow \infty} f(x^k) = \inf\{f(x) : x \in M\}$ . Furthermore, if the optimal

set is nonempty,  $\{x^k\}$  converges to an optimal solution of the problem. On the other hand, Papa Quiroz and Oliveira in [26] have proved that if  $f$  is continuous and quasiconvex in (1.2),  $\{\lambda_k\}$  satisfies  $\lambda_k \rightarrow 0$  and the optimal solution is nonempty then,  $\{x^k\}$  converges to an optimal solution of the problem. Observe that in [26] we assume that (1.2) is solved exactly, which is a strong assumption that will be relaxed in the current paper.

In this paper we are interested in extending the global convergence of the proximal point method to minimize quasiconvex functions on Hadamard manifolds. The motivation to study this subject 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  $X^{-2}(I - X)^{-2} = \text{diag}(x_1^{-2}(1 - x_1)^{-2}, \dots, x_n^{-2}(1 - x_n)^{-2})$ , see Theorems 3.1 and 3.2 of [25];
- (ii) the positive orthant  $\mathbb{R}_{++}^n$  with the Dikin metric  $X^{-2} = \text{diag}(1/x_1^2, \dots, 1/x_n^2)$ , see Section 4.1, Example 1, of [24];
- (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 [22] and Corollary 5.10 of [28];
- (iv) the cone  $K := \{z = (\tau, x) \in \mathbb{R}^{1+n} : \tau > \|x\|\}$  endowed with the Hessian of the barrier  $-\ln(\tau^2 - \|x\|^2)$ , see Corollary 3.1 of [22] and Corollary 5.10 of [28].

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 [6,10,15,18] for details.

The main difficulty we observed in extending the proximal method for nonconvex function 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 we propose the following iteration:

$$0 \in \widehat{\partial} (f(\cdot) + (\lambda_k/2)d^2(\cdot, x^{k-1})) (x^k) \tag{1.4}$$

where  $\widehat{\partial}$  is the regular subdifferential on Hadamard manifolds (see Sect. 3). Of course both (1.2) and (1.4) are equivalent when  $f$  is convex in  $M$ , but in the quasiconvex case these iterative schemes are quite different in nature. Convex problems can be addressed by conventional local algorithms since all critical points are global minimizers. The regularized subproblem being strongly convex when the objective function is convex, we expect the local algorithms to perform efficient enough to find a good approximate solution at reasonable execution time. Suppose now that  $f$  is quasiconvex, if the regularized subproblem turned to be quasiconvex, similar considerations as in the convex case would hold true: the only practical difficulty for global minimization occurring in the uncommon event that we must go throughout nonzero measure “plateaux” of critical points. But the class of quasiconvex functions is not closed by addition. Thus the augmented auxiliary function to be minimized in (1.2) may be nonquasiconvex and indeed multiple isolated local minimizers may occur. Therefore in our opinion the local stationary iteration (1.4) makes much more sense than the previously considered (1.2) for dealing with nonconvex problems. In this sense the paper improves, for nonconvex objective functions, the work of Ferreira and Oliveira [16].

Under some natural assumption on the function  $f$  we prove that the sequence  $\{x^k\}$  given by (1.4) there exists and converges to a generalized critical point of the problem. Of course, in the actual quasiconvex case the best that one can expect from (1.4) 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 Rem. 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  $\mathbb{R}_+^n$ . The approach consists in taking the interior of the positive orthant as the manifold, namely  $M = \mathbb{R}_{++}^n$  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  $\partial M$  of  $M$  in  $\mathbb{R}^n$ , while the constraint  $x \geq 0$  is irrelevant for those critical points in  $\mathbb{R}_{++}^n$ . 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 regular and generalized subdifferential on Hadamard manifolds are presented, providing some characterization and calculus rules. This section is not particularly difficult but should be interesting on its own for readers specialized on nonsmooth analysis. In Section 4, we extend and analyze the proximal point method with Riemannian distances to solve minimization problems on Hadamard manifolds for quasiconvex functions. Then, assuming that the objective function is bounded from below and continuous, we prove the full convergence of this method for a certain generalized 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 [14], Sakai [31], Udriste [35] and Rapcsák [27].

**Definition 2.1.** Let  $(X, \rho)$  be a complete metric space with metric  $\rho$ . 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  $(X, \rho)$ , 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  $\lim_{k \rightarrow +\infty} z^k = \bar{z}$ .

*Proof.* See, for example, Lemma 6.1 by Ferreira and Oliveira [16]. □

Let  $M$  be a differential manifold with finite dimension  $n$ . We denote by  $T_x M$  the tangent space of  $M$  at  $x$  and  $TM = \bigcup_{x \in M} T_x M$ .  $T_x M$  is a linear space and has the same dimension of  $M$ . Because we restrict ourselves to real manifolds,  $T_x M$  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  $u, v \in T_x M$  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  $v \in T_x M$  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  $\|\cdot\| = \|\cdot\|_x$ . The metrics can be used to define the length of a piecewise smooth curve  $\alpha : [t_0, t_1] \rightarrow M$  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  $\nabla$  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  $\alpha : [t_0, t_1] \rightarrow M$ , another vector field is obtained, denoted by  $DV/dt$ . The parallel transport along  $\alpha$  from  $\alpha(t_0)$  to  $\alpha(t_1)$ , denoted by  $P_{\alpha, t_0, t_1}$ , is an application  $P_{\alpha, t_0, t_1} : T_{\alpha(t_0)} M \rightarrow T_{\alpha(t_1)} M$  defined by  $P_{\alpha, t_0, t_1}(v) = V(t_1)$  where  $V$  is the unique vector field along  $\alpha$  so that  $DV/dt = 0$  and  $V(t_0) = v$ . Since  $\nabla$  is a Riemannian connection,  $P_{\alpha, t_0, t_1}$  is a linear isometry, furthermore  $P_{\alpha, t_0, t_1}^{-1} = P_{\alpha, t_1, t_0}$  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  $\gamma : I \rightarrow M$  is called a geodesic if  $D\gamma'/dt = 0$ .

A Riemannian manifold is complete if its geodesics are defined for any value of  $t \in \mathbb{R}$ . Let  $x \in M$ , the exponential map  $\exp_x : T_x M \rightarrow M$  is defined as  $\exp_x(v) = \gamma(1)$ , where  $\gamma$  is the geodesic such that  $\gamma(0) = x$  and  $\gamma'(0) = v$ . If  $M$  is complete, then  $\exp_x$  is defined for all  $v \in T_x M$ . 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  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  we denote its domain by  $\text{dom} f := \{x \in M : f(x) < +\infty\}$  and its epigraph by  $\text{epi} f := \{(x, \beta) \in M \times \mathbb{R} : f(x) \leq \beta\}$ .  $f$  is said to be proper if  $\text{dom} f \neq \emptyset$  and  $\forall x \in \text{dom} f$  we have  $f(x) > -\infty$ .  $f$  is a lower semicontinuous function if  $\text{epi} f$  is a closed subset of  $M \times \mathbb{R}$ .

The gradient of a differentiable function  $f : M \rightarrow \mathbb{R}$ ,  $\text{grad} f$ , is a vector field on  $M$  defined through  $df(X) = \langle \text{grad} f, X \rangle = X(f)$ , 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  $\mathbb{R}^n$ ,  $n = \dim M$ . More precisely, at any point  $x \in M$ , the exponential mapping  $\exp_x : T_x M \rightarrow M$  is a global diffeomorphism.*

*Proof.* See Sakai [31], Theorem 4.1, p. 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:*

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

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

*Proof.* See Sakai [31], Proposition 4.5, p. 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  $\gamma : [0, 1] \rightarrow M$  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  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper function.  $f$  is called convex if for all  $x, y \in M$  and  $t \in [0, 1]$ , it holds that

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

for the geodesic curve  $\gamma : [0, 1] \rightarrow M$  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  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper function.  $f$  is called quasiconvex if for all  $x, y \in M$ ,  $t \in [0, 1]$ , it holds that

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

for the geodesic  $\gamma : [0, 1] \rightarrow M$ , so that  $\gamma(0) = x$  and  $\gamma(1) = y$ .

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

*Proof.* See Udriste [35], p. 98, Theorem 10.2. □

**Definition 2.5.** Let  $M$  be a Hadamard manifold and  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper and differentiable function on the open convex set  $\text{dom}f$  of  $M$ .  $f$  is called pseudoconvex if, for every pair of distinct points  $x, y \in \text{dom}f$  we have

$$\langle \text{grad}f(x), \gamma'(0) \rangle \geq 0, \text{ then } f(y) \geq f(x),$$

for the geodesic  $\gamma$  joining  $x$  to  $y$ .

**Theorem 2.5.** Let  $M$  be a Hadamard manifold and  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper and differentiable pseudoconvex function on the open convex set  $\text{dom}f$  of  $M$ . If  $\text{grad}f(x^*) = 0$ , then  $x^*$  is a global minimum of  $f$ .

*Proof.* Immediate. □

**Theorem 2.6.** 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  $\text{grad}g(x) = -2 \exp_x^{-1} y$ .

*Proof.* See Ferreira and Oliveira [16], Proposition II.8.3. □

### 3. REGULAR AND GENERAL SUBGRADIENTS ON HADAMARD MANIFOLDS

In this section, we extend some definitions and results of regular and general subgradient from Euclidean spaces to Hadamard manifolds. The developed theory will be useful to define the proximal algorithm as well as in the convergence proofs. Our results are motivated from Rockafellar and Wets [30].

Along this section  $M$  will be a Hadamard manifold.

**Definition 3.1.** Let  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper function. Given  $x \in \text{dom}f$ , we say that  $s \in T_x M$  is a regular subgradient of  $f$  at  $x$  if the following is satisfied

$$\liminf_{y \neq x, y \rightarrow x} \frac{1}{d(x, y)} [f(y) - f(x) - \langle s, \gamma'(0) \rangle] \geq 0,$$

for the geodesic curve joining  $x$  to  $y$  ( $\gamma(0) = x$  and  $\gamma(1) = y$ ).

Observe that the above inequality is equivalent to

$$f(y) \geq f(x) + \langle s, \gamma'(0) \rangle + o(d(x, y)),$$

for the geodesic  $\gamma$  joining  $x$  to  $y$ , where  $\lim_{x \neq y, d(x, y) \rightarrow 0} \frac{o(d(x, y))}{d(x, y)} = 0$ .

The set of regular subgradient of  $f$  at  $x \in \text{dom}f$ , denoted by  $\widehat{\partial}f(x)$ , is called Fréchet subdifferential. If  $x \notin \text{dom}f$  then we define  $\widehat{\partial}f(x) = \emptyset$ .

The concept of Fréchet subdifferential is inadequate for the calculus covering some of the properties we need, so we introduce the following:

**Definition 3.2.** Let  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper function. Given a point  $x \in \text{dom}f$ , we say that  $s \in T_x M$  is a generalized subgradient of  $f$  at  $x$  if there exist  $\{x^l\}, \{s^l\}$ , with  $s^l \in \widehat{\partial}f(x^l)$ , satisfying  $x^l \rightarrow x, f(x^l) \rightarrow f(x)$ , and  $P_{\gamma_l, 0, 1} s^l \rightarrow s$ , for the geodesic  $\gamma_l$  joining  $x^l$  to  $x$ , where  $P_{\gamma_l, 0, 1}$  is the parallel transport along  $\gamma_l$ .

The limiting subdifferential of  $f$  at  $x \in M$ , denoted by  $\partial f(x)$ , is defined as the set of all the generalized subgradient of  $f$  at  $x$ , that is,

$$\partial f(x) := \left\{ s \in T_x M : \exists x^l \rightarrow x, f(x^l) \rightarrow f(x), \exists s^l \in \widehat{\partial}f(x^l) : P_{\gamma_l, 0, 1} s^l \rightarrow s \right\}.$$

**Proposition 3.1.** The following properties are true:

- a.  $\widehat{\partial}f(x) \subset \partial f(x)$ , for all  $x \in M$ ;
- b. If  $f$  is differentiable at  $\bar{x}$  then  $\widehat{\partial}f(\bar{x}) = \{\text{grad}f(\bar{x})\}$ , so  $\text{grad}f(\bar{x}) \in \partial f(\bar{x})$ ;

- c. If  $f$  is continuously differentiable in a neighborhood of  $\bar{x}$ , then  $\widehat{\partial}f(\bar{x}) = \partial f(\bar{x}) = \{\mathbf{grad}f(\bar{x})\}$ ;
- d. If  $g = f + h$ , with  $f$  finite at  $\bar{x}$  and  $h$  continuously differentiable on a neighborhood of  $\bar{x}$  then  $\widehat{\partial}g(\bar{x}) = \widehat{\partial}f(\bar{x}) + \mathbf{grad}h(\bar{x})$ , and  $\partial g(\bar{x}) = \partial f(\bar{x}) + \mathbf{grad}h(\bar{x})$ .

*Proof.*

- a. It is immediate from Definitions 3.1 and 3.2.
- b. The first part is a consequence of the differentiability of  $f$  and Definition 3.1. The second part is a direct consequence from Definitions 3.1 and 3.2.
- c. Let  $s \in \partial f(\bar{x})$ , then

$$\exists x^l \rightarrow \bar{x}, f(x^l) \rightarrow f(\bar{x}), \exists s^l \in \widehat{\partial}f(x^l) : P_{\gamma_l, 0, 1} s^l \rightarrow s,$$

for the geodesic  $\gamma_l$  joining  $x^l$  to  $\bar{x}$ , where  $P_{\gamma_l, 0, 1}$  is the parallel transport along the geodesic  $\gamma_l$ . From item b, we have that  $s^l = \mathbf{grad}f(x^l)$ , for  $l$  sufficiently large. From the continuity of  $\mathbf{grad}f$  and parallel transport  $P$ , and using the uniqueness of the limit point we have that  $s = \mathbf{grad}f(\bar{x})$ .

- d. The inclusion  $\subset$  is immediate. We get the inclusion  $\supset$  by applying this rule to the representation  $f = g + (-h)$ . □

In order to work with minimization problems we need the following definition.

**Definition 3.3.** Let  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper function. A point  $x \in \text{dom}f$  is said to be a generalized critical point of  $f$  if  $0 \in \partial f(x)$ .

**Theorem 3.1.** If a proper function  $g : M \rightarrow \mathbb{R} \cup \{+\infty\}$  has a local minimum at  $\bar{x}$  then  $0 \in \widehat{\partial}g(\bar{x})$  and therefore,  $0 \in \partial g(\bar{x})$ .

*Proof.* Immediate. □

We will prove that in the convex case the concepts of limiting and Fréchet subdifferential coincide with the classical one. Some definitions and properties are needed to achieve that aim.

**Definition 3.4.** Let  $C$  be a subset of  $M$ . Given  $\bar{x} \in C$ , a vector  $v \in T_{\bar{x}}M$  is called normal to  $C$  at  $\bar{x}$ , in the regular sense, denoted by  $v \in \widehat{N}_C(\bar{x})$ , if

$$\langle v, \gamma'(0) \rangle \leq o(d(\bar{x}, x)), \tag{3.5}$$

$\forall x \in C$  and for the geodesic  $\gamma$  joining  $\bar{x}$  to  $x$ , where  $o(d(\bar{x}, x))$  denotes a term with the property that

$$\lim_{x \neq \bar{x}, d(x, \bar{x}) \rightarrow 0} \frac{o(d(\bar{x}, x))}{d(\bar{x}, x)} = 0.$$

A vector  $v$  is normal to  $C$  at  $\bar{x}$  in a generalized sense or simply a normal vector, denoted by  $v \in N_C(\bar{x})$ , if there exist sequences  $\{x^l\} \subset C$  and  $\{v^l\} \subset \widehat{N}_C(x^l)$  so that  $x^l \rightarrow \bar{x}$  and for the geodesic  $\gamma_l$  such that  $\gamma_l(0) = x^l$  and  $\gamma_l(1) = \bar{x}$ ,  $P_{\gamma_l, 0, 1} v^l \rightarrow v$ , where  $P_{\gamma_l, 0, 1}$  is the parallel transport along the geodesic  $\gamma_l$ .

Observe that inequality (3.5) is equivalent to

$$\limsup_{x \in C, x \neq \bar{x}, d(x, \bar{x}) \rightarrow 0} \frac{\langle v, \gamma'(0) \rangle}{d(x, \bar{x})} \leq 0, \tag{3.6}$$

for the geodesic  $\gamma$  joining  $\bar{x}$  to  $x$ .

**Definition 3.5.** Let  $C$  be a subset of  $M$ . Given  $\bar{x} \in C$ , a vector  $w \in T_{\bar{x}}M$  is tangential to  $C$  at  $\bar{x}$ , denoted by  $w \in T_C(\bar{x})$ , if there are sequences  $\{x^l\} \subset C$ ,  $x^l \neq \bar{x}$ , and  $\{\tau^l\} \subset \mathbb{R}$  with  $x^l \rightarrow \bar{x}$  and  $\tau^l \rightarrow 0$  with  $\tau^l > 0$  so that for the geodesic  $\gamma_l$  joining  $\bar{x}$  to  $x^l$  we have

$$\lim_{l \rightarrow +\infty} \frac{\gamma'_l(0)}{\tau^l} = w.$$

From now on  $\tau^l \searrow 0$  will denote that  $\tau^l > 0$  and  $\tau^l \rightarrow 0$ , as  $l \rightarrow +\infty$ .

**Proposition 3.2.** Let  $C \subset M$ . For each  $\bar{x}$ , the sets  $\widehat{N}_C(\bar{x})$  and  $T_C(\bar{x})$  are closed cones. Furthermore,  $\widehat{N}_C(\bar{x})$  is convex in  $T_{\bar{x}}M$  and characterized by

$$v \in \widehat{N}_C(\bar{x}) \text{ if and only if } \langle v, w \rangle \leq 0, \forall w \in T_C(\bar{x}).$$

*Proof.* The properties whereby both sets are closed cones as well as  $\widehat{N}_C(\bar{x})$  is convex are obtained by using elementary analysis. We prove the characterization property. Let  $w \in T_C(\bar{x})$ , arbitrary. From Definition 3.5 there are sequences  $\{x^l\} \subset C$ ,  $x^l \neq \bar{x}$ , and  $\{\tau^l\} \subset \mathbb{R}$  with  $x^l \rightarrow \bar{x}$  and  $\tau^l \searrow 0$  such that for the geodesic  $\gamma_l$  joining  $\bar{x}$  to  $x^l$  we have  $\lim_{l \rightarrow +\infty} \frac{\gamma'_l(0)}{\tau^l} = w$ . Defining  $w^l = \frac{\gamma'_l(0)}{\tau^l}$  and using  $v \in \widehat{N}_C(\bar{x})$  we have  $\langle v, w^l \rangle \leq \frac{o(\|\tau^l w^l\|)}{\tau^l}$ . Taking  $l \rightarrow +\infty$  we obtain that  $\langle v, w \rangle \leq 0$ .

Reciprocally, suppose that  $\langle v, w \rangle \leq 0, \forall w \in T_C(\bar{x})$  and  $v \notin \widehat{N}_C(\bar{x})$ . From (3.6), for any  $\delta > 0$

$$\sup_{x \in B(\bar{x}, \delta), x \neq \bar{x}} \left( \frac{\langle v, \gamma'(0) \rangle}{d(x, \bar{x})} \right) \geq m > m_1 > 0,$$

where  $\gamma$  is the geodesic joining  $\bar{x}$  to  $x$ ,  $m := \limsup_{x \neq \bar{x}, d(x, \bar{x}) \rightarrow 0} \frac{\langle v, \gamma'(0) \rangle}{d(x, \bar{x})}$  and  $m_1$  is some positive number between 0 and  $m$ . Take  $\delta = 1/l$  and using the supreme property, there exists  $x^l \in B(\bar{x}, 1/l)$  with  $x^l \neq \bar{x}$  so that

$$\left\langle v, \frac{\gamma'_l(0)}{d(x^l, \bar{x})} \right\rangle \geq m_1,$$

for the geodesic  $\gamma_l$  joining  $\bar{x}$  to  $x^l$ . Defining  $w^l := \frac{\gamma'_l(0)}{d(x^l, \bar{x})}$  and taking  $\liminf$  we obtain

$$\liminf_{l \rightarrow +\infty} \langle v, w^l \rangle \geq m_1.$$

As  $\{w^l\}$  is bounded then there exist subsequences, without loss of generality, also denoted by  $\{x^l\}$  and  $\{w^l\}$ , so that  $x^l \in B(\bar{x}, 1/l) \cap C$  and  $w^l \rightarrow w$ , for some  $w$ . Defining  $\tau^l := d(x^l, \bar{x})$  and from the definition of  $w^l$ , we have that  $w \in T_C(\bar{x})$  and  $\langle v, w \rangle > 0$ . This is a contradiction with the hypothesis, and therefore the statement of the proposition is true.  $\square$

**Theorem 3.2.** Let  $C$  be a subset of  $M$  and  $\bar{x} \in C$ . If  $C$  is convex then

$$T_C(\bar{x}) = cl\{w : \text{there exists } \lambda > 0 \text{ with } \exp_{\bar{x}}(\lambda w) \in C\},$$

$$N_C(\bar{x}) = \widehat{N}_C(\bar{x}) = Z,$$

where

$$Z = \{v : \langle v, \gamma'(0) \rangle \leq 0, \forall x \in C, \text{ where } \gamma \text{ is the geodesic such that } \gamma(0) = \bar{x} \text{ and } \gamma(1) = x\}.$$

*Proof.* Define

$$K(\bar{x}) := \{w \in T_{\bar{x}}M : \text{there exists } \lambda > 0 \text{ with } \exp_{\bar{x}}(\lambda w) \in C\}.$$



As  $C$  is convex, it includes the geodesics joining the point  $\bar{x}$  and any of their points (see Def. 2.2), so the vectors of  $K(\bar{x})$  are the multiple by a positive scalar of the vectors  $\gamma'(0)$ , where  $\gamma$  is the geodesic curve joining  $\bar{x}$  and some  $x \in C$ . We prove that

$$K(\bar{x}) \subset T_C(\bar{x}) \subset clK(\bar{x}).$$

The first inclusion is immediate (give  $w \in K(\bar{x})$ , then there exists  $\lambda > 0$  such that  $\exp_{\bar{x}}(\lambda w) \in C$ , define  $x^l = \exp_{\bar{x}}(\tau^l w)$  with  $\tau^l \in (0, \lambda]$  and  $\tau^l \rightarrow 0$ ), so we prove the second inclusion. Let  $w \in T_C(\bar{x})$ , then from Definition 3.5, there exist sequences  $x^l \rightarrow \bar{x}$ ,  $x^l \in C$ ,  $x^l \neq \bar{x}$  and  $\tau^l \searrow 0$  such that  $\lim_{l \rightarrow +\infty} \frac{\gamma_l'(0)}{\tau^l} = w$ , where  $\gamma_l$  is the geodesic joining  $x^l$  and  $\bar{x}$ . Define  $w^l := \frac{\gamma_l'(0)}{\tau^l}$ , then from the convexity of  $C$ , there exists  $\lambda > 0$  so that  $\exp_{x^l}(\lambda w^l) \in C$ . Besides, as  $w^l \rightarrow w$ , we get  $w \in clK(\bar{x})$ . Now, as  $T_C(\bar{x})$  is closed (see Prop. 3.2) we have  $clK(\bar{x}) = T_C(\bar{x})$ , and therefore the first result is satisfied.

Now, we prove the second statement of the theorem. As, by definition,

$$Z \subset \widehat{N}_C(\bar{x}) \subset N_C(\bar{x}),$$

then it is sufficient to prove that

$$N_C(\bar{x}) \subset Z.$$

Let  $v \in N_C(\bar{x})$  then, there exist  $x^l \rightarrow \bar{x}$  with  $x^l \in C$  and  $P_{\gamma_l, 0, 1} v^l \rightarrow v$ , with  $v^l \in \widehat{N}_C(x^l)$ , where  $P_{\gamma_l, 0, 1}$  is the parallel transport along the geodesic  $\gamma_l$  such that  $\gamma_l(0) = x^l$  and  $\gamma_l(1) = \bar{x}$ .

As  $v^l \in \widehat{N}_C(x^l)$ , then from Proposition 3.2, it holds the characterization property,

$$\langle v^l, w \rangle \leq 0, \forall w \in T_C(x^l) = clK(x^l) \tag{3.7}$$

where  $K(x^l) := \{w : \text{there exists } \lambda > 0 \text{ with } \exp_{x^l}(\lambda w) \in C\}$ , the above equality being a consequence of the previous result. Let  $x \in C$ , arbitrary, and let  $\gamma$  be the geodesic such that  $\gamma(0) = \bar{x}$  and  $\gamma(1) = x$ . Defining  $w = P_{\gamma, 1, 0}(\gamma'(0))$ , we have that  $w \in K(x^l)$ , so using this fact in (3.7) we obtain that  $\langle v^l, P_{\gamma_l, 1, 0}(\gamma'(0)) \rangle \leq 0$ . Now, taking parallel transport and  $l \rightarrow +\infty$  we obtain that  $\langle v, \gamma'(0) \rangle \leq 0$ . Thus, the result is obtained.  $\square$

We define the difference quotient function as a function  $\Delta_\tau f(x) : T_x M \rightarrow \mathbb{R} \cup \{+\infty\}$  where

$$\Delta_\tau f(x)(w) = \frac{f(\exp_x(\tau w)) - f(x)}{\tau}, \text{ for } \tau > 0.$$

**Definition 3.6.** For a function  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  and a point  $\bar{x} \in M$  with  $f(\bar{x})$  finite, the subderivative function  $df(\bar{x}) : T_{\bar{x}} M \rightarrow \mathbb{R} \cup \{+\infty\}$  is defined as

$$df(\bar{x})(w) := \liminf_{\tau \searrow 0, w' \rightarrow w} \Delta_\tau f(x)(w').$$

**Lemma 3.1.**

$$\widehat{\partial}f(\bar{x}) = \{v \in T_{\bar{x}} M : \langle v, w \rangle \leq df(\bar{x})(w), \forall w \in T_{\bar{x}} M\}.$$

*Proof.* We prove the inclusion  $\subset$ . Let  $v \in \widehat{\partial}f(\bar{x})$  and  $w \in T_{\bar{x}} M$ . Define  $y = \exp_{\bar{x}}(\tau w')$  with  $w' \neq 0$  and  $\tau > 0$ . From Definition 3.1 and observing that  $d(\bar{x}, y) = \|\tau w'\|$  gives

$$\Delta_\tau f(\bar{x})(w') - \frac{o(\|\tau w'\|)}{\tau} \geq \langle v, w' \rangle.$$

Taking  $\liminf$  when  $\tau \searrow 0$  and  $w' \rightarrow w$  we obtain  $df(\bar{x})(w) \geq \langle v, w \rangle$ .

Now, we get the inclusion  $\supset$ . Let  $v \in T_{\bar{x}} M$  so that  $\langle v, w \rangle \leq df(\bar{x})(w), \forall w \in T_{\bar{x}} M$  and suppose that  $v \notin \widehat{\partial}f(\bar{x})$ , then from definition

$$S := \liminf_{y \neq \bar{x}, y \rightarrow \bar{x}} \frac{1}{d(\bar{x}, y)} [f(y) - f(\bar{x}) - \langle v, \gamma'(0) \rangle] < 0$$

where  $\gamma$  is the geodesic joining  $\bar{x}$  to  $y$ . From supreme and infimum properties, given  $\delta = 1/l$  there exists  $y^l \in B(\bar{x}, 1/l)$  so that

$$\frac{1}{d(\bar{x}, y^l)} [f(y^l) - f(\bar{x}) - \langle v, \gamma'_l(0) \rangle] \leq S,$$

where  $\gamma_l$  is a geodesic joining  $\bar{x}$  and  $y^l$ . Define  $w^l := \gamma'_l(0)/\tau^l$  where  $\tau^l := d(\bar{x}, y^l)$ , then gives

$$\frac{f(\exp_{\bar{x}}(\tau^l w^l)) - f(\bar{x})}{\tau^l} - \langle v, w^l \rangle \leq S.$$

As  $\{w^l\}$  is a bounded sequence then there is a point  $\bar{w} \in T_{\bar{x}}M$  and a subsequence, also denoted by  $\{w^l\}$ , so that  $w^l \rightarrow \bar{w}$  and as  $\tau^l \searrow 0$  we have

$$df(\bar{x})(\bar{w}) \leq S + \langle v, \bar{w} \rangle < \langle v, \bar{w} \rangle,$$

which is a contradiction. Therefore the proof is concluded. □

**Proposition 3.3.**  $v \in \widehat{\partial}f(\bar{x})$  if, and only if,  $\langle (v, -1), (w, \beta) \rangle \leq 0, \forall (w, \beta) \in \mathbf{epi}df(\bar{x})$ .

*Proof.* Let  $(w, \beta) \in \mathbf{epi}df(\bar{x})$ , then  $df(\bar{x})(w) \leq \beta$ . Using the previous lemma we obtain  $\langle v, w \rangle - \beta \leq 0$ . This implies that  $\langle (v, -1), (w, \beta) \rangle \leq 0, \forall (w, \beta) \in \mathbf{epi}df(\bar{x})$ .

Reciprocally, suppose that  $\langle (v, -1), (w, \beta) \rangle \leq 0, \forall (w, \beta) \in \mathbf{epi}df(\bar{x})$ . Let  $w \in T_{\bar{x}}M$  arbitrary, then  $(w, df(\bar{x})(w)) \in \mathbf{epi}df(\bar{x})$ . Using the above inequality we have  $\langle v, w \rangle \leq df(\bar{x})(w)$ . Finally from the previous lemma, we obtain that  $v \in \widehat{\partial}f(\bar{x})$ . □

**Remark 3.1.** Observe that

$$(w, \lambda) \in \mathbf{epi}\Delta_{\tau}f(\bar{x}) \text{ if and only if } (\exp_{\bar{x}}(\tau w), \lambda\tau + f(\bar{x})) \in \mathbf{epi}f.$$

**Proposition 3.4.**

$$\mathbf{epi}df(\bar{x}) = T_{\mathbf{epi}f(\bar{x}, f(\bar{x}))}.$$

*Proof.* We first prove the inclusion  $\subset$ . Let  $(w, \beta) \in \mathbf{epi}df(\bar{x})$ , then  $\forall \delta > 0$  and  $\gamma > 0$

$$\inf_{\tau < \gamma, w' \in B(w, \delta)} \Delta_{\tau}f(\bar{x})(w') \leq \beta.$$

Taking  $\delta = \gamma = 1/l$ , from the infimum property, there exists  $\tau^l < 1/l$  and  $w^l \in B(w, 1/l)$  so that

$$\Delta_{\tau^l}f(\bar{x})(w^l) < \beta + \frac{1}{l}.$$

From Remark 3.1 this implies that  $(\exp_{\bar{x}}(\tau^l w^l), (\beta + 1/l)\tau^l + f(\bar{x})) \in \mathbf{epi}f$ . Defining  $z^l = \exp_{\bar{x}}(\tau^l w^l)$  and  $\alpha^l = (\beta + 1/l)\tau^l + f(\bar{x})$  we obtain that  $(z^l, \alpha^l) \in \mathbf{epi}f$  with  $(z^l, \alpha^l) \rightarrow (\bar{x}, f(\bar{x})), \tau^l \searrow 0$ . Let  $\Psi_l$  be the geodesic joining  $(\bar{x}, f(\bar{x}))$  to  $(z^l, \alpha^l)$ , then

$$\lim_{l \rightarrow +\infty} \frac{\Psi'_l(0)}{\tau^l} = \lim_{l \rightarrow +\infty} \left( \frac{\gamma'_l(0)}{\tau^l}, \frac{\alpha^l - f(\bar{x})}{\tau^l} \right) = \lim_{l \rightarrow +\infty} (w^l, \beta + 1/l) = (w, \beta),$$

where  $\gamma_l$  is the geodesic joining  $\bar{x}$  and  $z^l$ . From Definition 3.5 the result is obtained.

Now, we prove the inclusion  $\supset$ . Let  $(w, \beta) \in T_{\mathbf{epi}f(\bar{x}, f(\bar{x}))}$  then there exist  $(z^l, \alpha_l) \in \mathbf{epi}f$  such that  $(z^l, \alpha_l) \rightarrow (\bar{x}, f(\bar{x})), \tau^l \searrow 0$  and for the geodesic  $\Psi$  joining  $(\bar{x}, f(\bar{x}))$  and  $(z^l, \alpha^l)$  we have

$$\lim_{l \rightarrow +\infty} \frac{\Psi'_l(0)}{\tau^l} = \lim_{l \rightarrow +\infty} \left( \frac{\gamma'_l(0)}{\tau^l}, \frac{\alpha^l - f(\bar{x})}{\tau^l} \right) = (w, \beta),$$

where  $\gamma_l$  is a geodesic joining  $\bar{x}$  and  $z^l$ . From the above equality we obtain that  $w^l := \frac{\gamma_l'(0)}{\tau^l} \rightarrow w$  and  $\beta^l := \frac{\alpha_l - f(\bar{x})}{\tau^l} \rightarrow \beta$ . Now, as  $(z^l, \alpha_l) \in \mathbf{epi}f$  then

$$\Delta_{\tau^l} f(\bar{x})(w^l) \leq \frac{\alpha_l - f(\bar{x})}{\tau^l}.$$

Taking  $\liminf$  when  $l \rightarrow +\infty$  we obtain  $\liminf_{l \rightarrow +\infty} \Delta_{\tau^l} f(\bar{x})(w^l) \leq \beta$ . So,  $df(\bar{x})(w) \leq \beta$ , and therefore  $(w, \beta) \in \mathbf{epid}f(\bar{x})$ .  $\square$

**Theorem 3.3.**

$$\widehat{\partial}f(\bar{x}) = \left\{ v \in T_{\bar{x}}M : (v, -1) \in \widehat{N}_{\mathbf{epi}f}(\bar{x}, f(\bar{x})) \right\}.$$

*Proof.* It is immediate from Propositions 3.3 and 3.4 and using  $C = \mathbf{epi}f$  in Proposition 3.2.  $\square$

**Theorem 3.4.**

$$\partial f(\bar{x}) \subset \{v \in T_{\bar{x}}M : (v, -1) \in N_{\mathbf{epi}f}(\bar{x}, f(\bar{x}))\}.$$

*Proof.* Let  $v \in \partial f(\bar{x})$ , then there exists  $\{x^l\}$  such that  $x^l \rightarrow \bar{x}$ ,  $f(x^l) \rightarrow f(\bar{x})$ ,  $v^l \in \widehat{\partial}f(x^l) : P_{\gamma_l, 0, 1} v^l \rightarrow v$  where  $P_{\gamma_l, 0, 1}$  is the parallel transport along the geodesic  $\gamma_l$  such that  $\gamma_l(0) = x^l$  and  $\gamma_l(1) = \bar{x}$ . As  $v^l \in \widehat{\partial}f(x^l)$  then, from the previous theorem,  $(v^l, -1) \in \widehat{N}_{\mathbf{epi}f}(x^l, f(x^l))$ . Now, define the following sequences:  $z^l = (x^l, f(x^l))$  and  $p^l = (v^l, -1)$ . Applying this definition, we obtain that there exists  $\{z^l\}$  such that  $z^l \in \mathbf{epi}f$  and  $p^l \in \widehat{N}_{\mathbf{epi}f}(x^l, f(x^l))$  so that  $z^l \rightarrow (\bar{x}, f(\bar{x}))$  and  $(P_{\gamma_l, 0, 1} p^l, -1) \rightarrow (v, -1)$ . Thus, from Definition 3.4 we conclude that  $(v, -1) \in N_{\mathbf{epi}f}(\bar{x}, f(\bar{x}))$ .  $\square$

Finally, we show that if  $f$  is convex then Fréchet and limiting subdifferential are the same set; besides, they coincide with the classical subdifferential in convex analysis. Let

$$\partial_F f(\bar{x}) := \{v \in T_{\bar{x}}M : f(x) \geq f(\bar{x}) + \langle v, \gamma'(0) \rangle, \text{ for the geodesic } \gamma \text{ joining } \bar{x} \text{ to } x\}.$$

**Theorem 3.5.** *If  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  is a convex function, then for each  $\bar{x} \in M$  we have*

$$\partial f(\bar{x}) = \widehat{\partial}f(\bar{x}) = \partial_F f(\bar{x}).$$

*Proof.* The implication  $\partial_F f(\bar{x}) \subset \widehat{\partial}f(\bar{x}) \subset \partial f(\bar{x})$  is trivial from their definitions. Then, it is sufficient to prove that  $\partial f(\bar{x}) \subset \partial_F f(\bar{x})$ . Let  $v \in \partial f(\bar{x})$ , then from Theorem 3.4,  $(v, -1) \in N_{\mathbf{epi}f}(\bar{x}, f(\bar{x}))$ . As  $\mathbf{epi}f$  is convex, because  $f$  is convex, we have from Theorem 3.2 that

$$\langle v, \gamma'(0) \rangle - (f(x) - f(\bar{x})) \leq 0,$$

for the geodesic  $\gamma$  joining  $\bar{x}$  to  $x$ . Taking  $r = f(x)$  we obtain that  $f(x) \geq f(\bar{x}) + \langle v, \gamma'(0) \rangle$ , and therefore,  $v \in \partial_F f(\bar{x})$ .  $\square$

#### 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  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper 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  $k = 1, 2, 3, \dots$

If  $0 \in \widehat{\partial}f(x^{k-1})$  stop.

Otherwise, find  $x^k \in M$  such that

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

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 [16]. In fact, as  $M$  is a Hadamard manifold, then  $d^2(\cdot, x^{k-1})$  is strictly convex and by the convexity of  $f$  and Theorem 3.5 then (4.9) becomes

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

which, from the convexity of  $f$ , is equivalent to

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

**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  $f(\cdot) + (\lambda_k/2)d^2(\cdot, x^{k-1})$ . So we believe that our algorithm is more practical than previous works in proximal methods with quasiconvex functions, see Cunha *et al.* [12], Chen and Pan [11], Souza *et al.* [32] and Papa Quiroz and Oliveira [26].

**Theorem 4.1.** *Let  $M$  be a Hadamard manifold. If  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper lower bounded and lower semicontinuous function on  $\text{dom}f$ , 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 on  $\text{dom}f$  then, the function  $f(\cdot) + (\lambda_{k+1}/2)d^2(\cdot, x^k)$  has compact level sets, and thus this function has a global minimum  $x^{k+1}$  and from Theorem 3.1 we have  $0 \in \widehat{\partial}(f(\cdot) + (\lambda_{k+1}/2)d^2(\cdot, x^k))(x^{k+1})$ .  $\square$

**Remark 4.3.** Under the assumptions of the preceding theorem and from (4.9), the smoothness of  $(\lambda_k/2)d^2(\cdot, x^{k-1})$ , Proposition 3.1, **d**, and Theorem 2.6, we have that there exists  $g^k \in \widehat{\partial}f(x^k)$  such that

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

We impose the following assumptions:

**Assumption A.**  $M$  is a Hadamard manifold and  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper function bounded from below on  $\text{dom}f$ .

**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  $0 \notin \widehat{\partial}f(x^k)$  which implies that  $x^k \neq x^{k-1}$ , for all  $k$ .

**Lemma 4.1.** *Let  $M$  be a Hadamard manifold and  $f : M \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower semicontinuous and quasiconvex function. If  $g \in \widehat{\partial}f(x)$  and  $f(y) \leq f(x)$  then  $\langle g, \exp_x^{-1} y \rangle \leq 0$ .*

*Proof.* Let  $t \in (0, 1)$ , then from the quasiconvexity of  $f$  and the assumption that  $f(y) \leq f(x)$  we have  $f(\gamma(t)) \leq \max\{f(x), f(y)\} = f(x)$ , where  $\gamma$  is the geodesic joining  $x$  to  $y$ . As  $g \in \widehat{\partial}f(x)$  we obtain

$$f(\gamma(t)) \geq f(x) + t \langle g, \exp_x^{-1} \gamma(t) \rangle + o(t),$$

where  $\lim_{t \rightarrow 0} \frac{o(t)}{t} = 0$ . From both inequalities we conclude that  $t \langle g, \exp_x^{-1} y \rangle + o(t) \leq 0$ . Dividing by  $t$  and taking  $t \rightarrow 0$  we obtain the desired result.  $\square$

**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  $f$ .  $\square$

Now, we define the following set

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

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  $\lim_{k \rightarrow +\infty} f(x^k) = \inf_{x \in M} f(x)$ , 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 Thm. 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  $g^k = \lambda_k(\exp_{x^k}^{-1} x^{k-1}) \in \widehat{\partial}f(x^k)$  (see Rem. 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  $U$ .  $\square$

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

*Proof.*

- a. From (4.12),  $\{d(x, x^k)\}$  is a bounded below nonincreasing sequence and hence convergent.
- b. Taking  $k \rightarrow +\infty$  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 Thm. 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  $\liminf_{j \rightarrow +\infty} f(x^{k_j}) \geq f(\bar{x})$ . As  $\{f(x^k)\}$  is a decreasing sequence and converges then  $f(\bar{x}) \leq \lim_{k \rightarrow +\infty} f(x^k) \leq f(x^k), \forall k \in \mathbb{N}$ . 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 a point of  $U$  and  $\lim_{k \rightarrow \infty} g^k = \lim_{k \rightarrow \infty} \lambda_k \exp_{x^k}^{-1} x^{k-1} = 0$ . Furthermore, if  $f$  is continuous on  $\text{dom}f$ , then it converges to a generalized critical point of  $f$ .*

*Proof.* The convergence was proved in Theorem 4.3. From Remark 4.3, using the fact that the parallel transport is an isometry and the boundedness of  $\lambda_k$  we have that

$$\|P_{\gamma_k, 0, 1} g^k\| = \|g^k\| \leq \bar{\lambda} \|\exp_{x^k}^{-1} x^{k-1}\| = \bar{\lambda} d(x^k, x^{k-1}),$$

where  $\gamma_k$  is the geodesic curve such that  $\gamma_k(0) = x^k$  and  $\gamma_k(1) = \bar{x}$ . Using Proposition 4.2, b, we get  $\lim_{k \rightarrow \infty} g^k = 0$ . Now, let  $f$  continuous, then there are sequences  $\{x^k\}$ ,  $\{f(x^k)\}$  and  $\{g^k\}$  with  $g^k \in \widehat{\partial}f(x^k)$  such that  $\lim_{k \rightarrow +\infty} x^k = \bar{x}$ ,  $\lim_{k \rightarrow \infty} f(x^k) = f(\bar{x})$  (from the continuity of  $f$ ) and  $\lim_{k \rightarrow \infty} P_{\gamma_{k_j}, 0, 1} g^{k_j} = 0$ . From Definition 3.2 it follows that  $0 \in \partial f(\bar{x})$ . □

As immediate particular cases of the above theorem we obtain the following results:

**Corollary 4.1.** *If the optimal set of the problem (p) is nonempty and  $f$  convex on  $M$ , then  $\{x^k\}$  converges to an optimal solution.*

**Corollary 4.2.** *If the optimal set of the problem (p) is nonempty,  $f$  differentiable and pseudoconvex on  $M$ , then  $\{x^k\}$  converges to an optimal solution of the problem (p).*

**Remark 4.4.** Observe that along this section we did not assume in  $M$  the existence of the optimal solution or generalized 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  $X \subset \mathbb{R}^n$  (probably implicit constraints),  $f : X \rightarrow \mathbb{R}$  is a nonconvex function. Assuming that the Hadamard manifold  $M$  is modelling the interior of  $X$ , i.e  $M = \text{int}(X)$ ,  $f$  is continuous quasiconvex on the manifold and there are no generalized critical points in  $M$ , then we have that  $U = \emptyset$  and so

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

Furthermore, if  $X$  is closed and convex in  $\mathbb{R}^n$ , we obtain that  $\inf_{x \in \text{int}(X)} f(x) = \inf_{x \in X} f(x)$  (see Kiwiel [20], Lem. 1), and thus

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

That is, if there are no generalized 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  $\partial X$  and  $f$  is convex as regards the manifold  $M$ , then using Theorem 3.5 we conclude that  $\{f(x^k)\}$  converges to the minimum of the original problem.

To finish this section we give some examples of proximal point methods on Hadamard manifolds. Example 4.1 shows proximal point methods for unconstrained minimization problems and Examples 4.2 to 4.5 show proximal point methods for constrained ones. Those algorithms are particularly useful when the objective function is not convex in the usual sense but becomes convex or quasiconvex on the manifold. The metric of Example 4.1 is motivated of a particular investigation of convexification of nonconvex functions and the metrics of Examples 4.2 to 4.5 are motivated from Hessian of self-concordant barriers in interior point methods, see Nesterov and Todd [22], Section 6.4, and Papa Quiroz and Oliveira [25], Section 3, for details. It should also be observed that Examples 4.2, 4.4 and 4.5 were already considered in Papa Quiroz and Oliveira [26].

**Example 4.1.** Consider the unconstrained minimization problem in  $\mathbb{R}^n$

$$\min\{f(x) : x \in \mathbb{R}^n\}.$$

$\mathbb{R}^n$  with the metric

$$G(x) = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & 0 & 1 + e^{2x_{n-1}} & -e^{x_{n-1}} \\ 0 & \dots & 0 & 0 & 0 & -e^{x_{n-1}} & 1 \end{bmatrix}$$

is isometric to the Euclidean space  $\mathbb{R}^n$  through the application  $\Phi : (\mathbb{R}^n, I) \rightarrow (\mathbb{R}^n, G(x))$ , defined by

$$\Phi(x_1, x_2, \dots, x_n) = (x_1, x_2, \dots, x_{n-1}, e^{x_{n-1}} - x_n).$$

$(\mathbb{R}^n, G(x))$  is a Hadamard manifold with null sectional curvature and the Riemannian distance between  $x = (x_1, \dots, x_n)$  and  $y = (y_1, y_2, \dots, y_n)$  is given by

$$d(x, y) = \left[ \sum_{i=1}^{n-1} (x_i - y_i)^2 + [(y_n - x_n) - (e^{y_{n-1}} - e^{x_{n-1}})]^2 \right]^{1/2}.$$

In fact, as  $\Phi$  is an isometry, then

$$d(x, y)^2 = \|\Phi^{-1}(x) - \Phi^{-1}(y)\|^2,$$

and therefore,

$$\begin{aligned} d(x, y)^2 &= \|(x_1 - y_1, x_2 - y_2, \dots, x_{n-1} - y_{n-1}, (y_n - x_n) - (e^{y_{n-1}} - e^{x_{n-1}}))\|^2 \\ &= \sum_{i=1}^{n-1} (x_i - y_i)^2 + [(y_n - x_n) - (e^{y_{n-1}} - e^{x_{n-1}})]^2. \end{aligned}$$

So, the proximal point method is given by  $x^0 \in \mathbb{R}^n$  and

$$0 \in \widehat{\partial} \left( f(x) + (\lambda_k/2) \left\{ \sum_{i=1}^{n-1} (x_i - x_i^{k-1})^2 + [(x_n - x_n^{k-1}) - (e^{x_{n-1}} - e^{x_{n-1}^{k-1}})]^2 \right\} \right) (x^k).$$

The metric  $G(x)$  was motivated by the function  $g(x_1, x_2) = x_1^2 + (e^{x_1} - x_2)^2$ , which is not convex in  $(\mathbb{R}^2, I)$  but it becomes convex on the manifold  $(\mathbb{R}^2, G_2(x))$ , where

$$G_2(x) = \begin{pmatrix} 1 + e^{2x_1} & -e^{x_1} \\ -e^{x_1} & 1 \end{pmatrix}.$$

Indeed,  $F(x_1, x_2) = (x_1, e^{x_1} - x_2)$  is a diffeomorphism between  $(\mathbb{R}^2, I)$  and  $(\mathbb{R}^2, G(x))$  and it is well known that the function  $f(x_1, x_2) = x_1^2 + x_2^2$  is convex in  $(\mathbb{R}^2, I)$ . Now, as  $g = f \circ F^{-1}$ , and using Theorem 4.5 we obtain that  $g$  is convex in  $(\mathbb{R}^2, G(x))$ .

**Example 4.2.** Consider the semidefinite optimization problem

$$\min\{f(X) : X \succeq 0\},$$

where the notation  $X \succeq 0$  means that  $X$  is a  $n \times n$  symmetric positive semidefinite matrix. Take  $\mathcal{S}_{++}^n$ , the set of  $n \times n$  symmetric positive definite matrices, as a smooth manifold and define the metric given by the Hessian of the barrier  $b(X) = -\log \det X$ .  $(\mathcal{S}_{++}^n, b''(X))$ . It is a Hadamard manifold with nonpositive sectional curvature, see Corollary 3.1 of Nesterov and Todd [22] and Corollary 5.10 of Rothaus [28], and the proximal method generates a sequence  $\{X^k\}$  given by  $X^0 \in \mathcal{S}_{++}^n$  and

$$0 \in \widehat{\partial} \left( f(X) + (\lambda_k/2) \sum_{i=1}^n \ln^2 \lambda_i \left( X^{-\frac{1}{2}} X^{k-1} X^{-\frac{1}{2}} \right) \right) (X^k),$$

where  $\lambda(Z)$  denotes the eigenvalue of the symmetric matrix  $Z$ .

**Example 4.3.** Consider the second order cone programming

$$\min\{f(x) : x \succeq_K 0\},$$

where  $x \succeq_K 0$  means that  $x \in K := \{(\tau, z) \in \mathbb{R}^{1+n} : \tau \geq \|z\|_2\}$ . If we endow  $K^0$ , the interior of  $K$ , with the Hessian of the barrier  $-\ln(\tau^2 - \|z\|_2^2)$ , we obtain a Hadamard manifold with a non-positive sectional curvature, see Corollary 3.1 of Nesterov and Todd [22] and Corollary 5.10 of Rothaus [28]. The proximal method generates a sequence  $\{x^k\}$  given by  $x^0 \in K^0$  and

$$0 \in \widehat{\partial} (f(x) + (\lambda_k/2)[\ln^2 \sigma(x, x^{k-1}) + \ln^2 \sigma(x^{k-1}, x)]) (x^k),$$

where  $\sigma(z_0, z_1) := \max\{\lambda : z_0 - \lambda z_1 \in K\}$ .

**Example 4.4.** Consider the problem on the nonnegative orthant

$$\min\{f(x) : x \geq 0\},$$

where  $x \geq 0$  means that  $x_i \geq 0, \forall i = 1, \dots, n$ . Take  $\mathbb{R}_{++}^n$ , the positive orthant, as a smooth manifold.  $(\mathbb{R}_{++}^n, X^{-2})$ , where  $X^{-2} = \text{diag}(1/x_1^2, 1/x_2^2, \dots, 1/x_n^2)$  is the Hessian of the  $-\log$  barrier, is a Hadamard manifold with null



sectional curvature, see Section 4.1, Example 1, of [24], and the proximal point method generates the sequence  $\{x^k\}$  given by  $x^0 \in \mathbb{R}_{++}^n$  and

$$0 \in \widehat{\partial} \left( f(x) + (\lambda_k/2) \sum_{i=1}^n \left( \ln \frac{x_i^{k-1}}{x_i} \right)^2 \right) (x^k).$$

**Example 4.5.** Consider the problem on the hypercube

$$\min\{f(x) : 0 \leq x \leq \bar{e}\},$$

where  $x \in \mathbb{R}^n$  and  $\bar{e} = (1, 1, \dots, 1) \in \mathbb{R}^n$ .  $((0, 1)^n, X^{-2}(I - X)^{-2})$ , where  $X^{-2}(I - X)^{-2}$  is the Hessian of the barrier  $\sum_{i=1}^n (2x_i - 1)[\ln x_i - \ln(1 - x_i)]$ , is a Hadamard manifold with null sectional curvature, see Papa Quiroz and Oliveira [25], Section 3. The proximal point method generates a sequence  $\{x^k\}$  given by  $x^0 \in (0, 1)^n$  and

$$0 \in \widehat{\partial} \left( f(x) + (\lambda_k/2) \sum_{i=1}^n \left( \ln \left( \frac{x_i^{k-1}}{1 - x_i^{k-1}} \right) - \ln \left( \frac{x_i}{1 - x_i} \right) \right)^2 \right) (x^k).$$

**Remark 4.5** (quasiconvexification of nonconvex functions). Convexification of nonconvex functions using some convenient metric was studied by some researchers, among whom Udriste [35], Rapcsák [27] and da Cruz Neto *et al.* [13]. A natural and evident question that arises in this paper is the following: is there a practical way to transform a general nonconvex problem into a quasiconvex one? This question is an open problem, but a contribution in this direction is given by the following result:

**Theorem 4.5.** *Let  $f : N_1 \rightarrow \mathbb{R}$  be a convex (quasiconvex) function in a complete Riemannian manifold  $(N_1, G_1)$  and  $F$  an isometry between  $(N_1, G_1)$  and  $(N_2, G_2)$ , then  $g = f \circ F^{-1}$  is convex (quasiconvex) in  $F(N_1)$ .*

*Proof.* See [35], Theorems 3.1 and 10.9. □

Based on the above result we obtain a technique to convexification or quasiconvexification of nonconvex function: Let  $(M_1, G_1)$  and  $(M_2, G_2)$  be two Riemannian manifolds such that  $M_2 \subset M_1$  and suppose that a function  $g$  is not quasiconvex in the totally convex set  $M_2$  (recalling that  $M_2$  is totally convex if given  $x, y \in M_2$ , every geodesic curve  $\gamma : [0, 1] \rightarrow M$  such that  $\gamma(0) = x$  and  $\gamma(1) = y$ , verifies  $\gamma(t) \in A$ , for all  $t \in [0, 1]$ ) with respect to the metric  $G_1$ . If there exists an isometry  $F : (M_1, G_1) \rightarrow (M_2, G_2)$  and a convex (quasiconvex) function  $f$  in  $(M_1, G_1)$  such that

$$g = f \circ F^{-1},$$

then  $g$  is convex (quasiconvex) in  $(M_2, G_2)$ .

## 5. CONCLUSION

We have obtained new results on the proximal point method for solving minimization problems with quasiconvex objective functions. Observe that for the proximal method the nonpositive sectional curvature is used to assure the smoothness of  $d^2(\cdot, y)$ , to obtain an explicit expression of  $\text{grad}^2$  and also to prove the boundedness of the proximal iterations. For that reason, an extension of the method for arbitrary Riemannian manifolds is still an open question.

*Acknowledgements.* We are grateful to the anonymous referees that suggested several important inclusions and changing on the paper.

## REFERENCES

- [1] P.A. Absil, R. Mahony and B. Andrews, Convergence of the iterates of descent methods for analytic cost function. *SIAM J. Optim.* **16** (2005) 531–547.
- [2] F. Alvarez, J. Bolte and O. Brahic, Hessian Riemannian gradient flows in convex programming. *SIAM J. Optim.* **43** (2004) 477–501.
- [3] H. Attouch and J. Bolte, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features. *Math. Program. B* **116** (2009) 5–16.
- [4] H. Attouch and A. Soubeyran, “Worthwhile-to-move” behaviors as temporary satisficing without too many sacrificing processes. Preprint [arXiv:0905.1238](https://arxiv.org/abs/0905.1238) (2009).
- [5] H. Attouch and M. Teboulle, Regularized Lotka-Volterra dynamical system as continuous proximal-like method in optimization. *J. Optim. Theory Appl.* **121** (2004) 541–580.
- [6] W. Ballmann, *Lectures on Spaces of Nonpositive Curvature*. Birkhäuser, Basel (1995).
- [7] N. Barron and W. Liu, Calculus of variation  $l^\infty$ . *Appl. Math. Opt.* **35** (1997) 237–263.
- [8] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis and Engineering Applications, MPS/SIAM Series on Optimization* **2**. SIAM (2001).
- [9] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge University Press, Cambridge (2004).
- [10] M.R. Bridson and A. Haefliger, *Metric Spaces of Non-Positive Curvature*. Springer-Verlag, Berlin (1999).
- [11] J.S. Chen and S.S.H. Pan, Proximal-like algorithm for a class of nonconvex programming. *Pacific Journal of Optimization* **4** (2008) 319–333.
- [12] G.F.M. Cunha, J.X. da Cruz Neto and P.R. Oliveira, A proximal point algorithm with  $\phi$ -divergence to quasiconvex programming. *Optimization* **59** (2010) 777–792.
- [13] J.X. da Cruz Neto, O.P. Ferreira, L. Lucambio Perez and S.Z. Németh, Convex-and monotone-transformable mathematical programming and a proximal-like point method. *J. Glob. Optim.* **35** (2006) 53–69.
- [14] M.P. do Carmo, *Riemannian Geometry*. Birkhäuser, Boston (1992).
- [15] P.B. Eberlein, *Geometry of Nonpositively Curved Manifolds*. University of Chicago Press, Chicago (1996).
- [16] O.P. Ferreira and P.R. Oliveira, Proximal point algorithm on Riemannian manifolds. *Optimization* **51** (2002) 257–270.
- [17] J. Gromicho, *Quasiconvex Optimization and Location Theory*. Kluwer Academic Publishers, Dordrecht (1998).
- [18] J. Jost, *Non Positive Curvature: Geometric and Analytic Aspects*. Lectures in Mathematics, Base; Boston, Birkhäuser (1997).
- [19] A. Kaplan and R. Tichatschke, Proximal point methods and nonconvex optimization. *J. Glob. Optim.* **13** (1998) 389–406.
- [20] K.C. Kiwiel, Convergence and efficiency of subgradient methods for quasiconvex minimization. *Math. Program. A* **90** (2001) 1–25.
- [21] B. Martinet, Brève communication. Régularisation d’inéquations variationnelles par approximations successives. *Revue Française D’Informatique et de Recherche Opérationnelle* **4** (1970) 154–158.
- [22] Y.E. Nesterov and M.J. Todd, On the Riemannian geometry defined by self-concordant barrier and interior-point methods. *Found. Comput. Math.* **2** (2002) 333–361.
- [23] S.H. Pan and J.S. Chen, Entropy-like proximal algorithms based on a second-order homogeneous distances function for quasiconvex programming. *J. Glob. Optim.* **39** (2007) 555–575.
- [24] E.A. Papa Quiroz and P.R. Oliveira, *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).
- [25] E.A. Papa Quiroz and P.R. Oliveira, A new self-concordant barrier for the hypercube. *J. Optim. Theory Appl.* **135** (2007) 475–490.
- [26] E.A. Papa Quiroz and P.R. Oliveira, Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds. *J. Convex Anal.* **16** (2009) 46–69.
- [27] T. Rapcsák, *Smooth Nonlinear Optimization*. Kluwer Academic Publishers (1997).
- [28] O.S. Rothaus, Domains of positivity. *Abh. Math. Sem. Univ. Hamburg* **24** (1960) 189–235.
- [29] R.T. Rockafellar, Monotone operations and the proximal point method. *SIAM J. Control Optim.* **14** (1976) 877–898.
- [30] R.T. Rockafellar and R. Wets, *Variational Analysis, Grundlehren der Mathematischen, Wissenschaften* **317**. Springer (1990).
- [31] T. Sakai, *Riemannian Geometry*. American Mathematical Society, Providence, RI (1996).
- [32] S.S. Souza, P.R. Oliveira, J.X. da Cruz Neto and A. Soubeyran, A proximal method with separable Bregman distance for quasiconvex minimization on the nonnegative orthant. *Eur. J. Oper. Res.* **201** (2010) 365–376.
- [33] A. Takayama, *Mathematical Economics*, 2nd Edition. Cambridge University Press, Cambridge (1995).
- [34] P. Tseng, Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* **109** (2001) 475–494.
- [35] C. Udriste, *Convex Function and Optimization Methods on Riemannian Manifolds*. Kluwer Academic Publishers (1994).
- [36] H. Wolkowicz, R. Saigal and L. Vanderberge, Eds., *Handbook of Semidefinite Programming Theory, Algorithms and Applications*, 1st Edition. Internat. Ser. Oper. Management Sci., Springer (2005).