Local Convergence Analysis of Proximal Point Method for a Special Class of Nonconvex Functions on Hadamard Manifolds

Glaydston de Carvalho Bento
Universidade Federal de Goiás
Instituto de Matemática e Estatística, Caixa Posta 131
CEP 74001-970, Goiânia, Brazil
glaydston@mat.ufg.br

Orizon Pereira Ferreira
Universidade Federal de Goiás
Instituto de Matemática e Estatística, Caixa Posta 131
CEP 74001-970, Goiânia, Brazil
orizon@mat.ufg.br

Paulo Roberto Oliveira
Universidade Federal do Rio de Janeiro
Departamento de Engenharia de Sistemas e Computação, Caixa Postal 68511
CEP 21945-970, Rio de Janeiro, Brazil
poliveir@cos.ufrj.br

Resumo

Neste artigo apresentamos o método de ponto proximal para uma classe especial de funções não-convexas em variedades de Hadamard. É garantida a boa definição das sequência gerada pelo método de ponto proximal. Além disso, é provado que cada ponto de acumulação da sequência satisfaz as condições necessárias de otimalidade e, sob hipóteses adicionais, a convergência para um minimizador é obtida.

Palavras Chave: Método de Ponto Proximal, funções não convexas, variedades de Hadamard.

Área principal: Programação matemática.

Abstract

In this paper we present the proximal point method for a special class of nonconvex function on a Hadamard manifold. The well definedness of the sequence generated by the proximal point method is guaranteed. Moreover, it is proved that each accumulation point of this sequence satisfies the necessary optimality conditions and, under additional assumptions, its convergence for a minimizer is obtained.

Key words: proximal point method, nonconvex functions, Hadamard manifolds.

Main area: Mathematical programming.

1 Introduction

The extension of the concepts and techniques of the Mathematical Programming from Euclidean space to Riemannian manifolds occurs naturally and has been frequently done in recent years, with a theoretical purpose as well as to obtain effective algorithms; see Absil et all [1], Attouch et all [2], Azagra et all [3], Alvarez et all [4], Bento et all [5], Bento et all [6], Bento et all [7], Bento et all [8], Barani and Pouryayevali [9] , Barani and Pouryayeval [10], da Cruz Neto et all [11], da Cruz Neto et all [12], da Cruz Neto et all [13], Ferreira and Oliveira [16], Ferreira and Oliveira [17], Ferreira and Svaiter [18], Ferreira [19], Ferreira and Silva [20], Ledyaev et all [25], Li and Wang [26], Li and Wang [27], Li et all [28], Li el all [29], Li et all [30], Nesterov and Todd [33], Németh [34], Papa Quiroz and Oliveira[35], Papa Quiroz and Oliveira [36], Rapcsák [37], Rapcsák [38], Smith [43], Tang [44], Wang at all [48], Wang at all [49], Wang at all [50] and Wang [51]. In particular, we observe that, these extensions allow the solving of some nonconvex constrained problems in Euclidean space. More precisely, nonconvex problems in the classic sense may become convex with the introduction of an adequate Riemannian metric on the manifold (see, for example da Cruz Neto et all [13] and Bento e Melo [7]. The proximal point algorithm, introduced by Martinet [31] and Rockafellar [39], has been extended to different contexts, see Ferreira and Oliveira [17], Li et all [29] and Papa Quiroz and Oliveira[36] and their references. In Ferreira and Oliveira [17] the proximal point method has been generalized in order to solve convex optimization problems of the form

$$(P) \quad \begin{array}{l} \min f(p) \\ \text{s.t. } \ p \in M, \end{array} \tag{1}$$

where M is a Hadamard manifold and $f : M \to \mathbb{R}$ is a convex function (in Riemannian sense). The method was described as follows:

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

with $p^{\circ} \in M$ an arbitrary point, d the intrinsic Riemannian distance (to be defined later on) and $\{\lambda_k\}$ a sequence of positive numbers. The authors also showed that this extension is natural. In Li at all [29] the important notion of maximal monotonicity from a multivalued operator defined on a Banach space to multivalued vector field defined on a Hadamard manifold has been extended. Beside the authors present a general proximal point method to finding singularity of a multivalued vector field. In particular, as an application of the convergence result obtained for the proposed algorithm, constrained optimization problems have been solved. With regards to Papa Quiroz and Oliveira[36] the authors generalized the proximal point method with Bregman distance for solving quasiconvex and convex optimization problems on Hadamard manifolds. In particular, in Spingarn [42] has been developed the proximal point method for the minimization of a certain class of nondifferentiable and noncovex functions, namely, lower-C^2 functions defined on the Euclidean space, see also Hare and Sagastizábal [22]. Kaplan and Tichatschke [23] also applied the proximal point method for the minimization of a similar class of the ones studied in Hare and Sagastizábal [22] and Spingarn [42], namely, the maximum of continuously differentiable functions.

Our goal is to study the same class objective functions studied in Kaplan and Tichatschke, [23], in the Riemannian context, applying the proximal point method (2) in order to solve the problem (1) with the objective function in that class. For this purpose, it is necessary to study the generalized directional derivative and subdifferential in the Riemannian manifolds context. Several works have studied such concepts and presented many useful results in the Riemannian nonsmooth optimization context, see for example Azagra at all [3], Ladyaev at all[25], Montreanu and Pavel [32] and Thamelt [45] .

The organization of our paper is as follows. In Section 1.1 we define the notations and list some results of Riemannian geometry to be used throughout this paper. In Section 2, we recall some facts of the convex analysis on Hadamard manifolds. In Section 3 we present some properties of the directional derivative of a convex function defined on a Hadamard manifold, including a characterization of the directional derivative and of the subdifferential of the maximum of a certain class of convex functions. Here we also present the definition for the generalized directional derivative of locally Lipschitz functions (not necessarily convex) and an important property of the subdifferential of the maximum of differentiable continuously functions. In Section 4 we present an application of the proximal point method (2) to solve the problem (1) in the case where the objective function is a real-valued function on a Hadamard manifold M (not necessarily convex) given by the maximum of a certain class of functions. The main results are the proof of well definition of the sequence generated by (2), the proof that each accumulation point of this sequence is a stationary point of the objective function and, under some additional assumptions, the proof of convergence of that sequence to a solution of the problem (1).

1.1 Notation and terminology

In this section, we introduce some fundamental properties and notations about Riemannian geometry. These basics facts can be found in any introductory book on Riemannian geometry, such as do Carmo [14] or Sakai [41].

Let M be a n-dimentional connected manifold. We denote by T_pM the n-dimentional *tangent space* of M at p, by $TM = \cup_{p\in M} T_pM$ *tangent bundle* of M and by $\mathcal{X}(M)$ the space of smooth vector fields over M. When M is endowed with a Riemannian metric $\langle\,,\,\rangle$, with the corresponding norm denoted by $\|\;\|$, then M is now a Riemannian manifold. Recall that the metric can be used to define the length of piecewise smooth curves $\gamma : [a, b] \to M$ joining p to q, i.e., such that $\gamma(a) = p$ and $\gamma(b) = q$, by:

$$l(\gamma) = \int_a^b \|\gamma'(t)\| dt,$$

and, moreover, by minimizing this length functional over the set of all such curves, we obtain a Riemannian distance $d(p, q)$ inducing the original topology on M. The metric induces a map $f \mapsto \operatorname{grad} f \in \mathcal{X}(M)$ which, for each function smooth over M, associates its gradient via the rule $\langle \operatorname{grad} f, X\rangle = df(X),\ X \in \mathcal{X}(M)$. Let ∇ be the Levi-Civita connection associated with $(M, \langle\,,\,\rangle)$. In each point $p \in M$, we have a linear map $A_X(p)\colon T_pM \to T_pM$ defined by:

$$A_X(p)v = \nabla_v X. \tag{3}$$

If $X = \operatorname{grad} f$, where $f\colon M \to \mathbb{R}$ is a twice differentiable function, then $A_X(p)$ is the *Hessian* of f at p and is denoted by $\operatorname{Hess}_p f$. A vector field V along γ is said to be *parallel* if $\nabla_{\gamma'} V = 0$. If γ' itself is parallel we say that γ is a *geodesic*. Given that the geodesic equation $\nabla_{\gamma'}\gamma' = 0$ is a second order nonlinear ordinary differential equation, we conclude that the geodesic $\gamma = \gamma_v(., p)$ is determined by its position p and velocity v at p. It is easy to check that $\|\gamma'\|$ is constant. We say that γ is *normalized* if $\|\gamma'\| = 1$. The restriction of a geodesic to a closed bounded interval is called a *geodesic segment*. A geodesic segment joining p to q in M is said to be *minimal* if its length is equals $d(p, q)$ and the geodesic in question is said to be a *minimizing geodesic*. If γ is a geodesic joining points p and q in M then, for each $t \in [a, b]$, ∇ induces a linear isometry, relative to $\langle\,,\,\rangle$, $P_{\gamma(a)\gamma(t)} : T_{\gamma(a)}M \to T_{\gamma(t)}M$, the so-called *parallel transport* along γ from $\gamma(a)$ to $\gamma(t)$. The inverse map of $P_{\gamma(a)\gamma(t)}$ is denoted by $P^{-1}_{\gamma(a)\gamma(t)} : T_{\gamma(t)}M \to T_{\gamma(a)}M$. In the particular case of γ to be the unique geodesic segment joining p and q, then the parallel transport along γ from p to q is denoted by $P_{pq} : T_pM \to T_qM$.

A Riemannian manifold is *complete* if the geodesics are defined for any values of t. Hopf-Rinow's theorem asserts that if this is the case then any pair of points, say p and q, in M can be joined by a (not necessarily unique) minimal geodesic segment. Moreover, (M, d) is a complete metric space so that bounded and closed subsets are compact. From the completeness of the

Riemannian manifold M, the *exponential map* $exp_p : T_pM \to M$ is defined by $exp_p v = \gamma_v(1, p)$, for each $p \in M$.

We denote by R *the curvature tensor* defined by $R(X,Y) = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[Y,X]} Z$, with $X, Y, Z \in \mathcal{X}(M)$, where $[X,Y] = YX - XY$. Moreover, the *sectional curvature* with respect to X and Y is given by $K(X,Y) = \langle R(X,Y)Y, X\rangle/(||X||^2||X||^2 - \langle X\,,\,Y\rangle^2)$, where $||X|| = \langle X, X\rangle^2$. If $K(X,Y) \leqslant 0$ for all X and Y, then M is called a *Riemannian manifold of nonpositive curvature* and we use the short notation $K \leqslant 0$.

Theorem 1.1. *Let M be a complete, simply connected Riemannian manifold with nonpositive sectional curvature. Then M is diffeomorphic to the Euclidean space $\mathbb{R}^n$, $n = dim M$. More precisely, at any point $p \in M$, the exponential map exp_p is a diffeomorphism.*

Proof. See Lemma 3.2 of do Carmo [14], p. 149 or Theorem 4.1 of Sakai [41], p. 221. □

A complete simply connected Riemannian manifold of nonpositive sectional curvature is called a *Hadamard manifold*. Thus Theorem 1.1 states that if M is a Hadamard manifold, then M has the same topology and differential structure of the Euclidean space $\mathbb{R}^n$. Furthermore, are known some similar geometrical properties of the Euclidean space $\mathbb{R}^n$, such as, given two points there exists an unique geodesic segment that joins them. *In this paper, all manifolds M are assumed to be Hadamard and finite dimensional.*

2 Convexity in Hadamard manifolds

In this section, we introduce some fundamental properties and notations of convex analysis on Hadamard manifolds. References of the convex analysis, on the Euclidean space $\mathbb{R}^n$ may be found in Hiriart-Urruty [21] and on Riemannian manifolds may be found in da Cruz Neto at all [12], Ferreira and Oliveira [17], Rapcsák [37], Sakai [41], Smith [43] and Udriste [46].

The set $\Omega \subset M$ is said to be *convex* if for any geodesic segment, with end points in Ω, is contained in Ω. Let $\Omega \subset M$ be an open convex set. A function $f : M \to \mathbb{R}$ is said to be *convex* (respectively, *strictly convex*) on Ω if for any geodesic segment $\gamma : [a,b] \to \Omega$ the composition $f \circ \gamma : [a,b] \to \mathbb{R}$ is convex (respectively, strictly convex). Moreover, a function $f : M \to \mathbb{R}$ is said to be *strongly convex* on Ω with constant $L > 0$ if, for any geodesic segment $\gamma : [a,b] \to \Omega$, the composition $f \circ \gamma : [a,b] \to \mathbb{R}$ is strongly convex with constant $L\|\gamma'(0)\|^2$. Take $p \in M$. A vector $s \in T_pM$ is said to be a *subgradient* of f at p, if:

$$f(q) \geq f(p) + \langle s, \exp_p^{-1} q\rangle,$$

for any $q \in M$. The set of all subgradients of f at p, $\partial f(p)$, is called the *subdifferential* of f at p.

The following result provides a characterization of convexity in the case of differentiable functions.

Proposition 2.1. *Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a differentiable function on Ω. We say that f is convex on Ω if, and only if, for any $p \in \Omega$:*

$$f(q) - f(p) \geq \langle \operatorname{grad} f(p), \exp_p^{-1} q\rangle, \qquad \forall q \in \Omega.$$

Proof. See Theorem 5.1 of Udriste [46], page 78. □

The most important consequence of the previous proposition is that with f being convex, any of its critical points are global minimum points. In particular, if M is compact, then f is constant. Moreover, $0 \in \partial f(p)$ if, and only if, p is a minimum point of f in M. See, for example, [46].

Definition 2.1. *Let $\Omega \subset M$ be an open convex set and X a vector field defined in M. X is said to be monotone on Ω, if:*

$$\left\langle \exp_q^{-1} p\,,\, P_{qp}^{-1} X(p) - X(q) \right\rangle \geq 0, \qquad p,\, q \in \Omega, \tag{4}$$

where P_{qp} is the parallel transport along the geodesic joining q to p. If (4) is satisfied with strict inequality for all $p, q \in \Omega$, $p \neq q$, then X is said to be strictly monotone. Moreover, X is strongly monotone if there exists $\lambda > 0$ such that:

$$\left\langle \exp_q^{-1} p\,,\, P_{qp}^{-1} X(p) - X(q) \right\rangle \geq \lambda d^2(p,q) \qquad p,\, q \in \Omega. \tag{5}$$

Remark 2.1. *In the particular case that $M = \mathbb{R}^n$ with the usual metric, inequality* (4) *and* (5) *becomes, respectively:*

$$\langle p - q\,,\, X(p) - X(q)\rangle \geq 0, \qquad \langle p - q\,,\, X(p) - X(q)\rangle \geq \lambda \|p - q\|^2,$$

because $\exp_q^{-1} p = p - q$ and $P_{qp}^{-1} = I$. Therefore the Definition 2.1 extends the concept of monotone operators from $\mathbb{R}^n$ to Riemannian manifolds.

Now we present an important example of strong monotone vector field being particularly useful in the remainder of this work.

Take $p \in M$ and let $exp_p^{-1} : M \to T_pM$ be the inverse of the exponential map. Note that $d(q\,,\,p) = ||exp_p^{-1}q||$, the map $d^2(\,.\,,p)\colon M \to \mathbb{R}$ is C^∞ and

$$\operatorname{grad} \frac{1}{2} d^2(q,p) = -exp_q^{-1}p,$$

(M is a Hadamard manifold). See, for example, Proposition 4.8 of [41], p. 108.

Proposition 2.2. *Take $p \in M$. The gradient vector field $\operatorname{grad}(d^2(\,.\,,p)/2)$ is strongly monotone with $\lambda =$ 1.*

Proof. See da Cruz Neto at all [12]. □

Proposition 2.3. *Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a differentiable function on Ω.*

(i) f is convex on Ω if and only if $\operatorname{grad} f$ is monotone on Ω;

(ii) f is strictly convex on Ω if and only if $\operatorname{grad} f$ is strictly monotone on Ω;

(iii) f is strongly convex on Ω if and only if $\operatorname{grad} f$ is strongly monotone on Ω.

Proof. See da Cruz Neto at all [12]. □

Remark 2.2. *Take $p \in M$. From Propositions 2.2 and 2.3 it follows that the map $d^2(\,.\,,p)/2$ is strongly convex.*

Proposition 2.4. *Let $\Omega \subset M$ be a convex set and $T \subset \mathbb{R}$ a compact set. Let $\psi : M \times T \to \mathbb{R}$ be a continuous function on $\Omega \times T$ such that $\psi_\tau := \psi(.,\tau) : M \to \mathbb{R}$ is strongly convex on Ω with constant $L > 0$ for all $\tau \in T$. Then, $\phi : M \to \mathbb{R}$ defined by:*

$$\phi(p) := \max_{\tau \in T} \psi(p,\tau),$$

is strongly convex on Ω with constant L. In particular, if ψ_τ is convex for all $\tau \in T$ then ϕ is convex on Ω.

Proof. See Bento at all [8]. □

Definition 2.2. *Let $\Omega \subset M$ be an open convex set. A function $f : M \to \mathbb{R}$ is said to be Lipschitz on Ω if there exists a constant $L := L(\Omega) \geq 0$ such that*

$$|f(p) - f(q)| \leq Ld(p, q), \qquad p, q \in \Omega. \tag{6}$$

Moreover, if for each $p_0 \in \Omega$ there exists $L(p_0) \geq 0$ and $\delta = \delta(p_0) > 0$ such that inequality (6) holds with $L = L(p_0)$ for all $p, q \in B_\delta(p_0) := \{p \in \Omega : d(p, p_0) < \delta\}$, then f is called locally Lipschitz on Ω.

Definition 2.3. *Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a continuously differentiable function on Ω. The gradient vector field of f, $\operatorname{grad} f$, is said to be Lipschitz with constant $\Gamma \geq 0$ on Ω whenever:*

$$\| \operatorname{grad} f(q) - P_{pq} \operatorname{grad} f(p)\| \leq \Gamma d(p, q), \qquad p, q \in \Omega,$$

where P_{pq} is the parallel transport along the geodesic joining p to q.

Proposition 2.5. *Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a twice continuously differentiable function on Ω. If Hess ${}_p f$ is bounded on Ω, then the gradient vector field $\operatorname{grad} f$ is Lipschitz on Ω.*

Proof. The proof is an immediate consequence of the fundamental theorem of calculus for vector fields, see for example Ferreira and Svaiter [18]. □

3 Directional derivatives

In this section we present some properties of the directional derivative of a convex function defined on a Hadamard manifold, including a characterization of the directional derivative and of the subdifferential of the maximum of a certain class of convex functions. We also give a definition of the generalized directional derivative of a locally Lipschitz function (not necessarily convex), see Azagra et.all [3], and an important property of the subdifferential of the maximum of continuously differentiable functions.

3.1 Directional derivatives of convex functions

In this subsection we present the definition of the directional derivative of a convex function defined on a Hadamard manifold and some properties involving its subdifferential, which allow us to obtain an important property of the subdifferential of the maximum of a certain class of convex functions.

Let $\Omega \subset M$ be an open convex set and $f \colon M \to R$ a convex function on Ω. Take $p \in \Omega$, $v \in T_pM$ and $\delta > 0$ and let $\gamma : [-\delta\, , \, \delta] \to \Omega$ be the geodesic segment such that $\gamma(0) = p$ and $\gamma'(0) = v$. Due to the convexity of $f \circ \gamma : [-\delta\, , \, \delta] \to \mathbb{R}$, the function $q_\gamma : (0\, , \, \delta] \to \mathbb{R}$, given by

$$q_\gamma(t) := \frac{f(\gamma(t)) - f(p)}{t}, \tag{7}$$

is nondecreasing. Moreover, since f is locally Lipschitzian, it follows that q_γ is bounded near zero. This leads to the following definition:

Definition 3.1. *Let $\Omega \subset M$ be an open convex set and $f \colon M \to R$ a convex function on Ω. Then the directional derivative of f at $p \in \Omega$ in the direction of $v \in T_pM$ is defined by*

$$f'(p, v) := \lim_{t \to 0^+} q_\gamma(t) = \inf_{t>0} q_\gamma(t), \tag{8}$$

where $\delta > 0$ and $\gamma : [-\delta\, , \, \delta] \to \Omega$ is the geodesic segment such that $\gamma(0) = p$ and $\gamma'(0) = v$.

Proposition 3.1. *Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a convex function on Ω. Then, for each fixed $p \in \Omega$, the subdifferential $\partial f(p)$ is convex.*

Proof. See Theorem 4.6 of Udriste [46], p. 74. □

Proposition 3.2. *Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a convex function on Ω. Then, for each point fixed $p \in \Omega$, the following statement holds:*

i) $f'(p, v) = \max_{s \in \partial f(p)} \langle s, v\rangle$, *for all* $v \in T_pM$;

ii) $\partial f(p) = \{s \in T_pM : f'(p, v) \geq \langle s, v\rangle, v \in T_pM\}$.

Proof. See da Cruz Neto at all [11]. □

Proposition 3.3. *Let T be a compact set, $\Omega \subset M$ an open convex set and $h : M \times T \to \mathbb{R}$ a continuous function on $\Omega \times T$ such that $h(.,\tau) : M \to \mathbb{R}$ is convex on Ω for all $\tau \in T$. If $f : M \to \mathbb{R}$ is given by $f(p) = \max_{\tau \in T} h(p, \tau)$, then f is convex on Ω and*

$$f'(p, v) = \max_{\tau \in T(p)} h'(p, \tau, v), \qquad p \in \Omega, \quad v \in T_pM,$$

where $T(p) = \{\tau \in T : f(p) = h(p, \tau)\}$. Moreover, if $h(., \tau)$ is differentiable on Ω for all $\tau \in T$ and $\operatorname{grad}_p h(p, .)$ is continuous for all $p \in \Omega$, then:

$$\partial f(p) = \operatorname{conv}\left\{\operatorname{grad}_p h(p, \tau) : \tau \in T(p)\right\}.$$

Proof. See Bento at all [8]. □

Corollary 3.1. *Let $\Omega \subset M$ be a open convex set and $h_i : M \to \mathbb{R}$ a differentiable convex function on Ω for $i \in I := \{1, ..., m\}$. If $h : M \to \mathbb{R}$ is defined by $h(p) := \max_{i \in I} h_i(p)$, then:*

$$\partial h(p) = \operatorname{conv}\{\operatorname{grad} h_i : i \in I(p)\} = \left\{ y \in T_pM : y = \sum_{i \in I(p)} \alpha_i \operatorname{grad} h_i(p), \sum_{i \in I(p)} \alpha_i = 1, \alpha_i \geq 0 \right\},$$

where $I(p) := \{i \in I : h(p) = h_i(p)\}$. In particular, p minimizes h on Ω, if and only if, there exist $\alpha_i \geq 0$, $i \in I(p)$, such that:

$$0 = \sum_{i \in I(p)} \alpha_i \operatorname{grad} h_i(p), \qquad \sum_{i \in I(p)} \alpha_i = 1.$$

Proof. It follows directly from Proposition 3.3. □

3.2 Directional derivatives of locally Lipschitz functions

In the sequel we present the definition of generalized directional derivative of a locally Lipschitz function (not necessarily convex) and an important property of the subdifferential of the maximum of continuously differentiable functions.

Definition 3.2. *Let $\Omega \subset M$ be an open convex set and $f : M \to \mathbb{R}$ a locally Lipschitz function on Ω. The generalized directional derivative of f at $p \in \Omega$ in the direction $v \in T_pM$ is defined by:*

$$f^\circ(p, v) := \limsup_{t \downarrow 0 \;\; q \to p} \frac{f(\exp_q t(D(\exp_p)_{\exp_p^{-1} q} v)) - f(q)}{t}. \tag{9}$$

It is worth noting that an equivalent definition has appeared in [3].

Remark 3.1. *The generalized directional derivative is well defined. Indeed, let $L_p > 0$ be the Lipschitz constant of f in p and $\delta = \delta(p) > 0$ such that*

$$|f(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - f(q)| \leq L_p\, d(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v, q), \quad q \in B_\delta(p), \quad t \in [0, \delta).$$

Since $d(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v, q) = t\|(D\exp_p)_{\exp_p^{-1} q} v\|$, the above inequality becomes:

$$|f(\exp_q t(D\exp_p)_{\exp_p^{-1} q} v) - f(q)| \leq L_p\, t\|(D\exp_p)_{\exp_p^{-1} q} v\|, \quad q \in B_\delta(p), \quad t \in [0, \delta).$$

Since $\lim_{q \to p} (D\exp_p)_{\exp_p^{-1} q} v = v$, our statement follows from the latter inequality.

Remark 3.2. *Note that, if* $M = \mathbb{R}^n$ *then* $\exp_p w = p + w$ *and*

$$D(\exp_p)_{\exp_p^{-1} q} v = v.$$

In this case, (9) *becomes:*

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

which is the Clarke's generalized directional derivative in the Euclidean case, see Clarke [15]. Therefore, the generalized differential derivative on Hadamard manifolds is a natural extension of the Clarke's generalized differential derivative.

Next we generalize the definition of subdifferential for locally Lipschitz functions defined on Hadamard manifolds, see Proposition 3.2 item ii.

Definition 3.3. *Let* $\Omega \subset M$ *be an open convex set and* $f : M \to \mathbb{R}$ *a locally Lipschitz function on* Ω*. The generalized subdifferential of* f *at* $p \in \Omega$*, denoted by* $\partial^{\circ} f(p)$*, is defined by:*

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

Remark 3.3. *If* f *is convex on* Ω*, then* $f^{\circ}(p, v) = f'(p, v)$ *(respectively,* $\partial^{\circ} f(p) = \partial f(p)$*) for all* $p \in \Omega$*, i.e., the directional derivatives (respectively, subdifferential) for Lipschitz functions is a generalization of the directional derivatives (respectively, subdifferential) for convex functions. See Azagra at all [3] Claim* 5.4 *in the proof of Theorem* 5.3.

Definition 3.4. *Let* $\Omega \subset M$ *be an open convex set and* $f : M \to \mathbb{R}$ *a locally Lipschitz function on* Ω*. A point* $p \in \Omega$ *is a stationary point of* f *if* $0 \in \partial^{\circ} f(p)$*.*

Lemma 3.1. *Let* $\Omega \subset M$ *be an open set. If* $f : M \to \mathbb{R}$ *is locally Lipschitz on* Ω *and* $g : M \to \mathbb{R}$ *is continuously differentiable on* Ω*, then:*

$$(f + g)^{\circ}(p, v) = f^{\circ}(p, v) + g'(p, v) \qquad p \in \Omega, \quad v \in T_pM. \tag{10}$$

As a consequence,

$$\partial^{\circ}(f + g)(p) = \partial^{\circ} f(p) + \operatorname{grad} g(p), \qquad p \in \Omega. \tag{11}$$

Proof. See Bento at all [8]. □

It is possible to prove that the next result holds with equality. However we will prove just the inclusion needed to prove our main result.

Proposition 3.4. *Let* $\Omega \subset M$ *be an open convex set and* $I = \{1, ..., m\}$*. Let* $f_i : M \to \mathbb{R}$ *be a continuously differentiable function on* Ω *for all* $i \in I$ *and* $f : M \to \mathbb{R}$ *defined by*

$$f(p) := \max_{i \in I} f_i(p).$$

Then f *is Lipschitz locally on* Ω *and for each* $p \in \Omega$ $\operatorname{conv}\{\operatorname{grad} f_i(p) : i \in I(p)\} \subset \partial^{\circ} f(p)$*, where* $I(p) := \{i \in I : f_i(p) = f(p)\}$*.*

Proof. See Bento at all [8]. □

4 Proximal Point Method for Nonconvex Problems

In this section we present an application of the proximal point method to minimize a real-valued function (not necessarily convex) given by the maximum of a certain class of continuously differentiable functions. Our goal is to prove the following theorem:

Theorem 4.1. *Let $\Omega \subset M$ be an open convex set, $q \in M$ and $I = \{1, ..., m\}$. Let $f_i : M \to \mathbb{R}$ be a continuously differentiable function on Ω, for all $i \in I$, and $f : M \to \mathbb{R}$ defined by*

$$f(p) := \max_{i\in I} f_i(p).$$

Assume that $-\infty < \inf_{p\in M} f(p)$, $\operatorname{grad} f_i$ is Lipschitz on Ω with constant L_i for each $i \in I$ and

$$L_f(f(q)) = \{p \in M : f(p) \leq f(q)\} \subset \Omega, \qquad \inf_{p\in M} f(p) < f(q).$$

Take $0 < \bar{\lambda}$ and a sequence $\{\lambda_k\}$ satisfying $\max_{i\in I} L_i < \lambda_k \leq \bar{\lambda}$ and $\hat{p} \in L_f(f(q))$. Then the proximal point method

$$p^{k+1} := \operatorname{argmin}_{p\in M} \left\{ f(p) + \frac{\lambda_k}{2} d^2(p, p^k) \right\}, \qquad k = 0, 1, \ldots, \tag{12}$$

with starting point $p^0 = \hat{p}$ is well defined, the generated sequence $\{p^k\}$ rest in $L_f(f(q))$ and satisfies only one of the following statements

i) $\{p^k\}$ is finite, i.e., $p^{k+1} = p^k$ for some k and, in this case, p^k is a stationary point of f,

ii) $\{p^k\}$ is infinite and, in this case, any accumulation point of $\{p^k\}$ is a stationary point of f.

Moreover, assume that the minimizer set of f is non-empty, i. e.,

h1) $U^* = \{p : f(p) = \inf_{p\in M} f(p)\} \neq \emptyset$.

Let $c \in (\inf_{p\in M} f(p),\ f(q))$. If, in addition, the following assumptions hold:

h2) *$L_f(c)$ is convex, f is convex on $L_f(c)$ and f_i is continuous on $\bar{\Omega}$ the closure of Ω for $i \in I$;*

h3) *For all $p \in L_f(f(q)) \setminus L_f(c)$ and $y(p) \in \partial^\circ f(p)$ we have $\|y(p)\| > \delta > 0$,*

then the sequence $\{p^k\}$ generated by (12) with

$$\max_{i\in I} L_i < \lambda_k \leq \bar{\lambda}, \qquad k = 0, 1, \ldots \tag{13}$$

converge to a point $p^* \in U^*$.

Proof. See Bento at all [8]. □

Remark 4.1. *The continuity of each function f_i on $\bar{\Omega}$ in* **h2** *guarantees that the level sets of the fuction f, in particular the solution set U^*, are closed in the topology of the manifold.*

In the next remark we show that if Ω is bounded and f_i is convex on Ω and continuous on $\bar{\Omega}$ for all $i \in I$, then f satisfies the assumptions **h2** and **h3**.

Remark 4.2. *If f_i is also a convex function on Ω for each $i \in I$ then by the Proposition 2.4, the function f is convex on Ω and the assumption* **h2** *is satisfied for all $c \leq f(q)$. Moreover, from Remark 3.3,*

$$\partial f^\circ(p) = \partial f(p), \quad \forall\, p \in \Omega. \tag{14}$$

Take $c \in (\inf_{p\in M} f(p),\ f(q))$ *and let us suppose that* **h1** *hold and* Ω *is a bounded set. Then, we have*

$$0 < \sup\{d(p^*, p) : p^* \in U^*, p \in L_f(f(q)) \setminus L_f(c)\} = \epsilon < +\infty. \tag{15}$$

Let $p^* \in U^*$ *be fixed,* $p \in L_f(f(q)) \setminus L_f(c)$ *and* $y(p) \in \partial f(p)$*. The convexity of* f *on* Ω *implies that:*

$$\langle y(p)\ ,\ -\exp_p^{-1} p^*\rangle \geq f(p) - f(p^*).$$

Since $\|y(p)\|\|\exp_p^{-1} p^*\| \geq \langle y(p), -\exp_p^{-1} p^*\rangle$, $d(p^*, p) = \|\exp_p^{-1} p^*\|$, $p \in L_f(f(q)) \setminus L_f(c)$ *and* U^* *is a proper subset of* $L_f(c)$*, from the above inequality, we obtain* $\|y(p)\|d(p^*, p) > c - f(p^*) > 0$*. Thus, from* (15) *and latter inequality* $\|y(p)\|\epsilon > c - f(p^*) > 0$*. Therefore, choosing* $\delta = (c - f(p^*))/\epsilon$*, we have:* $\|y(p)\| > \delta > 0$*, which, combined with* (14)*, shows that* f *satisfies* **h3***.*

Proof. See Bento at all [8]. □

References

[1] Absil, P. -A., Baker, C. G., Gallivan, K. A.: Trust-region methods on Riemannian manifolds, Found. Comput. Math. **7** (3), 303-330 (2007)

[2] Attouch, H., Bolte, J., Redont, P., Teboulle, M.: Singular Riemannian barrier methods and gradient-projection dynamical systems for constrained optimization, Optimization. **53** (5-6), 435-454 (2004)

[3] Azagra, D., Ferrera, J. López-Mesas, M.: Nonsmooth analysis and Hamilton-Jacobi equations on Riemannian manifolds. J. Funct. Anal. **220**, 304-361 (2005)

[4] Alvarez, F., Bolte, J., Munier, J.: A Unifying Local Convergence Result for Newton's Method in Riemannian Manifolds. Found. Comput. Math. **8**, 197-226 (2008)

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

[6] Bento, G. C., Ferreira, O. P., Oliveira, P. R.: Unconstrained steepest descent method for multicriteria optimization on Riemannian manifolds. J. Optim. Theory Appl. DOI: 10.1007/s10957-011-9984-2 (2012)

[7] Bento, G. C., Melo, J. G.: A subgradient method for convex feasibility on Riemannian Manifolds. J. Optim. Theory Appl. DOI: 10.1007/s10957-011-9921-4 (2011)

[8] Bento, G. C., Ferreira, O. P., Oliveira, P. R.: Proximal Point Method for a Special Class of Nonconvex Functions on Hadamard Manifolds, 2010. http://arxiv.org/abs/0809.2594.

[9] Barani, A., Pouryayevali, M.R.: Invex sets and preinvex functions on Riemannian manifolds. J. Math. Anal. Appl. **328** (2), 767-779 (2007)

[10] Barani, A., Pouryayevali, M.R.: Invariant monotone vector fields on Riemannian manifolds. Nonlinear Anal. **70**(5), 1850-1861 (2009)

[11] da Cruz Neto, J. X., Ferreira, O. P., Lucambio Pérez, L. R.: Monotone point-to-set vector fields. Balkan J. Geom. Appl. **5**, (1), 69-79 (2000)

[12] da Cruz Neto, J. X., Ferreira, O. P., Lucâmbio Pérez, L. R.: Contribution to the study of monotone vector fields. Acta Math. Hungar. **94** (4), 307-320 (2002)

[13] da Cruz Neto, J. X., Ferreira, O. P., Lucâmbio Pérez, L. R., Németh, S. Z.: Convex- and Monotone-Transformable Mathematical Programming Problems and a Proximal-Like Point Method. J. Glob. Optim. **35**, 53-69, (2006)

[14] do Carmo, M.P.: Riemannian Geometry, Boston, Birkhauser, (1992)

[15] Clarke, F. H.: Optimization and Nonsmooth Analysis. Classics in applied mathematics. **5**, SIAM, New York, (1983).

[16] Ferreira, O. P., Oliveira, P. R.: Subgradient algorithm on Riemannian manifolds. J. Optim. Theory Appl. **97** (1), 93-104 (1998)

[17] Ferreira, O. P., Oliveira, P. R.: Proximal point algorithm on Riemannian manifolds. Optimization]**51** (2), 257-270 (2000)

[18] Ferreira, O. P., Svaiter, B. F.: Kantorovich's Theorem on Newton's Method in Riemannian manifolds. J. Complexity. **18**, 304-329 (2002)

[19] Ferreira, O.P.: Proximal subgradient and a characterization of Lipschitz function on Riemannian manifolds, J. Math. Anal. Appl. **313**, 587-597 (2006)

[20] Ferreira, O. P., Silva, R. C. M.: Local convergence of Newton's method under majorant condition in Riemannian Manifolds. IMA J. Numer. Anal. To appear, (2012)

[21] Hiriart-Urruty, J.-B, Lemaréchal, C.: Convex analysis and minimization algorithms I and II, Springer-Verlag, (1993).

[22] Hare, W., Sagastizábal, C.: Computing proximal points of nonconvex functions. Math. Program. Ser. B **116**, 221-258 (2009)

[23] Kaplan, A., Tichatschke, R.: Proximal point methods and nonconvex optimization. J. Global Optim. **13** (4), 389-406 (1998)

[24] Lang, S.: Fundamentals of Differential Geometry. Springer - Verlag, (1998).

[25] Ledyaev, Yu. S., Zhu, Q.J.: Nonsmooth analysis on smooth manifolds, Trans. Amer. Math. Soc. **359**, 3687-3732 (2007)

[26] Li, C., Wang, J. H.: Newton's method on Riemannian manifolds: Smale's point estimate theory under the γ-condition. IMA J. Numer. Anal. **26**(2), 228-251 (2006)

[27] Li, C., Wang, J. H.: Newton's Method for Sections on Riemannian Manifolds: Generalized Covariant α-Theory. J. Complexity, **24**, 423-451 (2008)

[28] Li, S.L., Li, C., Liou, Y.C., Yao, J.C.: Existence of solutions for variational inequalities on Riemannian manifolds, Nonlinear Anal. **71**, 5695-5706 (2009)

[29] Li, C., López, G., Martín-Márquez, V.: Monotone vector fields and the proximal point algorithm on Hadamard manifolds, J. London Math. Soc. **79**(2), 663-683 (2009)

[30] Li, C., Mordukhovich, B. S., Wang, J. H., Yao, J. C.: Weak Sharp Minima on Riemannian Manifolds. SIAM J. Optim.**21**, 1523 (2011)

[31] Martinet, B.: Régularisation, d'inéquations variationelles par approximations successives. (French) Rev. Française Informat. Recherche Opérationnelle **4**, Ser. R-3, 154-158 (1970)

[32] Motreanu, D., Pavel, N. H.: Quasitangent vectors in flow-invariance and optimization problems on Banach manifolds. J. Math. Anal. Appl. **88** (1), 116-132 (1982)

[33] Nesterov, Y. E. and Todd, M. J.: On the Riemannian Geometry Defined by Self-Concordant Barriers and Interior-Point Methods. Found. Comput. Math. **2** (4), 333-361 (2002)

[34] Németh, S. Z.: Variational inequalities on Hadamard manifolds. Nonlinear Anal. **52**(5), 1491-1498 (2003)

[35] Papa Quiroz, E. A., Quispe, E. M., Oliveira, P. R.: Steepest descent method with a generalized Armijo search for quasiconvex functions on Riemannian manifolds. J. Math. Anal. Appl. **341**(1), 467-477 (2008)

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

[37] Rapcsák, T.: Smooth nonlinear optimization in R^n, Kluwer Academic Publishers, Dordrecht, (1997)

[38] Rapcsák, T.: Local Convexity on Smooth Manifolds. J. Optim. Theory Appl. **127**(1), 165-176 (2005)

[39] Rockafellar, R. T.: Monotone operators and the proximal point algorithm. SIAM J. Control. Optim. **14**, 877-898 (1976)

[40] Rothaus, O. S.: Domains of positivity. Abh. Math. Sem. Univ. Hamburg. **24**, 189-235 (1960)

[41] Sakai, T.: Riemannian geometry. Translations of mathematical monographs, **149**, Amer. Math. Soc., Providence, R.I. (1996).

[42] Spingarn, J. E.: Submonotone mappings and the proximal point algorithm. Numer. Funct. Anal. Optim. **4** (2), 123-150 (1981/82)

[43] Smith, S. T.: Optimization techniques on Riemannian Manifolds. Fields Institute Communications, Amer. Math. Soc., Providence, R.I. 3, 113-146, (1994)

[44] Tang, G., Huang, N.: Korpelevich's method for variational inequality problems on Hadamard manifolds. J. Glob. Optim. DOI 10.1007/s10898-011-9773-3 (2011)

[45] Thämelt, W.: Directional derivatives and generalized gradients on manifolds. Optimization **25** (2-3), 97-115 (1992)

[46] Udriste, C.: Convex functions and optimization Algorithms on Riemannian manifolds, in: Mathematics and its Applications, **297**, Kluwer Academic Publishers (1994)

[47] Wang, J. H., Li, C.: Uniqueness of the singular points of vector fields on Riemannian manifolds under the γ-condition. J. Complexity **22**(4), 533-548 (2006)

[48] Wang, J.H., Huang, S.C., Li, C.: Extended Newton's Algorithm for mappings on Riemannian manifolds with values in a cone. Taiwanese J. Math. **13**, 633-656 (2009)

[49] Wang, J.H., Lopez, G., Martin-Marquez, V., Li, C.: Monotone and accretive vector fields on Riemannian manifolds. J. Optim. Theory Appl. **146**, 691-708 (2010)

[50] Wang, J. H., Yao, J. C., Li, C.: Gauss-Newton method for convex composite optimizations on Riemannian manifolds. J. Glob. Optim. DOI 10.1007/s10898-010-9638-1 (2011)

[51] Wang, J. H.: Convergence of Newton's method for sections on Riemannian manifolds. J. Optim. Theory Appl. **148** (1), 125-145 (2011)