Proximal Point Methods for Functions Involving Lojasiewicz, Quasiconvex and Convex Properties on Hadamard Manifolds

E. A. Papa Quiroz and P. Roberto Oliveira

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

Abstract

This paper extends the full convergence of the proximal point method with Riemannian, Semi-Bregman and Bregman distances to solve minimization problems on Hadamard manifolds. For the unconstrained problem, under the assumptions that the optimal set is nonempty and the objective function is continuous and either quasiconvex or satisfies a generalized Lojasiewicz property, we prove the full convergence of the sequence generated by the proximal point method with Riemannian distances to certain generalized critical point of the problem. For the constrained case, under the same assumption on the optimal set and the quasiconvexity or convexity of the objective function, we study two methods. One of them is the proximal method with semi-Bregman distance, obtaining that any cluster point of the sequence is an optimal solution. The other one, is the proximal method with Bregman distance where we obtain the global convergence of the method to an optimal solution of the problem. In particular, our work recovers some interesting optimization problems, for example, convex and quasiconvex minimization problems in $I\!R^n$, semidefinite problems (SDP), second order cone problems (SOCP), in the same way that, extends the applications of the proximal point methods for solving constrained minimization problems with nonconvex objective functions in Euclidian spaces when the objective function is convex or quasiconvex on the manifold.

Keywords: Proximal Point Methods, Riemannian Distances, Semi-Bregman Functions, Bregman Functions, Quasiconvex Functions, Limiting and Fréchet Subdifferential.

1 Introduction

In this paper we treat with proximal point methods to solve the optimization problem

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

where $f : M \to I\!R \cup \{+\infty\}$ is a proper lower semicontinuous function and X is a nonempty closed convex set on a Hadamard manifold M (recall that a Hadamard manifold is a simply connected Riemannian manifold with nonpositive sectional curvature).

A motivation to study optimization problems on Riemannian manifolds is that it permits an unified analysis of the theory and methods of diverse kinds of problems defined in not needily vectorial spaces. Another one is that the geometry of the space is dependent of the Riemannian metric (arbitrarily introduced), for example, we can consider the intrinsic geometry of the manifold, and constrained problems can be seen as unconstrained ones, or, certain non convex functions in the usual sense become convex or quasiconvex on the Riemannian manifold, so we can use more efficient optimization techniques, see [13, 17, 20, 25, 32, 35, 37], and references therein. Another motivation is the relation between Riemannian metrics and interior point methods. Indeed, it has been proved, see Nesterov and Todd [27], that the continuous path of each polynomial primal-dual interior point method is near to the geodesic of the Riemannian manifold induced by the Hessian of self-concordant barrier. Then, we can interpret that the trajectory of the primal-dual method is a good approximation of the geodesic between the initial point and an ϵ-solution of the problem. Furthermore, we can use Riemannian metrics to introduce new algorithms in interior point methods, see [15, 28, 31, 33].

When $X = M$ in (1.1) (the unconstrained minimization problem) the proximal point method to solve this problem generates a sequence $\{x^k\}$ given by $x^0 \in M$, and

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

where λ_k is certain positive parameter and d is the Riemannian distance in M. Observe that, in particular, when $M = I\!R^n$ we obtain the classic proximal method introduced by Martinet [26] and further developed by Rockafellar [38] (in a general framework): $x^k \in \texttt{arg}\min\{f(x) + (\lambda_k/2)||x - x^{k-1}||^2 : x \in I\!R^n\}$, where $||.||$ is the Euclidian norm, i.e, $||x||^2 = \sqrt{\langle x, x\rangle}$. It is well known, see Ferreira and Oliveira [19], that if f is convex in (1.2) and $\{\lambda_k\}$ satisfies $\sum_{k=1}^{n}(1/\lambda_k) = +\infty$, then $\lim_{k\to\infty} f(x^k) = \inf\{f(x) : x \in M\}$. Furthermore, if the optimal set is nonempty we obtain that $\{x^k\}$ converges to an optimal solution of the problem.

When $X \subseteq M$ in (1.1), the proximal point method with Semi-Bregman and Bregman distances to solve (1.1), generates a sequence $\{y^k\}$ such that $y^0 \in \text{int}X \cap S$ (int$X$ denotes the interior of X and S is an open convex set) and

$$y^k \in \texttt{arg}\min\{f(y) + \lambda_k D_h(y, y^{k-1}) : y \in X \cap \bar{S}\}, \tag{1.3}$$

where $D_h(x, y) := h(x) - h(y) - \langle \texttt{grad}h(y), \exp_y^{-1} x\rangle_y$, h is a Semi-Bregman or Bregman function in S, respectively (see Section 5.1 for a formal definition of those classes of functions), $\texttt{grad}h$ is the gradient vector field of h, $\exp^{-1}$ is the inverse of the exponential function exp and λ_k is a positive parameter. Observe that if $M = I\!R^n$ we obtain that $\exp_y^{-1} x = x - y$ and $\texttt{grad}h = \nabla h$ (classic gradient of h) and so $D_h(x, y) = h(x) - h(y) - \langle \nabla h(y), x - y\rangle$, where $\langle , \rangle$ is the Euclidian inner product, furthermore, if h is a Bregman function we obtain the classic proximal method with Bregman distances in $I\!R^n$.

In this work we treat (1.1) with functions with Lojasiewicz, quasiconvex and convex properties. Our motivation to study minimization problems with Lojasiewicz features becomes from the fact that this class of functions has recently received much attention in optimization, see [1, 6, 7, 8] and references therein. The motivation to study minimization problems with quasiconvex objective functions comes from two fields. One of them, is the broad range of applications of quasiconvex optimizations in diverse areas of sciences and engineering, for example, in location theory [23], control theory [5] and economic theory [36]. The other motivation, is to solve more general nonconvex constrained optimization problems. Now, the Riemannian techniques can transform those problems in quasiconvex ones, under appropriate metrics on the manifolds. We believe that this kind of application could permit more examples, if we compare with the convex case. Finally, our motivation to work with convex objective functions is to generalize known methods, besides to extend the applications of the convex framework in Riemannian manifolds to nonconvex functions in Euclidian spaces.

Proximal methods for solving minimization problems within Lojasiewicz features in Euclidian spaces was studied by Attouch and Bolte in [3]. Smooth quasiconvex problems are considered by Attouch and Teboulle [4], and Godou and Munier [22] with a dynamical method in Euclidian and Hilbert spaces. For the proximal method with $\phi-$divergence, see Cunha et al. [11], with a class of separated Bregman distances, see Souza et al. [40], and, finally, Quiroz and Oliveira, [29] with the classical and logarithmic quadratic proximal methods. When f is nonsmooth but continuous and quasiconvex on the manifold M we have proved in [30], for $X = M$ in (1.1), the convergence of the proximal point methods with Riemannian and Bregman distances to a minimal point when $\lim_{k\to 0} \lambda_k = 0$. Observe that the above condition is very restrictive. In fact, suppose that f in (1.1) is ill conditioned, then for very small λ_k, the regularize function in (1.2) and (1.3) (respectively) will be numerically almost ill behaved as f. Furthermore, the case $X \neq M$ was not considered.

The main result of this paper is the extension of the convergence properties of the proximal methods with Riemannian and Bregman distances to solve minimization problems for objective functions with Lojasiewicz, quasiconvex and convex properties on Hadamard manifolds as well, the introduction of a proximal method with Semi-Bregman distances, that permits recover some important optimization methods, for example, proximal methods for semidefinite optimization and second order cone program studied by Doljanski and Teboulle [16] and Chen [10], respectively.

The paper is divided as follows. In Section 2 we give some results of Fejér convergence theory and convex analysis on Hadamard manifolds. In section 3, some results of Fréchet subdifferential and limiting subdifferential on Hadamard manifolds are presented. In Section 4, we analyze the proximal point method with Riemannian distances for solving unconstrained optimization problems on Hadamard manifolds for objective Lojasiewicz and quasiconvex functions. Then, assuming continuity of the objective function and nonemptyness of the global optimal minimizer set, we prove the full convergence of this method to a certain limiting critical point of the problem. It is also given, for the quasiconvex case, a localization of the limit point with respect to the initial date. In Section 5, we consider the constrained optimization problem for convex and quasiconvex objective functions. In the first subsection we introduce the definitions of Semi-Bregman and Bregman functions for an arbitrary open convex domain and give some properties. In subsection 5.2 we give the Féjer convergence theory for Semi-Bregman and Bregman distances. In subsection 5.3 we study existence properties for a regularized function with Semi-Bregman and Bregman distances. In subsection 5.4 we present the proximal point method with Semi-Bregman distances and present their convergence properties. We show that this approach recovers, in particular, proximal methods for SDP and SOCP. In subsection 5.5

we study proximal point methods with Riemannian distances and prove their convergence. This approach recovers in particular, convex optimization problems in $I\!R^n$, constrained minimization problems with quasiconvex objective functions, as well, constrained problems with convex and quasiconvex objective functions. Finally, in Section 6 is given some conclusions and future research.

2 Some Basic Facts on Metric Spaces and Hadamard Manifolds

In this section we recall some fundamental properties and notation on Fejér convergence in metric spaces and convex analysis on Hadamard manifolds.

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

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

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

Proof. See, for example, Lemma 6.1 of Ferrreira and Oliveira [19]. ∎

Let M be a differential manifold of finite dimension n. We denote by T_xM the tangent space of M at x and $TM = \bigcup_{x\in M} T_xM$. T_xM is a linear space and has the same dimension of M. Because we restrict ourselves to real manifolds, T_xM is isomorphic to $I\!R^n$. If M is endowed with a Riemannian metric g, then M is a Riemannian manifold and we denote it by (M, G) or only by M when no confusion can arise, where G denotes the matrix representation of the metric g. The inner product of two vectors $u, v \in T_xS$ is written $\langle u, v\rangle_x := g_x(u, v)$, where g_x is the metric at the point x. The norm of a vector $v \in T_xS$ is defined by $||v||_x := \langle v, v\rangle_x^{1/2}$. If there is not confusion we denote $\langle , \rangle = \langle , \rangle_x$ and $||.|| = ||.||_x$. The metric can be used to define the length of a piecewise smooth curve $\alpha : [t_0, t_1] \to S$ 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 (a vector field V is an application of M in TM), the covariant derivative of W in the direction V is denoted by $\nabla_V W$. In this paper ∇ is the Levi-Civita connection associated to (M, G). This connection defines an unique covariant derivative D/dt, where for each vector field V, along a smooth curve $\alpha : [t_0, t_1] \to M$, another vector field is obtained, denoted by DV/dt. The parallel transport along α from $\alpha(t_0)$ to $\alpha(t_1)$, denoted by P_{α,t_0,t_1}, is an application $P_{\alpha,t_0,t_1} : T_{\alpha(t_0)}M \to T_{\alpha(t_1)}M$ defined by $P_{\alpha,t_0,t_1}(v) = V(t_1)$ where V is the unique vector field along α such that $DV/dt = 0$ and $V(t_0) = v$. Since that ∇ is a Riemannian connection, P_{α,t_0,t_1} is a linear isometry, furthermore $P^{-1}_{\alpha,t_0,t_1} = P_{\alpha,t_1,t_0}$ and $P_{\alpha,t_0,t_1} = P_{\alpha,t,t_1} P_{\alpha,t_0,t}$, for all $t \in [t_0, t_1]$. A curve $\gamma : I \to M$ is called a geodesic if $D\gamma'/dt = 0$. A Riemannian manifold is complete if its geodesics are defined for any value of $t \in I\!R$. Let $x \in M$, the exponential map $\exp_x : T_xM \to M$ is defined as $\exp_x(v) = \gamma(1)$. If M is complete, then $\exp_x$ is defined for all $v \in T_xM$. Besides, there is a minimal geodesic (its length is equal to the distance between the extremes).

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

Now, the sectional curvature with respect to X and Y is defined by

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

The complete simply connected Riemannian manifolds with non positive curvature are denominated *Hadamard manifolds*.

Theorem 2.2 *Let M be a Hadamard manifold. Then M is diffeomorphic to the Euclidian space $I\!R^n$, $n = \dim M$. More precisely, at any point $x \in M$, the exponential mapping $\exp_x : T_xM \to M$ is a global diffeomorphism.*

Proof. See [34], Theorem 4.1, page 221. ■

A consequence of the preceding theorem is that Hadamard manifolds have the property of uniqueness of geodesic between any two points. Another useful property is the following: let $[x, y, z]$ 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:*

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

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

Proof. See [34], Proposition 4.5, page 223. ■

The gradient of a differentiable function $f : M \to I\!R$, $\mathtt{grad} f$, is a vector field on M defined through $df(X) = \langle \mathtt{grad} f, X\rangle = X(f)$, where X is also a vector field on M. The Hessian of a twice differentiable function f at x with direction $v \in T_xM$ is given by $H_x^f(v) = \frac{D}{dt}(\mathtt{grad}\, f)(x) = \nabla_v \mathtt{grad}\, f\,(x)$.

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

Definition 2.3 *Let C be a nonempty set of M and $y \in M$. A point $P_Cy \in C$ for which*

$$d(P_Cy, y) = \min_{x\in C} d(x,y)$$

is called a projection *of the point y on the set C.*

Theorem 2.4 *Let C be a nonempty closed convex set in a Hadamard manifold M. Take $y \in M$ arbitrary, then there exists a unique projection $z = P_Cy$. Furthermore, for all $x \in C$ the following inequality holds*

$$\langle \exp_z^{-1} y, \exp_z^{-1} x\rangle \leq 0.$$

Proof. See [19], Propositions 3.1 and 3.2. ■

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

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

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

where $\gamma : [0, 1] \to M$ is the geodesic curve such 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 said to be strictly convex.

Theorem 2.5 *Let $f : M \to I\!R \cup \{+\infty\}$ with* **domf** *convex in a Hadamard manifold M. The function $f : M \to I\!R$ is convex in M if and only if $\forall\, x, y \in M$ and $\gamma : [0, 1] \to M$ (the geodesic joining x to y) the function $f(\gamma(t))$ is convex in $[0, 1]$.*

Proof. See [37], page 61, Theorem 2.2. ■

A function $f : M \to I\!R \cup \{+\infty\}$ with `domf` convex is called concave if $-f$ is convex. Furthermore, if f is both convex and concave then f is said to be linear affine in M. Observe that a twice differentiable function f in M is linear affine if and only if $\langle H^f_x(v), v\rangle_x = 0$, for all $x \in M$ and $v \in T_xM$.

Theorem 2.6 *Let M be a Hadamard manifold and let y fixed. Then the function $g(x) = d^2(x, y)$ is strictly convex and* $\textbf{grad}g(x) = -2 \exp^{-1}_x y$.

Proof. See [18], Proposition II.8.3. ■

Proposition 2.1 *Let M be a Hadamard manifold and $h : M \to I\!R$ a differentiable function. Let $y \in M$ and define $g(x) = \langle \textbf{grad}h(y), \exp^{-1}_y x\rangle_y$. Then the following statements are true:*

i. $\textbf{grad}\, g(x) = P_{\gamma,0,1}\textbf{grad}\, h(y)$, *where $\gamma : [0, 1] \to M$ is the geodesic curve such that $\gamma(0) = y$ and $\gamma(1) = x$.*

ii. *g is an affine linear function in M.*

Proof. See [30], Proposition 3.1. ■

Definition 2.5 *Let $f : M \to I\!R \cup \{+\infty\}$ be a proper function with* **domf** *convex in a Hadamard manifold M. f is called quasiconvex on M if for all $x, y \in M$, $t \in [0, 1]$, it holds that*

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

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

Theorem 2.7 *Let $f : M \to I\!R \cup \{+\infty\}$ be a proper function with* **domf** *convex in a Hadamard manifold M. f is quasiconvex in M if and only if the set $\{x \in M : f(x) \leq c\}$ is convex for each $c \in I\!R$.*

Proof. See [37], page 98, Theorem 10.2. ■

Definition 2.6 *Let $f : M \to I\!R \cup \{+\infty\}$ be a proper and differentiable function on the open convex set* **domf** *. f is called pseudoconvex if, for every pair of distinct points $x, y \in$* **domf** *we have*

$$\langle \textbf{grad}f(x), \exp^{-1}_x y\rangle \geq 0, \quad then \quad f(y) \geq f(x).$$

Theorem 2.8 *Let $f : M \to I\!R \cup \{+\infty\}$ be a proper and differentiable pseudoconvex function on the open convex set* **domf**. *If $\textbf{grad}f(x^*) = 0$, then x^* is a global minimum of f.*

Proof. Immediate. ■

3 Fréchet and Limiting Subdifferential on Hadamard Manifolds

In this section, we extend some definitions and results of Fréchet and limiting subdifferential from Euclidian spaces to Hadamard manifolds. Our results are motivated by [39].

Definition 3.1 *Let $f : M \to I\!R \cup \{+\infty\}$ be a proper function on a Hadamard manifold M. For each $x \in \texttt{dom}f$, the Fréchet subdifferential of f at x, denoted by $\widehat{\partial} f(x)$, is the set of vectors $s \in T_xM$ such that*

$$\liminf_{y \neq x, y \to x} \frac{1}{d(x,y)}[f(y) - f(x) - \langle s, \exp_x^{-1} y\rangle] \geq 0.$$

If $x \notin \texttt{dom}f$ then $\widehat{\partial} f(x) = \phi$.

Observe that the above definition is equivalent to

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

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

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

Definition 3.2 *Let $f : M \to I\!R \cup \{+\infty\}$ be a proper function on a Hadamard manifold M. The limiting subdifferential of f at $x \in M$, denoted by $\partial f(x)$, is defined as follows*

$$\partial f(x) := \{s \in T_xM : \exists x^n \to x, f(x^n) \to f(x), s^n \in \widehat{\partial} f(x^n) : P_{\gamma_n,0,1} s^n \to s\},$$

where $P_{\gamma_n,0,1}$ is the parallel transport along the geodesic γ_n such that $\gamma_n(0) = x^n$ and $\gamma_n(1) = x$.

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 $\widehat{\partial} f(\bar{x}) = \{\texttt{grad}f(\bar{x})\}$, so $\texttt{grad}f(\bar{x}) \in \partial f(\bar{x})$;*

c. *If f is differentiable in a neighborhood of x, then $\widehat{\partial} f(x) = \partial f(x) = \{\texttt{grad}f(x)\}$;*

d. *If $g = f + h$ with f finite at $\bar{x}$ and h is differentiable on a neighborhood of $\bar{x}$ then $\widehat{\partial} g(\bar{x}) = \widehat{\partial} f(\bar{x}) + \texttt{grad}h(\bar{x})$, and $\partial g(\bar{x}) = \partial f(\bar{x}) + \texttt{grad}h(\bar{x})$.*

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

$$\exists x^n \to x, f(x^n) \to f(x), s^n \in \widehat{\partial} f(x^n) : P_{\gamma_n,0,1} s^n \to s,$$

where $P_{\gamma_n,0,1}$ is the parallel transport along the geodesic γ_n such that $\gamma_n(0) = x^n$ and $\gamma_n(1) = x$. From item b, we have that $s^n = \texttt{grad}f(x^n)$, for n sufficient large. From the continuity of $\texttt{grad}f$ and the Parallel transport P, and using uniqueness of the limit point we have that $s = \texttt{grad}f(\bar{x})$.
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 generalized definition.

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

Definition 3.4 *Let X be a convex set of M and $f : X \to I\!R$ a real function. A point $x \in X$ is called limiting critical point of f on X if $0 \in \partial(f + I_X)(x)$, where $I_X(x) = 0$ if $x \in X$ and $I_X(x) = +\infty$, otherwise.*

Theorem 3.1 *If a proper function $g : M \to I\!R \cup \{+\infty\}$ 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 ∎

Theorem 3.2 *If a function $g : X \to I\!R$ has a local minimum at $\bar{x} \in X$ then $0 \in \widehat{\partial}(g + I_X)(\bar{x})$ and therefore,*

$$0 \in \partial(g + I_X)(\bar{x}).$$

Proof. Immediate ∎

Definition 3.5 *Let M be a Hadamard manifold and $C \subset M$ a subset of M. Given $\bar{x} \in C$, a vector $v \in T_{\bar{x}}M$ is called a normal to C at $\bar{x}$, in the regular sense, denoted by $v \in \widehat{N}_C(\bar{x})$, if*

$$\langle v, \exp_{\bar{x}}^{-1} x \rangle \leq o(d(\bar{x}, x)), \quad \forall x \in C \tag{3.4}$$

where $o(d(\bar{x}, x))$ denotes a term with the property that $\lim_{x \neq \bar{x}, d(x,\bar{x}) \to 0} \frac{o(d(\bar{x},x))}{d(\bar{x},x)} = 0$.
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 $\{x^n\} \subset C$ and $\{v^n\} \subset \widehat{N}_C(x^n)$ such that $x^n \to \bar{x}$ and $P_{\gamma_n,0,1} v^n \to v$, where $P_{\gamma_n,0,1}$ is the parallel transport along the geodesic γ_n such that $\gamma_n(0) = x^n$ and $\gamma_n(1) = \bar{x}$.

Observe that the inequality (3.4) is equivalent to

$$\limsup_{x \neq \bar{x}, d(x,\bar{x}) \to 0} \frac{\langle v, \exp_{\bar{x}}^{-1} x \rangle}{d(x, \bar{x})} \leq 0. \tag{3.5}$$

Definition 3.6 *Let M be a Hadamard manifold and $C \subset M$ a subset of M. Given $\bar{x} \in C$, a vector $w \in T_{\bar{x}}M$ is tangent to C at $\bar{x}$, denoted by $w \in T_C(\bar{x})$, if there exist sequences $\{x^n\} \subset C$, $x^n \neq \bar{x}$, and $\{\tau^n\} \subset I\!R$ with $x^n \to \bar{x}$ and $\tau^n \to 0$ with $\tau^n > 0$ such that*

$$\lim_{n \to +\infty} \frac{\exp_{\bar{x}}^{-1} x^n}{\tau^n} = w.$$

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

Proposition 3.2 *Let M be a Hadamard manifold and $C \subset M$ a subset of 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 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 that both set are closed cones as well that $\widehat{N}_C(\bar{x})$ is convex are obtained using elementary analysis. We prove the characterization property. Let $w \in T_C(\bar{x})$, arbitrary. From Definition 3.6 there exist sequences $\{x^n\} \subset C$, $x^n \neq \bar{x}$, and $\{\tau^n\} \subset I\!R$ with $x^n \to \bar{x}$ and $\tau^n \searrow 0$ such that $\lim_{n\to+\infty} \frac{\exp_{\bar{x}}^{-1} x^n}{\tau^n} = w$. Defining $w^n = \frac{\exp_{\bar{x}}^{-1} x^n}{\tau^n}$ and using that $v \in \widehat{N}_C(\bar{x})$ we have

$$\langle v, w^n \rangle \leq \frac{o(||\tau^n w^n||)}{\tau^n}.$$

Taking $n \to +\infty$ 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 (3.5), for any $\delta > 0$

$$\sup_{x \in B(\bar{x},\delta), x \neq \bar{x}} \left(\frac{\langle v, \exp_{\bar{x}}^{-1} x \rangle}{d(x, \bar{x})} \right) \geq m > 0,$$

where $m := \limsup_{x \neq \bar{x}, d(x,\bar{x}) \to 0} \frac{\langle v, \exp_{\bar{x}}^{-1} x \rangle}{d(x,\bar{x})}$. Take $\delta = 1/n$ and using the supreme property, there exists $x^n \in B(\bar{x}, 1/n)$ with $x^n \neq \bar{x}$ such that

$$\left\langle v, \frac{\exp_{\bar{x}}^{-1} x^n}{d(x^n, \bar{x})} \right\rangle \geq m > 0.$$

Defining $w^n := \frac{\exp_{\bar{x}}^{-1} x^n}{d(x^n,\bar{x})}$ and taking lim inf we obtain

$$\liminf_{n \to +\infty} \langle v, w^n \rangle > 0.$$

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

Theorem 3.3 *Let M be a Hadamard manifold and C a subset of M. If C is convex then*

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

$$N_C(\bar{x}) = \widehat{N}_C(\bar{x}) = \{v : \langle v, \exp_{\bar{x}}^{-1} x \rangle \leq 0, \forall x \in C\}.$$

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 geodesic segment joining the point $\bar{x}$ and any of their points (See Definition 2.2), so the vectors of K are the multiple by a positive scalar of the vectors $\exp_{\bar{x}}^{-1} x$ for each $x \in C$. 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 $x^n = \exp_{\bar{x}}(\tau^n w)$ with $\tau^n \in (0, \lambda]$ and $\tau^n \to 0$), so we prove the second inclusion. Let $w \in T_C(\bar{x})$, then, from Definition 3.6, there exist sequences $x^n \to \bar{x}$, $x^n \in C$, $x^n \neq \bar{x}$ and $\tau^n \searrow 0$ such that $\lim_{n\to+\infty} \frac{\exp_{\bar{x}}^{-1} x^n}{\tau^n} = w$. Define $w^n := \frac{\exp_{\bar{x}}^{-1} x^n}{\tau^n}$, then from the convexity of C, there exists $\lambda > 0$ such that $\exp(\lambda w^n) \in C$. Besides, as $w^n \to w$, we get $w \in clK(\bar{x})$.
Now, as $T_C(\bar{x})$ is closed (See Proposition 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, $\{v : \langle v, \exp_{\bar{x}}^{-1} x\rangle \leq 0, \forall x \in C\} \subset \widehat{N}_C(\bar{x}) \subset N_C(\bar{x})$, then it is sufficient to prove that

$$N_C(\bar{x}) \subset \{v : \langle v, \exp_{\bar{x}}^{-1} x\rangle \leq 0, \forall x \in C\}.$$

Let $v \in N_C(\bar{x})$ then, there exist $x^n \to \bar{x}$ with $x^n \in C$ and $P_{\gamma_n,0,1}v^n \to v$, with $v^n \in \widehat{N}_C(x^n)$, where $P_{\gamma_n,0,1}$ is the parallel transport along the geodesic γ_n such that $\gamma_n(0) = x^n$ and $\gamma_n(1) = \bar{x}$. As $v^n \in \widehat{N}_C(x^n)$, then from Proposition 3.2, it holds the characterization property,

$$\langle v^n, w\rangle \leq 0, \forall w \in T_C(x^n) = clK(x^n)$$

where $K(x^n) := \{w : \text{ there exists } \lambda > 0 \text{ with } \exp_{x^n}(\lambda w) \in C\}$, the above equality being a consequence of the previous result. Let $x \in C$, arbitrary, then $w = \exp_{x^n}^{-1} \in K(x^n)$, so using this fact in the above inequality we obtain $\langle v^n, \exp_{x^n}^{-1} x\rangle \leq 0$. Now, taking parallel transport and $n \to +\infty$ we obtain that $\langle v, \exp_{\bar{x}}^{-1} x\rangle \leq 0$. Therefore the result is obtained. ■

We define the difference quotient function as a function $\Delta_\tau f(x) : T_xM \to I\!R \cup \{+\infty\}$ where

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

Definition 3.7 *For a function $f : M \to I\!R \cup \{+\infty\}$ and a point $\bar{x} \in M$ with $f(\bar{x})$ finite, the Subderivative function $df(\bar{x}) : T_{\bar{x}}M \to I\!R \cup \{+\infty\}$ is defined by*

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

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 lim inf when $\tau \searrow 0$ and $w' \to w$ we obtain $df(\bar{x})(w) \geq \langle v, w\rangle$.
Now, we get the inclusion $\supset$. Let $v \in T_{\bar{x}}M$ such 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

$$S := \liminf_{y \neq \bar{x}, y \to \bar{x}} \frac{1}{d(\bar{x}, y)}[f(y) - f(\bar{x}) - \langle v, \exp_{\bar{x}}^{-1} y\rangle] < 0.$$

From supreme and infimum properties, given $\delta = 1/n$ there exists $y^n \in B(\bar{x}, 1/n)$ such that

$$\frac{1}{d(\bar{x}, y^n)}[f(y^n) - f(\bar{x}) - \langle v, \exp_{\bar{x}}^{-1} y^n\rangle] \leq S.$$

Define $w^n := \exp_{\bar{x}}^{-1} y^n/\tau^n$ where $\tau^n := d(\bar{x}, y^n)$, then gives

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

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

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

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

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

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

Remark 3.1 *Observe that*

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

Proposition 3.4

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

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

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

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

$$\Delta_{\tau^n} f(\bar{x})(w^n) \le \beta.$$

From the Remark 3.1 this implies that $(\exp_{\bar{x}}(\tau^n w^n), \beta\tau^n + f(\bar{x})) \in \mathtt{epi} f$. Defining $z^n = \exp_{\bar{x}}(\tau^n w^n)$ and $\alpha^n = \beta\tau^n + f(\bar{x})$ we obtain that $(z^n \alpha^n) \in \mathtt{epi} f$ with $(z^n \alpha^n) \to (\bar{x}, f(\bar{x}))$, $\tau^n \searrow 0$ and

$$\lim_{n\to+\infty} \frac{\exp^{-1}_{(\bar{x},f(\bar{x}))}(z^n,\alpha^n)}{\tau^n} = \lim_{n\to+\infty} \left(\frac{\exp^{-1}_{\bar{x}} x^n}{\tau^n}, \frac{\alpha^n - f(\bar{x})}{\tau^n} \right) = \lim_{n\to+\infty} (w^n,\beta) = (w,\beta).$$

From the Definition 3.6 the result is obtained.
Now, we prove the inclusion $\supset$. Let $(w,\beta) \in T_{\mathtt{epi} f}(\bar{x}, f(\bar{x}))$ then there exist $(z^n,\alpha_n) \in \mathtt{epi} f$ such that $(z^n,\alpha_n) \to (\bar{x}, f(\bar{x}))$, $\tau^n \searrow 0$ and

$$\lim_{n\to+\infty} \frac{\exp^{-1}_{(\bar{x},f(\bar{x}))}(z^n,\alpha^n)}{\tau^n} = (w,\beta).$$

From the above equality we obtain that $w^n := \frac{\exp^{-1}_{\bar{x}} z^n}{\tau^n} \to w$ and $\beta^n := \frac{\alpha_n - f(\bar{x})}{\tau^n} \to \beta$. Now, as $(z^n,\alpha_n) \in \mathtt{epi} f$ then

$$\Delta_{\tau^n} f(\bar{x})(w^n) \le \frac{\alpha_n - f(\bar{x})}{\tau^n}.$$

Taking $\liminf$ when $n \to +\infty$ we obtain

$$\liminf_{n\to+\infty} \Delta_\tau f(\bar{x}) \le \beta.$$

So, $df(\bar{x})(w) \le \beta$, and therefore $(w,\beta) \in \mathtt{epi} df(\bar{x})$. ■

Theorem 3.4

$$\widehat{\partial} f(\bar{x}) = \{v \in T_{\bar{x}}M : (v,-1) \in \widehat{N}_{\textbf{\textit{epi}} f}(\bar{x}, f(\bar{x}))\}$$

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

Theorem 3.5

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

Proof. Let $v \in \partial f(\bar{x})$, then there exist $x^n \to \bar{x}, f(x^n) \to f(\bar{x}), v^n \in \widehat{\partial} f(x^n) : P_{\gamma_n,0,1}v^n \to v$ where $P_{\gamma_n,0,1}$ is the parallel transport along the geodesic γ_n such that $\gamma_n(0) = x^n$ and $\gamma_n(1) = \bar{x}$. As $v^n \in \widehat{\partial} f(x^n)$ then, from the previous theorem, $(v^n, -1) \in \widehat{N}_{\texttt{epi}f}(x^n, f(x^n))$. Now, define the following sequences: $z^n = (x^n, f(x^n))$ and $p^n = (v^n, -1)$. Applying this definition we obtain that there exist $z^n \in \texttt{epi}f$ and $p^n \in \widehat{N}_{\texttt{epi}f}(x^n, f(x^n))$ such that $z^n \to (\bar{x}, f(\bar{x}))$ and $(P_{\gamma_n,0,1}v^n, -1) \to (v, -1)$. Thus, from the Definition 3.5 we conclude that $(v, -1) \in N_{\texttt{epi}f}(\bar{x}, f(\bar{x}))$. ■

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

$$\partial_F f(\bar{x}) := \{v \in T_{\bar{x}}M : f(x) \geq f(\bar{x}) + \langle v, \exp_{\bar{x}}^{-1} x, \forall x \in M\rangle\}$$

Theorem 3.6 *If $f : M \to I\!R \cup \{+\infty\}$ 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. Tehn 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 3.5, $(v, -1) \in N_{\texttt{epi}f}(\bar{x}, f(\bar{x}))$. As $\texttt{epi}f$ is convex, because f is convex, we have from Theorem 3.3 that

$$\langle (v, -1), \exp^{-1}_{(\bar{x}, f(\bar{x}))}(x, r)\rangle \leq 0, \forall (x, r) \in \texttt{epi}f.$$

Taking $r = f(x)$ we obtain $\langle v, \exp_{\bar{x}}^{-1} x\rangle - (f(x) - f(\bar{x})) \leq 0$. This implies that

$$f(x) \geq f(\bar{x}) + \langle v, \exp_{\bar{x}}^{-1} x\rangle,$$

and therefore, $v \in \partial f_F(\bar{x})$. ■

4 Proximal Method for Unconstrained Minimization

Consider the problem:

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

where $f : M \to I\!R \cup \{+\infty\}$ is a proper function on a Hadamard manifold M. The proximal point method to solve (p) generates a sequence $\{x^k\}$ given by

$$x^0 \in M, \tag{4.6}$$

$$x^k \in \texttt{arg}\min\{f(x) + (\lambda_k/2)d^2(x, x^{k-1}) : x \in M\} \tag{4.7}$$

where λ_k is some positive parameter. We assume the following assumptions.

Assumption A. $f : M \to I\!R \cup \{+\infty\}$ is bounded below.

Assumption B. $f : M \to I\!R \cup \{+\infty\}$ is lower semicontinuous on $\texttt{dom}f$.

Theorem 4.1 *Assume assumptions A and B. Then, the sequence $\{x^k\}$, generated by the proximal algorithm, is well defined (the solution of (4.7) exists) and the set of points that satisfies (4.7) is compact.*

Proof. Immediate ■

Lemma 4.1 *The iterates $\{x^k\}$ do not cycle.*

Proof. Consider two cases: contiguous and non contiguous iterations. Clearly if $x^l = x^{l-1}$ for some $l > 0$ we have from (4.7) that $0 \in \partial f(x^l)$. Therefore, x^l is a critical point and the method stops in finitely many iterations. Otherwise, let $l > j+1$ such that $x^l = x^j$ and $x^k \neq x^{k-1}$, for all k. From (4.7), the minimum condition implies that, for all k,

$$f(x^k) + (\lambda_k/2)d^2(x^k, x^{k-1}) \leq f(x^{k-1}). \tag{4.8}$$

As $x^{k-1} \neq x^k$ we have from (4.8) that $f(x^k) < f(x^{k-1})$ ($\{f(x^k)\}$ is a strictly nonincreasing sequence). Now, this implies that $f(x^l) < f(x^{l-1}) < ... < f(x^j)$, which contradicts that $f(x^l) = f(x^j)$. ■

Proposition 4.1 *Under assumptions A and B, the following facts are true*

a. $\{f(x^k)\}$ *is nonincreasing.*

b. $\sum\limits_{k=1}^{\infty} \lambda_k d^2(x^k, x^{k-1}) < +\infty$

Proof.
a. Showed in the preceding Proposition.
b. From inequality (4.8) and Assumption A we obtain $\sum\limits_{k=1}^{n} (\lambda_k/2)d^2(x^k, x^{k-1}) \leq f(x^0) - \alpha$, where α is the lower bounded of f. Taking $n \to \infty$ we obtain the result. ■

Theorem 4.2 *Suppose that assumption A is satisfied and that f is continuous on* $\mathtt{dom}f$*. Let* $\lim_{k\to\infty} d(x^k, x^{k-1}) = 0$ *and* $0 < \lambda_k < \bar{\lambda}$*, for some* $\bar{\lambda}$*. Then any cluster point of the sequence* $\{x^k\}$ *is a limiting critical point of f.*

Proof. Suppose that there exist $\bar{x}$ and a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ converging to $\bar{x}$. We will prove that $0 \in \partial f(\bar{x})$. From (4.7) and Theorem 3.1 we have $0 \in \hat{\partial}\left(f(.) + (\lambda_{k_j}/2)d^2(., x^{k_j-1})\right)(x^{k_j})$ From the smoothness of $(\lambda_k/2)d^2(., x^{k-1})$, Proposition 3.1, **d**, and Theorem 2.6 this imply that

$$g^{k_j} := \lambda_{k_j} \exp^{-1}_{x^{k_j}} x^{k_j-1} \in \hat{\partial} f(x^{k_j}).$$

Using the fact that the parallel transport is an isometry and using the boundedness of λ_k we have that

$$||P_{\gamma_{k_j},0,1} g^{k_j}|| = ||g^{k_j}|| \leq \bar{\lambda}|| \exp^{-1}_{x^{k_j}} x^{k_j-1}|| = \bar{\lambda} d(x^{k_j}, x^{k_j-1}),$$

where γ_{k_j} is the geodesic curve such that $\gamma_{k_j}(0) = x^{k_j}$ and $\gamma_{k_j}(1) = \bar{x}$. Now, using the fact that $\lim_{k\to\infty} d(x^k, x^{k-1}) = 0$, we have $\lim_{j\to\infty} P_{\gamma_{k_j},0,1} g^{k_j} = 0$. Therefore, there are sequences $\{x^{k_j}\}$, $\{f(x^{k_j})\}$ and $\{g^{k_j}\}$ with $g^{k_j} \in \hat{\partial} f(x^{k_j})$ such that $\lim_{j\to+\infty} x^{k_j} = \bar{x}$, $\lim_{j\to\infty} f(x^{k_j}) = f(\bar{x})$ (from the continuity of f) and $\lim_{j\to\infty} P_{\gamma_{k_j},0,1} g^{k_j} = 0$. From Definition 3.2 it follows that $0 \in \partial f(\bar{x})$. ■

4.1 The Generalized Lojasiewicz Case

Functions that satisfies a Lojasiewicz property has recently received much attention in optimization see for example [1, 3, 6, 7, 8]. In this subsection we generalize to Hadamard manifolds the result convergence in $\mathbb{R}^n$ of [3], where it has been proved the convergence of the method for functions in $\mathbb{R}^n$ satisfying the Lojasiewicz property.

In this subsection, we let $0 < \lambda' < \bar{\lambda}$ fixed, and utilize the following condition for the parameters λ_k :

$$\lambda' < \lambda_k < \bar{\lambda}, \tag{4.9}$$

Proposition 4.2 *Let $\{x^k\}$ the sequence generated by the proximal point algorithm, with λ_k satisfying (4.9). Under assumptions A and the continuity of f, the following facts are true*

a. $\sum_{k=1}^{\infty} d^2(x^k, x^{k-1}) < +\infty$

b. *Any cluster point of $\{x^k\}$ is a generalized critical point of f.*

In addition, if $\{x^k\}$ is bounded then

c. *f is finite and constant in $W(x^0)$, where $W(x^0)$ denotes the set of cluster points of $\{x^k\}$.*

d. *$W(x^0)$ is connected, compact and*

$$d(x^k, W(x^0)) \to 0, k \to +\infty.$$

Proof.
a. It is immediate from Proposition 4.1, item b, and the condition $\lambda' < \lambda_k$.
b. From item a, we obtain that $\lim_{k\to+\infty} d(x^k, x^{k-1}) = 0$. Now, using this in Theorem 4.2, we obtain the result.
c. The sequence $\{f(x^k)\}$ is non increasing and bounded from below, then it converges to a point β. Now, let x' any cluster point of $\{x^k\}$, then there exists $\{x^{k_j}\}$ such that $\lim_{j\to+\infty} x^{k_j} = x'$. From the continuity of f we obtain $f(x') = \lim_{j\to+\infty} f(x^{k_j}) = \beta$. As x' is arbitrary the proof is concluded.
d. The connectedness and compactness are trivial, so we only prove that $d(x^k, W(x^0)) \to 0, k \to +\infty$. In fact, by contradiction suppose that there exist $\beta > 0$ and $n_0 \in \mathbb{N}$ such that:

$$d(x^k, W(x^0)) \geq 2\beta > 0, \quad \forall k \geq n_0.$$

As $\{x^k\}$ is bounded, then there exist $\bar{x} \in M$ and a subsequence, denoted by $\{x^{k_j}\}$, such that $\lim_{j\to+\infty} x^{k_j} = \bar{x}$, so $d(x^{k_j}, \bar{x}) \to 0$, when $j \to +\infty$. Therefore

$$d(x^{k_j}, W(x^0)) \to 0, j \to +\infty.$$

From the definition of convergent subsequence, there exists $n_1 := n_1(\beta) \in \mathbb{N}$ such that for all $j \geq n_1$ we have

$$d(x^{k_j}, W(x^0)) < \beta.$$

Now, consider $j \geq n_1$ such that $k_j \geq n_0$ (such j exists because $\{x^{k_j}\}$ is a subsequence of $\{x^k\}$) then

$$2\beta \leq d(x^{k_j}, W(x^0)) < \beta,$$

which is a contradiction. therefore we obtain the aimed result.

■

As was observed in the Euclidian case by Absil et.al [1], the convergence of the whole sequence $\{x^k\}$ may fail, so we introduce the following

Generalized Lojasiewicz Property

For each $\hat{x} \in \texttt{Crit}f$ (the set of the generalized critical points of f) there exist ϵ, C and $\theta \in [0, 1)$ such that

$$|f(x) - f(\hat{x})|^{\theta} \leq C \|x^*\|,$$

for all $x \in B(\hat{x}, \epsilon)$ and for all $x^* \in \partial f(x)$.

Assumption C. $f : M \to I\!R \cup \{+\infty\}$ satisfies the Generalized Lojasiewicz Property.

Examples of functions in $I\!R^n$ satisfying the Assumption C has been done by Attouch and Bolte [3]. Now, we introduce a Lemma that will be used to prove the convergence of $\{x^k\}$.

Lemma 4.2 *Assume that f verifies the Lojasiewicz property, then*

i. *If K is a nonempty connected set of* Crit$f$, *then f is constant on K.*

ii. *Furthermore, if K is compact, then there exist ϵ, C and $\theta \in [0, 1)$ such that for all $x \in M$ with $d(x, K) \leq \epsilon$,*

$$|f(x) - f(\hat{x})|^{\theta} \leq C \|x^*\|,$$

for all $x^* \in \partial f(x)$.

Proof. It is a trivial generalization from Euclidian spaces to Riemannian manifold setting, see Attouch and Bolte [3]. ■

Theorem 4.3 *Suppose that f is continuous on* dom$f$, *and assumptions A and C are satisfied. If $\{x^k\}$ is bounded, then*

$$\sum_{k=1}^{\infty} d(x^k, x^{k-1}) < +\infty.$$

In particular, $\{x^k\}$ converge to some generalized critical point of f.

Proof. As $\{f(x^k)\}$ is a nonincreasing and bounded below sequence then, it is convergent and furthermore

$$f(x^k) \to \inf f(x^k) \text{ as } k \to +\infty.$$

Define

$$g(x^k) = f(x^k) - \inf f(x^k)$$

Then we obtain that $g(x^k) \geq 0$ and

$$g(x^k) \to 0 \text{ as } k \to +\infty.$$

We analyze two cases:
If $x^l = x^{l-1}$ for some $l > 0$ then, from (4.7) we have that $0 \in \partial f(x^l)$ and the algorithm finishes in a finite number of iterations. On the other hand, consider $x^k \neq x^{k-1}$ for all k, then

$d(x^k, x^{k+1}) > 0$ so $\{g(x^k)\}$ is a positive strictly decreasing sequence.
Now, we define the function $h : I\!R^{++} \to I\!R$ such that $h(s) = -s^{1-\theta}$, $\theta \in [0, 1)$. This function is differentiable and strictly convex, so

$$h(s) - h(s_0) \geq h'(s_0)(s - s_0) \tag{4.10}$$

Taking $s_0 = g(x^k)$ and $s = g(x^{k+1})$ in (4.10) we have

$$g(x^k)^{1-\theta} - g(x^{k+1})^{1-\theta} \geq (1 - \theta) g(x^k)^{-\theta}(g(x^k) - g(x^{k+1})).$$

As $(\lambda_k/2)d^2(x^{k+1}, x^k) \leq g(x^k) - g(x^{k+1})$ then

$$g(x^k)^{1-\theta} - g(x^{k+1})^{1-\theta} \geq (\lambda_k/2)(1 - \theta) g(x^k)^{-\theta} d^2(x^{k+1}, x^k) \tag{4.11}$$

Using Lemma 4.2 for $K = W(x^0)$ we obtain that exist $\hat{x} \in W(x^0)$, $I\!N_0$, ϵ, C and $\theta \in [0, 1)$ such that for all $k \geq I\!N_0$,

$$\left| f(x^k) - f(\hat{x}) \right|^\theta \leq C||g^k||,$$

for all $x^* \in \partial f(x)$. As f is constant in K, we can substitute $f(\hat{x})$ by $f(\bar{x})$, where $\bar{x}$ is a limiting point of $\{x^k\}$. Thus $\forall k \geq I\!N_0$

$$\left| f(x^k) - f(\bar{x}) \right|^\theta \leq C||g^k||.$$

As $\{f(x^k)\}$ converge to $\inf f(x^k)$ we get $\forall k \geq I\!N_0$:

$$0 < g(x^k)^\theta \leq C\lambda_k d(x^k, x^{k-1}) \tag{4.12}$$

We observe that $\theta \neq 0$. Indeed, if $\theta = 0$ then $\forall k \geq I\!N_0 : 1 \leq C\lambda_k d(x^k, x^{k-1})$. Taking a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ such that $x^{k_j} \to \bar{x}$ we obtain that $1 \leq 0$. Therefore, $\theta \in (0, 1)$. From (4.12), also because $\theta \neq 0$, we obtain

$$d(x^k, x^{k-1})^{-1}/(C\lambda_k) \leq g(x^k)^{-\theta}$$

$\forall k \geq I\!N_0$.
Using the above inequality in (4.11) we obtain that there exists $M > 0$ such that $\forall k \geq I\!N_0$:

$$\frac{d(x^{k+1}, x^k)^2}{d(x^k, x^{k-1})} \leq M(g(x^k)^{1-\theta} - g(x^{k+1})^{1-\theta}) \tag{4.13}$$

Let $r \in (0, 1)$ and take $k \geq N_0$. If $d(x^{k+1}, x^k) \geq r d(x^k, x^{k-1})$, (4.13) implies that

$$d(x^{k+1}, x^k) \leq (M/r)(g(x^k)^{1-\theta} - g(x^{k+1})^{1-\theta}),$$

and therefore $\forall k \geq I\!N_0$:

$$d(x^{k+1}, x^k) \leq r d(x^k, x^{k-1}) + (M/r)(g(x^k)^{1-\theta} - g(x^{k+1})^{1-\theta}).$$

Now, if $I\!N \geq I\!N_0$ then

$$\sum_{k=I\!N_0}^{I\!N} d(x^{k+1}, x^k) \leq \frac{r}{1-r} d(x^{I\!N_0}, x^{I\!N_0 - 1}) + (M/(r(1-r)))(g(I\!N_0)^{1-\theta} - g(x^{I\!N+1})^{1-\theta}),$$

and the conclusion follows from the fact that f is bounded from below. ∎

4.2 The Quasiconvex Case

In this subsection we substitute the assumption A and B by the following

Assumption A'. The set of global minimizer points, denoted by X^*, is nonempty.

Assumption B'. f is continuous and quasiconvex on $\texttt{dom}f$, supposed convex.

Now, we define the following sets

$$U := \{x \in M : f(x) \leq \inf_j f(x^j)\} \subset V_k := \{x \in M : f(x) \leq f(x^k)\}.$$

From Assumptions A' and B' those sets are nonempty closed and convex (see Theorem 2.7 for the convex property).

Theorem 4.4 *Under assumptions A' and B', the sequence $\{x^k\}$, generated by the Proximal algorithm, is Fejér convergent to U.*

Proof. From (4.7) we have

$$f(x^k) + (\lambda_k/2)d^2(x^k, x^{k-1}) \leq f(x) + (\lambda_k/2)d^2(x, x^{k-1}), \forall x \in M. \tag{4.14}$$

Taking $x \in V_k$ we obtain

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

Therefore x^k is the unique projection of x^{k-1} on V_k. From Theorem 2.4, taking $y = x^{k-1}$ and $z = x^k$ we obtain

$$\langle \exp^{-1}_{x^k} x^{k-1} \exp^{-1}_{x^k} x \rangle \leq 0. \tag{4.15}$$

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

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

Now, the last inequality and (4.15) imply, in particular,

$$0 \leq d^2(x^k, x^{k-1}) \leq d^2(x, x^{k-1}) - d^2(x, x^k) \tag{4.16}$$

for every $x \in U$. Thus

$$d(x, x^k) \leq d(x, x^{k-1}). \tag{4.17}$$

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

Proposition 4.3 *Under the assumptions of the precedent theorem, the following facts are true*

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

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

c. *If* $\lim_{j \to +\infty} x^{k_j} = \bar{x}$ *then,* $\lim_{j \to +\infty} x^{k_j+1} = \bar{x}$.

Proof.
a. From (4.17), $\{d(x, x^k)\}$ is a bounded below nonincreasing sequence and hence convergent.
b. Taking limit when k goes to infinity in (4.16) and using the previous result we obtain

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

c. Take the triangular inequality property, applied to the Riemannian distance d, which gives, particularly

$$d(x^{k_j+1}, \overline{x}) \leq d(x^{k_j+1}, x^{k_j}) + d(x^{k_j}, \overline{x}).$$

Taking $j \to \infty$ and using **b**, we obtain the result. ■

We will use the following parameter condition to the algorithm:

$$0 < \lambda_k < \bar{\lambda}. \tag{4.18}$$

Theorem 4.5 *Suppose that Assumptions A' and B are satisfied. Then the sequence $\{x^k\}$, generated by the Proximal algorithm with (4.18), converges to a limiting critical point of f.*

Proof. From previous theorem, $\{x^k\}$ is Fejér convergent to U, thus $\{x^k\}$ is bounded (see Theorem 2.1). Then, there exist $\bar{x}$ and a subsequence $\{x^{k_j}\}$ of $\{x^k\}$ converging to $\bar{x}$. From continuity of f and Proposition 4.1, **a**, we obtain $\bar{x} \in \mathtt{dom} f$ and

$$\lim_{j \to +\infty} f(x^{k_j}) = f(\bar{x}).$$

As $\{f(x^k)\}$ is a non increasing sequence with a subsequence converging to $f(\bar{x})$, the overall sequence converges to $f(\bar{x})$ and therefore

$$f(\bar{x}) \leq f(x^k), \; \forall k \in I\!N.$$

This implies that $\bar{x} \in U$. Now, from Theorem 2.1 we conclude that $\{x^k\}$ converges to $\bar{x}$. Finally, from Proposition 4.3, item b, and Theorem 4.2, $\bar{x}$ is a limiting critical point of f. ■As immediate particular cases of the above theorem we obtain the following results

Corollary 4.1 *Under Assumption A' and f convex on M, then $\{x^k\}$ converges to an optimal solution of the problem (p).*

Corollary 4.2 *Under Assumption A', f differentiable and pseudoconvex on M, then $\{x^k\}$ converges to an optimal solution of the problem (p).*

To finish this subsection we give a localization of the limit point. Let $P_U : M \to M$ such that

$$P_U x^0 = \arg\min\{d(x, x^0) : x \in U\}.$$

Note that $P_U x^0$ exists and is unique because U is a nonempty closed convex set. We denote

$$\rho(x^0, U) := \min\{d(x, x^0) : x \in U\} = d(P_U x^0, x^0).$$

Proposition 4.4 *Let $\{x^k\}$ the sequence generated by the proximal point algorithm and $\bar{x}$ the limit point, then*

$$d(x^0, \bar{x}) \leq 2\rho(x^0, U).$$

Proof. Setting $x = P_U x^0$ in (4.17) we obtain $d(P_U x^0, x^k) \leq d(P_U x^0, x^0)$. Taking $k \to \infty$ and using the continuity of d gives

$$d(P_U x^0, \bar{x}) \leq d(P_U x^0, x^0), \tag{4.19}$$

Now, from the triangular inequality

$$\begin{aligned} d(x^0, \bar{x})^2 &\leq \left(d(x^0, P_U x^0) + d(P_U x^0, \bar{x})\right)^2 \\ &\leq 2\left(d(x^0, P_U x^0)^2 + d(P_U x^0, \bar{x})^2\right) \\ &\leq 2\left(d(x^0, P_U x^0)^2 + d(x^0, P_U x^0)^2\right) \\ &= 4d(x^0, P_U x^0)^2. \end{aligned}$$

where the third inequality is due to (4.19). Thus, the proof is completed. ■

Remark 4.1 *If $\bar{x}$ is a global minimum of (p), for example when f is convex, or differentiable and pseudoconvex, the above result remain true if we substitute U by X^*.*

4.3 Examples of Proximal Methods on Hadamard Manifolds

Examples 4.1 and 4.2 show proximal point methods for unconstrained minimization problems and examples 4.3 to 4.6 show proximal point methods for constrained ones. Those algorithms find interior solutions and they are particularly useful when the objective function is not convex in the usual sense but becomes convex or quasiconvex on the manifold. We recall that proximal methods without the convex condition on the objective function very are not studied.

Example 4.1 *Consider the unconstrained minimization problem in $I\!R^n$*

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

Take $I\!R^n$ as a smooth manifold and consider the following metrics:

i. *$M = (I\!R^n, \langle , \rangle)$, where $\langle u, v\rangle = \sum_{i=1}^{n} u_i v_i$, for $u, v \in I\!R^n$, (the Euclidian space) is a Hadamard manifold with null sectional curvature and the proximal point method is the classical one, that generates a sequence $\{x^k\}$ given by $x^0 \in I\!R^n$ and*

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

where $||.||$ is the Euclidian norm, i.e, $||x||^2 = \sqrt{\langle x, x\rangle}$.

ii. *$M = (I\!R^n, G(x))$ where*

$$G(x) = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & 0 & 0 & 0 & 1 + 4x_{n-1}^2 & -2x_{n-1} \\ 0 & \ldots & 0 & 0 & 0 & -2x_{n-1} & 1 \end{bmatrix}.$$

is a Hadamard manifold with null sectional curvature, see [12], and the proximal point method is given by $x^0 \in \mathbb{R}^n$ *and*

$$x^k \in \textit{arg}\min\left\{f(x) + (\lambda_k/2)\left(\sum_{i=1}^{n-1}(x_i - x_i^{k-1})^2 + (x_{n-1}^2 - x_n - (x_{n-1}^{k-1})^2 + x_n^{k-1})^2\right) : x \in \mathbb{R}^n\right\}$$

Example 4.2 *Consider the matricial minimization problem*

$$\min\{f(X) : X \in \mathbb{R}^{n\times n}\},$$

where $\mathbb{R}^{n\times n}$ *is the set of* $n \times n$ *matrices. Take* $\mathbb{R}^{n\times n}$ *as a smooth manifold and define* $\langle U, V\rangle = Tr(UV)$ *(the trace operation) for* $U, V \in \mathbb{R}^{n\times n}$*.* $M = (\mathbb{R}^{n\times n}, \langle , \rangle)$*, is a Hadamard manifold and the proximal point method to solve this problem generates a sequence* $\{X^k\}$ *given by* $X^0 \in \mathbb{R}^{n\times n}$ *and*

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

where $||.||_F$ *is the Frobenius norm, i.e,* $||X||_F^2 = Tr(X^2)$*.*

Example 4.3 *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* $b(X) = -\log\det X$*.* $(\mathcal{S}_{++}^n, b''(X))$ *is a Hadamard manifold with non positive sectional curvature, see [27], and the proximal method generates a sequence* $\{X^k\}$ *given by* $X^0 \in \mathcal{S}_{++}^n$ *and*

$$X^k \in \textit{arg}\min\{f(X) + (\lambda_k/2)\sum_{i=1}^{n}\ln^2\lambda_i(X^{-\frac{1}{2}}X^{k-1}X^{-\frac{1}{2}}) : X \in \mathcal{S}_{++}^n\},$$

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

Example 4.4 *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 non positive sectional curvature, see [27]. The proximal method generates a sequence* $\{x^k\}$ *given by* $x^0 \in K$ *and*

$$x^k \in \textit{arg}\min\{f(x) + (\lambda_k/2)[\ln^2\sigma(x, x^{k-1}) + \ln^2\sigma(x^{k-1}, x)] : x \in K^0\},$$

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

Example 4.5 *Consider the problem on the positive orthant*

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

where $x \geq 0$ *means that* $x_i \geq 0$*,* $\forall i = 1, ..., n$*. Take* $\mathbb{R}_{++}^n$ *the positive orthant as a smooth manifold.* $(\mathbb{R}_{++}^n, X^{-2})$*, where* X^{-2} *is the Hessian of the* $-\log$ *barrier, is a Hadamard manifold* *with null sectional curvature and the proximal point method generates the sequence* $\{x^k\}$ *given by* $x^0 \in I\!R^n_{++}$ *and*

$$x^k \in \boldsymbol{arg}\min\left\{ f(x) + (\lambda_k/2)\sum_{i=1}^{n}\left(\ln\frac{x_i^{k-1}}{x_i}\right)^2 : x \in I\!R^n_{++}\right\}$$

Example 4.6 *Consider the problem on the hypercube*

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

where $x \in I\!R^n$ *and* $e = (1, 1, ..., 1) \in I\!R^n$. $((0,1)^n, X^{-2}(I-X)^{-2})$, *where* $X^{-2}(I-X)^{-2}$ *is the Hessian of the barrier* $\sum_{i=1}^{n}(2x_i - 1)[\log x_i - \log(1-x_i)]$, *is a Hadamard manifold with null sectional curvature, see [28]. The proximal point method generates a sequence* $\{x^k\}$ *given by* $x^0 \in (0,1)^n$ *and*

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

5 Proximal Method for Constrained Minimizaton

In this section we are interested in solve the constrained minimization problem:

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

where X is a closed convex set with interior, $intX$, non null on a Hadamard manifold M and $f : X \to I\!R$ is a continuous quasiconvex or convex function on X.

To solve the problem (P), we will use proximal point methods with generalized Bregman distances and Bregman distances defined on Hadamard manifolds. The generalized Bregman distances permit generalize some important methods in optimization, such as proximal methods for semidefinite optimization problems and second order cone programs.

5.1 Generalized Bregman and Bregman Distances on Hadamard Manifolds

Starting from Censor and Lent [9] definition, we propose the following. Let M be a Hadamard manifold and S a nonempty open convex set of M with a topological closure $\bar{S}$. Let $h : M \to I\!R$ be a strictly convex function on $\bar{S}$ and differentiable in S. Define the funcion $D_h(.,.) : \bar{S} \times S \to I\!R$ such that

$$D_h(x,y) := h(x) - h(y) - \langle \mathtt{grad}h(y), \exp_y^{-1} x\rangle_y. \tag{5.20}$$

Let us adopt the following notation for the partial level sets of D_h. For $\alpha \in I\!R$, take

$$\Gamma_1(\alpha, y) := \{x \in \bar{S} : D_h(x,y) \leq \alpha\},$$

$$\Gamma_2(x, \alpha) := \{y \in S : D_h(x,y) \leq \alpha\}.$$

Definition 5.1 *Let* M *be a Hadamard manifold. A real function* $h : M \to I\!R$ *is called a* Generalized Bregman function, *denoted by* $h \in \mathcal{GB}$, *if there exists a nonempty open convex set* S *such that*

a. *h is continuous on $\bar{S}$;*

b. *h is strictly convex on $\bar{S}$;*

c. *h is continuously differentiable in S;*

d. *For all $\alpha \in I\!R$ the partial level sets $\Gamma_1(\alpha, y)$ and $\Gamma_2(x, \alpha)$ are bounded for every $y \in S$ and $x \in \bar{S}$, respectively.*

In this case D_h is called generalized Bregman distance *from x to y.*

Definition 5.2 *Let M be a Hadamard manifold. A real function $h : M \to I\!R$ is called a* Bregman function*, denoted by $h \in B$, if there exists a nonempty open convex set S such that h satisfies the conditions of the above definition and, furthermore, satisfies:*

e. *If $\lim_{k\to+\infty} y^k = y^* \in \bar{S}$, then $\lim_{k\to+\infty} D_h(y^*, y^k) = 0$, and*

f. *If $\lim_{k\to+\infty} D_h(z^k, y^k) = 0$, $\lim_{k\to+\infty} y^k = y^* \in \bar{S}$ and $\{z^k\}$ is bounded then $\lim_{k\to+\infty} z^k = y^*$.*

In this case D_h is called Bregman distance *from x to y.*

From above definitions we obtain, obviously, that $\mathcal{B} \subseteq \mathcal{GB}$ and the equality is satisfied when $S = M$. In both definitions, the set S is called the *zone* of the function h. Some examples of Bregman distance to different Hadamard manifolds has been done in [30], Section 8.

Lemma 5.1 *Let $h \in \mathcal{GB}$ with zone S. Then*

i. *$\mathtt{grad}\, D_h(.,y)(x) = \mathtt{grad}\, h(x) - P_{\gamma,0,1}\mathtt{grad}\, h(y)$, for all $x, y \in S$, where $\gamma : [0,1] \to M$ is the geodesic curve such that $\gamma(0) = y$ and $\gamma(1) = x$.*

ii. *$D_h(.,y)$ is strictly convex on $\bar{S}$ for all $y \in S$.*

iii. *For all $x \in \bar{S}$ and $y \in S$, $D_h(x,y) \geq 0$ and $D_h(x,y) = 0$ if and only if $x = y$.*

Proof. Analogous to Lemma 4.1 of [30]. ∎

Observe that D_h is not a distance in the usual sense of the term. In general, the triangular inequality is not valid, as the symmetry property.
From now on, we use the notation $\mathtt{grad}\, D_h(x,y)$ to mean $\mathtt{grad}\, D_h(.,y)(x)$. So, if γ is the geodesic curve such that $\gamma(0) = y$ and $\gamma(1) = x$, from Lemma 5.1, i, we obtain

$$\mathtt{grad} D_h(x,y) = \mathtt{grad}\, h(x) - P_{\gamma,0,1}\mathtt{grad}\, h(y).$$

Definition 5.3 *Let $\Omega \subset M$, S an open convex set, and let $y \in S$. A point $Py \in \Omega \cap \bar{S}$ for which*

$$D_h(Py, y) = \min_{x \in \Omega \cap \bar{S}} D_h(x,y) \tag{5.21}$$

*is called a D_h-*projection *of the point y on the set Ω.*

The next Lemma furnishes the existence and uniqueness of D_h-projection for a Bregman function, under an appropriate assumption on Ω.

Lemma 5.2 *Let $\Omega \subset M$ a closed convex set and $h \in \mathcal{GB}$ with zone S. If $\Omega \cap \bar{S} \neq \phi$ then, for any $y \in S$, there exists a unique D_h-projection Py of the point y on Ω.*

Proof. For any $x \in \Omega \cap \bar{S}$, the set

$$B := \{z \in \bar{S} : D_h(z, y) \leq D_h(x, y)\}$$

is bounded (from Definition 5.1, d) and closed (because $D_h(., y)$ is continuous in $\bar{S}$, due to Definition 5.1, a). Therefore, the set

$$T := (\Omega \cap \bar{S}) \cap B$$

is nonempty, because $x \in B \cap \Omega$, and bounded. Now, as the intersection of two closed sets is closed, then T is also closed, hence compact. Consequently, $D_h(z, y)$, a continuous function in z, takes its minimum on the compact set T at some point, let denote it by x^*. For every $z \in \Omega \cap \bar{S}$ which lies outside B

$$D_h(x, y) < D_h(z, y);$$

hence, x^* satisfies (5.21). The uniqueness follows from the strict convexity of $D_h(., y)$, therefore

$$x^* = Py. \quad \blacksquare$$

Lemma 5.3 *Let $h \in \mathcal{GB}$ with zone S and $y \in S$. Suppose that $Py \in S$, where Py is the D_h-projection on some closed convex set Ω such that $\Omega \cap \bar{S} \neq \phi$. Then, the function*

$$G(x) := D_h(x, y) - D_h(x, Py)$$

is linear affine on $\bar{S}$.

Proof. From (5.20)

$$G(x) = h(Py) - h(y) + \langle \mathtt{grad} h(Py), \exp^{-1}_{Py} x\rangle_{Py} - \langle \mathtt{grad} h(y), \exp^{-1}_{y} x\rangle_{y}.$$

Due to the affine linearity of the functions $\langle \mathtt{grad} h(Py), \exp^{-1}_{Py} x\rangle_{Py}$ and $\langle \mathtt{grad} h(y), \exp^{-1}_{y} x\rangle_{y}$ in x the result follows. $\blacksquare$

Proposition 5.1 *Let $h \in \mathcal{GB}$ with zone S and $\Omega \subset M$ a closed convex set such that $\Omega \cap \bar{S} \neq \phi$. Let $y \in S$ and assume that $Py \in S$, where Py denotes the D_h-projection of y on Ω. Then, for any $x \in \Omega \cap \bar{S}$, the following inequality is true*

$$D_h(Py, y) \leq D_h(x, y) - D_h(x, Py).$$

Proof. Let $\gamma : [0, 1] \to M$ be the geodesic curve such that $\gamma(0) = Py$ and $\gamma(1) = x$. Due to Lemma 5.3 the function

$$G(x) = D_h(x, y) - D_h(x, Py)$$

is linear affine on $\bar{S}$. Then in particular $G(\gamma(t))$ is convex for $t \in (0, 1)$ (see Theorem 2.5). Thus,

$$G(\gamma(t)) \leq tG(x) + (1 - t)G(Py),$$

which gives,

$$D_h(\gamma(t), y) - D_h(\gamma(t), Py) \leq t(D_h(x, y) - D_h(x, Py)) + D_h(Py, y) - tD_h(Py, y),$$

where we took in account that $D_h(Py,Py)=0$. The above inequality is equivalent to

$$(1/t)\left(D_h(\gamma(t),y)-D_h(Py,y)\right)-(1/t)D_h(\gamma(t),Py)\leq D_h(x,y)-D_h(x,Py)-D_h(Py,y). \quad (5.22)$$

As $\Omega\cap\bar{S}$ is convex, and $x, Py\in\Omega\cap\bar{S}$, we have $\gamma(t)\in\Omega\cap\bar{S}$ for all $t\in(0,1)$. Then, use the fact that Py is the projection to get

$$(1/t)(D_h(\gamma(t),y)-D_h(Py,y))\geq 0.$$

Using this inequality in (5.22) we obtain

$$-(1/t)D_h(\gamma(t),Py)\leq D_h(x,y)-D_h(x,Py)-D_h(Py,y).$$

Now, as $D_h(.,z)$ is differentiable for all $z\in S$, we can take the limit in t, obtaining

$$-\langle \mathtt{grad} D_h(Py,Py),\exp^{-1}_{Py}x\rangle_{Py}\leq D_h(x,y)-D_h(x,Py)-D_h(Py,y).$$

Clearly, the left side is null, leading to the aimed result. ■

Proposition 5.2 *Let $h\in\mathcal{GB}$ with zone S and $\Omega\subset M$ a closed convex set such that $\Omega\cap\bar{S}\neq\phi$. Let $y\in S$ and assume that $Py\in S$, where Py denotes the D_h-projection of y on Ω. Then, for any $x\in\Omega\cap\bar{S}$, the following is true*

$$\langle \mathtt{grad} D_h(Py,y),\exp^{-1}_{Py}x\rangle = D_h(x,y)-D_h(x,Py)-D_h(Py,y).$$

Proof. Let $\gamma:[0,1]\to M$ be the geodesic curve such that $\gamma(0)=Py$ and $\gamma(1)=x$. Due to Lemma 5.3 the function

$$G(x)=D_h(x,y)-D_h(x,Py)$$

is linear affine on $\bar{S}$. Then, from Theorem 2.5 we have

$$G(\gamma(t))=tG(x)+(1-t)G(Py),$$

which gives,

$$D_h(\gamma(t),y)-D_h(\gamma(t),Py)=t(D_h(x,y)-D_h(x,Py))+D_h(Py,y)-tD_h(Py,y),$$

where we took in account that $D_h(Py,Py)=0$. The above equality is equivalent to

$$(1/t)\left(D_h(\gamma(t),y)-D_h(Py,y)\right)-(1/t)D_h(\gamma(t),Py)=D_h(x,y)-D_h(x,Py)-D_h(Py,y).$$

Taking limit when $t\to 0$ we obtain

$$\langle \mathtt{grad} D_h(Py,y),\exp^{-1}_{Py}x\rangle = D_h(x,y)-D_h(x,Py)-D_h(Py,y).$$

■

5.2 Fejér Convergence with Bregman Distances

Definition 5.4 *Let M be a Hadamard manifold. A sequence $\{y^k\}$ of M is D_h-*Fejér convergent *to a nonempty set $U \subset M$, if*

$$D_h(u, y^{k+1}) \le D_h(u, y^k),$$

for every $u \in U$.

Theorem 5.1 *Let M be a Hadamard manifold and $h \in \mathcal{GB}$ with zone S. If $\{y^k\}$ is D_h-Fejér convergent to a nonempty set $U \subset M$, then $\{y^k\}$ is bounded. If, furthermore, $h \in \mathcal{B}$ and a cluster point $\bar{y}$ of $\{y^k\}$ belongs to U, then $\{y^k\}$ converges and $\lim_{k\to+\infty} y^k = \bar{y}$.*

Proof. From above definition

$$0 \le D_h(u, y^k) \le D_h(u, y^0), \tag{5.23}$$

for all $u \in U$. Thus, $y^k \in \Gamma_2(u, \alpha)$ with $\alpha = D_h(u, y^0)$. We can now apply Definition 5.1, **d**, to see that $\{y^k\}$ is bounded.
Let $\bar{y}$ a cluster point of $\{y^k\}$, with $\bar{y} \in U$, then there exists a subsequence $\{y^{k_j}\}$ such that $\lim_{j\to\infty} y^{k_j} = \bar{y}$. From Definition 5.1, **e**, it is true that $\lim_{j\to+\infty} D_h(\bar{y}, y^{k_j}) = 0$. From (5.23), $\{D_h(\bar{y}, y^k)\}$ is a nonincreasing bounded below sequence with a subsequence converging to 0, hence the overall sequence converges to 0, that is,

$$\lim_{k\to+\infty} D_h(\bar{y}, y^k) = 0. \tag{5.24}$$

To prove that $\{y^k\}$ has a unique limit point, let y' be another limit point of $\{y^k\}$. From (5.24) $\lim_{l\to+\infty} D_h(\bar{y}, y^{k_l}) = 0$ with $\lim_{l\to+\infty} y^{k_l} = y'$. Using Definition 5.1, **f**, we have $y' = \bar{y}$. It follows that $\{y^k\}$ cannot have more than one limit point and therefore, $\lim_{k\to+\infty} y^k = \bar{y}$. ■

5.3 Regularization

Let M be a Hadamard manifold and $f : X \subset M \to I\!R$ a real function. Let S an open convex set and $h : \bar{S} \to I\!R$ a differentiable function in S. For $\lambda > 0$, the Moreau-Yosida regularization $f_\lambda : S \to I\!R$ of f is defined by

$$f_\lambda(y) = \inf_{x \in X \cap \bar{S}} \{f(x) + \lambda D_h(x, y)\} \tag{5.25}$$

where $D_h(x, y)$ is given in (5.20). In order to prove that the function f_λ is well defined, h and f should satisfy some conditions.

Proposition 5.3 *If $f : X \subset M \to I\!R$ is a bounded below and lower semicontinuous function over a closed convex set X, and $h \in \mathcal{GB}$ with zone S such that $X \cap \bar{S} \ne \phi$, then, for every $y \in S$ and $\lambda > 0$ there exists a point, denoted by $x_f(y, \lambda)$, such that*

$$f_\lambda(y) = f(x_f(y, \lambda)) + \lambda D_h(x_f(y, \lambda), y). \tag{5.26}$$

Proof. Let β a lower bound for f on X, then

$$f(x) + \lambda D_h(x, y) \ge \beta + \lambda D_h(x, y),$$

for all $x \in X \cap \bar{S}$. It follows from Definition 5.1, d, that the level sets of the function $f(.) + \lambda D_h(., y)$ are bounded. Also, this function is lower semicontinuous on $X \cap \bar{S}$, due to Definition 5.1, a, and the hypothesis on f. So, the level sets of $(f(.) + \lambda D_h(., y))$ are closed, hence compact. Now, from lower semicontinuity and compactness arguments, $f(.) + \lambda D_h(., y)$ has a (not need unique) global minimum $x_f(y, \lambda) \in X \cap \bar{S}$. Then the equality (5.26) follows from (5.25). ■

5.4 The Proximal Method with Generalized Bregman Distances

The Proximal Point Method with Generalized Bregman Distances to solve the problem (P), henceforth abbreviated PGBD method, is defined as

$$y^0 \in intX \cap S, \tag{5.27}$$

$$y^k \in \texttt{arg}\min\{f(y) + \lambda_k D_h(y, y^{k-1}) : y \in X \cap \bar{S}\}, \tag{5.28}$$

where $h \in \mathcal{GB}$ (h is a Generalized Bregman function) with zone S, such that $X \cap \bar{S} \neq \phi$, D_h is as in (5.20) and λ_k is a positive parameter.

Along this section we assume the following assumption

Assumption D. The sequence $\{y^k\}$ exists and satisfies $y^k \in intX \cap S$, for all $k \in I\!N$.

Some examples of the PGBD method are the following

Example 5.1 (Semidefinite Optimization) .
Consider the problem

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

where the notation $X \succeq 0$ means that X is a $n \times n$ symmetric semidefinite matrix. Let $\mathcal{S}^{n\times n}$ the $n \times n$ symmetic matrix space and $\langle U, V\rangle = Tr(UV)$ (the trace operation) for $U, V \in I\!R^{n\times n}$. $M = (\mathcal{S}^{n\times n}, \langle , \rangle)$, is a Hadamard manifold where $\exp_X^{-1} Y = Y - X$, $\texttt{grad}f = \nabla f$ (the usual gradient in matrix spaces) and $\langle U, V\rangle_X = \langle U, V\rangle$. Then, the PGBD method generates a sequence $\{Y^k\}$ given by

$$Y^k \in \texttt{arg}\min\left\{f(Y) + \lambda_k D_h(Y, Y^{k-1}) : Y \succeq 0\right\},$$

where $D_h(Y, Y^{k-1}) = h(Y) - h(Y^{k-1}) - \texttt{Tr}(\nabla h(Y^{k-1})(Y - Y^{k-1}))$ and $h \in \mathcal{GB}$ with zone $S = \mathcal{S}^n_{++}$ and λ_k is some positive parameter.
If f is lower bounded and furthermore convex or continuously differentiable (not needly convex), then some examples of $h \in \mathcal{GB}$ satisfying assumption D, see Teboulle for the conve case, are

i. $h_1 = \texttt{tr}(X \ln X)$ *with the generalized distance given by*

$$D_{h_1}(X, Y) = \texttt{Tr}(X \ln X - X \ln Y + Y - X),$$

ii. $h_2 = -\ln \texttt{det}X$ *with the generalized distance given by*

$$D_{h_2}(X, Y) = +\texttt{Tr}(XY^{-1}) - \ln det(YX^{-1}) - n,$$

Example 5.2 (Second Order Cone Programming) .
Consider the problem

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

where $x \succeq_K 0$ means that $x \in K := \{(x_1, x_2) \in I\!R^{1+n} : x_1 \geq ||x_2||\}$. The Second Order Cone Program (SOCP) has recently received much attention in optimization ....
Given $x, y \in I\!R^{1+n}$, the Jordan product is defined as

$$x \, o \, y = (x_2^T y_2, y_1 x_2 + x_1 y_2).$$

This product satisfies the following properties: $e\,o\,x = x$, $x\,o\,y = y\,o\,x$ *and* $(x+y)\,o\,z = x\,o\,z + y\,o\,z$. *For each* $x = (x_1, x_2) \in I\!R^{1+n}$ *we define* $\mathbf{Tr}x := 2x_1$ *and*

$$\ln x := \frac{1}{2}\left(\ln(x_1^2 - ||x_2||^2), \ln\left(\frac{x_1 + ||x_2||}{x_1 - ||x_2||}\right)\frac{x_2}{||x_2||}\right)$$

if $x_2 \neq 0$ *and* $\ln x := (\ln x_1, 0)$ *otherwise. From the trace definition it is easy of chek that* $\langle x, y\rangle = \frac{1}{2}(x\,o\,y)$.
Now, consider the space $I\!R^{1+n}$ *as a smooth manifold and define for* $x, y \in I\!R^{1+n}$, $\langle x, y\rangle = x^T y$. $(I\!R^n, \langle , \rangle)$ *is a Hadamard manifold where* $\exp_x^{-1} y = y - x$, $\mathbf{grad}f = \nabla f$ *(the usual gradient) and* $\langle u, v\rangle_x = \langle u, v\rangle$. *Then, the PGBD method generates a sequence* $\{y^k\}$ *given by*

$$y^k \in \mathbf{arg}\min\left\{f(y) + \lambda_k D_h(y, y^{k-1}) : y \succeq_K 0\right\},$$

where $D_h(y, y^{k-1}) = h(y) - h(y^{k-1}) - (1/2)\mathbf{Tr}\left(\nabla h(y^{k-1})\,o\,(y - y^{k-1})\right)$ *and* $h \in \mathcal{GB}$ *with zone* $S = intK$ *and* λ_k *is some positive parameter.*
An example of $h \in \mathcal{GB}$*, studied recently by Jein-Shan Chen [10], is* $h(x) = \mathbf{Tr}(x\,o\,\ln x)$ *and the distance is given by*

$$D_h(x, y) = \mathbf{Tr}(x\,o\,\ln x - x\,o\,\ln y + y - x),$$

for all $y \in intK$ *and* $x \in K$. *When* f *is lower bounded and convex then, the above function satisfies the assumption D.*

Proposition 5.4 *Under the Assumption D, the iterates* y^k *do not cycle.*

Proof. Consider two cases: contiguous and non contiguous iterations. Clearly if $y^l = y^{l-1}$ for some $l > 0$ we have from (5.28) that $0 \in \partial f(y^l)$. Therefore, y^l is a critical point and the method stops in finitely many iterations. Otherwise, let $l > j+1$ such that $y^l = y^j$ and $y^k \neq y^{k-1}$, for all k. From (5.28), the minimum condition implies that, for all k,

$$f(y^k) + \lambda_k D_h(y^k, y^{k-1}) \leq f(y^{k-1}). \tag{5.29}$$

As $y^{k-1} \neq y^k$ we have from (5.29) that $f(y^k) < f(y^{k-1})$ ($\{f(y^k)\}$ is a strictly nonincreasing sequence). Now, this implies that $f(y^l) < f(y^{l-1}) < \ldots < f(y^j)$, which contradicts that $f(y^l) = f(y^j)$. ■

Proposition 5.5 *Under assumptions D the following fact is true*

a. $\{f(y^k)\}$ *is nonincreasing.*

Furthermore, if f *is lower bounded then*

b. $\sum_{k=1}^{\infty} \lambda_k D_h(y^k, y^{k-1}) < +\infty$

Proof.
a. Shows in the preceding Proposition.
b. From inequality (5.29) and lower boundedness of f we obtain

$$\sum_{k=1}^{n} \lambda_k D_h(y^k, y^{k-1}) \leq f(y^0) - \alpha,$$

where α is the lower bounded of f. Taking $n \to \infty$ we obtain the result. ■

5.4.1 The Quasiconvex Case

I this subsection we assume that f is continuous and quasiconvex in X. Furthermore, we assume the following assumption:

Assumption E. $X^* \cap \bar{S} \neq \phi$, where X^* is the set of global minimizer points.

Now, we define the following sets

$$U := \{x \in X : f(x) \leq \inf_j f(y^j)\} \subset V_k := \{x \in X : f(x) \leq f(y^k)\}.$$

From Assumptions E and continuous quasiconvexity of f those sets are nonempty closed and convex (see Theorem 2.7 for the convex property).

Theorem 5.2 *Under assumptions D and E, and continuous quasiconvexity of f, the sequence $\{y^k\}$, generated by the PGBD algorithm, is Fejér convergent to $U \cap \bar{S}$.*

Proof. Since y^k satisfies (5.28) we have

$$f(y^k) + \lambda_k D_h(y^k, y^{k-1}) \leq f(x) + \lambda_k D_h(x, y^{k-1}),\ \forall x \in X \cap \bar{S}. \tag{5.30}$$

Hence, $\forall x \in V_k \cap \bar{S}$ is true that

$$D_h(y^k, y^{k-1}) \leq D_h(x, y^{k-1}).$$

Therefore y^k is the unique D_h-projection from y^{k-1} on the convex set V_k (see Definition 5.3). Using Proposition 5.1 and in particular for all $x \in U \cap \bar{S}$ we have

$$0 \leq D_h(y^k, y^{k-1}) \leq D_h(x, y^{k-1}) - D_h(x, y^k) \tag{5.31}$$

Thus for every $x \in U \cap \bar{S}$:

$$D_h(x, y^k) \leq D_h(x, y^{k-1}). \tag{5.32}$$

The last inequality tell us that $\{y^k\}$ is D_h- Fejér convergent to $U \cap \bar{S}$. ■

Corollary 5.1 *Under the assumptions of the precedent theorem, the sequence $\{y^k\}$, generated by the PGBD algorithm, is bounded.*

Proof. See Theorem 5.1. ■

Proposition 5.6 *Under the assumptions of the precedent theorem, the following facts are true*

a. *For all $x \in U \cap \bar{S}$ the sequence $\{D_h(x, y^k)\}$ is convergent;*

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

Proof.
a. From (5.32), $\{D_h(x, y^k)\}$ is a bounded below nonincreasing sequence and hence convergent.
b. Taking limit when k goes to infinity in (5.31) and using the previous result we obtain $\lim_{k\to\infty} D_h(y^k, y^{k-1}) = 0$, as desired. ■

Theorem 5.3 *Suppose that Assumptions D and E are satisfied and that f is continuous and quasiconvex. If λ_k satisfies*

$$\lim_{k\to\infty} \lambda_k = 0,$$

then, any cluster point of $\{y^k\}$ is an optimal solution of the problem (P).

Proof. Given $x^* \in X^* \cap \bar{S}$. As y^k is a solution of (5.28) we have

$$f(y^k) + \lambda_k D_h(y^k, y^{k-1}) \leq f(x^*) + \lambda_k D_h(x^*, y^{k-1}).$$

As y^k is D_h-Fejér convergent and $x^* \in U \cap \bar{S}$, we have

$$f(y^k) + \lambda_k D_h(y^k, y^{k-1}) \leq f(x^*) + \lambda_k D_h(x^*, y^0).$$

Let $\bar{y}$ a cluster point of $\{y^k\}$ (assured by Corollary 5.1) with $\lim_{j \to +\infty} y^{k_j} = \bar{y}$. Taking $j \to +\infty$ in the above inequality and using Proposition 5.6, b, and continuity of f we obtain $f(\bar{y}) \leq f(x^*)$. ■

5.4.2 The Convex Case

Now, we assume that f is convex then, from strictly convexity of $D_h(., y)$, we obtain that for each k there exists an unique y^k in (5.28).

Lemma 5.4 *Suppose that assumptions D and E are satisfied and that f is convex, then we obtain*

$$(1/\lambda_k)(f(y^k) - f(y)) \leq D_h(y, y^{k-1}) - D_h(y, y^k) - D_h(y^k, y^{k-1}),$$

for all $y \in X \cap \bar{S}$.

Proof. From (5.28) we have

$$0 \in \partial \left(f + \lambda_k D_h(., y^{k-1}) \right) (y^k).$$

From smoothness of $D_h(., y^{k-1})$ this implies that

$$-\lambda_k \texttt{grad} D_h(y^k, y^{k-1}) \in \partial f(y^k).$$

Equivalently, for all $y \in X \cap \bar{S}$.

$$\frac{1}{\lambda_k}(f(y^k) - f(y)) \leq \langle \texttt{grad} D_h(y^k, y^{k-1}), \exp^{-1}_{y^k} y \rangle_{y^k}$$

Using Proposition 5.2 for $Py = y^k$, $y = y^{k-1}$ and $x = y$ we have the aimed result. ■

Lemma 5.5 *Suppose that assumptions D and E are satisfied and that f is convex, then we obtain $\forall y \in X \cap \bar{S}$:*

$$\sigma_n(f(y^n) - f(y)) \leq D_h(y, y^0) - D_h(y, y^n) - \sum_{k=1}^{n} \sigma_k \lambda_k D_h(y^k, y^{k-1}),$$

where $\sigma_k = \sum_{n=1}^{k} (1/\lambda_n)$. Therefore,

$$\sigma_k(f(y^k) - f(y)) \leq D_h(y, y^0). \tag{5.33}$$

Proof. Define $\sigma_k = (1/\lambda_k) + \sigma_{k-1}$ with $\sigma_0 = 0$. From (5.28) we have

$$\begin{aligned}\lambda_k \sigma_{k-1} D_h(y^k, y^{k-1}) &\le \sigma_{k-1}(f(y^{k-1}) - f(y^k)) \\ &= \sigma_{k-1} f(y^{k-1}) - (\sigma_k - (1/\lambda_k)) f(y^k) \\ &= \sigma_{k-1} f(y^{k-1}) - \sigma_k f(y^k) + (1/\lambda_k) f(y^k).\end{aligned}$$

Taking the sum over k, from $k = 1$ to n gives

$$\sum_{k=1}^{n} \lambda_k \sigma_{k-1} D_h(y^k, y^{k-1}) \le -\sigma_n f(y^n) + \sum_{k=1}^{n} (1/\lambda_k) f(y^k) \tag{5.34}$$

On the other hand, from the above Lemma

$$\sum_{k=1}^{n} (1/\lambda_k)(f(y^k) - f(y)) \le \sum_{k=1}^{n} \left(D_h(y, y^{k-1}) - D_h(y^k, y^{k-1}) - D_h(y, y^k) \right).$$

This implies that

$$\sum_{k=1}^{n} (1/\lambda_k) f(y^k) \le \sigma_n f(y) + D_h(y, y^0) - D_h(y, y^n) - \sum_{k=1}^{n} D_h(y^k, y^{k-1}).$$

The above inequality and (5.34) give

$$\sum_{k=1}^{n} \lambda_k \sigma_{k-1} D_h(y^k, y^{k-1}) + \sigma_n f(y^n) \le \sigma_n f(y) + D_h(y, y^0) - D_h(y, y^n) - \sum_{k=1}^{n} D_h(y^k, y^{k-1}).$$

Thus

$$\sigma_n(f(y^n) - f(y)) \le D_h(y, y^0) - D_h(y, y^n) - \sum_{k=1}^{n} (1 + \lambda_k \sigma_{k-1}) D_h(y^k, y^{k-1}).$$

As $1 + \lambda_k \sigma_{k-1} = \lambda_k \sigma_k$ then

$$\sigma_n(f(y^n) - f(y)) \le D_h(y, y^0) - D_h(y, y^n) - \sum_{k=1}^{n} \sigma_k \lambda_k D_h(y^k, y^{k-1}).$$

Putting as k iteration we have

$$\sigma_k(f(y^k) - f(y)) \le D_h(y, y^0), \forall y \in X \cap \bar{S}.$$

where

$$\sigma_k = \sum_{n=1}^{k} (1/\lambda_n).$$

Theorem 5.4 *Suppose that assumptions D and E are satisfied and that f is convex. Let y^* is an optimal point of (p), if λ_k satisfies*

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

then

$$f(y^n) - f(y^*) \le \frac{D_h(y^*, y^0)}{\sum_{k=1}^{n} (1/\lambda_k)}.$$

Furthermore, any cluster point of $\{y^k\}$ is an optimal solution of the problem (P).

Proof. From (5.33) we have

$$(f(y^k) - f(y^*)) \leq (1/\sigma_k) D_h(y^*, y^0),$$

$\forall y^* \in X^* \cap \bar{S}$.
Now, let $\bar{y} \in X \cap \bar{S}$ be a cluster point of $\{y^k\}$ with

$$\lim_{j \to +\infty} y^{k_j} = \bar{y}.$$

Take $k = k_j$ in the above inequality. Taking $j \to +\infty$ and using $\lim_{j\to\infty} \sigma_{k_j} = +\infty$ we have

$$f(y^*) \geq f(\bar{y}).$$

Therefore, any cluster point is an optimal solution of (P). ■

Remark 5.1 *Observe that if we substitute, in the convex case, the assumption E by:*
There exists c tal que

$$-\infty < c < f^* = \inf\{f(x) : x \in X\}.$$

Then, Lemma 5.4 and Lemma 5.5 are also true. Thus by definition of f^* there exists a v such that

$$f(v) < f^* + \epsilon.$$

Let $y = v \in X \cap \bar{S}$ in (5.33), then

$$f(y^k) - f(v) \leq \frac{D_h(v, y^0)}{\sum_{k=1}^{n} (1/\lambda_k)}.$$

Taking $k \to +\infty$ we obtain

$$\lim_{k \to +\infty} f(y^k) \leq f(v) < f^* + \epsilon.$$

Since $\epsilon > 0$ is arbitrary we obtain

$$\lim_{k \to +\infty} f(y^k) = f^*.$$

5.5 Proximal Method with Bregman Distances

The Proximal Point Method with Bregman distances to solve the problem (P), henceforth abbreviated PBD method, generates the sequence $\{y^k\}$ given by (5.27) and (5.28), but with $h \in \mathcal{B}$ (h is a Bregman function).

We will assume the same assumptions D and E of the subsection 5.4.

Example 5.3 (Unconstrained Problems on Hadamard Manifolds) .
Consider the problem

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

where M is a Hadamard manifold and $f : M \to I\!R$ is a convex or continuous quasiconvex function. Defining $X = S = M$, we obtain the PBD sequence $\{y^k\}$, studied and analyzed in [30], given by $y^0 \in M$ and

$$y^k \in \arg\min\{f(y) + \lambda_k D_h(y, y^{k-1}) : y \in M\}.$$

Observe that any Bregman function satisfies the assumption D, where the existence of $\{y^k\}$ is assured by Proposition 5.3.

Example 5.4 (Convex Minimization Problems in $I\!R^n$) .
Let consider the problem

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

where $f : I\!R^n \to I\!R$ is a convex function on a closed convex set X of $I\!R^n$. The PBD sequence $\{y^k\}$ is given by

$$\textit{Given } \; y^0 \in S,$$

$$y^k = \textbf{\textit{arg}} \min_{x\in X\cap\bar{S}} \{f(x) + \lambda_k D_h(x, y^{k-1})\},$$

where h is a Bregman function with zone S, such that $X \cap \bar{S} \neq \phi$, λ_k is a positive parameter and D_h is a Bregman distance defined as

$$D_h(x, y) = h(x) - h(y) - \langle \nabla f(y), x - y\rangle,$$

where $\langle , \rangle$ denotes the usual inner product on $I\!R^n$.

a. *;*

b. *;*

Example 5.5 (Minimization of Quasiconvex Functions in $I\!R^n_+$) .
Consider the problem

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

where f is a continuously differentiable quasiconvex function in $I\!R^n$ bounded from below and $x \geq 0$ means that $x_i \geq 0$, $\forall i = 1, ..., n$. Let $X = I\!R^n_+$ (the non negative orthant), $S = I\!R^n_{++}$ (the positive orthant), $M = I\!R^n$ and consider the separated Bregman functions $h(x) = \sum_{i=1}^{n} h_i(x_i)$, such that h is zone coercive ($\textbf{\textit{Img}}\nabla h = I\!R^n$), in particular $h(x) = \sum_{i=1}^{n} x_i \ln x_i$. The PBD sequence $\{y^k\}$ is given by $y^0 \in I\!R^n_{++}$, and

$$y^k \in \textbf{\textit{arg}} \min\{f(x) + \lambda_k D_h(x, y^{k-1}) : x \geq 0\}.$$

It can be proved, see Souza et al. [40], Proposition 3.1, that $y^k \in I\!R^n_{++}$, for each k. Thus, the assumption D is assured.

5.5.1 The Quasiconvex Case

Theorem 5.5 *Suppose that Assumptions D and E are satisfied and f is a continuous quasiconvex function. Then the sequence $\{y^k\}$, generates by the PBD method, converges to some point of $U \cap \bar{S}$.*

Proof. From Corollary 5.1, $\{y^k\}$ is bounded. Then, there exist $\bar{y}$ and a subsequence $\{y^{k_j}\}$ of $\{y^k\}$ converging to $\bar{y}$. From continuity of f we have

$$\lim_{j\to+\infty} f(y^{k_j}) = f(\bar{y}).$$

As $\{f(y^k)\}$ is a monotone nonincreasing sequence then it converges to $f(\bar{y})$ and

$$f(\bar{y}) \leq f(y^k), \;\forall k.$$

That implies that $\bar{y} \in U \cap \bar{S}$. Now, as $h \in B$ we obtain from Theorem 5.1 that $\{y^k\}$ converges and

$$\lim_{k\to+\infty} y^k = \bar{y} \in U \cap \bar{S}.$$

■

Theorem 5.6 *Suppose that Assumptions D and E are satisfied and that f is continuous and quasiconvex. If* λ_k *satisfies*

$$\lim_{k\to\infty} \lambda_k = 0,$$

then $\{y^k\}$ *converges to an optimal solution of the problem* (P).

Proof. From theorems 5.3 and 5.5 we obtain the result. ■

Theorem 5.7 *Suppose that assumptions D and E are satisfied and f is a continuous quasiconvex function. Let* λ_k *such that*

$$0 < \lambda_k < \bar{\lambda}.$$

If $||P_{\alpha_k,0,1}\textit{grad}\, h(y^{k-1}) - \textit{grad}\, h(y^k)|| \to 0, k \to +\infty,$ *or* $\{y^k\}$ *converges to an interior point of* $U \cap \bar{S}$ *then, the sequence* $\{y^k\}$ *converges to a limiting critical point of f on X.*

Proof. We will prove that $0 \in \partial(f + I_X)(\bar{y})$. From (5.28) and Theorem 3.1 we have

$$0 \in \widehat{\partial}\left((f + I_X) + \lambda_k D_h(., y^{k-1})\right)(y^k).$$

From smoothness of $D_h(., y^{k-1})$ and Proposition 3.1, item d, this implies that

$$-\lambda_k \mathtt{grad} D_h(y^k, y^{k-1}) \in \widehat{\partial}(f + I_X)(y^k).$$

Using Lemma 5.1,**i,** there exists $g^k \in \widehat{\partial}(f + I_X)(y^k)$ such that

$$g^k = \lambda_k[P_{\alpha_k,0,1}\mathtt{grad}\, h(y^{k-1}) - \mathtt{grad}\, h(y^k)],$$

where α_k is the geodesic curve such that $\alpha_k(0) = y^{k-1}$ and $\alpha_k(1) = y^k$. Taking parallel transport and using the above boundedness of λ_k we obtain

$$||P_{\gamma_k,0,1}g^k|| = ||g^k|| \leq \bar{\lambda}||P_{\alpha_k,0,1}\mathtt{grad}\, h(y^{k-1}) - \mathtt{grad}\, h(y^k)||$$

where γ_k is the geodesic curve such that $\gamma_k(0) = y^k$ and $\gamma_k(1) = \bar{y}$. Now, if if $\bar{y} \in intX \cap S$, taking $k \to +\infty$ and using continuity of the gradient and parallel transport, we have

$$\lim_{k\to+\infty} ||P_{\alpha_k,0,1}\mathtt{grad}\, h(y^{k-1}) - \mathtt{grad}\, h(y^k)|| = 0,$$

Thus,

$$\lim_{k\to\infty} P_{\gamma_k,0,1}g^k = 0.$$

Therefore, there are sequences $\{y^k\}$, $\{f(y^k)\}$ and $\{g^k\}$ with $g^k \in \widehat{\partial}(f + I_X)(y^k)$ such that $\lim_{k\to+\infty} y^k = \bar{y}$, $\lim_{k\to\infty}(f+I_X)(y^k) = f(\bar{y})$ (from the continuity of $(f+I_X)$) and $\lim_{k\to\infty} P_{\gamma_k,0,1}g^k = 0$. Using Definition 3.4 it follows that $0 \in \partial(f + I_X)(\bar{y})$. ■

Corollary 5.2 *Let* $X = S = M$ *(the unconstrained optimization problem). If Assumption E is satisfied and f is a continuous and quasiconvex function on M. Then, the sequence* $\{y^k\}$ *converges to a limiting critical point of f on X.*

Proof. Immediate from above theorem. ■

Corollary 5.3 *Let* $X = S = M$ *(the unconstrained optimization problem). If Assumption E is satisfied and f is a differentiable and pseudoconvex function on M. Then, the sequence* $\{y^k\}$ *converges to an optimal solution of the problem* (p).

Proof. Analogous to Corollary 4.2. ■

5.5.2 The convex case

Theorem 5.8 *Suppose that assumptions D and E are satisfied and that f is convex. If λ_k satisfies*

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

then

$$f(y^n) - f(y^*) \leq \frac{D_h(y^*, y^0)}{\sum\limits_{k=1}^{n}(1/\lambda_k)},$$

where y^* is an optimal point of (P). Furthermore, the sequence $\{y^k\}$ converges to a solution of the problem (P).

Proof. It is immediate from theorems 5.4 and 5.5. ∎

6 Conclusion

Observe that Theorem 4.5 is new in the theory of the proximal point methods.

References

[1] Absil, P.A., Mahony, R., and Andrews, B. Convergence of the iterates of descent methods for analytic cost functions, SIAM opti, 2005.

[2] Arrow, K. J. and Debreu, G. (1954), Existence of an Equilibrium for a Competitive Economy, Econometrica 22, 265-290.

[3] H. Attouch, and J. Bolte, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, to appear in Mathematical Programming.

[4] H. Attouch, and M. Teboulle, Regularized Lotka-Volterra dynamical system as continuous proximal-like method in optimization, Journal Optimiation Theory and Applications, 121 (2004) 541-580.

[5] Barron, N. and Liu, W. (1997) Calculus of Variation l^∞, Applied Math. Optim. 35, 237-263.

[6] J. Bolte, A. Daniilis and A. Lewis, The Lojasiewicz Inequality for Nonsmooth Subanalytic Functions with Applications to Subgradient Dynamical Systems, to appear in SIAM Optimization.

[7] J. Bolte, A. Daniilis and A. Lewis, The Morse-Sard theorem for non-differentiable subanalytic functions, Mathematical Analysis and Applications, v. 31, n. 2, 2006, pp 729-740.

[8] J. Bolte, A. Daniilis, A. Lewis and S. Masahiro,Annales Polonici Mathematici, 87, 2005, pp. 13-25. Clarke critical values of subanalytic Lipschitz continuous functions

[9] Censor, Y. and Lent, A. (1981), An Iterative Row-Action Method for Interval Convex Programming, Journal of Optimization Theory and applications, 34, 3, 321-353.

[10] J. S. Chen, A Proximal-Like Algorithm for Convex Second-Order Cone Programming, Preprint 2006.

[11] Cunha, G. F. M., da Cruz Neto, J. X. and Oliveira, P. R. (2006), A Proximal Point Algorithm with ϕ-Divergence to Quasiconvex Programming, Optimization Online, Preprint.

[12] da Cruz Neto, J. X., Ferreira, O. P., Lucambio Perez, L. and Nmeth, S. Z. Convex-and Monotone-Transformable Mathematical Programming and a proximal-Like Point Method, to apper in Journal Of Global Optimization.

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

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

[15] den Hertog, D., Roos, C. and Terlaky, T. (1991), Inverse Barrier Methods for Linear Programming, Report 91-27, Faculty of Technical Mathematics and Informatics, Delf University of Technology, The Netherlands.

[16] M. Doljansky, and M. Teboulle, An Interior Proximal Algorithm and the Exponential Multiplier Method for Semidefinite programming, SIAM J. Optim., Vol 9, No 1, 1998.

[17] Ferreira, O. P. and Oliveira, P. R. Sub gradient Algorithm on Riemannian Manifolds (1998), Journal of Optimization Theory and Application, 97, 1, 93-104.

[18] Ferreira, O. P. (1997), Programação Matemática em Variedades Riemannianas: Algoritmos Subgradiente e Ponto Proximal, PhD These (supervisor, Paulo Roberto Oliveira).

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

[20] Gabay, D. Minimizing a Differentiable Function over a Differential Manifold (1982) Journal of Optimization Theory and Applications, 37, 2, 177-219.

[21] Ginsberg W. (1973) Concavity and Quasiconcavity in Economics, Journal of Economic Theory 6, 596-605.

[22] Goudou, X. and Munier J. (2005), The Heavy Ball with Friction Method: the quasiconvex case, Submmited.

[23] J. Gromicho, Quasiconvex optimization and location theory, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1998.

[24] Jein-Shan Chen, A proximal-like algorithm for quasiconvex programming, Department of Mathematics, National Taiwan Normal University, Taipe Taiwan, submmited paper 2006.

[25] Luenberger, D. G. The Gradient Projection Method Along Geodesics (1972), *Management Science*, 18, 11, 620-631.

[26] Martinet, B., (1970), Régularisaton, d'inéquations variationelles par approximations successives, Revue Française d'Informatique et de Recherche operationelle, pp. 154-159.

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

[28] Papa Quiroz, E. A. and Oliveira, P. R., A New Self-concordant Barrier for the Hypercube, to appear in Journal of Optimization Theory and Applications JOTA.

[29] Papa Quiroz, E. A. and Oliveira, P. R., Classic and Logarithmic-Quadratic Proximal Point Method for Quasiconvex Minimization. Optimization Online, $http://www.optimization-online.org/DB_HTML/2006/06/1417.html$

[30] E.A. Papa Quiroz and P. Roberto Oliveira, Proximal point methods for quasiconvex and convex functions with Bregman distances on Hadamard manifolds, to appear in Journal of Convex Analysis, 2006.

[31] Pereira, G. J. and Oliveira, P. R. (2005), A New Class of Proximal Algorithms for the Nonlinear Complementary Problems, In: Qi, Liqun; Teo Koklay; Yang, Xiaoqi, Optimization and Control with Applications 1 ed, 69, 552-562.

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

[33] Saigal, R. (1993), The Primal Power Affine Scaling Method, Tech. Report 93-21, Dep. Ind. and Oper. Eng., University of Michigan.

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

[35] Smith, S. T. (1994), Optimization Techniques on Riemannian Manifolds, Fields Institute Communications, AMS, Providence, RI, 3, 113-146.

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

[37] Udriste, C. (1994), Convex Function and Optimization Methos on Riemannian Manifolds, Kluwer Academic Publishers.

[38] Rockafellar, R.T. (1976), Monotone operations and the Proximal Point Method, SIAM J. Control Optimization, 14, 877-898.

[39] Rockafellar, R.T., Wets, R., (1990), Variational Analysis. Grundlehren der Mathematischen, Wissenschaften, 317, Springer.

[40] S.S. Souza, P.R. Oliveira, J.X. Cruz Neto and A. Soubeyran, A proximal Point Algorithm with Bregman Distances for Quasiconvex Optimization over the Positive Orthant, Prepint, 2006.