Journal of Convex Analysis
Volume 16 (2009), No. 1, 49–69

# Proximal Point Methods for Quasiconvex and Convex Functions with Bregman Distances on Hadamard Manifolds

**E. A. Papa Quiroz**
*Universidad Nacional de Callao, Lima, Peru*
*erikpapa@gmail.com*

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

Received: October 25, 2005
Revised manuscript received: November 16, 2006

This paper generalizes the proximal point method using Bregman distances to solve convex and quasiconvex optimization problems on Hadamard manifolds. We will proved that the sequence generated by our method is well defined and converges to an optimal solution of the problem. Also, we obtain the same convergence properties for the classical proximal method, applied to quasiconvex problems. Finally, we give some examples of Bregman distances in non-Euclidean spaces.

*Keywords:* Proximal point algorithms, Hadamard manifolds, Bregman distances, Bregman functions

## 1. Introduction

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 proximal point algorithm with Bregman distance, henceforth abbreviated PBD algorithm, generates a sequence $\{x^k\}$ defined by

$$\text{Given } x^0 \in S,$$

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

where $h$ is a Bregman function with zone $S$, such that $X \cap \bar{S} \neq \phi$, $\lambda_k$ 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 $I\!R^n$. Convergence and rate of convergence results, under appropriate assumptions on the problem, have been proved by several authors for certain choices of the regularization parameters $\lambda_k$, (see, for example, [5, 7, 19, 20]). That algorithm has also been generalized for variational inequalities

ISSN 0944-6532 / \$ 2.50 

problems in Hilbert and Banach spaces, see [3, 4, 16]. Variational Inequalities Problems arise naturally in several Engineering applications and recover optimization problems as a particular case.
On the other hand, generalization of known methods in optimization from Euclidean space to Riemannian manifolds is in a certain sense natural, and advantageous in some cases. For example, we can consider the intrinsic geometry of the manifold, and constrained problems can be seen as unconstrained ones. Another advantage is that certain non convex optimization problems become convex or quasiconvex through the introduction of an adequate Riemannian metric on the manifold, so we can use more efficient optimization techniques, see [10, 13, 15, 21, 27, 31, 33], and the references therein. Besides we can use Riemannian metrics to introduce new algorithms in interior point methods, (see, for example, [8, 12, 26, 28, 30]).

In this paper we generalize the PBD algorithm to solve quasiconvex and convex optimization problems on Hadamard manifolds. Our approach is new but it is related to the work of Ferreira and Oliveira [14]. In that paper, the authors generalized the proximal point method using the intrinsic Riemannian distance for those manifolds. Here, we work with Bregman distances and consider the following regularization parameter conditions

$$\lim_{k\to+\infty} \lambda_k = 0, \quad \text{with } \lambda_k > 0, \tag{1}$$

$$0 < \lambda_k < \bar{\lambda}. \tag{2}$$

For $\lambda_k$ satisfying (1), we obtain the convergence of the PBD algorithm for quasiconvex optimization problems and for $\lambda_k$ satisfying (2) we obtain the convergence of the PBD algorithm for convex optimization problems. The notion of quasiconvexity appears in the value theory in economics [1, 17, 32], in control theory [2] and, recently, in dynamical systems [18].

The paper is divided as follows. In Section 2 we give the notation and some results on Riemannian geometry that we will use along the paper. In Section 3, we recall some facts on convex analysis on Hadamard manifolds. In Section 4 the definition of Bregman function is introduced, besides some properties. Section 5 presents the Moreau-Yosida regularization applied to continuous quasiconvex functions, by considering Bregman distances. In Section 6 we introduce the PDB algorithm with Bregman Distances to solve minimization problems on Hadamard manifolds; we prove the convergence of the sequence generated by the algorithm, for continuous quasiconvex functions, under the condition (1) and for convex functions, when the regularization parameter verifies (2). Section 7 is an application of the precedent developments to the *classical* proximal algorithm for continuous quasiconvex functions, based on the Riemannian distance, thus extending the convex results of [14]. In Section 8 are presented some examples of Bregman distances in non Euclidean spaces, in the following section we give our conclusions and future works.

## 2. Some Tools of Riemannian Geometry

In this section we recall some fundamental properties and notation on Riemannian manifolds. Those basic facts can be seen, for example, in [11] and [29].

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 $||v||_x := \langle v, v\rangle_x^{1/2}$. 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 $L(\alpha) = \int_{t_0}^{t_1} \|\alpha'(t)\| dt$. 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 $\nabla$ is the Levi-Civita connection associated to $(M, G)$. This connection defines an unique covariant derivative $D/dt$, where for each vector field $V$, along a smooth curve $\alpha : [t_0, t_1] \to M$, another vector field is obtained, denoted by $DV/dt$. The parallel transport along $\alpha$ from $\alpha(t_0)$ to $\alpha(t_1)$, denoted by $P_{\alpha,t_0,t_1}$, is an application $P_{\alpha,t_0,t_1} : T_{\alpha(t_0)}M \to T_{\alpha(t_1)}M$ defined by $P_{\alpha,t_0,t_1}(v) = V(t_1)$ where $V$ is the unique vector field along $\alpha$ such that $DV/dt = 0$ and $V(t_0) = v$. Since that $\nabla$ is a Riemannian connection, $P_{\alpha,t_0,t_1}$ 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 2.1.** *Let $M$ be a Hadamard manifold. Then $M$ is diffeomorphic to the Euclidian space $I\!R^n$, $n = dim M$. More precisely, at any point $x \in M$, the exponential mapping $\exp_x : T_xM \to M$ is a global diffeomorphism.*

**Proof.** See [29], 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 2.2.** *Given a geodesic triangle $[x, y, z]$ in a Hadamard manifold, it holds that:*

$$d^2(x, z) + d^2(z, y) - 2\langle \exp_z^{-1} x, \exp_z^{-1} y\rangle_z \le d^2(x, y) \tag{3}$$

**Proof.** See [29], 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

$$H_x^f(v) = \frac{D}{dt}\left(\mathtt{grad}\, f\right)(x) = \nabla_v\, \mathtt{grad}\, f\,(x).$$

## 3. Convex Analysis on Hadamard Manifolds

In this section we give some definitions and results of Convex Analysis on Hadamard manifolds. We refer the reader to [14] and [33] for more details.

**Definition 3.1.** Let $M$ be a Hadamard manifold. A subset $A$ is said convex in $M$ if, for any pair of points the geodesic joining these points is contained in $A$, that is, given $x, y \in A$ and $\gamma : [0,1] \to M$, the geodesic curve such that $\gamma(0) = x$, $\gamma(1) = y$ verifies $\gamma(t) \in A$, for all $t \in [0,1]$.

**Definition 3.2.** Let $A$ be a convex set in a Hadamard manifold $M$ and $f : A \to I\!R$ be a real function. $f$ is called convex on $A$ if for all $x, y \in A$, $t \in [0,1]$, it holds that

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

where $\gamma : [0,1] \to I\!R$ 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 3.3.** *Let $M$ be a Hadamard manifold and $A$ be a convex set in $M$. The function $f : A \to I\!R$ is convex if and only if $\forall\, x, y \in A$ and $\gamma : [0,1] \to M$ (the geodesic joining $x$ to $y$) the function $f(\gamma(t))$ is convex on $[0,1]$.*

**Proof.** See [33], page 61, Theorem 2.2. □

A function $f : A \to I\!R$ is called concave if $-f$ is convex. Furthermore, if $f$ is both convex and concave then $f$ is said to be linear affine on $A$. Observe that a twice differentiable function $f$ on an open convex set $A$ is linear affine if and only if $\langle H_x^f(v), v \rangle_x = 0$, for all $x \in A$ and $v \in T_xM$. Indeed, $\langle H_x^f(v), v \rangle_x = 0$, if and only if $\langle H_x^f(v), v \rangle_x \geq 0$ and $\langle H_x^f(v), v \rangle_x \leq 0$, equivalently $f$ is convex and concave. In other words, $f$ is linear affine if and only if the vector field $\mathtt{grad}\, f$ is parallel.

**Proposition 3.4.** *Let $M$ be a Hadamard manifold and $h : M \to I\!R$ a differentiable function. Let $y \in M$ and define $g : M \to I\!R$ such that*

$$g(x) = \langle \mathtt{grad}\, h(y), \exp_y^{-1} x \rangle_y,$$

*for $x \in M$. Then the following statements are true:*

*i.* $\mathtt{grad}\, g(x) = P_{\gamma,0,1}\, \mathtt{grad}\, h(y)$, *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.** *i.* Let $v \in T_xM$ (arbitrary). Consider the variation of the geodesic $\alpha : [0,1] \times (-\epsilon, \epsilon) \to M$ such that

$$\alpha(t,s) = \exp_y(tu(s)),$$

where $u(s) = \exp_y^{-1} x + sP_{\gamma,1,0}\, v$. Then

$$\begin{aligned}(dg)_x v &= \frac{d}{ds}\left(g(\alpha(1,s))\right)|_{s=0}\\ &= \frac{d}{ds}\langle \operatorname{\mathtt{grad}} h(y), u(s)\rangle\ |_{s=0}\\ &= \langle \operatorname{\mathtt{grad}} h(y), u'(0)\rangle\\ &= \langle \operatorname{\mathtt{grad}} h(y), P_{\gamma,1,0} v\rangle\\ &= \langle P_{\gamma,0,1} \operatorname{\mathtt{grad}} h(y), v\rangle.\end{aligned}$$

Therefore,

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

*ii.* The result follows from *i.* □

Let $M$ be a Hadamard manifold and let $f : M \to I\!R$ be a convex function. Take $y \in M$, the vector $s \in T_yM$ is said to be a subgradient of $f$ at $y$ if

$$f(x) \geq f(y) + \langle s, \exp_y^{-1} x\rangle_y, \tag{4}$$

for all $x \in M$. The set of all subgradients of $f$ at $y$ is called the subdifferential of $f$ at $y$ and is denoted by $\partial f(y)$.

**Theorem 3.5.** *Let $M$ be a Hadamard manifold and let $f : M \to I\!R$ be a convex function. Then, for any $y \in M$, there exists $s \in T_yM$ such that $\forall x \in M$ (4), is true.*

**Proof.** See [14], Theorem 3.3. □

From the previous theorem the subdifferential $\partial f(x)$ of a convex function $f$ at $x \in M$ is nonempty.

**Theorem 3.6.** *Let $M$ be a Hadamard manifold and $f : M \to I\!R$ be a convex function. Then $0 \in \partial f(x)$ if and only if $x$ is a minimum point of $f$ in $M$.*

**Proof.** Immediate. □

**Definition 3.7.** Let $A$ be a convex set in a Hadamard manifold $M$ and $f : A \to I\!R$ be a real function. $f$ is called quasiconvex on $A$ if for all $x, y \in A$, $t \in [0,1]$, 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 3.8.** *Let $A$ be a convex set in a Hadamard manifold $M$. The function $f : A \to I\!R$ is quasiconvex if and only if the set $\{x \in A : f(x) \leq c\}$ is convex for each $c \in I\!R$.*

**Proof.** See [33], page 98, Theorem 10.2. □

**Theorem 3.9.** *Let $C$ be a closed convex set in a Hadamard manifold $M$. Take $y \in M$ arbitrary, then there exists a unique projection $z = P_C y$. Furthermore, the following inequality holds*

$$\langle \exp_z^{-1} y, \exp_z^{-1} x \rangle \leq 0, \tag{5}$$

*for all $x \in C$.*

**Proof.** See in [14], Propositions 3.1 and 3.2. □

## 4. Bregman Distances and Functions on Hadamard Manifolds

To construct generalized proximal point algorithms with Bregman distances for solving optimization problems on Hadamard manifolds, it is necessary to extend the definitions of Bregman distances and Bregman functions to that framework. Starting from Censor and Lent [6] definition, we propose the following.
Let $M$ be a Hadamard manifold and $h : M \to I\!R$ be a strictly convex and differentiable function. Then, the *Bregman distance* associated to $h$, denoted by $D_h$, is defined as a function $D_h(.,.) : M \times M \to I\!R$ such that

$$D_h(x, y) := h(x) - h(y) - \langle \mathtt{grad}\, h(y), \exp_y^{-1} x \rangle_y. \tag{6}$$

Notice that the expression of the Bregman distance depends on the definition of the metric. Some examples for different manifolds will be given in Section 8. Let us adopt the following notation for the partial level sets of $D_h$. For $\alpha \in I\!R$, take

$$\Gamma_1(\alpha, y) := \{x \in M : D_h(x, y) \leq \alpha\},$$
$$\Gamma_2(x, \alpha) := \{y \in M : D_h(x, y) \leq \alpha\}.$$

**Definition 4.1.** Let $M$ be a Hadamard manifold. A real function $h : M \to I\!R$ is called a *Bregman function* with zone $M$ if:

a. $h$ is continuous differentiable on $M$;
b. $h$ is strictly convex on $M$;
c. For all $\alpha \in I\!R$ the partial level sets $\Gamma_1(\alpha, y)$ and $\Gamma_2(x, \alpha)$ are bounded for every $y \in M$ and $x \in M$, respectively.

We denote the family of Bregman functions by $\mathcal{B}$.

**Lemma 4.2.** *Let $h \in \mathcal{B}$. Then*

*i.* $\mathtt{grad}\, D_h(., y)(x) = \mathtt{grad}\, h(x) - P_{\gamma,0,1}\, \mathtt{grad}\, h(y)$, *for all $x, y \in M$, 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 $M$ for all $y \in M$.*
*iii.* *For all $x, y \in M$, $D_h(x, y) \geq 0$ and $D_h(x, y) = 0$ if and only if $x = y$.*

**Proof.** *i.* From Proposition 3.4, *i*, we obtain the result.

*ii.* As $h$ is strictly convex in $M$ and $\langle \mathtt{grad}\, h(y), \exp_y^{-1} x \rangle_y$, is a linear affine function (Proposition 3.4 *ii*) then the result follows.

*iii.* Use, again, the strict convexity of $h$. □

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 $\texttt{grad}\, D_h(x,y)$ to mean $\texttt{grad}\, D_h(.,y)(x)$. So, if $\gamma$ is the geodesic curve such that $\gamma(0) = y$ and $\gamma(1) = x$, from Lemma 4.2 *i*, we obtain

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

**Proposition 4.3.** *Let $h \in \mathcal{B}$, then*

*i.* *If $\lim_{k\to+\infty} y^k = y^* \in M$, then $\lim_{k\to+\infty} D_h(y^*, y^k) = 0$;*

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

**Proof.** *i.* Let $y, y^* \in M$, then $D_h(y^*, y^k) = h(y^*) - h(y^k) - \langle \texttt{grad}\, h(y^k), \exp_{y^k}^{-1} y^* \rangle_{y^k}$. Taking $k \to +\infty$ and using Definition 4.1 a, we obtain the result.

*ii.* Let $\bar{z} \in M$ an arbitrary limit point of $\{z^k\}$, then there exists $\{z^{k_j}\}$ such that $z^{k_j} \to \bar{z}$, when $j \to +\infty$. Therefore $D_h(z^{k_j}, y^{k_j}) = h(z^{k_j}) - h(y^{k_j}) - \langle \texttt{grad}\, h(y^{k_j}), \exp_{y^{k_j}}^{-1} z^{k_j} \rangle_{y^{k_j}}$. Taking $j \to +\infty$ and using Definition 4.1 a, we obtain that $D_h(\bar{z}, y^*) = 0$. From Lemma 4.2 *iii*, we get $\bar{z} = y^*$. Thus, there exists an unique limit point of $\{z^k\}$ and therefore the aimed result is obtained. □

**Definition 4.4.** Let $\Omega \subset M$, and let $y \in M$. A point $Py \in \Omega$ for which

$$D_h(Py, y) = \min_{x\in\Omega} D_h(x, y) \tag{7}$$

is called a $D_h-$*projection* of the point $y$ on the set $\Omega$.

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

**Lemma 4.5.** *Let $\Omega \subset M$ a nonempty closed convex set and $h \in \mathcal{B}$. Then, for any $y \in M$, there exists a unique $D_h-$projection $Py$ of the point $y$ on $\Omega$.*

**Proof.** For any $x \in \Omega$, the set

$$B := \{z \in M : D_h(z, y) \le D_h(x, y)\}$$

is bounded (from Definition 4.1 c) and closed (because $D_h(., y)$ is continuous in $M$, due to Definition 4.1 a). Therefore, the set

$$T := \Omega \cap B$$

is nonempty 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 $z \in \Omega$ which lies outside $B$

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

hence, $x^*$ satisfies (7). The uniqueness follows from the strict convexity of $D_h(., y)$, therefore

$$x^* = Py.$$

□

**Lemma 4.6.** *Let $h \in \mathcal{B}$ and $y \in M$. Let $Py$ the $D_h-$projection on some nonempty closed convex set $\Omega$. Then, the function*

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

*is linear affine on $M$.*

**Proof.** From (6)

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

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

**Proposition 4.7.** *Let $h \in \mathcal{B}$ and $\Omega \subset M$ a nonempty closed convex set. Let $y \in M$ and $Py$ denotes the $D_h-$projection of $y$ on $\Omega$. Then, for any $x \in \Omega$, 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 4.6 the function

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

is linear affine on $M$. Then in particular $G(\gamma(t))$ is convex for $t \in (0, 1)$ (see Theorem 3.3). 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

$$\begin{aligned} &(1/t)\,(D_h(\gamma(t), y) - D_h(Py, y)) - (1/t)D_h(\gamma(t), Py) \\ &\leq D_h(x, y) - D_h(x, Py) - D_h(Py, y). \end{aligned} \tag{8}$$

As $\Omega$ is convex, and $x, Py \in \Omega$, we have $\gamma(t) \in \Omega$ 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 (8) 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 $z \in M$, we can take the limit in $t$, obtaining

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

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

## 5. Regularization

Let $M$ be a Hadamard manifold and $f : M \to I\!R$ a real function. Let $h : M \to I\!R$ a differentiable function in $M$. For $\lambda > 0$, the Moreau-Yosida regularization $f_\lambda : M \to I\!R$ of $f$ is defined by

$$f_\lambda(y) = \inf_{x \in M}\{f(x) + \lambda D_h(x, y)\} \tag{9}$$

where $D_h(x, y)$ is given in (6). Now, we prove the existence of a (non necessarily unique) solution to (9), under some conditions on $h$ and $f$.

**Proposition 5.1.** *If $f : M \to I\!R$ is a bounded below quasiconvex and continuous function and $h \in \mathcal{B}$ then, for every $y \in M$ and $\lambda > 0$ there exists a point, denoted by $x_f(y, \lambda)$, such that*

$$f_\lambda(y) = f(x_f(y, \lambda)) + \lambda D_h(x_f(y, \lambda), y). \tag{10}$$

**Proof.** Let $\beta$ a lower bound for $f$ on $M$, then

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

for all $x \in M$. It follows from Definition 4.1 c, that the level sets of the function $f(.) + \lambda D_h(., y)$ are bounded. Also, this function is continuous on $M$, due to Definition 4.1 a, and the hypothesis on $f$. So, the level sets of $(f(.)+\lambda D_h(., y))$ are closed, hence compact. Now, from continuity and compactness arguments, $f(.) + \lambda D_h(., y)$ has a minimum on $M$. The equality (10) follows from (9). □

Now, we let a different condition on $h$ that also ensures the existence of solution to (9). Let us introduce the following definition.

**Definition 5.2.** A function $g : M \to I\!R$ is 1−coercive at $y \in M$ if

$$\lim_{d(x,y)\to+\infty} \frac{g(x)}{d(x, y)} = +\infty.$$

Note that if $g : M \to I\!R$ is a continuous 1−coercive function at $y \in M$, then it is easy to show that the minimizer set of $g$ on $M$ is nonempty.

**Lemma 5.3.** *If $f : M \to I\!R$ is bounded below, $\lambda > 0$, and $h : M \to I\!R$ is 1−coercive at $y \in M$, then the function $f(.) + \lambda D_h(., y) : M \to I\!R$ is 1−coercive at $y \in M$.*

**Proof.** As above, let $\beta$ a lower bound for $f$. Then:

$$\begin{aligned}\frac{f(x) + \lambda D_h(x, y)}{d(x, y)} &\geq \frac{\beta}{d(x, y)} + \lambda\frac{D_h(x, y)}{d(x, y)}\\ &= \frac{\beta}{d(x, y)} + \lambda\frac{h(x)}{d(x, y)} - \lambda\frac{h(y)}{d(x, y)} - \lambda\left\langle \mathtt{grad}\, h(y), \frac{\exp_y^{-1} x}{d(x, y)}\right\rangle_y\\ &\geq \frac{\beta}{d(x, y)} + \lambda\frac{h(x)}{d(x, y)} - \lambda\frac{h(y)}{d(x, y)} - \lambda||\,\mathtt{grad}\, h(y)||,\end{aligned}$$

where the equality comes from the definition of $D_h$, and the last inequality results from the application of Cauchy inequality, and the fact that $||\exp_y^{-1} x|| = d(x,y)$. Taking $d(x,y) \to +\infty$, we use the 1-coercivity assumption of $h$ at $y$, to get

$$\lim_{d(x,y)\to+\infty} \frac{(f(.) + \lambda D_h(.,y))(x)}{d(x,y)} = +\infty.$$

□

**Proposition 5.4.** *Let $h : M \to I\!R$ be a $1-$coercive strictly convex function at $y \in M$, $f : M \to I\!R$ a continuous quasiconvex function and bounded below. Then, there exists some point $x_f(y,\lambda)$ such that*

$$f_\lambda(y) = f(x_f(y,\lambda)) + \lambda D_h(x_f(y,\lambda), y)$$

**Proof.** The result follows from the Lemma above. □

## 6. Proximal Point Algorithm with Bregman Distances

We are interested in solving the optimization problem:

$$(p) \ \min_{x\in M} f(x)$$

where $M$ is a Hadamard manifold. The main convergence results will be given when $f$ is a continuous quasiconvex or a convex function on $M$. The PBD algorithm is defined as

$$x^0 \in M, \tag{11}$$
$$x^k \in \texttt{arg}\min_{x\in M}\{f(x) + \lambda_k D_h(x, x^{k-1})\}, \tag{12}$$

where $h$ is a Bregman function with zone $M$, $D_h$ is as in (6) and $\lambda_k$ is a positive parameter. Observe that if $f$ is a bounded below quasiconvex and continuous function, the above iteration exists, see Proposition 5.1.
In the particular case where $M$ is the Euclidean space $I\!R^n$, and $h(x) = (1/2)x^T x$, we have

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

Therefore, the PBD algorithm is another natural generalization of the proximal point algorithm on $I\!R^n$, see [14].
We will use the following parameter conditions to the PBD algorithm:

$$0 < \lambda_k < \bar{\lambda}, \tag{13}$$
$$\text{or} \quad \lim_{k\to+\infty} \lambda_k = 0, \quad \text{with } \lambda_k > 0, \tag{14}$$

Note that (13) implies

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

Observe that the above condition is a minimal assumption in Euclidean and in Banach spaces in proximal point methods. Next, we assume the following assumption.

**Assumption 6.1.** The optimal set of the problem $(p)$, denoted by $X^*$, is nonempty.

### 6.1. Convergence Results

In this subsection we prove the convergence of the proposed algorithm. Our results are motivated by the works [19, 7, 5].

#### 6.1.1. The quasiconvex case

**Theorem 6.2.** *Assume Assumption* 6.1 *and that* $f$ *is a continuous quasiconvex function. Then, the sequence* $\{x^k\}$*, generated by the PBD algorithm, is bounded.*

**Proof.** Since $x^k$ satisfies (12) we have

$$f(x^k) + \lambda_k D_h(x^k, x^{k-1}) \leq f(x) + \lambda_k D_h(x, x^{k-1}), \quad \forall x \in M. \tag{15}$$

Hence, $\forall x \in M$ such that $f(x) \leq f(x^k)$ is true that

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

Therefore $x^k$ is the unique $D_h-$projection of $x^{k-1}$ on the closed convex set (see Theorem 3.8)

$$\Omega := \{x \in M : f(x) \leq f(x^k)\}.$$

Using Proposition 4.7 and the fact that $X^* \subset \Omega$ we have

$$0 \leq D_h(x^k, x^{k-1}) \leq D_h(x^*, x^{k-1}) - D_h(x^*, x^k) \tag{16}$$

for every $x^* \in X^*$. Thus

$$D_h(x^*, x^k) \leq D_h(x^*, x^{k-1}). \tag{17}$$

This means that $\{x^k\}$ is $D_h-$ Fejér monotone with respect to set $X^*$. We can now apply Definition 4.1 d, to see that $x^k$ is bounded, because

$$x^k \in \Gamma_2(x^*, \alpha),$$

with $\alpha = D_h(x^*, x^0)$. □

**Proposition 6.3.** *Under the assumptions of the precedent theorem, the following facts are true*

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

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

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

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

**Proof.** *a.* From (17), $\{D_h(x^*, x^k)\}$ is a bounded below nondecreasing sequence and hence convergent.

*b.* Taking limit when $k$ goes to infinity in (16) and using the previous result we obtain $\lim_{k\to+\infty} D_h(x^k, x^{k-1}) = 0$, as desired.

*c.* Considering $x = x^{k-1}$ in (15) it follows that

$$0 \leq D_h(x^k, x^{k-1}) \leq (1/\lambda_k)(f(x^{k-1}) - f(x^k)),$$

since $D_h(x^{k-1}, x^{k-1}) = 0$. Thus $\{f(x^k)\}$ is nondecreasing. *d.* Taking $z^k = x^{k_j+1}$ and $y^k = x^{k_j}$ in Proposition 4.3 *ii*, we obtain the result. □

**Theorem 6.4.** *Under Assumption* 6.1 *and that $f$ is a continuous quasiconvex function, any limit point of* $\{x^k\}$ *generated by the PBD algorithm with* $\lambda_k$ *satisfying* (14) *is an optimal solution of* $(p)$.

**Proof.** Let $x^* \in X^*$ and let $\bar{x} \in M$ be a cluster point of $\{x^k\}$ then, there exists a subsequence $\{x^{k_j}\}$ such that

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

As $x^{k_j}$ is a solution of (12) we have

$$f(x^{k_j}) + \lambda_{k_j} D_h(x^{k_j}, x^{k_j - 1}) \leq f(x^*) + \lambda_{k_j} D_h(x^*, x^{k_j - 1}).$$

This rewrites

$$\lambda_{k_j}(D_h(x^{k_j}, x^{k_j-1}) - D_h(x^*, x^{k_j-1})) \leq f(x^*) - f(x^{k_j}).$$

Using the differential characterization of convex functions for $D_h(., x^{k_j-1})$ gives:

$$\lambda_{k_j} \langle \texttt{grad}\, D_h(x^*, x^{k_j-1}), \exp_{x^*}^{-1} x^{k_j} \rangle_{x^*} \leq f(x^*) - f(x^{k_j}).$$

Taking, above, $j \to +\infty$, considering the hypothesis (14) and using the continuity of $\texttt{grad}\, D_h(x^*, .)$ and $\exp_{x^*}^{-1}$ we obtain

$$f(\bar{x}) \leq f(x^*).$$

Therefore, any limit point is an optimal solution of the problem $(p)$. □

**Theorem 6.5.** *Under Assumption* 6.1 *and that $f$ is a quasiconvex and continuous function, the sequence* $\{x^k\}$ *generated by the PBD algorithm, with* $\lambda_k$ *satisfying* (14), *converges to an optimal solution of* $(p)$.

**Proof.** From Theorem 6.2 $\{x^k\}$ is bounded so there exists a convergent subsequence. Let $\{x^{k_j}\}$ be a subsequence of $\{x^k\}$ such that $\lim_{j \to +\infty} x^{k_j} = x^*$. From Proposition 4.3 *i*, it is true that

$$\lim_{j \to +\infty} D_h(x^*, x^{k_j}) = 0.$$

Now, from Theorem 6.4, $x^*$ is an optimal solution of $(p)$, so from Proposition 6.3 *a*, $D_h(x^*, x^k)$ is a convergent sequence, with the subsequence converging to 0, hence the overall sequence converges to 0, that is,

$$\lim_{k \to +\infty} D_h(x^*, x^k) = 0.$$

To prove that $\{x^k\}$ has a unique limit point let $\bar{x} \in X^*$ be another limit point of $\{x^k\}$. Then $\lim_{l \to +\infty} D_h(x^*, x^{k_l}) = 0$ with $\lim_{l \to +\infty} x^{k_l} = \bar{x}$. So, from Proposition 4.3 *ii*, $x^* = \bar{x}$. It follows that $\{x^k\}$ cannot have more than one limit point and therefore,

$$\lim_{k \to +\infty} x^k = x^* \in X^*.$$

□

#### 6.1.2. The convex case

**Theorem 6.6.** *Suppose that Assumption* 6.1 *is satisfied and that* $f$ *is convex. If* $\lambda_k$ *satisfies* (13) *then, any limit point of* $\{x^k\}$ *is an optimal solution of the problem* $(p)$.

**Proof.** Let $\bar{x} \in M$ be a limit point of $\{x^k\}$ then, there exists a subsequence $\{x^{k_j}\}$ such that

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

From (12) and Theorem 3.6

$$0 \in \partial[f(.) + \lambda_{k_j+1} D_h(., x^{k_j})](x^{k_j+1}),$$

or,

$$-\lambda_{k_j+1}\, \texttt{grad}\, D_h(x^{k_j+1}, x^{k_j}) \in \partial f(x^{k_j+1}).$$

Let $\gamma_{k_j}$ be the geodesic curve such that $\gamma_{k_j}(0) = x^{k_j}$ and $\gamma_{k_j}(1) = x^{k_j+1}$. By Lemma 4.2 *i*, we obtain

$$\lambda_{k_j+1}[P_{\gamma_{k_j},0,1}\, \texttt{grad}\, h(x^{k_j}) - \texttt{grad}\, h(x^{k_j+1})] \in \partial f(x^{k_j+1}).$$

Let $x^*$ be an optimal solution of $(p)$. Using (4) for $x = x^*$ and $y = x^{k_j+1}$ we have

$$f(x^*) - f(x^{k_j+1}) \geq \langle y^{k_j}, \exp^{-1}_{x^{k_j+1}} x^* \rangle_{x^{k_j+1}} \tag{18}$$

where,

$$y^{k_j} := \lambda_{k_j+1}[P_{\gamma_{k_j},0,1}\, \texttt{grad}\, h(x^{k_j}) - \texttt{grad}\, h(x^{k_j+1})].$$

On the other hand, from Cauchy inequality

$$|\langle y^{k_j}, \exp^{-1}_{x^{k_j+1}} x^* \rangle_{x^{k_j+1}}| \leq ||y^{k_j}||_{x^{k_j+1}} || \exp^{-1}_{x^{k_j+1}} x^*||_{x^{k_j+1}}.$$

We have $|| \exp^{-1}_{x^{k_j+1}} x^*||_{x^{k_j+1}} = d(x^*, x^{k_j+1})$, also, from Theorem 6.2, there exists $M > 0$ such that

$$|\langle y^{k_j}, \exp^{-1}_{x^{k_j+1}} x^* \rangle_{x^{k_j+1}}| \leq M ||y^{k_j}||_{x^{k_j+1}}.$$

Using this fact in the inequality (18) we obtain

$$f(x^*) - f(x^{k_j+1}) \geq -M ||y^{k_j}||_{x^{k_j+1}}. \tag{19}$$

To conclude the proof we will show that

$$\lim_{j\to+\infty} ||y^{k_j}||_{x^{k_j+1}} = 0.$$

Indeed, using the continuity of the parallel transport, continuity of the gradient field, Proposition 6.3 *d*, and the boundedness of $\{\lambda_k\}$ we obtain $\lim_{j\to+\infty} ||y^{k_j}||_{x^{k_j+1}} = 0$, as wanted. Finally, taking $j \to +\infty$ in (19), use the continuity of $f$ to get

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

Therefore, any limit point is an optimal solution of the problem $(p)$. □

**Theorem 6.7.** *Under Assumptions* 6.1 *and that* $f$ *is convex, the sequence* $\{x^k\}$ *generated by the PBD algorithm, with* $\lambda_k$ *satisfying* (13), *converges to an optimal solution of* $(p)$.

**Proof.** Analogous to the proof of Theorem 6.5. □

## 7. Proximal Methods with Riemannian Distances

In this section we adapt the previous results to (classical) proximal methods on Hadamard manifolds, based on the Riemannian distance function. Observe that in these manifolds the distance function is convex (see [29]). Our results are an extension of [14], who considered the convex case.

### 7.1. The quasiconvex case

We consider, at first, the Moreau-Yosida regularization. For $\lambda > 0$, let:

$$\varphi_\lambda(y) = \inf_{x \in M}\{f(x) + \frac{\lambda}{2} d^2(x, y)\}$$

We have:

**Proposition 7.1.** *If $f : M \to I\!R$ is a bounded below continuous quasiconvex function, then, for every $y \in M$ and $\lambda > 0$, there exists some point, denoted by $x(y, \lambda)$, such that*

$$\varphi_\lambda(y) = f(x(y, \lambda)) + \frac{\lambda}{2} d^2(x(y, \lambda), y).$$

**Proof.** Clearly, the function $d^2(.,.)$ is 1-coercive. Therefore, Lemma 5.3 and Proposition 5.4 are easily adaptable. □

Now, we will present the convergence results, for the PPA algorithm, defined by

$$x^0 \in M, \tag{20}$$

$$x^k \in \texttt{arg}\min_{x \in M}\{f(x) + (\lambda_k/2) d^2(x, x^{k-1})\}. \tag{21}$$

Observe, analogously to the PBD algorithm, that if $f$ is a bounded below quasiconvex and continuous function, the above iteration exists.

**Theorem 7.2.** *Assume Assumption* 6.1 *and that $f$ is a continuous quasiconvex function. Then, the sequence $\{x^k\}$, generated by the PPA algorithm, is bounded.*

**Proof.** Regarding Theorem 6.2, and letting $D_h(.,.)$ as $d^2(.,.)$, we see that all the steps of the proof can be copied, the unique point that deserves a justification being the inequality (16), which writes here

$$0 \leq d^2(x^k, x^{k-1}) \leq d^2(x^*, x^{k-1}) - d^2(x^*, x^k) \tag{22}$$

for every $x^* \in X^*$. Indeed, it is a consequence of Theorem 2.2 and Theorem 3.9: take, in (3), $x = x^*$, $y = x^{k-1}$ and $z = x^k$, and using (5) we obtain (22).

**Proposition 7.3.** *Under the assumptions of the precedent theorem, the following facts are true*

*a.* *For all $x^* \in X^*$ the sequence $\{d^2(x^*, x^k)\}$ is convergent;*
*b.* $\lim_{k\to+\infty} d^2(x^k, x^{k-1}) = 0$;
*c.* *$\{f(x^k)\}$ is non increasing;*
*d.* *If $\lim_{j\to+\infty} x^{k_j} = \bar{x}$ then, $\lim_{j\to+\infty} x^{k_j+1} = \bar{x}$.*

**Proof.** For $a$, $b$ and $c$, see, respectively, the proof of $a$, $b$ and $c$, in Proposition 6.3, with the obvious substitution of $D_h(.,.)$ by $d^2(.,.)$. For $d$, 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. □

**Theorem 7.4.** *Under Assumption* 6.1 *and that $f$ is a continuous quasiconvex function, any limit point of $\{x^k\}$ generated by the PPA algorithm with $\lambda_k$ satisfying* (14) *is an optimal solution of* $(p)$.

**Proof.** The adaptation of the corresponding proof of Theorem 6.4 is immediate. □

**Theorem 7.5.** *Under Assumption* 6.1 *and that $f$ is a continuous quasiconvex function, the sequence $\{x^k\}$ generated by the PPA algorithm, with $\lambda_k$ satisfying* (14)*, converges to an optimal solution of* $(p)$.

**Proof.** As in Theorem 6.5. □

## 7.2. The convex case

Convergence results for the convex case, using $\lambda_k$ such that $\sum_{k=1}^{n}(1/\lambda_k) = +\infty$, have been proved in [14]. Here we present a rate estimative for the residual $f(x^k) - f(x^*)$, where $x^*$ is an optimal solution of $(p)$.

**Theorem 7.6.** *Let $f$ a convex function. Under Assumption* 6.1 *and $\lambda_k$ such that $\sum_{k=1}^{n}(1/\lambda_k) = +\infty$. Then the sequence $\{x^k\}$, generated by the PPA algorithm, converges to a solution of $(p)$ and satisfies*

$$f(x^n) - f(x^*) \leq \frac{d(x^*, x^0)}{2\sum_{k=1}^{n}(1/\lambda_k)}.$$

**Proof.** The convergence proof has been given in [14], so we will prove the second part. From (21) we have

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

for all $x \in M$. Taking $x = x^{k-1}$ gives

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

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

$$\begin{aligned}(\lambda_k/2)\sigma_{k-1}d^2(x^k, x^{k-1}) &\leq \sigma_{k-1}(f(x^{k-1}) - f(x^k))\\ &= \sigma_{k-1}f(x^{k-1}) - (\sigma_k - (1/\lambda_k))f(x^k)\\ &= \sigma_{k-1}f(x^{k-1}) - \sigma_k f(x^k) + (1/\lambda_k)f(x^k).\end{aligned}$$

Taking the sum over $k$, from 1 to $n$ and multiplying by 2 we have

$$\sum_{k=1}^{n} \lambda_k \sigma_{k-1} d^2(x^k, x^{k-1}) \leq -2\sigma_n f(x^n) + \sum_{k=1}^{n} (2/\lambda_k) f(x^k) \tag{23}$$

On the other hand, it can be proved (see [14], Lemma 6.2) that

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

Then,

$$\sum_{k=1}^{n} (2/\lambda_k)(f(x^k) - f(x)) \leq \sum_{k=1}^{n} \left( d^2(x, x^{k-1}) - d^2(x^k, x^{k-1}) - d^2(x, x^k) \right).$$

This implies that

$$\sum_{k=1}^{n} (2/\lambda_k) f(x^k) \leq 2\sigma_n f(x) + d^2(x, x^0) - d^2(x, x^n) - \sum_{k=1}^{n} d^2(x^k, x^{k-1}).$$

The above inequality and (23) give

$$\sum_{k=1}^{n} \lambda_k \sigma_{k-1} d^2(x^k, x^{k-1}) + 2\sigma_n f(x^n)$$
$$\leq 2\sigma_n f(x) + d^2(x, x^0) - d^2(x, x^n) - \sum_{k=1}^{n} d^2(x^k, x^{k-1}).$$

Thus

$$2\sigma_n (f(x^n) - f(x)) \leq d^2(x, x^0) - d^2(x, x^n) - \sum_{k=1}^{n} (1 + \lambda_k \sigma_{k-1}) d^2(x^k, x^{k-1}).$$

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

$$2\sigma_n (f(x^n) - f(x)) \leq d^2(x, x^0) - d^2(x, x^n) - \sum_{k=1}^{n} \lambda_k \sigma_k d^2(x^k, x^{k-1}).$$

Therefore

$$f(x^n) - f(x) \leq \frac{d^2(x, x^0)}{2\sigma_n}.$$

Taking $x^* \in X^*$ in the previous inequality we conclude the proof. □

## 8. Examples

Examples 8.1 to 8.4 are Hadamard manifolds with zero sectional curvature. In Example 8.5, it is negative. In all examples, we present the general Bregman distance formulation, depending on the $h$ function, and, due to possible using of the classical proximal method, the Riemannian distance.

**Example 8.1.** The Euclidean space is a Hadamard manifold with the metric $G(x) = I$ (its sectional curvature is null). Its geodesic curves are the straight lines and the Bregman distance has the form

$$D_h(x, y) = h(x) - h(y) - \sum_{i=1}^{n} (x_i - y_i) \frac{\partial h(y)}{\partial y_i}.$$

The distance is given by

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

**Example 8.2.** Let $I\!R^n$ with the metric

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

Thus $(I\!R^n, G(x))$ is a Hadamard manifold isometric to $(I\!R^n, I)$ through the application $\phi : I\!R^n \to I\!R^n$ defined by $\Phi(x) = (x_1, x_2, \dots, x_{n-1}, x_{n-1}^2 - x_n)$, see [9]. The geodesic curve, joining the points $\gamma(0) = y$ and $\gamma(1) = x$ is $\gamma(t) = (\gamma_1, \dots, \gamma_n)$, where $\gamma_i(t) = y_i + t(x_i - y_i), \forall i = 1, \dots n-1$ and $\gamma_n(t) = y_n + t((x_n - y_n) - 2(x_{n-1} - y_{n-1})^2) + 2t^2(x_{n-1} - y_{n-1})^2$. Then the Bregman distance is

$$D_h(x, y) = h(x) - h(y) - \sum_{i=1}^{n} (x_i - y_i) \frac{\partial h(y)}{\partial y_i} + 2 \frac{\partial h(y)}{\partial y_n} (x_n - y_n).$$

The Riemannian distance, see [9], is

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

**Example 8.3.** $M = I\!R_{++}^n$ with the Dikin metric $X^{-2}$ is a Hadamard manifold. Defining $\pi : M \to I\!R^n$ such that $\pi(x) = (-\ln x_1, \dots, -\ln x_n)$, it can be proved that $\pi$ is an isometry. It is well known, see for example [27], that the geodesic curve joining the points $\gamma(0) = y$ and $\gamma(1) = x$ is

$$\gamma(t) = \left( x_1^t y_1^{1-t}, \dots, x_n^t y_n^{1-t} \right),$$

with

$$\gamma'(t) = \left( x_1^t y_1^{1-t} (\ln x_1 - \ln y_1), \dots, x_n^t y_n^{1-t} (\ln x_n - \ln y_n) \right).$$

Then the Bregman distance is:

$$D_h(x, y) = h(x) - h(y) - \sum_{i=1}^{n} y_i \ln(x_i / y_i) \frac{\partial h(y)}{\partial y_i}.$$

The Riemannian distance is

$$d(x,y) = \left[ \sum_{i=1}^{n} \left( \ln \frac{y_i}{x_i} \right)^2 \right]^{1/2}$$

**Example 8.4.** Let $M = (0,1)^n$. We will consider three metrics.

1. $(M, X^{-2}(I-X)^{-2})$ is a Hadamard manifold; it is isometric to $\mathbb{R}^n$ through the function $\pi(x) = \left( \ln\left(\frac{x_1}{1-x_1}\right), \ldots, \ln\left(\frac{x_n}{1-x_n}\right) \right)$. The geodesic curve, see [24], joining the points $\gamma(0) = y$ and $\gamma(1) = x$ is $\gamma(t) = (\gamma_1, \ldots, \gamma_n)$ such that

$$\gamma_i(t) = \frac{1}{2} + \frac{1}{2} \tanh \left[ (1/2) \left\{ \ln\left( \frac{x_i}{1-x_i} \right) - \ln\left( \frac{y_i}{1-y_i} \right) \right\} t + (1/2) \ln\left( \frac{y_i}{1-y_i} \right) \right],$$

with

$$\gamma_i'(t) = \frac{\ln\left(x_i/(1-x_i)\right) - \ln\left(y_i/(1-y_i)\right)}{4\cosh\left((1/2)t + (1/2)\ln(y_i/(1-y_i))\right)}.$$

Then, the Bregman distance is

$$D_h(x,y) = h(x) - h(y) - \sum_{i=1}^{n} \frac{(1-y_i)^2}{4y_i^2 \cosh^2(1/2)} \left\{ \ln\left( \frac{x_i}{1-x_i} \right) - \ln\left( \frac{y_i}{1-y_i} \right) \right\} \frac{\partial h(y)}{\partial y_i}.$$

The Riemannian distance is given by:

$$d(x,y) = \left[ \sum_{i=1}^{n} \left( \ln\left( \frac{y_i}{1-y_i} \right) - \ln\left( \frac{x_i}{1-x_i} \right) \right)^2 \right]^{\frac{1}{2}}.$$

2. $(M, csc^4(\pi x))$ is a Hadamard manifold, isometric to $\mathbb{R}^n$, through the function $\pi(x) = \frac{1}{\pi}\left(\cot(\pi x_1), \ldots, \cot(\pi x_n)\right)$. The geodesic curve, see [23], joining the points $\gamma(0) = y$ and $\gamma(1) = x$ is $\gamma(t) = (\gamma_1, \ldots, \gamma_n)$ such that

$$\gamma_i(t) = \frac{1}{\pi} \arg\cot[(\cot \pi x_i - \cot \pi y_i)t + \cot(\pi y_i)],$$

with

$$\gamma_i'(t) = (1/\pi)\left(\cot(\pi y_i) - \cot(\pi x_i)\right) \sin^2(\pi \gamma_i(t)),$$

and the Bregman distance is

$$D_h(x,y) = h(x) - h(y) - \sum_{i=1}^{n} \frac{1}{\pi} (\cot(\pi y_i) - \cot(\pi x_i)) \sin^2(\pi y_i) \frac{\partial h(y)}{\partial y_i}.$$

The Riemannian distance is

$$d(x,y) = \left[ \sum_{i=1}^{n} \left[ \cot(\pi y_i) - \cot(\pi x_i) \right]^2 \right]^{1/2}.$$

3. Finally, we consider $(M, \csc^2(\pi x))$. It is a Hadamard manifold isometric to $\mathbb{R}^n$, see [22]. The geodesic curve joining the points $\gamma(0) = y$ and $\gamma(1) = x$ is $\gamma(t) = (\gamma_1, \ldots, \gamma_n)$ such that

$$\gamma_i(t) = \psi^{-1}\left(\psi(y_i) + t(\psi(x_i) - \psi(y_i))\right),$$

where

$$\psi(\tau) := \ln\left(\csc(\pi\tau) - \cot(\pi\tau)\right).$$

So,

$$\gamma_i'(t) = (1/\pi)\ln\left(\frac{\csc(\pi x_i) - \cot(\pi x_i)}{\csc(\pi y_i) - \cot(\pi y_i)}\right)\sin(\pi\gamma_i(t)).$$

Therefore, the Bregman distance is

$$D_h(x, y) = h(x) - h(y) - \frac{1}{\pi}\sum_{i=1}^{n}\ln\left(\frac{\csc(\pi x_i) - \cot(\pi x_i)}{\csc(\pi y_i) - \cot(\pi y_i)}\right)\sin(\pi y_i)\frac{\partial h(y)}{\partial y_i}.$$

The Riemannian distance is

$$d(x, y) = \left[\sum_{i=1}^{n}\left(\psi(y_i) - \psi(x_i)\right)^2\right]^{1/2}.$$

**Example 8.5.** $M = \mathcal{S}^n_{++}$, the set of the $n \times n$ positive definite symmetric matrices, with the metric given by the Hessian of $-\ln det(X)$, is a Hadamard manifold with nonpositive curvature. The geodesic curve joining the points $\gamma(0) = Y$ and $\gamma(1) = X$, see [22], is given by

$$\gamma(t) = Y^{1/2}(Y^{-1/2}XY^{-1/2})^t Y^{1/2},$$

with

$$\gamma'(t) = Y^{1/2}\ln(Y^{-1/2}XY^{-1/2})(Y^{-1/2}XY^{-1/2})^t Y^{1/2}.$$

Then, the Bregman distance is

$$D_h(X, Y) = h(X) - h(Y) - \mathtt{tr}[\nabla h(Y)Y^{1/2}\ln(Y^{-1/2}XY^{-1/2})Y^{1/2}].$$

The Riemannian distance is

$$d^2(X, Y) = \sum_{i=1}^{n}\ln^2\lambda_i(X^{-\frac{1}{2}}YX^{-\frac{1}{2}}),$$

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

## 9. Conclusion and Future Works

We generalize the PBD algorithm to solve optimization problems defined on Hadamard manifolds. We observe that none of our proofs need further than the uniqueness of the minimal geodesic (which is true in Hadamard manifolds). So, we conclude that our approach can be extended to more general manifolds, specifically to manifolds without focal points. The generalization of this method to solve zeros of monotone operators on those manifolds are in order in our working paper [25].

**Acknowledgements.** We are grateful to Professor Maria Fernanda Elbert, from FURJ, RJ, and the reviewer for their valuable comments on this work.

## References

[1] K. Arrow, G. Debreu: Existence of an equilibrium for a competitive economy, Econometrica 22 (1954) 265–290.

[2] N. Barron, W. Liu: Calculus of variations $l^\infty$, Appl. Math. Optimization 35 (1997) 237–263.

[3] R. S. Burachik, A. Iusem: Generalized proximal point algorithm for the variational inequality problem in Hilbert space, SIAM J. Optim. 8 (1998) 197–216.

[4] R. S. Burachik, S. Scheimberg: A proximal point method for the variational inequality problem in Banach spaces, SIAM J. Control Optim. 39(5) (2000) 1633–1649.

[5] Y. Censor, A. Zenios: Proximal minimization algorithm with D-functions, J. Optim. Theory Appl. 73(3) (1992) 451–464.

[6] Y. Censor, A. Lent: An iterative row-action method for interval convex programming, J. Optim. Theory Appl. 34(3) (1981) 321–353.

[7] G. Chen, M. Teboulle: Convergence analysis of the proximal-like minimization algorithm using Bregman functions, SIAM J. Optim. 3 (1993) 538–543.

[8] G. F. M. Cunha, A. M. Pinto, P. R. Oliveira, J. X. da Cruz Neto: Generalization of the primal and dual affine scaling algorithms, Technical Report ES 675/05, PESC/COPPE, Federal University of Rio de Janeiro, Rio de Janeiro (2005).

[9] J. X. da Cruz Neto, O. P. Ferreira, L. Lucambio Perez, S. Z. Németh: Convex- and monotone-transformable mathematical programming and a proximal-like point method, J. Glob. Optim. 35(1) (2006) 53–69.

[10] J. X. da Cruz Neto, L. L. Lima, P. R. Oliveira: Geodesic algorithms in Riemannian geometry, Balkan J. Geom. Appl. 3(2) (1998) 89–100.

[11] M. P. do Carmo: Riemannian Geometry, Birkhäuser, Boston (1992).

[12] D. den Hertog, C. Roos, T. Terlaky: Inverse barrier methods for linear programming, Report 91-27, Faculty of Technical Mathematics and Informatics, Delft University of Technology, Delft (1991).

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

[14] O. P. Ferreira, P. R. Oliveira: Proximal point algorithm on Riemannian manifolds, Optimization 51(2) (2002) 257–270.

[15] D. Gabay: Minimizing a differentiable function over a differential manifold, J. Optim. Theory Appl. 37(2) (1982) 177–219.

[16] O. Garciga, A. Iusem: Proximal methods with penalization effects in Banach spaces, Numer. Funct. Anal. Optimization 25(1-2) (2004) 69–91.

[17] W. Ginsberg: Concavity and quasiconcavity in economics, J. Econ. Theory 6 (1973) 596–605.

[18] Y. Goudou, J. Munier: The heavy ball with friction method: the quasiconvex case, submitted (2005).

[19] A. Iusem: Métodos de Ponto Proximal em Otimização, IMPA, Rio de Janeiro (1995).

[20] K. C. Kiwiel: Proximal minimization methods with generalized Bregman functions, SIAM J. Control Optim. 35 (1997) 1142–1168.

[21] D. G. Luenberger: The gradient projection method along geodesics, Manage. Sci., Theory 18(11) (1972) 620–631.

[22] Y. E. Nesterov, M. J. Todd: On the Riemannian geometry defined by self-concordant barriers and interior-point methods, Found. Comput. Math. 2 (2002) 333–361.

[23] E. A. Papa Quiroz, P. R. Oliveira: New results on linear optimization through diagonal metrics and Riemannian geometry tools, Technical Report ES-645/04, PESC/COPPE, Federal University of Rio de Janeiro, Rio de Janeiro (2004).

[24] E. A. Papa Quiroz, P. R. Oliveira: A new self-concordant barrier for the hypercube, J. Optim. Theory Appl. 135 (2007) 475–490.

[25] E. A. Papa Quiroz, P. R. Oliveira: Proximal methods to solve zeros of monotone operators with Bregman function on Riemannian manifolds, working paper.

[26] G. J. Pereira, P. R. Oliveira: A new class of proximal algorithms for the nonlinear complementary problems, in: Optimization and Control with Applications, L. Qi et al. (ed.), Applied Optimization 96, Springer, New York (2005) 552–562.

[27] T. Rapcsák: Smooth Nonlinear Optimization of $\mathbb{R}^n$, Kluwer, Dordrecht (1997).

[28] R. Saigal: The primal power affine scaling method, Ann. Oper. Res. 62(1) (1996) 375–417.

[29] T. Sakai: Riemannian Geometry, AMS, Providence (1996).

[30] S. Souza, G. L. de Oliveira, J. X. da Cruz Neto, P. R. Oliveira: A new class of interior proximal methods with variable metric for optimization over the positive orthant, working paper.

[31] S. T. Smith: Optimization techniques on Riemannian manifolds, in: Hamiltonian and Gradient Flows, Algorithms and Control, A. Bloch (ed.), Fields Inst. Commun. 3, AMS, Providence (1994) 113–136.

[32] A. Takayama: Mathematical Economics. 2nd Ed., Cambrigde University Press, Cambridge (1995).

[33] C. Udriste: Convex Functions and Optimization Methods on Riemannian Manifolds, Kluwer, Dordrecht (1994).