Proximal Point Method for a Class of Bregman Distances on Hadamard Manifolds

E. A. Papa Quiroz and P. Roberto Oliveira

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

September 2005

Abstract

This paper generalizes the proximal point method using a class of Bregman distances to solve optimization problems on Hadamard manifolds. We will prove, under standard assumptions, that the sequence generated by our method is well defined and converges to an optimal solution of the problem. Also, we give some examples of Bregman distances in non-Euclidean spaces.

Keywords: Proximal point algorithms, Hadamard manifolds, Bregman distances, diagonal Riemannian metrics.

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 abreviated PBD algorithm, generates a sequence $\{x^k\}$ defined by

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

$$x^k = \texttt{arg} \min_{x \in X \cap \bar{S}} \{f(x) + \lambda_k D_h(x, x^{k-1})\}, \tag{1.1}$$

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

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

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 λ_k, see for example [2], [4], [15], [16]. That algorithm has also been generalized for variational inequalities problems in Hilbert and Banach spaces, see [1], [14]. 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 application is that certain non convex optimization problems become convex through the introduction of an adequate Riemannian metric on the manifold, so we can use more efficient optimization techniques, see, Gabay [8], da Cruz Neto et.al [6], Ferreira and Oliveira [12], Luenberger [11], Rapcsák [20], Smith [22], Udriste [23], and the references therein. Another advantage of that approach is that we can use Riemannian metric to introduce new algorithms in interior point methods, see for example, Saigal [21], Cunha et al. [5], den Hertog [9], Pereira and Oliveira [19].

In this paper we generalize the PBD algorithm to solve optimization problems on Hadamard manifolds. Our approach is new and it is related to the work of Ferreira and Oliveira [13]. In that paper, the authors have been generalized the proximal point method using the intrinsic Riemannian distance. Here, we consider the following regularization parameters

$$0 < \lambda_k < \bar{\lambda}, \tag{1.2}$$

$$\lim_{k \to +\infty} \lambda_k = 0, \;\; \text{with} \;\; \lambda_k > 0. \tag{1.3}$$

For λ_k satisfying (1.2), we obtain the convergence of the PBD algorithm using a certain class of Bregman functions. That class satisfies either a Lipchitz condition on its gradient or a relation between the gradient and the geodesic triangle of the manifold. In particular, the second condition is satisfied by any Bregman function in $I\!R^n$. For λ_k satisfying (1.3), we obtain the convergence of the PBD algorithm with no assumptions on the Bregman function.

The paper is divided as follows. In Section 2 we give the notation 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, by considering Bregman distances. In Section 6 we introduce the PDB algorithm to solve minimization problems on Hadamard manifolds and we prove the convergence of the sequence generated by the algorithm. In Section 7 are presented some explicit examples of Bregman distances in non Euclidean spaces and 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 do Carmo [10].

Let M be a differential manifold. 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 denoted 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 ∇ 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 conection, 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 $\gamma : [0, 1] \to M$ starting from x on the direction $v \in T_pM$ is called a geodesic if

$$\frac{d^2\gamma_k}{dt^2} + \sum_{i,j=1}^{n} \Gamma^k_{ij} \frac{d\gamma_i}{dt}\frac{d\gamma_j}{dt} = 0, \quad k = 1, ..., n, \tag{2.4}$$

$$\gamma(0) = x \text{ and } \gamma'(0) = v,$$

where γ_i are the coordinates of γ and Γ^k_{ij} are the Christoffel's symbols expressed by

$$\Gamma^m_{ij} = \frac{1}{2} \sum_{k=1}^{n} \left\{ \frac{\partial}{\partial x_i} g_{jk} + \frac{\partial}{\partial x_j} g_{ki} - \frac{\partial}{\partial x_k} g_{ij} \right\} g^{km}, \tag{2.5}$$

(g^{ij}) denotes the inverse matrix of the metric $G = (g_{ij})$, and x_i are the coordinates of x. 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 $\exp_x : T_xM \to M$, defined as $\exp_x(v) = \gamma(1)$. If M is complete, then $\exp_x$ is defined for all $v \in T_xM$.

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

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

If $K(X,Y) \leq 0$, for all X and Y, then M is called a Riemannian manifold of nonpositive sectional curvature. 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. A complete simply connected Riemannian manifold of nonpositive sectional curvature is called a Hadamard manifold. The Hadamard's Theorem says that this manifold is diffeomorphic to the Euclidean space. More precisely, at any point $x \in M$ the exponential mapping $\exp_x : T_xM \to M$ is a diffeomorphism. Therefore, it is clear that the exponential map has a inverse $\exp_x^{-1}$ and that there exists an unique geodesic joining two arbitrary points of M.

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 Ferreira and Oliveira [13] and Udriste [23] 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 (unique) 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-valued function. f is called a convex function on A if for all $x, y \in A$ and $t \in [0,1]$*

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

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

Theorem 3.1 *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 in $[0,1]$.*

Proof. See Udriste [23]. ■

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

$$f(x) \geq f(y) + \langle s, \exp_y^{-1} x\rangle_y, \tag{3.6}$$

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 3.2 *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 $\forall x \in M$ (3.6) is true.*

Proof. See Ferreira and Oliveira [13] and Udriste [23] for a geometric and an analytic proof, respectively. ■

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

Theorem 3.3 *Let M be a Hadamard manifold and $f : M \to I\!R$ be a convex function. $0 \in \partial f(x)$ if and only if x is a minimum point of f in M.*

Proof. see Udriste [23].

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 [3] 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. The *Bregman distance* associated to f, denoted by D_h, is defined as a function $D_h(.,.) : \bar{S} \times S \to I\!R$ such that

$$D_h(x,y) := h(x) - h(y) - \langle \mathtt{grad} h(y), \exp_y^{-1} x\rangle_y. \tag{4.7}$$

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

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

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

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

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

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 $\lim_{k\to\infty} z^k = y^*$.*

We denote the family of Bregman functions by $\mathcal{B}$ and refer to the set S as the *zone* of the function h.

Lemma 4.1 *Let $h \in \mathcal{B}$ 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 $y \in S$, $D_h(x,y) \ge 0$ and $D_h(x,y) = 0$ if and only if $x = y$.*

Proof.
i. From definition (4.7), we have

$$p(t) := D_h(\gamma(t), \gamma(0)) = h(\gamma(t)) - h(\gamma(0)) - t\langle \mathtt{grad}\, h(\gamma(0)), \gamma'(0)\rangle_{\gamma(0)}.$$

Derivation with respect to t gives

$$p'(t) = \langle \mathtt{grad}\, h(\gamma(t)), \gamma'(t)\rangle_{\gamma(t)} - \langle \mathtt{grad}\, h(\gamma(0)), \gamma'(0)\rangle_{\gamma(0)}.$$

Using the fact that $P_{\gamma,0,t}\gamma'(0) = \gamma'(t)$ we obtain

$$p'(t) = \langle \mathtt{grad}\, h(\gamma(t)) - P_{\gamma,0,t}\mathtt{grad}\, h(\gamma(0)), \gamma'(t)\rangle_{\gamma(t)}. \tag{4.8}$$

On the other hand,

$$p'(t) = \frac{d}{dt}(D_h(\gamma(t), \gamma(0))) = d(D_h(., \gamma(0)))_{\gamma(t)}(\gamma'(t)) = \langle \mathtt{grad}\, D_h(., \gamma(0))(\gamma(t)), \gamma'(t)\rangle_{\gamma(t)}. \tag{4.9}$$

From (4.8) and (4.9) follows

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

Taking $t = 1$ gives

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

ii. As h is a strictly convex function on $\bar{S}$ and $\langle \mathtt{grad}\, h(y), \exp_y^{-1} x\rangle_y$ is a linear function in $x \in M$, then $D_h(., y)$ is strictly convex on $\bar{S}$.
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 $\mathtt{grad}\, D_h(x, y)$ to mean $\mathtt{grad}\, D_h(., y)(x)$. So, if γ is a geodesic curve such that $\gamma(0) = y$ and $\gamma(1) = x$, from Lemma 4.1, i, we obtain

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

Definition 4.2 *Let $\Omega \subset M$, S an open convex set, and let $y \in S$. A point $Py \in \Omega \cap \bar{S}$ for which*

$$D_h(Py, y) = \min_{x \in \Omega \cap \bar{S}} D_h(x, y) \tag{4.10}$$

*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 4.2 *Let $\Omega \subset M$ a closed convex set, and $h \in \mathcal{B}$ with zone S. If $\Omega \cap \bar{S} \neq \phi$ then, for any $y \in S$, there exists a unique D_h-projection Py 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 4.1, d) and closed (because $D_h(z, y)$ is continuous in $z \in \bar{S}$, due to Definition 4.1, 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 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 (4.10). The uniqueness follows from the strict convexity of $D_h(., y)$, therefore

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

Lemma 4.3 *Let $h \in \mathcal{B}$ 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 convex on $\bar{S}$.

Proof. From (4.7)

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

Due to the 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. $\blacksquare$

Proposition 4.1 *Let $h \in \mathcal{B}$ with zone S; S and $\Omega \subset M$ are closed convex sets 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 a geodesic curve such that $\gamma(0) = Py$ and $\gamma(1) = x$. Due to Lemma 4.3 the function

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

is convex on $\bar{S}$. Then in particular $G(\gamma(t))$ is convex for $t \in (0, 1)$ (see Theorem 3.1). 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 used the fact that $D_h(Py, Py) = 0$. The above inequality is equivalent to

$$(1/t)(D_h(\gamma(t), y) - D_h(Py, y)) - (1/t)D_h(\gamma(t), Py) \leq D_h(x, y) - D_h(x, Py) - D_h(Py, y). \quad (4.11)$$

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 yo get

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

Using this inequality in (4.11) 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

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

Clearly, the left side is null, leading to the aimed result. $\blacksquare$

5 Regularization

Let M a Hadamard manifold and $f : X \subset M \to I\!R$ a real-valued 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

$$f_\lambda(y) = \inf_{x \in X \cap \bar{S}} \{f(x) + \lambda D_h(x, y)\} \tag{5.12}$$

where $D_h(x, y)$ is given in (4.7). In order to prove that the function (5.12) is well defined, h and f should satisfy some conditions.

Proposition 5.1 *If $f : X \subset M \to I\!R$ is a bounded below convex and continuous function over a closed convex set X, and $h \in \mathcal{B}$ with zone S such that $X \cap \bar{S} \neq \phi$, then, for every $y \in S$ and $\lambda > 0$ there exists a unique point, denoted by $x_f(y, \lambda)$, such that*

$$f_\lambda(y) = f(x_f(y, \lambda)) + \lambda D_h(x_f(y, \lambda), y). \tag{5.13}$$

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

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

for all $x \in X \cap \bar{S}$. It follows from Definition 4.1, d, that the level sets of the function $f(.) + \lambda D_h(., y)$ are bounded. Also, this function is continuous on $X \cap \bar{S}$, due to Definition 4.1, a. 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 $X \cap \bar{S}$. Furthermore $f(.) + \lambda D_h(., y)$ is strictly convex, due to the convexity of f and Lemma 4.1, ii. Therefore, $x_f(y, \lambda)$ is unique, the equality (5.13) following from (5.12). ■

Now, we let different conditions on h, that also ensure that (5.12) is well defined. We consider the unconstrained case $X = S = M$, so (5.12) is reduced to

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

Let us introduce the following definition.

Definition 5.1 *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 $1-$coercive function at $y \in M$, then it is easy to show that the minimizer set of g on M is nonempty.

Lemma 5.1 *If $f : M \to I\!R$ is convex, $\lambda > 0$, 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. Since f is convex, from Theorem 3.2, it follows that there exists $s \in T_x M$ such that

$$\frac{f(x) + \lambda D_h(x, y)}{d(x, y)} \geq \frac{f(y)}{d(x, y)} + \left\langle s, \frac{\exp_y^{-1} x}{d(x, y)} \right\rangle_y + \lambda \frac{D_h(x, y)}{d(x, y)}$$

$$= \frac{f(y)}{d(x,y)} + \left\langle s, \frac{\exp_y^{-1} x}{d(x,y)} \right\rangle_y + \lambda \frac{h(x)}{d(x,y)} - \lambda \frac{h(y)}{d(x,y)} - \lambda \left\langle \mathtt{grad} h(y), \frac{\exp_y^{-1} x}{d(x,y)} \right\rangle_y$$

$$\geq \tfrac{f(y)}{d(x,y)} - ||s|| + \lambda \tfrac{h(x)}{d(x,y)} - \lambda \tfrac{h(y)}{d(x,y)} - \lambda ||\mathtt{grad} h(y)||,$$

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

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

Proposition 5.2 *Let $h : M \to I\!R$ be a $1-$coercive strictly convex function at $y \in M$, $f : M \to I\!R$ a convex function. Then, there exists an unique 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 and the strict convexity of $D_h(.,y)$. $\blacksquare$

6 Proximal Point Algorithm

We are interested in solving the optimization problem:

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

where M is a Hadamard manifold and f is a convex function on M. The PBD algorithm is defined as

$$x^0 \in M, \tag{6.14}$$

$$x^k = \mathtt{arg}\min_{x\in M}\{f(x) + \lambda_k D_h(x, x^{k-1})\}, \tag{6.15}$$

where h is a Bregman function with zone M, D_h is as in (4.7) and λ_k is a positive parameter. In the particular case where M is the Euclidean space $I\!R^n$, and $h(x) = (1/2)x^T x$, we have

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

Therefore, the PBD algorithm is a natural generalization of the proximal point algorithm on $I\!R^n$.

We will use the following parameter conditions to the PBD algorithm:

$$0 < \lambda_k < \bar{\lambda}, \quad \text{or} \tag{6.16}$$

$$\lim_{k\to+\infty} \lambda_k = 0, \ \text{ with } \ \lambda_k > 0. \tag{6.17}$$

Note that (6.16) implies that $\sum_{k=1}^{\infty}(1/\lambda_k) = +\infty$. Next, we set the Hypothesis to be assumed from now on:

Assumption A1. The optimal set of the problem (p), denoted by X^*, is nonempty.

Assumption A2. $h \in \mathcal{B}$ satisfies one of the following conditions:

A2.1 $\texttt{grad}h$ is a Lipschitz function on M, that is, for any $x, y \in M$:

$$||\texttt{grad}h(x) - P_{\gamma,0,1}\texttt{grad}h(y)|| \leq Ld(x,y),$$

where $L > 0$, $\gamma(t)$ is the geodesic curve such that $\gamma(0) = y$ and $\gamma(1) = x$, and d is the Riemannian distance.

A2.2 Given $x, y, z \in M$ and letting $\gamma : [0,1] \to M$ the geodesic curve joining $\gamma(0) = y$ and $\gamma(1) = z$, then h satisfies

$$\langle P_{\gamma,0,1}\texttt{grad}h(y), \exp_z^{-1} x - \exp_z^{-1} y - P_{\gamma,0,1} \exp_y^{-1} x\rangle_z \geq 0.$$

Remark 1.
The alternative Assumptions A2 will be used to prove that any limit point of the sequence $\{x^k\}$, with λ_k as in (6.16), is an optimal solution of the problem (p). When λ_k verifies (6.17), we will show that Assumption A2 is not necessary to prove convergence of the PBD algorithm. On the other hand, it is worthwhile observe that when M is the Euclidean space $\mathbb{R}^n$, any Bregman function satisfies the Assumption $A2.2$. Clearly, in the Euclidean space geodesic curves are straight lines, so

$$\exp_z^{-1} y + P_{\gamma,0,1} \exp_y^{-1} x = (y - z) + (x - y) = x - z = \exp_z^{-1} x,$$

Therefore,

$$\langle P_{\gamma,0,1}\texttt{grad}h(y), \exp_z^{-1} x - \exp_z^{-1} y - P_{\gamma,0,1} \exp_y^{-1} x\rangle_z = 0.$$

Furthermore, that hypothesis implies

$$D_h(x,y) - D_h(z,y) - D_h(x,z) \geq \langle \texttt{grad}h(z) - P_{\gamma,0,1}\texttt{grad}h(y), \exp_z^{-1} x\rangle_z. \tag{6.18}$$

Indeed, using the definition of D_h, it gets

$$D_h(x,y) - D_h(z,y) - D_h(x,z) = \langle \texttt{grad}h(y), \exp_y^{-1} z - \exp_y^{-1} x\rangle_y + \langle grad\, h(z), \exp_z^{-1} x\rangle_z$$

Taking Parallel transport from $\gamma(0) = y$ to $\gamma(1) = z$ in the first term on the right side gives

$$D_h(x,y) - D_h(z,y) - D_h(x,z) =$$

$$= \langle P_{\gamma,0,1}\texttt{grad}\, h(y), -\exp_z^{-1} y - P_{\gamma,0,1} \exp_y^{-1} x\rangle_z + \langle \texttt{grad}\, h(z), \exp_z^{-1} x\rangle_z,$$

Now, add and subtract the term $\langle P_{\gamma,0,1}\texttt{grad}h(y), \exp_z^{-1} x\rangle_z$, to produce

$$D_h(x,y) - D_h(z,y) - D_h(x,z) = \langle \texttt{grad}\, h(z) - P_{\gamma,0,1}\texttt{grad}\, h(y), \exp_z^{-1} x\rangle_z +$$

$$\langle P_{\gamma,0,1}\texttt{grad}h(y), \exp_z^{-1} x - \exp_z^{-1} y - P_{\gamma,0,1} \exp_y^{-1} x\rangle_z.$$

The aimed result is evident after using the assumption $A2.2$.

Remark 2.
In order to verify that $h : M \to \mathbb{R}$ is a Bregman function, with zone M, it is sufficient to show that conditions **a** to **d** in Definition 4.1 are satisfied, as **e** and **f** are an immediate consequence of **a**, **b**, **c** and **d**.

6.1 Convergence Results

In this subsection we prove the convergence of the proposed algorithm. Our results are motivated by the works of Iusem [15], Chen and Teboulle [4] and Censor and Zenios [2].

Theorem 6.1 *The sequence $\{x^k\}$ generated by the PBD algorithm is well defined.*

Proof. The proof proceeds by induction. It holds for $k = 0$, due to (6.14). Assume that x^k is well defined; from Proposition 5.1, for $X = S = M$, we have that x^{k+1} exists and it is unique. ■

Theorem 6.2 *Under Assumption A1 the sequence $\{x^k\}$, generated by the PBD algorithm, is bounded.*

Proof. Since x^k satisfies (6.15) we have

$$f(x^k) + \lambda_k D_h(x^k, x^{k-1}) \leq f(x) + \lambda_k D_h(x, x^{k-1}),\ \forall x \in M. \tag{6.19}$$

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 convex set

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

Using Proposition 4.1 and the fact that $X^* \subset \Omega$ we have

$$0 \leq D_h(x^k, x^{k-1}) \leq D_h(x^*, x^{k-1}) - D_h(x^*, x^k) \tag{6.20}$$

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

$$D_h(x^*, x^k) \leq D_h(x^*, x^{k-1}). \tag{6.21}$$

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 6.1 *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 noincreasing;*

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 (6.21), $D_h(x^*, x^k)$ is a bounded below nonincreasing sequence and hence convergent.
b. Taking limit when k goes to infinity in (6.20) 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 (6.19) it follows that

$$0 \leq D_h(x^k, x^{k-1}) \leq (1/\lambda_k)(f(x^{k-1}) - f(x^k)), \tag{6.22}$$

since $D_h(x^{k-1}, x^{k-1}) = 0$. Thus $\{f(x^k)\}$ is noincreasing.
d. Taking $z^k = x^{k_j+1}$ and $y^k = x^{k_j}$ in Definition 4.1, f, we obtain the result. ■

Theorem 6.3 *Suppose that assumptions A1 and A2.1 are satisfied. If λ_k satisfies (6.16) 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 (6.15) and Theorem 3.3

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

or,

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

Let γ_{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 4.1, i, we obtain

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

Let x^* be an optimal solution of (p). Using (3.6) for $x = x^*$ and $y = x^{k_j+1}$ we have

$$f(x^*) - f(x^{k_j+1}) \geq \langle y^{k_j}, \exp^{-1}_{x^{k_j+1}} x^*\rangle_{x^{k_j+1}} \tag{6.23}$$

where,

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

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 (6.23) we obtain

$$f(x^*) - f(x^{k_j+1}) \geq -M||y^{k_j}||_{x^{k_j+1}}. \tag{6.24}$$

To conclude the proof we will show that

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

Indeed, from the assumption A2.1

$$0 \leq ||y^{k_j}||_{x^{k_j+1}} \leq \lambda_{k_{j+1}} L d(x^{k_j+1}, x^{k_j}).$$

Now, Taking $j \to +\infty$ and using the boundedness of λ_k and Proposition 6.1, d, we obtain $\lim_{j\to\infty} ||y^{k_j}||_{x^{k_j}} = 0$, as wanted. Finally, taking $j \to +\infty$ in (6.24), 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 6.4 *Suppose that assumption A1 and A2.2 are satisfied. If λ_k satisfies (6.16) then, any limit point of $\{x^k\}$ is an optimal solution of the problem (p).*

Proof. Analogously o the previous theorem, given $x \in M$ and using (6.15), and (3.6) for x and $y = x^k$ we have

$$-\frac{1}{\lambda_k}\langle y^k, \exp^{-1}_{x^k}, x\rangle_{x^k} \geq \frac{1}{\lambda_k}(f(x^k) - f(x)) \tag{6.25}$$

where,

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

and γ_k is the geodesic curve joining $\gamma_k(0) = x^{k-1}$ to $\gamma_k(1) = x^k$.
On the other hand, taking $y = x^{k-1}$ and $z = x^k$ in (6.18) we have

$$D_h(x, x^{k-1}) - D_h(x^k, x^{k-1}) - D_h(x, x^k) \geq -\frac{1}{\lambda_k}\langle y^k, \exp^{-1}_{x^k}, x\rangle_{x^k}.$$

In that inequality, use (6.25) and the boundedness of λ_k, giving

$$D_h(x, x^{k-1}) - D_h(x^k, x^{k-1}) - D_h(x^*, x^k) \geq \frac{1}{\bar{\lambda}}(f(x^k) - f(x)) \tag{6.26}$$

Now, 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 (6.26) for $x = x^*$ we have

$$D_h(x^*, x^{k_j-1}) - D_h(x^{k_j}, x^{k_j-1}) - D_h(x^*, x^{k_j}) \geq \frac{1}{\bar{\lambda}}(f(x^{k_j}) - f(x^*)).$$

Take $j \to \infty$ and apply Proposition 6.1, a and b, to obtain

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

Therefore, any limit point of $\{x^k\}$ is an optimal solution of (p). ■

Theorem 6.5 *Under Assumption A1, any limit point of $\{x^k\}$ generated by the PBD algorithm with λ_k satisfying (6.17) is an optimal solution of (p).*

Proof. Let $x^* \in M$ be an optimal solution of $(\bar{p})$ 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 (6.15) 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:

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

Taking, above, $j \to \infty$ and considering the hypothesis (6.17) we obtain

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

Therefore, any limit point is a solution of the problem (p). ■

Theorem 6.6 *Under Assumptions A1 and A2 the sequence $\{x^k\}$ generated by the PBD algorithm, with λ_k satisfying (6.16), 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 Definition 4.1, e, it is true that

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

Now, from Theorems 6.3 and 6.4 x^* is an optimal solution of (p), so from Proposition 6.1, 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 Deinition 4.1, f, $x^* = \bar{x}$. It follows that $\{x^k\}$ cannot have more than one limit point and therefore,

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

Theorem 6.7 *Under Assumption A1 the sequence $\{x^k\}$ generated by the PBD algorithm, with λ_k satisfying (6.17), converges to an optimal solution of (p).*

Proof. It is analogous to the previous theorem. $\blacksquare$

7 Some Examples of Bregman Distances

Example 7.1 *The Euclidean space is a Hadamard manifold with the metric $G(x) = I$. Its geodesic curves are the straight lines and the Bregman distances have the form*

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

Example 7.2 *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 $\Phi(x) = (x_1, x_2, ..., x_{n-1}, x_{n-1}^2 - x_n)$, see [7]. The unique geodesic curve, joining the points $\gamma(0) = y$ and $\gamma(1) = x$ is $\gamma(t) = (\gamma_1, ..., \gamma_n)$, where $\gamma_i(t) = y_i + t(x_i - y_i), \forall i = 1, ...n-1$ 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

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

Example 7.3 *$M = I\!R^n_{++}$ with the Dikin metric X^{-2} is a Hadamard manifold. Defining $\pi : M \to I\!R^n$ such that $\pi(x) = (-\ln x_1, ..., -\ln x_n)$, it can be proved that π is an isometry. It is well known, see for example [20], that the unique geodesic curve joining the points $\gamma(0) = y$ and $\gamma(1) = x$ is*

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

with

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

Then the Bregman distance is:

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

Example 7.4 *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 $I\!R^n$ through the function $\pi(x) = \left(\ln\left(\frac{x_1}{1-x_1}\right), ..., \ln\left(\frac{x_n}{1-x_n}\right)\right)$. The unique geodesic curve, see Section 9, joining the points $\gamma(0) = y$ and $\gamma(1) = x$ is $\gamma(t) = (\gamma_1, ..., \gamma_n)$ 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

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

2. *$(M, csc^4(\pi x))$ is a Hadamard manifold, isometric to $I\!R^n$, through the function $\pi(x) = \frac{1}{\pi}\left(\cot(\pi x_1), ..., \cot(\pi x_n)\right)$. The unique geodesic curve, see Section 9, joining the points $\gamma(0) = y$ and $\gamma(1) = x$ is $\gamma(t) = (\gamma_1, ..., \gamma_n)$ 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

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

3. *Finally, we consider $(M, \csc^2(\pi x))$. It is a Hadamard manifold isometric to $I\!R^n$, see [17]. The unique geodesic curve joining the points $\gamma(0) = y$ and $\gamma(1) = x$ is $\gamma(t) = (\gamma_1, ..., \gamma_n)$ 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

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

Example 7.5 *$M = \mathcal{S}^n_{++}$, 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 [17], is given by*

$$\gamma(t) = X^{1/2}(X^{-1/2}YX^{-1/2})^t X^{1/2},$$

with

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

Then, the Bregman distance is

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

8 Conclusion and Future Works

We generalize the PBD algorithm to solve optimization problems defined on Hadamard manifolds. When the PBD algorithm works with λ_k satisfying $\lim_{k\to+\infty}\lambda_k = 0$, with $\lambda_k > 0$, the sequence generated by this algorithm converges to an optimal solution of the problem without any assumption on the Bregman functions. On the other hand when λ_k satisfies $0 < \lambda_k < \bar{\lambda}$, we use a class of Bregman function that guarantees the convergence results. We are working to remove that assumption in a general framework of Variational Inequality Problems on Hadamard manifolds.

9 Appendix: Geodesic Equation for a Special Diagonal Metric

In this Appendix we give a simplified equation to obtain geodesic curves in closed form for a special class of diagonal Riemannian metric defined in manifolds that are open subset of $I\!R^n$. This approach recover known metrics that are associated to some interior point algorithms, for example, the Dikin metric and some metrics obtained by the Hessian of separated barrier functions. Those results are extracted from [18].

Let $M = M_1 \times M_2 ... \times M_n$ be a manifold in $I\!R^n$ where M_i are open subsets of $I\!R$, $\forall i = 1, ..., n$. Given $x \in M$ and $u, v \in T_xM \equiv I\!R^n$ we induce in M the metric $\langle u, v\rangle_x = u^T G(x)v$, where $G(x) = \texttt{diag}\left(\frac{1}{p_1(x_1)^2}, \frac{1}{p_2(x_2)^2}, ..., \frac{1}{p_n(x_n)^2}\right)$ and $p_i : M_i \to I\!R_{++}$ are differentiable functions. In this case we have

$$g_{ij}(x) = \frac{\delta_{ij}}{p_i(x_i)p_j(x_j)}.$$

Considering the Riemannian manifold $(M, G(x))$ we obtain the following results:

1. Christoffel's Symbols.
   We use the relation between metrics and Christoffel's symbols, given in (2.5).
   If $k \neq m$ then $g^{mk} = 0$, and the expression is reduced to:

$$\Gamma_{ij}^{m} = \frac{1}{2}\left\{\frac{\partial}{\partial x_i} g_{jm} + \frac{\partial}{\partial x_j} g_{mi} - \frac{\partial}{\partial x_m} g_{ij}\right\} g^{mm}.$$

   We consider two cases.
   First case: $i = j$

$$\Gamma_{ii}^{m} = \frac{1}{2}\left\{\frac{\partial}{\partial x_i} g_{im} + \frac{\partial}{\partial x_i} g_{mi} - \frac{\partial}{\partial x_m} g_{ii}\right\} g^{mm}.$$

   If $m = i$ then

$$\Gamma_{ii}^{m} = -\frac{1}{p_i(x_i)}\frac{\partial p_i(x_i)}{\partial x_i},$$

   otherwise,

$$\Gamma_{ii}^{m} = 0.$$

   Second case: $i \neq j$:

$$\Gamma_{ij}^{m} = \frac{1}{2}\left\{\frac{\partial}{\partial x_i} g_{im} + \frac{\partial}{\partial x_j} g_{mi}\right\} g^{mm}.$$

   If $m = i$ then $m \neq j$ and:

$$\Gamma_{ij}^{i} = 0.$$

   If $m = j$ then $m \neq i$ and:

$$\Gamma_{ij}^{j} = 0.$$

   If $m \neq i$ and $m \neq j$ then:

$$\Gamma_{ij}^{m} = 0.$$

   In both cases we have

$$\Gamma_{ij}^{m} = -\frac{1}{p_i(x_i)}\frac{\partial p_i(x_i)}{\partial x_i}\delta_{im}\delta_{ij}. \tag{9.27}$$

2. Geodesic Equation.

   Substituting the Christoffel's symbols (9.27) in the equation (2.4) we have that the unique geodesic curve γ starting from $x = (x_1, x_2, ..., x_n) \in M$, with direction $v = (v_1, v_2, ..., v_n) \in T_p(M) \equiv \mathbb{R}^n$ is obtained by solving the following equation:

$$\frac{d^2\gamma_i}{dt^2} - \frac{1}{p_i(\gamma_i)}\frac{\partial p_i(\gamma_i)}{\partial \gamma_i}\left(\frac{d\gamma_i}{dt}\right)^2 = 0, i = 1, ..., n \tag{9.28}$$

   with initial conditions:

$$\begin{aligned} \gamma_i(0) &= x_i, \qquad i = 1, ..., n, \\ \gamma_i'(0) &= v_i, \qquad i = 1, ..., n. \end{aligned}$$

   We can see that the above equation has the following equivalent form:

$$\frac{d}{dt}\left(\frac{\gamma'(t)}{p_i(\gamma_i(t))}\right) = 0.$$

   Then, the geodesic equation (9.28) can be solved through:

$$\int \frac{1}{p_i(\gamma_i)} d\gamma_i = a_i t + b_i \qquad i = 1, ..., n \tag{9.29}$$

where a_i and b_i are real constants such that $\gamma_i(0) = x_i$ and $\gamma_i'(0) = v_i$.

With the equation (9.29) we can reproduce some known geodesic curves and obtain new geodesic curves. Next, we present some examples:

(a) Consider the Riemannian manifold $(\mathbb{R}^n_{++}, X^{-2})$ with $p_i(x_i) = x_i$. Using (9.29) the geodesic curves $\gamma(t) = (\gamma_1(t), \gamma_2(t), ..., \gamma_n(t))$ such that $\gamma(0) = x$ and $\gamma'(0) = v$ are:

$$\gamma_i(t) = x_i \exp\left(\frac{v_i}{x_i} t\right), \ i = 1, 2..., n.$$

(b) Consider the manifold $((0,1)^n, \csc^4(\pi x))$, where $p_i(x_i) = \sin^2(\pi x_i)$. The geodesic curve $\gamma(t) = (\gamma_1(t), \gamma_2(t), ..., \gamma_n(t))$ starting in $\gamma(0) = x$ on the direction $\gamma'(0) = v$ is:

$$\gamma_i(t) = \frac{1}{\pi} \arctan\Big(-\pi \csc^2(\pi x_i) v_i t + \cot(\pi x_i)\Big), \ i = 1, 2..., n.$$

The geodesic curve is well defined for all $t \in \mathbb{R}$. Therefore this manifold is complete with null curvature. The Riemannian distance from $x = \gamma(0)$ to $y = \gamma(t_0)$, $t_0 > 0$, is given by:

$$d(x,y) = \int_0^{t_0} \|\gamma'(t)\| dt = \left\{ \sum_{i=1}^n [\cot(\pi y_i) - \cot(\pi x_i)]^2 \right\}^{\frac{1}{2}}.$$

(c) Consider $((0,1)^n, X^{-2}(I-X)^{-2})$ with $p_i(x_i) = x_i(1-x_i)$. The geodesic curve $\gamma(t) = (\gamma_1(t), \gamma_2(t), ..., \gamma_n(t))$ defined in this manifold, such that $\gamma(0) = x$ and $\gamma'(0) = v$ is:

$$\gamma_i(t) = \frac{1}{2}\left\{1 + \tanh\left(\frac{1}{2}\frac{v_i}{x_i(1-x_i)} t + \frac{1}{2} \ln \frac{x_i}{1-x_i}\right)\right\} \ i = 1, 2..., n.$$

where $\tanh(z) = (e^z - e^{-z})/(e^z + e^{-z})$ is the hyperbolic tangent function. The geodesic curve is well defined for all $t \in \mathbb{R}$. Therefore, this manifold is complete with null curvature. The Riemannian distance from $x = \gamma(0)$ to $y = \gamma(t_0)$, $t_0 > 0$, is given by:

$$d(x,y) = \int_0^{t_0} \|x'(t)\| dt = \left\{ \sum_{i=1}^n \left[\ln\left(\frac{y_i}{1-y_i}\right) - \ln\left(\frac{x_i}{1-x_i}\right)\right]^2 \right\}^{\frac{1}{2}}$$

The geodesic curves and Riemannian distances in (b) and (c) are new to our knowledge. Therefore, new explicit geodesics, metric variables and proximal point algorithms can be defined to solve non convex unconstrained optimization problems on $(0,1)^n$, whose convergence results are known for general metrics, see da Cruz Neto et.al [6], Ferreira and Oliveira [12], [13] and Rápcsak [20].

References

[1] R. S. Burachik and A. Iusem, Generalized Proximal Point algorithm for the Variational Inequality Problem in Hilbert Space, SIAM, J. Optim., 8, 1998, 197-216.

[2] Y. Censor and A. Zenios, Proximal minimization algorithms with D-functions, journal of Optimization Theory and Aplication, 73, no 3, 1992.

[3] Y. Censor and Lent, An Iterative Row-Action Method for Interval Convex Programming, Journal of Optimization Theory and applications, vol. 34, No. 3, July 1981.

[4] G. Chen and M. Teboulle, Convergence Analysis of the Proximal-Like Minimization Algorithm Using Bregman Functions. SIAM J. Optimization, vol. 3, pp. 538-543, August 1993.

[5] G.F.M. Cunha, A. M. Pinto, P. R. Oliveira, and 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, 2005.

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

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

[8] D. Gabay, Minimizing a Differentiable Function over a Differential Manifold, Journal of Optimization Theory and Applications, vol.37, 2, 1982.

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

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

[11] D. G. Luenberger, The Gradient Projection Method Along Geodesics, *Management Science*, v. 18, n. 11, 1972, pp. 620-631.

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

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

[14] O.Garciga and A. Iusem, Proximal Methods with Penalization Effects in Banach Spaces, IMPA, Technical Report, Serie A-119 Janeiro, 2002.

[15] A. Iusem, Métodos de Ponto Proximal em Otimização, Colóquio Brasileiro de Matemática, IMPA, Brazil, 1995.

[16] K. C. Kiwiel. Proximal Minimization Methods with Generalized Bregman Functions, SIAM Journal of Control and Optimization, 35, 1142-1168, 1997.

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

[18] 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, 2004.

[19] G.J. Pereira and P. R. Oliveira, 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, 2005, v 69, p 552-562.

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

[21] R. Saigal, The primal power affine scaling method, Tech. Report, No 93-21, Dep. Ind. and Oper. Eng., University of Michigan, 1993.

[22] S. T. Smith, Optimization Techniques on Riemannian Manifolds, fields Institute Communications,vol 3, 1994.

[23] C. Udriste, Convex Function and Optimization Methos on Riemannian Manifolds, Kluwer academic Publishers, 1994.