



HAL
open science

On the nature of Bregman functions

Edouard Pauwels

► **To cite this version:**

| Edouard Pauwels. On the nature of Bregman functions. 2023. hal-03974132

HAL Id: hal-03974132

<https://hal.science/hal-03974132>

Preprint submitted on 5 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the nature of Bregman functions

Edouard Pauwels*

Draft of February 5, 2023

Abstract

Let $C \subset \mathbb{R}^n$ be convex, compact, with nonempty interior and h be Legendre with domain C , continuous on C . We prove that h is Bregman if and only if it is strictly convex on C and C is a polytope. This provides insights on sequential convergence of many Bregman divergence based algorithm: abstract compatibility conditions between Bregman and Euclidean topology may equivalently be replaced by explicit conditions on h and C . This also emphasizes that a general convergence theory for these methods (beyond polyhedral domains) would require more refinements than Bregman's conditions.

1 Introduction

Let $C \subset \mathbb{R}^n$ be convex compact with nonempty interior, $h: C \mapsto \mathbb{R}$ be convex. We assume throughout the article that h is Legendre:

- h is essentially smooth: continuously differentiable on $\text{int}C$, such that $\|\nabla h\|$ goes to $+\infty$ when approaching the boundary of C .
- h is strictly convex on $\text{int}C$.

We will also assume throughout the text that h is continuous on C . The Bregman divergence associated to h is then for all $y \in C$ and $x \in \text{int}C$,

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

1.1 Fejérian sequences:

Let $f: C \rightarrow \mathbb{R}$ be convex proper lower semi-continuous. Consider the problem

$$\min_{x \in C} f(x) \tag{1}$$

*IRIT, CNRS, Université de Toulouse, Toulouse, France. Institut universitaire de France (IUF).

and denote by $S \subset C$, the solution set of (1). The following definition is adapted from the Euclidean setting [11].

Definition 1. Let $(x_k)_{k \in \mathbb{N}}$ be a sequence in $\text{int}C$, it is called *Fejérian*, or *Fejér*, for problem (1) if

- All accumulation points are in S .
- For all $y \in S$, $D_h(y, x_k)$ has a finite limit as $k \rightarrow \infty$.

Actually, Definition 1 may be given for a general closed convex set $S \subset C$ (see *e.g.* [11] which may be directly adapted). We stick to the optimization problem (1) for simplicity since it already cover a wide range of algorithms.

1.2 Algorithmic examples

We list below some of the most common algorithms generating Fejérian sequences in convex optimization.

Mirror descent: the algorithm is due to Nemirovsky [13]. Assume that f is Lipschitz. Initialize $x_0 \in \text{int}C$ and, given sequence of positive step sizes $(\alpha_k)_{k \in \mathbb{N}}$, iterate

$$x_{k+1} = \nabla h^*(\nabla h(x_k) - \alpha_k v_k), \quad (2)$$

$$v_k \in \partial f(x_k). \quad (3)$$

Assuming $\sum_{k=0}^{+\infty} \alpha_k = +\infty$ and $\sum_{k=0}^{+\infty} \alpha_k^2 < +\infty$ the resulting algorithmic sequence is Fejér. In this case equation (4.22) in [6] ensures that accumulation points are in S and equation (4.21) combined with [Lemma 2, Section 2.2][14] ensures that $D(y, x_k)$ has a limit for all $y \in S$.

Bregman gradient and NoLips algorithm: Assume that f is C^1 on an open set containing C and $h - \alpha f$ is convex for some $\alpha > 0$. Initialize $x_0 \in \text{int}C$ and iterate

$$x_{k+1} = \nabla h^*(\nabla h(x_k) - \alpha \nabla f(x_k)). \quad (4)$$

In this case, for all $y \in S$, $D_h(y, x_k)$ is non increasing as k grows [4, Lemma 5]. This extends to the NoLips algorithm proposed for composite objectives in [4]. Similarly, the work presented in [1] presents a continuous time variant of Fejér property in the context of Hessian-Riemannian gradient flows.

Proximal minimization with D -functions: Assume that f is proper lower-semicontinuous. Initialize $x_0 \in \text{int}C$ and, given sequence of positive step sizes $(\alpha_k)_{k \in \mathbb{N}}$, iterate

$$x_{k+1} = \arg \min_{x \in C} \alpha_k f(x) + D_h(x, x_k). \quad (5)$$

In this case, for all $y \in S$, $D_h(y, x_k)$ is non increasing as k grows [10, Lemma 3.3] and if $\sum_{k=0}^{+\infty} \alpha_k = +\infty$, then the resulting sequence has all its accumulation point in S [10, Theorem 3.4].

Alternating projection: Beyond optimization problem (1), in his foundational paper [7], Bregman considers an alternating projection algorithm generating a Fejér sequence to find an element in the intersection of convex sets.

1.3 Sequential convergence analysis

The Fejérian property can be used to prove convergence of the sequence similarly as in the Euclidean case, in relation to Opial's Lemma (see [11] for an overview). This argument is valid provided that the topology encoded by D_h is equivalent to the Euclidean topology. This is always true in the interior of C by strict convexity, but further assumptions need to be made to ensure that this also holds at the boundary. The first condition is

$$(x_k)_{k \in \mathbb{N}} \subset \text{int}C, y \in C \quad D_h(y, x_k) \xrightarrow[k \rightarrow \infty]{} 0 \quad \Rightarrow \quad x_k \xrightarrow[k \rightarrow \infty]{} y \quad (\text{A})$$

We know from continuity of h on its domain that strict convexity of h on the whole domain C (not only the interior) is sufficient for (A) [12, Lemma 2.16]. If we assume the opposite implication

$$(x_k)_{k \in \mathbb{N}} \subset \text{int}C, y \in C, \quad x_k \xrightarrow[k \rightarrow \infty]{} y \quad \Rightarrow \quad D_h(y, x_k) \xrightarrow[k \rightarrow \infty]{} 0 \quad (\text{B})$$

then we have the following Féjer argument: assume that (A) and (B) hold true, then any Fejérian sequence $(x_k)_{k \in \mathbb{N}}$ (Definition 1) converges. Indeed, an accumulation point \bar{x} exists by compactity, it must be in S by Definition 1, from property (A), up to a subsequence, $D_h(\bar{x}, x_k) \rightarrow 0$, but Definition 1 ensures that $D_h(\bar{x}, x_k)$ converges so its limit must be 0 and property (B) ensures that $x_k \rightarrow \bar{x}$.

Note that by continuity of h on its domain, condition (B) is equivalent to $\langle \nabla h(x_k), y - x_k \rangle \rightarrow 0$ as $x_k \rightarrow y$. Conditions (A) and (B) date back to Bregman [7] and have been extensively considered in the literature [8, 9, 10, 5, 12, 1, 2, 4] in the same abstract form or with adaptation to broader settings that considered here. Given h , a Legendre function, continuous on its compact domain C , if conditions (A) and (B) are satisfied, then h is called a Bregman function.

1.4 Main results

We are interested in the following question: How restrictive are conditions (A) and (B)? In other words, how much does it take for a continuous Legendre function to be Bregman? We provide the following answer.

Theorem 1. *Let $C \subset \mathbb{R}^n$ be convex, compact, with nonempty interior and h be Legendre with domain C , continuous on C . Then:*

- (A) holds if and only if h is strictly convex on C .
- (B) holds if and only if C is a polytope.

Theorem 1, provides an explicit sufficient condition on h ensuring convergence of all algorithms described in Section 1.2. Indeed if C is a polytope and h is strictly convex on C , then any Fejérian sequence converges. Furthermore, Theorem 1 illustrates the fact that existing Fejér arguments for sequential convergence are only valid for polytopic domains, and convergence analysis for more general domains will require different arguments.

Theorem 1 can be obtained by combining Proposition 1, Proposition 2, Lemma 5 and Lemma 6. We make crucial use of [15, Lemma 1]. Let us mention that Theorem 1 and all our presentation is limited to compact C , but most proof arguments relate to possibly unbounded closed C so that the intermediate results convey information of independent interest for this more general case. The link between strict convexity and (A) is natural and one implication is due to Kiwiel [12]. The connection between (B) and the polytopic nature of C is the most interesting part of our results and we will start with it.

2 Condition (B)

Let us first illustrate failure of (B) and its relation with curvature with a simple example, which was described independently in [3, Remark 2].

2.1 Intuition: incompatibility with curvature

Set $h: x \mapsto -\sqrt{1 - \|x\|^2}$ on $C \subset \mathbb{R}^2$, the unit Euclidean ball in the plane. We have $\nabla h: x \mapsto x \frac{1}{\sqrt{1 - \|x\|^2}}$. Considering polar coordinates in the plane, $x = (r \cos(\theta), r \sin(\theta))$, with e_1 the first basis vector $(1, 0)$, we have

$$h(e_1) - h(x) - \langle \nabla h(x), e_1 - x \rangle = -\sqrt{1 - r^2} + \frac{r(r - \cos(\theta))}{\sqrt{1 - r^2}}$$

Choosing $\theta(r) = \arccos(r - \sqrt{1 - r^2})$ for $r \geq 0$, as $r \rightarrow 1$, $x(r) = (r \cos(\theta(r)), r \sin(\theta(r)))$ goes to e_1 and $D_h(e_1, x) \rightarrow 1$.

Actually, the curve $r \rightarrow x(r)$ converges to e_1 with a vertical tangent as shown in Figure 1. This illustrates the main mechanism of failure of condition B. Because of curvature, chords all meet the interior of C so that the directional derivative of h along chords explodes at their endpoints. Here the curve remains on a chord segment in a region of negative directional derivative along the chord.

From this illustration, one intuition is that the boundary of the compact domain C of a Bregman function should not have too much curvature. More precisely it should not have curvature accumulating anywhere. This intuition is actually correct, the key mechanism is that around extreme points, there is a lot of curvature which allows to generate behavior similar to the circle example above, see Figure 1. This cannot happen too densely otherwise this contradicts (B), and as a result, the extreme points of C have to be isolated, this is expressed in Proposition 1. On the other hand, polyhedra have a very strong structure which will enforce (B) this is Proposition 2.

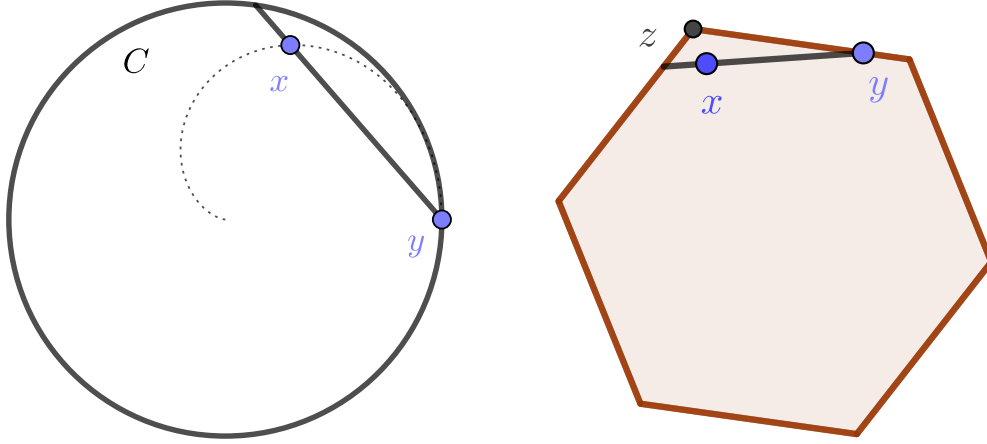


Figure 1: Left: Illustration of the tangential phenomenon, closedness to the boundary and to the opposite extremity of the chord. Right: in general extreme points constitute a region of high curvature and it is possible to find neighboring points in a similar configuration. This cannot happen too much under condition (B) and in particular extreme points should not accumulate anywhere.

2.2 First implication

In this section we prove the following

Proposition 1. *Let $C \subset \mathbb{R}^n$ be convex closed with non-empty interior and h be Legendre on C , continuous on C and satisfy (B). Then the extreme points of C are locally finite. In particular if C is bounded it is a polytope.*

We start with two lemmas

Lemma 1. *Under the assumptions of Proposition 1, let $x \in \text{bd}C$ and $y \in C$ be such that $\frac{x+y}{2} \in \text{int}C$. Then $D_h(y, (1-\lambda)x + \lambda y) \rightarrow +\infty$ as $\lambda \rightarrow 0$.*

Proof : Note that in this case the open segment (x, y) is contained in the interior of C . For any $\lambda \in (0, 1)$, setting $z_\lambda = (1-\lambda)x + \lambda y$, we have

$$D_h(y, (1-\lambda)x + \lambda y) = h(y) - h(z_\lambda) - (1-\lambda) \langle \nabla h(z_\lambda), y - x \rangle.$$

Letting $\lambda \rightarrow 0$, consider any accumulation point, $d \in \mathbb{R}^n$ of $\frac{\nabla h(z_\lambda)}{\|\nabla h(z_\lambda)\|}$. We have $d \in N_C(x)$ by [1, Lemma 4.2] and $y - x \in \text{int}T_C(x)$ (since $\frac{x+y}{2} \in \text{int}C$), so that $\langle d, y - x \rangle < 0$. Therefore, up to a subsequence as $\lambda \rightarrow 0$,

$$\begin{aligned} D_h(y, (1-\lambda)x + \lambda y) &= h(y) - h(z_\lambda) - (1-\lambda) \|\nabla h(z_\lambda)\| \left\langle \frac{\nabla h(z_\lambda)}{\|\nabla h(z_\lambda)\|}, y - x \right\rangle, \\ &\rightarrow +\infty \end{aligned}$$

since $h(z_\lambda) \rightarrow h(x)$, $\|\nabla h(z_\lambda)\| \rightarrow \infty$ and $\langle d, y - x \rangle < 0$. The result follows because d was arbitrary. This concludes the proof. \square

Lemma 2. *Under the hypotheses of Proposition 1, let $a \in \text{bd}C$ be an extreme points and $y \in C$, different from a . Then for any $\epsilon, K > 0$, there exists $c \in \text{int}C$ such that*

$$\begin{aligned} \|c - a\| &\leq \epsilon \\ D_h(y, c) &\geq K. \end{aligned}$$

Proof : First if $[a, y] \cap \text{int}C \neq \emptyset$, then we are in the conditions of Lemma 1, which provides the desired result (with $x = a$). We may therefore assume that $[a, y] \subset \text{bd}C$.

Claim 1. *There exists $\epsilon_1 > 0$ such that for any $b \in \text{int}C \cap B_{\epsilon_1}(a)$, the line from y to b crosses the boundary of C at x such that $\|x - a\| \leq \epsilon$.*

Proof of the claim. We may assume that $\epsilon_1 < \|y - a\|$, so that, since $a \neq y$, the points y and b define a unique line for any $b \in B_{\epsilon_1}(a)$. We also impose that $\epsilon_1 < \epsilon$. Assume toward a contradiction that for all such $\epsilon_1 > 0$, there is $b \in \text{int}C \cap B_{\epsilon_1}(a)$, such that the line from y to b either does not cross the boundary of C or it crosses it at a point at distance greater than ϵ . In both cases, since $b \in B_{\epsilon_1}(a)$ (as we assumed $\epsilon_1 < \epsilon$), the line from y to b exits $B_{\epsilon_1}(a)$ at a point $x \in C$. In this case, using a vanishing sequence of values for ϵ_1 , we can produce a sequence $(b_k)_{k \in \mathbb{N}}$ converging to a such that each segment $[y, b_k]$ can be extended up to $x_k \in C$ where $b_k \in [x_k, y]$ and $\|x_k - a\| = \epsilon$. Passing to the limit, up to subsequences $b_k \rightarrow a$ and $x_k \rightarrow x \in C$, we have $a \in [x, y]$ where $a \neq y$ and $a \neq x$ which contradicts the fact that a is an extreme point of C . This proves the claim.

We may choose $\epsilon_1 \leq \epsilon$, now consider any point $b \in \text{int}C \cap B_{\epsilon_1}(a)$, b belongs to a segment of the form $[x, y]$ for some $x \in \text{bd}C \cap B_{\epsilon_1}(a)$. Note that $[b, x] \subset B_{\epsilon_1}(a)$ so that we can apply Lemma 1 and obtain $c \in [b, x]$ with $D_h(y, c)$ arbitrarily large. This proves the desired result. □

Proof of Proposition 1: Assume that the set of extreme points of C it is not locally finite. This means that we can find a bounded sequence $(z_k)_{k \in \mathbb{N}}$ of pairwise distinct extreme points. This sequence has a converging subsequence, let y be its limit. Since the sequence has pairwise distinct elements, there is at most one $k \in \mathbb{N}$ such that $y = z_k$ and we may remove it from the sequence. In other words $z_k \rightarrow y \in C$ as $k \rightarrow \infty$ and $z_k \neq y$ for all $k \in \mathbb{N}$. We will show that in this case, condition B is violated.

Fix $k \in \mathbb{N}$, using Lemma 2 with $a = z_k$, we may find $x_k \in \text{int}C$ such that $\|x_k - z_k\| \leq \|z_k - y\|$ and $D_h(y, x_k) \geq k$. Now as $k \rightarrow \infty$, we have $x_k \rightarrow y$ and $D_h(y, x_k) \rightarrow +\infty$ which contradicts condition B. This concludes the proof. □

2.3 Reverse implication

Proposition 2. *Let $C \subset \mathbb{R}^n$ be convex closed with non-empty interior and locally polyhedral (represented locally by finitely many affine inequalities, for example a polytope), let h be Legendre on C , continuous on C , then (B) holds.*

The result follows from the following ([15, Lemma 1]).

Lemma 3 (Tseng and Bertsekas). *Let $h: \mathbb{R}^p \mapsto \mathbb{R} \cup +\infty$, be lower semicontinuous and continuous on its domain C . Then*

- For any $y \in C$, there exists a nondegenerate closed ball centered at y such that $C \cap B$ is closed.
- For any $y \in C$, and d such that $y + d \in C$ and sequences $x_k \rightarrow y$ and $d_k \rightarrow d$ such that $x_k + d_k \in C$ for all k , we have

$$\limsup_{k \rightarrow \infty} h'(x_k, d_k) \leq h'(y, d).$$

Lemma 4. *Let C be a polyhedra and $y \in C$. Then there exists $\epsilon > 0$ such that for any $x \in C \cap B_\epsilon(y)$, $2x - y \in C$.*

Proof : Let $a_1, \dots, a_m \in \mathbb{R}^n$ and $b_1, \dots, b_m \in \mathbb{R}$ such that $C = \{x \mid \langle a_i, x \rangle \leq b_i, i = 1, \dots, m\}$. Fix $y \in C$, and $I \subset \{1, \dots, m\}$ the set of active indices, such that $\langle a_i, y \rangle = b_i$ if and only if $i \in I$. By continuity of linear functions, there exists $\epsilon > 0$ such that for all d with $\|d\| \leq 2\epsilon$, and all $i \notin I$, $\langle a_i, y + d \rangle < b_i$. Now for any $x \in C$ such that $\|x - y\| \leq \epsilon$, we have

$$\begin{aligned} \langle a_i, 2x - y \rangle &= \langle a_i, y + 2(x - y) \rangle < b_i && \forall i \notin I, \\ \langle a_i, 2x - y \rangle &= 2\langle a_i, x \rangle - \langle a_i, y \rangle \leq 2b_i - \langle a_i, y \rangle = 2b_i - b_i = b_i && \forall i \in I, \end{aligned}$$

which shows that $2x - y \in C$ and concludes the proof. \square

Proof of Proposition 2: By continuity of h on C , it suffices to show that

$$\lim_{x \rightarrow y, x \in \text{int}C} \langle \nabla h(x), y - x \rangle = 0$$

for all $y \in C$.

Fix $y \in C$ and consider $(x_k)_{k \in \mathbb{N}}$, converging to y . C can be locally represented by a polyhedra and therefore Lemma 4 can be applied to y and C . Let $\epsilon > 0$ be given by Lemma 4 and assume without loss of generality that $\|y - x_k\| \leq \epsilon$ for all $k \in \mathbb{N}$. We have for all $k \in \mathbb{N}$, $x_k + (y - x_k) = y \in C$ so that applying Lemma 4 with $d_k = y - x_k$, which converges to 0

$$\limsup_{k \rightarrow \infty} \langle \nabla h(x_k), y - x_k \rangle \leq h'(y, 0) = 0.$$

Furthermore, by Lemma 4, for all $k \in \mathbb{N}$, we have $2x_k - y = x_k + (x_k - y) \in C$. Therefore, one may apply Lemma 4 with $d_k = x_k - y$, which also converges to 0 to obtain

$$\limsup_{k \rightarrow \infty} \langle \nabla h(x_k), x_k - y \rangle \leq h'(y, 0) = 0.$$

which is equivalent to

$$\liminf_{k \rightarrow \infty} \langle \nabla h(x_k), y - x_k \rangle \geq 0.$$

We have shown that

$$0 \leq \liminf_{k \rightarrow \infty} \langle \nabla h(x_k), y - x_k \rangle \leq \limsup_{k \rightarrow \infty} \langle \nabla h(x_k), x_k - y \rangle \leq 0,$$

so that the limit is 0. This concludes the proof. \square

3 Condition (A)

One implication was given by [12, Lemma 2.16].

Lemma 5 (Kiwiel). *Let $C \subset \mathbb{R}^n$ be convex closed with non-empty interior, let h be Legendre on C , continuous on C and strictly convex on C , then (A) holds.*

The reverse implication follows from Lemma 3 again.

Lemma 6. *Let $C \subset \mathbb{R}^n$ be convex closed with non-empty interior, let h be Legendre on C , continuous on C such that (A) holds, then h is strictly convex on C .*

Proof : Toward a contradiction, assume that h is not strictly convex. This means that there exists $x, y \in C$ such that

$$h\left(\frac{x+y}{2}\right) = \frac{h(x)+h(y)}{2}, \quad (6)$$

which implies that h is affine along the segment $[x, y]$, that is

$$h\left(\frac{x+y}{2} + t(x-y)\right) = h\left(x\left(t + \frac{1}{2}\right) + y\left(\frac{1}{2} - t\right)\right) = \frac{h(x)+h(y)}{2} + t(h(x) - h(y))$$

for all $t \in [-1/2, 1/2]$. In particular, setting $z = \frac{x+y}{2}$, we have

$$h'(z, x-z) = -h'(z, y-z) = \frac{h(x) - h(y)}{2}. \quad (7)$$

Now consider $z_0 \in \text{int}C$ and the sequence $(z_k)_{k \in \mathbb{N}}$ in $\text{int}C$, such that for all $k \in \mathbb{N}$, $k \geq 1$, $z_k = \frac{1}{k}z_0 + \frac{k-1}{k}z \in C$.

We set for all $k \in \mathbb{N}$, $k \geq 1$, $y_k = \frac{2}{k+1}z_0 + \frac{k-1}{k+1}y \in C$. We have for all $k \geq 1$,

$$\begin{aligned} z_k - x &= \frac{1}{k}z_0 + \frac{k-1}{2k}y + \left(\frac{k-1}{2k} - \frac{2k}{2k}\right)x = \frac{2}{2k}z_0 + \frac{k-1}{2k}y - \frac{k+1}{2k}x \\ y_k - x &= \frac{2}{k+1}z_0 + \frac{k-1}{k+1}y - \frac{k+1}{k+1}x = (z_k - x)\frac{2k}{k+1} \\ y_k - z_k &= y_k - x + x - z_k = (z_k - x)\left(\frac{2k}{k+1} - 1\right) = \frac{k-1}{k+1}(z_k - x). \end{aligned}$$

Using Lemma 3, we have

$$\begin{aligned} \limsup_{k \rightarrow \infty} \langle \nabla h(z_k), y_k - z_k \rangle &= \limsup_{k \rightarrow \infty} \frac{k-1}{k+1} \langle \nabla h(z_k), z_k - x \rangle \\ &= -\liminf_{k \rightarrow \infty} \langle \nabla h(z_k), x - z_k \rangle \\ &\leq h'(z, y-z) = -h'(z, x-z), \end{aligned}$$

so that

$$\limsup_{k \rightarrow \infty} \langle \nabla h(z_k), x - z_k \rangle \leq h'(z, x - z) \leq \liminf_{k \rightarrow \infty} \langle \nabla h(z_k), x - z_k \rangle$$

and $\langle \nabla h(z_k), x - z_k \rangle \rightarrow h'(z, x - z)$ as $k \rightarrow \infty$. We deduce by continuity of h using (6) and (7)

$$\lim_{k \rightarrow \infty} h(x) - h(z_k) - \langle \nabla h(z_k), x - z_k \rangle = h(x) - \frac{h(x) + h(y)}{2} - \frac{h(x) - h(y)}{2} = 0$$

So we have that $z_k \rightarrow z \neq x$ but $D_h(x, z_k) \rightarrow 0$ which shows that condition (A) does not hold. This proves the result by contraposition. \square

Acknowledgements

The author would like to thank Jérôme Bolte for his continuous support and fruitful interactions. This work supported by the AI Interdisciplinary Institute ANITI, ANR-19-PI3A-0004, Air Force Office of Scientific Research, Air Force Material Command, USAF, FA9550-19-1-7026 FA8655-22-1-7012, and ANR MaSDOL 19-CE23-0017-01.

References

- [1] Alvarez, F., Bolte, J. and Brahic, O. (2004). Hessian Riemannian gradient flows in convex programming. *SIAM journal on control and optimization*, 43(2), 477–501.
- [2] Auslender, A., Teboulle, M. (2006). Interior gradient and proximal methods for convex and conic optimization. *SIAM Journal on Optimization*, 16(3), 697–725.
- [3] Azizian, W., Iutzeler, F., Malick, J., and Mertikopoulos, P. (2022). On the rate of convergence of Bregman proximal methods in constrained variational inequalities. *arXiv preprint arXiv:2211.08043*.
- [4] Bauschke, H. H., Bolte, J. and Teboulle, M. (2016). A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. *Mathematics of Operations Research*, 42(2), 330–348.
- [5] Bauschke, H. H., and Borwein, J. M. (1997). Legendre functions and the method of random Bregman projections. *Journal of convex analysis*, 4(1), 27–67.
- [6] Beck, A. and Teboulle, M. (2003). Mirror descent and nonlinear projected sub-gradient methods for convex optimization. *Operations Research Letters*, 31(3), 167–175.
- [7] Bregman, L. M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3), 200–217.

- [8] Censor, Y., Lent, A. (1981). An iterative row-action method for interval convex programming. *Journal of Optimization theory and Applications*, 34(3), 321-353.
- [9] Censor, Y., Zenios, S. A. (1992). Proximal minimization algorithm withd-functions. *Journal of Optimization Theory and Applications*, 73(3), 451-464.
- [10] Chen, G., Teboulle, M. (1993). Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM Journal on Optimization*, 3(3), 538-543.
- [11] Combettes, P. L. (2001). Fejér-monotonicity in convex optimization. *Encyclopedia of optimization*, 2, 106-114.
- [12] Kiwiel, K. C. (1997). Free-steering relaxation methods for problems with strictly convex costs and linear constraints. *Mathematics of Operations Research*, 22(2), 326-349.
- [13] A. S. Nemirovsky and D. B. Yudin (1983). *Problem Complexity and Method Efficiency in Optimization*. Wiley-Interscience, New York.
- [14] Polyak, B. T. (1987). *Introduction to optimization*. Optimization software Inc., Publications Division, New York, 1, 32.
- [15] Tseng, P., Bertsekas, D. P. (1991). Relaxation methods for problems with strictly convex costs and linear constraints. *Mathematics of operations research*, 16(3), 462-481.