Journal of Convex Analysis
Volume 18 (2011), No. 3, 601–619

On the Cutting Plane Property and the Bregman Proximal Point Algorithm

Nils Langenberg
Department of Mathematics, University of Trier,
54286 Trier, Germany
langenberg@uni-trier.de

Received: January 7, 2010
Revised manuscript received: May 6, 2010

The Bregman-function-based Proximal Point Algorithm (BPPA) for solving variational inequalities is considered. In this framework the customary assumption of the cutting plane property (CPP) of the related operator is investigated. Since this property cannot be expected in saddle-point-problems, it should be considered as rather restrictive. This paper contributes to the situation when the CPP fails to hold. For this situation, interior proximal(-like) methods have only been constructed for polyhedral sets up to now.

Under very mild assumptions (not implying the CPP) we show that whenever the sequence of iterates (generated by the BPPA) is convergent, its limit can only be a solution of the given problem. Further, using a (known) slight modification of Bregman functions, we show – still without using the CPP – that the sequence generated by the BPPA is convergent to a solution when the feasible set has some special nonlinear structure like a ball.

Keywords: Pseudomonotone operators, variational inequalities, Bregman distances, Proximal Point Algorithm, cutting plane property

1991 Mathematics Subject Classification: 47J20, 65J20, 65K10, 90C26, 90C30

1. Introduction

Let us consider the following variational inequality problem. Given a set $K \subset \mathbb{R}^n$ and a multi-valued operator $F : \mathbb{R}^n \to 2^{\mathbb{R}^n}$, the problem $VI(K, F)$ is to

$$\text{find } x^* \in K \text{ and } f^* \in F(x^*) : \quad \langle f^*, x - x^* \rangle \geq 0 \ \ \forall\, x \in K. \tag{1}$$

Here, $\langle a, b\rangle = b^T a$ denotes the canonical inner product in $\mathbb{R}^n$; the set of solutions of $VI(K, F)$ will be denoted by $SOL(K, F)$.

For the further discussion, we will make use of the following assumption.

Assumption A.

(A.1) Existence of solutions: Throughout this article, $x^* \in K$ is an arbitrary solution and $f^* \in F(x^*)$ is a vector fulfilling (1).

(A.2) The set K admits the following representation:

$$K = \{x \in \mathbb{R}^n : \ g_i(x) \leq 0, \ i \in I_1 \cup I_2\},$$

ISSN 0944-6532 / \$ 2.50

where the functions $g_i : \mathbb{R}^n \to \mathbb{R}$ are affine for $i \in I_1$ and convex and continuously differentiable (but not affine) for $i \in I_2$. Further assume that the set

$$M = \{y \in K : \exists j \in I_2 : g_j(y) = 0\}$$

contains no line segments.

(A.3) There is some $\tilde{x} \in K$ such that $g_j(\tilde{x}) < 0$ for all $j \in I_1 \cup I_2$ (strong version of Slater's constraint qualification).

(A.4) $\mathrm{dom}(F) \cap \mathrm{int}(K) \neq \emptyset$.

For the sake of generality, the operator F will be assumed to have some pseudomonotonicity properties (in the sense of Karamardian [16]) only; see e.g. [6, 9, 17] for sufficient conditions for assumption (A.1) in this situation. We shall give an exemplary result.

Theorem 1.1 (See [6], Theorem 2). *Let $K \subset \mathbb{R}^n$ be closed convex and $F(x)$ be nonempty, convex and compact for each $x \in K$. If F is upper semicontinuous (see e.g.* [9] *for the definition) and pseudomonotone in the sense of Karamardian, then $SOL(K, F)$ is nonempty and compact if and only if $K_\infty \cap \big(F(K)\big)^\circ = \{0\}$, where K_∞ is the recession cone of the set K and $\big(F(K)\big)^\circ$ is the polar cone of $F(K)$.*

Of course, some proofs and assumptions (both with respect to existence results as well as the results below) simplify when F is maximal monotone or also single-valued and continuous.

Assumptions (A.2) and (A.3) represent the state of the art in the existence theory of Bregman-like functions [15]. Property (A.3) here is equivalent to the requirement $\mathrm{int}(K) \neq \emptyset$, which is useful for the discussion of methods providing an interior-point-effect. For example, if all non-affine constraints are strictly convex and the Slater CQ holds, the described set satisfies (A.2) and (A.3). Assumptions of type (A.4) are usual.

For ease of reference, we give a definition of some classes of monotonicity. We restrict ourselves to the classes which are most important on the one hand in the existing literature concerning proximal-like methods and which are most important for the following discussion on the other hand. The reader is referred to [7] for a more detailed overview.

Definition 1.2 (Notions of (pseudo)monotonicity). In the situation $\emptyset \neq D \subset S \subset \mathbb{R}^n$, the mapping $F : S \to 2^{\mathbb{R}^n}$ is said to be

- **monotone on** D if $\langle f^x - f^y, x - y\rangle \geq 0$ for all $x, y \in D$, $f^x \in F(x)$ and $f^y \in F(y)$, and **maximal monotone** if the graph of F is not properly contained in the graph of another monotone operator,
- **paramonotone on** D or also **monotone$^+$ on** D if it is monotone on D and $\langle f^x - f^y, x - y\rangle = 0$ implies $f^x \in F(y)$ and $f^y \in F(x)$,
- **pseudomonotone on** D if for any $x, y \in D$ and any $f^x \in F(x)$ and $f^y \in F(y)$ the inequality $\langle f^y, x - y\rangle \geq 0$ implies $\langle f^x, x - y\rangle \geq 0$,
- **pseudomonotone* on** D if it is pseudomonotone and for $x, y \in D$ and any

$f^x \in F(x)$, $f^y \in F(y)$ the following implication holds true:

$$\langle f^x,\, y - x\rangle = 0 = \langle f^y,\, y - x\rangle \quad \Rightarrow \quad \exists\, k > 0 :\; k \cdot f^x \in F(y). \tag{2}$$

- **pseudomonotone with respect to** $SOL(K, F)$ **on** D if for arbitrary $x \in D$, $f^x \in F(x)$ and any $x^* \in SOL(K, F)$ the following holds: $\langle f^x,\, x^* - x\rangle \le 0$.

It is very easy to establish some relations between the above notions. For example,

paramonotonicity ⇒ monotonicity ⇒ pseudomonotonicity.

We will now define another central operator property.

Definition 1.3 (Cutting plane property / cut property / CPP, see [7, 10]). The operator F is said to have the cutting plane property (for short: cut property / CPP) on a set K, when the following implication holds:

$$\left.\begin{aligned} x^* \in SOL(K, F),\\ x^{**} \in K,\\ f^{**} \in F(x^{**}),\\ \langle f^{**},\, x^* - x^{**}\rangle \ge 0, \end{aligned}\right\} \quad \Rightarrow x^{**} \in SOL(K, F).$$

It is well-known that paramonotone and pseudomonotone* operators have the cut property (see e.g. [11], Proposition 2.3 and [7], Proposition 5). More general, among the pseudomonotone mappings, pseudomonotone* operators are – under very mild additional assumptions – exactly those which have the cut property, see Theorem 4.1 in [10].

However, following Remark 1.2 in [12], the cut property cannot be expected in the discussion of saddle-point-problems of the Lagrangian of a convex program. In this sense, it is a rather restrictive assumption and thus worth a discussion.

The Bregman Proximal Point Algorithm described below is a powerful tool for the stable solution of variational inequalities. Ill-posed (and possibly constrained) problems are solved by means of well-posed unconstrained ones. However, paramonotonicity and other properties implying the cut property are very customary for the discussion of this method, take e.g. [4, 5, 13, 19, 21], where also other remarks on the BPPA can be found. In [18] an extended BPPA method for (in general) not even pseudomonotone problems with composed operators (and a composed sort of cut property) is studied.

Up to now, proximal-like methods not requiring the cut property but still providing an interior-point-effect have only been constructed for polyhedral sets K (see e.g. [1, 8]).

Our central purposes are the following:

- Without using the cut property we will show that if the generated sequence of iterates is convergent, then its limit has to be a solution.

- The geometrical structure (convexity) of the solution set is investigated also for operators without the cut property.
- We give a large class of feasible sets for which the method generates iterates that converge to a solution of the considered problem without using the cut property.

2. Bregman-like functions and the BPPA

2.1. Bregman-like functions

Definition 2.1 (Bregman-like functions, see [13]). Let $S \subset \mathbb{R}^n$ be a nonempty set. A function $h: \ \mathrm{cl}(S) \to \mathbb{R}$ is said to be a Bregman-like function with zone S, when the following holds:

(B.1) S is an open and convex set.
(B.2) h is continuous and strictly convex on $\mathrm{cl}(S)$.
(B.3) $h \in C^1(S)$.
(B.4) The set $\mathcal{M}(x,\alpha) := \{y \in S : \ D_h(x,y) \leq \alpha\}$ is bounded for all fixed $\alpha \in \mathbb{R}$ and $x \in \mathrm{cl}(S)$, where the Bregman distance is defined by

$$D_h(x,y) := h(x) - h(y) - \langle \nabla h(y),\, x-y\rangle, \tag{3}$$

when $x \in \mathrm{cl}(S)$, $y \in S$.
(B.5) If $\{z_k\}_{k\in N}$ is a sequence in S, converging to $z \in \mathrm{cl}(S)$, at least one of the following statements holds:
(a) $D_h(z, z_k) \to 0$ for $k \to \infty$.
(b) If $\overline{z} \neq z$ is another point in $\mathrm{cl}(S)$, then $D_h(\overline{z}, z_k) \to \infty$ $(k \to \infty)$.
(B.6) Let $\{z_k\} \subset \mathrm{cl}(S)$ and $\{y_k\} \subset S$ be two sequences and assume that one of these sequences is convergent. If further $D_h(z_k, y_k) \to 0\,(k \to \infty)$ holds, then the other sequence converges to the same limit as well.

A Bregman-like function h is said to be zone-coercive, if additionally

(B.7) $\nabla h(S) = \mathbb{R}^n$

and boundary coercive, if the following implication holds:

(B.8) When $\{y^k\} \subset S$ and $y^k \to y \in \partial S$, then

$$\lim_{k\to\infty} \langle \nabla h(y^k),\, x - y^k\rangle = -\infty \ \ \forall x \in S,$$

equivalently $D_h(x, y^k) \to +\infty$ for $k \to \infty$.

One readily recognizes that the difference between such Bregman-like functions and standard Bregman functions can be found in condition (B.5). Property (B.6) follows from (B.2) and (B.3) (see Theorem 2.4 in [21]), and zone-coercive functions are also boundary coercive (see Theorem 4.5 in [2] for some detail). Furthermore it is well-known that D_h is a non-negative function and $D_h(x,y) = 0$ if and only if $x = y$ (since h is strictly convex), but D_h is not a distance function in general.

Now let us discuss the existence of a (zone-coercive) Bregman-like function with zone $\mathrm{int}(K)$ when $K \subset \mathbb{R}^n$ admits a description by (A.2) and (A.3). Kaplan and

Tichatschke [13, 14, 15] considered for fixed $\kappa > 0$

$$h(x) := \sum_{i=1}^{m} \phi(g_i(x)) + \frac{\kappa}{2} \cdot ||x||^2, \tag{4}$$

where g_i are the constraints describing K as in Assumption (A.2).

Constructing ϕ according to the following construction assignments, one gets a broad class of Bregman-like functions.

Lemma 2.2 (Construction of Bregman-like functions) (see [15]). *Let the function $\phi : \mathbb{R} \to \mathbb{R}$ used in (4) be constructed with the following properties:*

(C.1) *ϕ is strictly convex, continuous, increasing with $\operatorname{dom}\phi = (-\infty, 0]$ and continuously differentiable on $(-\infty, 0)$.*

(C.2) *It holds $t \cdot \phi'(t) \to 0$ and $\phi'(t) \to \infty$ for $t \uparrow 0$.*

Then the function h, defined by (4), is a strongly convex (with modulus κ) and zone-coercive Bregman-like function with zone $\operatorname{int}(K)$, where K has the properties described in Assumption A.

Note that there are functions ϕ satisfying (C.1)–(C.2), e.g. the potential-like function $\phi(t) = -(-t)^p$ with $p \in (0, 1)$ fixed. However, even if ϕ is chosen as above, the standard Bregman condition (B.5)a) is not always fulfilled (see Example 1 in [13]).

In the following we assume that h is a zone-coercive Bregman-like function with zone $\operatorname{int}(K)$, which is known to imply $\operatorname{dom}(\nabla h) = \operatorname{int}(K)$. Doing so, $K \subset \mathbb{R}^n$ is an arbitrary set which admits a description like in (A.2) and (A.3). Especially, in the absence of affine constraints the boundary ∂K cannot contain any line segment.

Algorithm 2.3: BPPA for $VI(K, F)$.

1. *Let some $x^0 \in \operatorname{int}(K)$ be given. Choose $\chi_0 > 0$ and set $k := 0$.*
2. *If x^k solves the problem $VI(K, F) \to$ STOP.*
3. *Find the next iterate $x^{k+1} \in \operatorname{int}(K)$ and $f^k \in F(x^{k+1})$ such that:*

$$\langle f^k + \chi_k\big(\nabla h(x^{k+1}) - \nabla h(x^k)\big),\, x - x^{k+1}\rangle \geq 0 \quad \forall x \in k, \tag{5}$$

being equivalent to

$$f^k + \chi_k\big(\nabla h(x^{k+1}) - \nabla h(x^k)\big) = 0. \tag{6}$$

4. *Choose $\chi_{k+1} > 0$, set $k := k + 1$ and go to Step 2.*

Algorithm 2.3 describes the method discussed in the sequel. Concerning F, the assumption of maximal monotonicity is quite natural – however, inspired by [19], we shall only assume pseudomonotonicity properties of F and some additional assumptions (related to closedness of the graph etc.) that are more or less implied by maximality, at least usual requirements even in the monotone case. With respect to well-definedness of this method we refer to the discussion in [19] or, for the maximal monotone case, to [4].

For the sake of simplicity, we leave out inexactness concerning the solution of the subproblems as well as ε-enlargements (defined for maximal monotone operators). Nevertheless, these aspects could be integrated as usual without any further restrictions on the given problem, cf. [15, 19, 21].

3. Customary convergence analysis

Let us begin with the convergence analysis of Algorithm 2.3. This section is essentially dedicated to an illustration of the usual convergence proofs using the cut property. To do so, we need some additional assumptions.

Assumption A (continuation).

(A.5) There are $\underline{\chi}, \overline{\chi} > 0$ such that $\underline{\chi} \leq \chi_k \leq \overline{\chi} < \infty$ for all $k \in \mathbb{N}$ (boundedness of regularization parameters).

(A.6) Whenever $\{z^k\} \to z$ and $f^k \in F(z^k)$, then there is a subsequence $f^{k_l} \to \overline{f} \in F(z)$.

(A.7) $\mathrm{dom}(F) \cap K$ is a closed convex set.

There are some methods not requiring that χ_k has to be bounded from above. However, since $\overline{\chi}$ can be chosen arbitrarily, this is not very restrictive. The choice of $\underline{\chi}$ is more important: For example, when F is not monotone, there are situations in which $\underline{\chi}$ could be chosen such that the regularized subproblems nevertheless have strongly monotone operators. Such a situation occurs e.g. when F is Lipschitz continuous (and thus weakly monotone), see Theorems 1 and 2 in [19].

Example 3.1. Consider the pseudomonotone operator $F : \mathbb{R} \to 2^{\mathbb{R}}$,

$$F(x) = \begin{cases} 1, & x < 0 \\ [\frac{1}{2}, 1], & x = 0 \\ \frac{1}{2}, & x > 0. \end{cases}$$

Although this operator obviously has Lipschitz continuous minorants (e.g. the zero mapping), there is no $\chi > 0$ such that the regularized operators $F + \chi(\nabla h - \nabla h(x^k))$ are strongly monotone, where h is any Bregman(-like) function. However, e.g. when using the classical prox-regularization given by $h(x) = \frac{1}{2}||x||^2$, the regularized operators are at least coercive for any $\chi > 0$ which guarantees the existence of at least one solution of each subproblem in the BPPA.

Assumption (A.6) is necessary when passing to the limit and also appears in e.g. [13]. When F is maximal monotone with $K \subset \mathrm{int}(\mathrm{dom}(F))$, then F is locally bounded on K [20] and has a closed graph, of course. Thus property (A.6) is valid in this situation. However, assumption (A.6) can be completely omitted in the maximal monotone case, see Lemma 4.5 and Proposition 4.6 in [21].

The following theorem gives several auxiliary results.

Theorem 3.2 (See [19], Theorem 5). *Suppose that* (A.1)–(A.5) *hold true and that* $x^* \in SOL(K, F)$ *is an arbitrary solution. Assume that the operator* F *is pseudomonotone with respect to the solution set* $SOL(K, F)$ *and let again* $\{x^k\}$ *denote the sequence generated by Algorithm* 2.3. *Then the following statements hold true:*

1. *The sequence* $\{D_h(x^*, x^k)\}$ *is convergent.*
2. *The sequence* $\{x^k\}$ *is bounded.*
3. *The series* $\sum_{k=0}^{\infty} D_h(x^{k+1}, x^k)$ *is convergent.*
4. *The series* $\sum_{k=0}^{\infty}\langle f^k, x^* - x^{k+1}\rangle$ *is convergent.*
5. $\chi_k\langle\nabla h(x^{k+1}) - \nabla h(x^k),\, x^* - x^{k+1}\rangle \to 0$ *for* $k \to \infty$.

Corollary 3.3 (See [19], Corollary 1). *As a direct consequence of the above theorem, we obtain the following:*

1. $\{x^k\}$ *has cluster points, and each cluster point belongs to* K.
2. *Since the sequence* $\{x^k\}$ *is bounded, it is convergent iff it has exactly one cluster point.*
3. *If* $\{x^{k_l}\} \to \overline{x}$ *denote a convergent subsequence and a cluster point, respectively, then we also have* $\{x^{k_l+1}\} \to \overline{x}$.

In view of Theorem 3.2 it is clear that

$$\langle f^{k_l},\, x^* - x^{k_l+1}\rangle \to 0 \tag{7}$$

has to hold. Surely, $x^{k_l+1} \to \overline{x}$ is convergent. The customary line of analysis is as follows: Whenever F has the cut property, one can show that each cluster point of $\{x^k\}$ belongs to $SOL(K, F)$. Afterwards, usually a result is applied which states that if every cluster point of $\{x^k\}$ is a solution, then the entire sequence converges to its only cluster point, i.e. to a solution.

We point this out in an exemplary way.

Lemma 3.4 (See [19], Lemma 8). *Let* $F : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *be a locally bounded operator. If the graph of* F *is closed, then we have that* $\{f^{k_l}\} \to \overline{f}$ *for some* $\overline{f} \in F(\overline{x})$ *with the property that* $\langle\overline{f},\, x^* - \overline{x}\rangle = 0$ *for any* $x^* \in SOL(K, F)$.

Theorem 3.5 (See [19], Theorem 6). *Assume that the assumptions of Lemma* 3.4 *hold true and that* F *either is pseudomonotone** *or Clarke's subdifferential of a pseudoconvex function* f. *Then each cluster point of* $\{x^k\}$ *is a solution of the considered problem.*

The next theorem stems from the discussion of monotone problems, but analyzing the proof we see that no monotonicity properties of F are required. This result will be extended in Theorem 4.6 below.

Theorem 3.6 (See [13], Lemma 1). *If each cluster point of* $\{x^k\}$ *is a solution of* $VI(K, F)$, *then the sequence* $\{x^k\}$ *generated by Algorithm* 2.3 *converges to some* $x^* \in SOL(K, F)$.

It is easy to see that the previous analysis proves the convergence of $\{x^k\}$ to a solution of $VI(K, F)$.

Since Theorem 3.2 only requires pseudomonotonicity with respect to the solution set, but no additional assumption like the cut property, we will make use of Theorem 3.2 and Corollary 3.3 for the rest of the present paper.

Beginning with the next section, we present new results including a new line of convergence analysis of the BPPA.

4. A fundamental lemma and its consequence

Let $\{x^k\}$ denote the sequence generated by Algorithm 2.3. The following lemma will be crucial for the subsequent analysis and only requires (A.6) and the assumptions of Theorem 3.2.

Lemma 4.1. *If the assumptions of Theorem* 3.2 *and property* (A.6) *hold true, then for arbitrary, but fixed* $x \in K$ *there is a cluster point* $\overline{x} = \overline{x}(x)$ *of* $\{x^k\}$ *such that*

$$\langle \overline{f},\, x - \overline{x} \rangle \geq 0, \tag{8}$$

where $\overline{f} \in F(\overline{x})$ *is appropriately chosen.*

Proof. Consider the related sequence $\{D_h(x, x^k)\}$ of Bregman distances. Two cases may occur.

Assume that $\{D_h(x, x^k)\}$ *converges.* Then we have in view of the well-known three-point-formula and Theorem 3.2:

$$\begin{aligned} 0 &= \lim_{k\to\infty} D_h(x, x^k) - D_h(x, x^{k+1}) \\ &= \lim_{k\to\infty} \langle \nabla h(x^{k+1}) - \nabla h(x^k),\, x - x^{k+1} \rangle + \lim_{k\to\infty} D_h(x^{k+1}, x^k) \\ &= \lim_{k\to\infty} \langle \nabla h(x^{k+1}) - \nabla h(x^k),\, x - x^{k+1} \rangle, \end{aligned} \tag{9}$$

which somehow generalizes some parts of Theorem 3.2. Now consider any convergent subsequence $\{x^{k_l}\} \to \overline{x}$ and assume $\{f^{k_l}\} \to \overline{f} \in F(\overline{x})$ in view of (A.6). Passing to the limit in the scheme

$$\langle f^{k_l} + \chi_{k_l}\big(\nabla h(x^{k_l+1}) - \nabla h(x^{k_l})\big),\, z - x^{k_l+1} \rangle \geq 0 \quad \forall\, z \in K, \tag{10}$$

we obtain in view of (9) for $z = x$:

$$\langle \overline{f},\, x - \overline{x} \rangle \geq 0, \tag{11}$$

which, as stated above, holds true for any cluster point $\overline{x}$.

Now assume that $\{D_h(x, x^k)\}$ *does not converge.* Then this sequence cannot be monotonically decreasing, since it would otherwise be convergent as a non-negative sequence.

Thus, there are infinitely many $k \in \mathbb{N}$ such that $D_h(x, x^{k+1}) \geq D_h(x, x^k)$. Let $\{k_l\} \subset \mathbb{N}$ be a subsequence such that

$$D_h(x, x^{k_l+1}) \geq D_h(x, x^{k_l}) \quad \forall\, l \in \mathbb{N}. \tag{12}$$

Then we have, again using the three-point-formula and Theorem 3.2:

$$\begin{aligned} 0 &\geq \limsup_{l\to\infty} D_h(x, x^{k_l}) - D_h(x, x^{k_l+1}) \\ &= \limsup_{l\to\infty} \langle \nabla h(x^{k_l+1}) - \nabla h(x^{k_l}),\, x - x^{k_l+1} \rangle + \lim_{l\to\infty} D_h(x^{k_l+1}, x^{k_l}) \\ &= \limsup_{l\to\infty} \langle \nabla h(x^{k_l+1}) - \nabla h(x^{k_l}),\, x - x^{k_l+1} \rangle. \end{aligned} \tag{13}$$

Since the corresponding subsequence $\{x^{k_l}\}$ is bounded, there is a convergent subsequence $\{x^{k_{l_j}}\} \to \overline{x}$. Theorem 3.2 and property (B.6) now imply that also $\{x^{k_{l_j}+1}\} \to \overline{x}$. Thus, passing to the limit $j \to \infty$ in the iteration scheme and remembering (13) and (A.6), we obtain

$$\begin{aligned} 0 &\leq \liminf_{j\to\infty} \langle f^{k_{l_j}} + \chi_{k_{l_j}} \big(\nabla h(x^{k_{l_j}+1}) - \nabla h(x^{k_{l_j}})\big),\, x - x^{k_{l_j}+1}\rangle \\ &\leq \limsup_{j\to\infty} \langle \nabla h(x^{k_{l_j}+1}) - \nabla h(x^{k_{l_j}}),\, x - x^{k_{l_j}+1}\rangle + \limsup_{j\to\infty} \langle f^{k_{l_j}},\, x - x^{k_{l_j}+1}\rangle \\ &\leq \limsup_{j\to\infty} \langle f^{k_{l_j}},\, x - x^{k_{l_j}+1}\rangle \\ &= \langle \overline{f},\, x - \overline{x}\rangle, \end{aligned}$$

where again, in view of (A.6), without loss of generality $\{f^{k_l}\} \to \overline{f} \in F(\overline{x})$ is assumed. □

The previous result shall be related to the cut property.

Remark 4.2. 1. It is important to note that the above proof unfortunately does not exclude that the cluster point $\overline{x}$ in (8) can depend on x. Also $\overline{f}$ can depend on x, at least as long as F is not single-valued.

2. Of course, if F has the cut property, convergence of $\{x^k\}$ is well-known. Thus, $\{x^k\}$ only has one cluster point; in consequence, this cluster point $\overline{x}$ would not depend on x.

3. In fact, we have shown that if $\{D_h(x, x^k)\}$ converges, property (8) holds true for any convergent subsequence of $\{x^k\}$, that is, any cluster point $\overline{x}$ fulfills (8).

4. If $\{D_h(x, x^k)\}$ does not converge, property (8) holds true for any cluster point $\overline{x}$ of $\{x^k\}$ such that there is a subsequence $\{x^{k_n}\} \to \overline{x}$ fulfilling $D_h(x, x^{k_n+1}) \geq D_h(x, x^{k_n})$ for all $n \geq n_0$.
If one knew that the entire sequences $\{D_h(x, x^k)\}$ necessarily is monotonically increasing for any $x \in K$, (8) would be valid for each cluster point $\overline{x}$ and any $x \in K$, in other words, each cluster point would be a solution. This establishes a relation between monotonicity of the sequences $\{D_h(x, x^k)\}$ and the convergence of $\{x^k\}$.

Lemma 4.1 enables us to prove the following central theorem. Recall that a set-valued mapping $F : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is called upper hemicontinuous on K if for all $u, v \in K$ the mapping

$$\lambda \mapsto F(\lambda u + (1-\lambda)v) \tag{14}$$

is upper semicontinuous on $[0, 1]$, that is, F is upper semicontinuous restricted to the line segments of K. (Especially, maximal monotone operators are upper hemicontinuous; see [20] for continuity properties of set-valued mappings.)

Theorem 4.3. *Let the sequence $\{x^k\}$ be generated by Algorithm 2.3 and F be pseudomonotone (with respect to the solution set) with non-empty, convex and compact values and upper hemicontinuous on $K \cap \operatorname{dom}(F)$.*[1] *If $\{x^k\}$ is convergent, then its limit point is a solution of $VI(K, F)$.*

[1]These assumptions hold for F maximal monotone with $K \cap \operatorname{dom}(F) \subset \operatorname{int}(\operatorname{dom}(F))$.

Proof. Since $\{x^k\}$ is convergent, it has exactly one cluster point $\overline{x}$ (thus, if $F(\overline{x})$ is a singleton, the assertion is already proved). By Lemma 4.1 we know that for each $y \in K$ there is some $\overline{f} = \overline{f}(y) \in F(\overline{x})$ with

$$\langle \overline{f},\, y - \overline{x}\rangle \geq 0. \tag{15}$$

Formulated in other terms, for every $y \in K$,

$$\max_{f \in F(\overline{x})} p(f, y) \geq 0, \tag{16}$$

where $p(f, y) = \langle f,\, y - \overline{x}\rangle$. Obviously, p is bilinear, the set K is convex, and since F is assumed to have convex and compact values, we can apply Lemma 1 in [3] to obtain the existence of $f^\circ \in F(\overline{x})$ with

$$0 \leq p(f^\circ, y) = \langle f^\circ,\, y - \overline{x}\rangle \quad \forall\, y \in K, \tag{17}$$

which just means that $\overline{x} \in SOL(K, F)$. $\square$

An analysis similar to (15)–(17) is helpful to prove the following lemma which extends Theorem 2.3.5 in [9]. Again, the maximal monotone case is included as a special situation.

Lemma 4.4. *If F is pseudomonotone with respect to the solution set, has non-empty convex and compact values on K and, further, is upper hemicontinuous on K, then the set $SOL(K, F)$ is convex.*

Proof. It will be sufficient to establish the following equations:

$$\begin{aligned} SOL(K, F) &= M_1 := \{x \in K : \langle f,\, y - x\rangle \geq 0\ \forall y \in K,\ \forall f \in F(y)\} \\ &= M_2 := \bigcap_{y \in K} \{x \in K : \langle f,\, y - x\rangle \geq 0\ \forall f \in F(y)\} \\ &= M_3 := \{x \in K : \exists f \in F(y) : \langle f,\, y - x\rangle \geq 0\ \forall y \in K\} \\ &= M_4 := \bigcap_{y \in K} \{x \in K : \exists f \in F(y) : \langle f,\, y - x\rangle \geq 0\} \end{aligned}$$

It is easy to see that $SOL(K, F) \subset M_1$ due to pseudomonotonicity (w.r.t. $SOL(K, F)$) of F. Further, M_2 is just another notation for M_1; the same holds for M_3 and M_4. Further, $M_2 \subset M_4$ ($M_1 \subset M_3$ respectively) is trivial to see. Hence,

$$SOL(K, F) \subset M_1 = M_2 \subset M_3 = M_4 \tag{18}$$

is already proven and it remains to show that $M_4 \subset SOL(K, F)$.

Let $x^* \in M_4$ and $z \in K$, $\lambda \in [0, 1]$ be arbitrary. Let $y(\lambda) := \lambda z + (1 - \lambda)x^*$, thus $y(\lambda) \in K$ due to the convexity of K.

Since $x^* \in M_4$ we know that for any $\lambda \in (0, 1)$ there is $f = f(\lambda) \in F(y(\lambda)$ such that

$$0 \leq \langle f(\lambda),\, y(\lambda) - x^*\rangle = \langle f(\lambda),\, \lambda(z - x^*)\rangle,$$

consequently, $\langle f(\lambda),\, z - x^* \rangle \geq 0$.

Due to upper hemicontinuity of F, the operator F is upper semicontinuous along the line segment $[x^*, z]$. Thus, for any $n \in \mathbb{N}$ there exists some number $\lambda_n > 0$ such that

$$F(y(\lambda)) \subset F(x^*) + B_{\frac{1}{n}}(0) \quad \forall\, \lambda \in (0, \lambda_n], \tag{19}$$

where $B_r(x)$ denotes the closed ball with radius r around some x.

Hence, for $\lambda \in (0, \lambda_n]$ there exists a decomposition of the form

$$f(\lambda) = f^*(\lambda) + \gamma(\lambda)$$

where $f^*(\lambda) \in F(x^*)$ and $\gamma(\lambda) \in B_{\frac{1}{n}}(0)$.

Of course we might assume $\lambda_n < 1$ for every $n \in \mathbb{N}$. We have

$$\begin{aligned} 0 \leq \langle f(\lambda),\, z - x^* \rangle &= \langle f^*(\lambda),\, z - x^* \rangle + \langle \gamma(\lambda),\, z - x^* \rangle \\ &\leq \langle f^*(\lambda),\, z - x^* \rangle + ||\gamma(\lambda)|| \cdot ||z - x^*|| \\ &\leq \langle f^*(\lambda),\, z - x^* \rangle + \frac{1}{n}||z - x^*|| \end{aligned}$$

by means of the Cauchy-Schwarz inequality.

Choosing in particular $\lambda = \lambda_n$ this just means

$$\langle f^*(\lambda_n),\, z - x^* \rangle \geq -\frac{1}{n}||z - x^*||.$$

Now let $n \to \infty$. Then $\lambda_n \to 0$ and $y(\lambda_n) \to x^*$ and, according to Assumption (A.6), we can assume that $f^*(\lambda_n) \to \overline{f} \in F(x^*)$ holds.

Since $z \in K$ was arbitrary, F has nonempty, convex and compact values, K is convex and the mapping

$$p(f, y) = \langle f,\, z - x^* \rangle$$

is bilinear and fulfils

$$\max_{f \in F(x^*)} p(f, z) \geq 0$$

for every $z \in K$, it again follows from Lemma 1 in [3] that there is some $f^\circ \in F(x^*)$ such that

$$p(f^\circ, z) \geq 0 \quad \forall\, z \in K,$$

that is, $\langle f^\circ,\, z - x^* \rangle \geq 0$ for every $z \in K$, i.e. $x^* \in SOL(K, F)$. Consequently, $M_4 \subset SOL(K, F)$, and since each M_i is convex (e.g. M_2 as an intersection of convex sets), $SOL(K, F)$ has to be convex as well. □

Remark 4.5. It is easy to see that when F is maximal monotone (with $\operatorname{int}(\operatorname{dom}(F)) \subset K$) or even single-valued and continuous, both the formulation of Theorem 4.3 and a fortiori the proof get significantly simpler. However, we chose this way to cover a rather general case.

We close this section with another useful result.

The following theorem generalizes Theorem 3.6 since no assumption on every cluster point is required to obtain the same result. Using this theorem, it will be sufficient to show that there is one cluster point of $\{x^k\}$ that is a solution, it will no longer be necessary to show that every cluster point of $\{x^k\}$ is a solution.

Theorem 4.6. *Assume that h is a Bregman-like function with zone* $\mathrm{int}(K)$ *and* $\{z^k\} \subset \mathrm{int}(K)$ *be a sequence such that* $\{D_h(\overline{z}, z^k)\}$ *converges for at least one cluster point* $\overline{z}$*, being the limit of some subsequence* $\{z^{k_l}\}$ *of* $\{z^k\}$.

Then we have $\{z^k\} \to \overline{z}$*, i.e. the entire sequence converges to its (only) cluster point.*

Proof. Suppose that the first alternative in (B.5) holds (i.e. we are discussing the case of a standard Bregman function). Due to the convergence $z^{k_l} \to \overline{z}$ we obtain from (B.5)(a) that $D_h(\overline{z}, z^{k_l}) \to 0$ for $l \to \infty$, i.e. $\overline{D} = 0$ has to be valid. Now $D_h(\overline{z}, z^k) \to 0$ for $k \to \infty$ and (B.6) yield the assertion $z^k \to \overline{z}$ for $k \to \infty$.

On the other hand, if (B.5)(b) holds, consider two convergent subsequences

$$z^{k_l} \to \overline{z} \quad \text{and} \quad z^{k_m} \to z',$$

and suppose $\overline{z} \neq z'$. Then (B.5)(b) yields $D_h(\overline{z}, z^{k_m}) \to \infty$, but that is a contradiction to the assumed convergence of the latter sequence. Therefore, there can only be one cluster point; hence, convergence is proved. □

In consequence, it suffices to show that for one cluster point the corresponding sequence of Bregman distances is convergent, for example, since this (single) cluster point is a solution. Note that in the latter proof no special property of F is required.

The following section deals with the investigation of some conditions permitting to conclude that $\{x^k\}$ indeed is convergent.

5. Convergence analysis

We first show that deducing convergence of $\{x^k\}$ is no problem if there is a solution in $\mathrm{int}(K)$.

Theorem 5.1. *Let the assumptions of Theorem* 3.2 *as well as* (A.6), (A.7) *hold true. Then the following are equivalent:*

1. $SOL(K,F) \cap \mathrm{int}(K) \neq \emptyset$.
2. $\{x^k\}$ *has all cluster points in* $\mathrm{int}(K)$.
3. $\{x^k\}$ *has one cluster point in* $\mathrm{int}(K)$.
4. $\{x^k\}$ *converges to a solution in* $\mathrm{int}(K)$.

Proof. It will be sufficient to prove "*1.* ⇒ *2.*" and "*3.* ⇒ *4.*".

"*1.* ⇒ *2.*" Let again $\{x^{k_l}\} \to \overline{x}$ denote a convergent subsequence. In view of Corollary 3.3 we directly obtain $\{x^{k_l+1}\} \to \overline{x}$.

Now for some $x^{**} \in SOL(K,F) \cap \mathrm{int}(K)$ we know that $D_h(x^{**}, x^{k_l})$ converges, which directly implies the convergence of $\langle \nabla h(x^{k_l}), x^{**} - x^{k_l} \rangle$.

On the other hand, since a zone-coercive Bregman-like function especially is *boundary coercive*, we can conclude that $\overline{x}$ and therefore each cluster point of the generated sequences has to belong to $\operatorname{int}(K)$.

"3. ⇒ 4." Now let $\{x^{k_l}\} \to \overline{x}$ denote a convergent subsequence such that $\overline{x} \in \operatorname{int}(K)$. As a consequence, we have

$$\nabla h(x^{k_l+1}) - \nabla h(x^{k_l}) \to 0, \quad l \to \infty.$$

Now pass to the limit ($l \to \infty$) in the iteration scheme (5). In view of $x - x^{k_l+1} \to x - \overline{x}$ and $\overline{x} \in \operatorname{int}(K)$ it follows $f^{k_l} \to 0$. Now (A.6) implies $0 \in F(\overline{x})$, i.e. $\overline{x} \in SOL(K, F)$. Thus, one cluster point of $\{x^k\}$ is a solution and by Theorem 4.6 the entire sequence $\{x^k\}$ converges to this solution. □

Note that the original version of Theorem 4.6, i.e. Theorem 3.6, would not be sufficient for the preceding proof.

Besides some sufficient condition for the existence of a solution in $\operatorname{int}(K)$, e.g. in [18], also convexity of the solution set (cf. Lemma 4.4) may be useful, see below.

As a consequence of the above results, we can concentrate on the situation that convergence of $\{x^k\}$ is unknown. The following lemma is helpful in this situation.

Lemma 5.2. *Consider the situation of Theorem* 3.2.

1. *If $\{x^k\}$ has a cluster point $\overline{x}$ in $\operatorname{int}(K)$, then $\{x^k\} \to \overline{x} \in SOL(K, F)$.*
2. *If no cluster point of $\{x^k\}$ belongs to $\operatorname{int}(K)$, it holds that*

$$\lim_{k\to\infty} \operatorname{dist}(x^k, \partial K) = 0. \tag{20}$$

Proof. The first assertion has already been proved above.

Thus, consider the second assertion and assume the contrary. Then there is some $\varepsilon > 0$ and a subsequence $\{x^{k_l}\}$ such that $\operatorname{dist}(x^{k_l}, \partial K) \geq \varepsilon$ holds true for each $l \in \mathbb{N}$. Since $\{x^{k_l}\}$ is bounded, we can without loss of generality assume that it is convergent to some $\overline{x}$.

Due to $\operatorname{dist}(\overline{x}, \partial K) = \lim_{l\to\infty} \operatorname{dist}(x^{k_l}, \partial K) \geq \varepsilon$ we obtain $\overline{x} \in \operatorname{int}(K)$, that is, $\{x^k\}$ has a cluster point in $\operatorname{int}(K)$. As one prefers to conclude: This is contradictory to the assumption, and on the other hand, it implies convergence of $\{x^k\}$ in view of Theorem 5.1. □

Thus, for the remaining discussion we can assume that the sequence $\{x^k\}$ has to approach the boundary ∂K. Convergence of $\{x^k\} \to \overline{x}$ would imply (via boundary coerciveness)

$$\lim_{k\to\infty} \langle \nabla h(x^k),\, x - x^k \rangle \to -\infty \quad \forall\, x \in \operatorname{int}(K), \tag{21}$$

which is by definition of D_h and continuity of h equivalent to

$$\lim_{k\to\infty} D_h(x, x^k) \to +\infty \quad \forall\, x \in \operatorname{int}(K). \tag{22}$$

However, we can prove this result – which may be considered as a necessary condition for convergence of $\{x^k\}$ – even without known convergence of $\{x^k\}$.

Lemma 5.3. *If in the situation of Theorem* 3.2 *the sequence* $\{x^k\}$ *has no cluster points in* $\mathrm{int}(K)$*, then for the generated sequence* $\{x^k\}$ *it holds*

$$\lim_{k\to\infty} D_h(x, x^k) \to +\infty \quad \forall\, x \in \mathrm{int}(K). \tag{23}$$

Proof. In view of Lemma 5.2 we know $||\nabla h(x^k)|| \to +\infty$ for $k \to \infty$. Since h is strictly convex, we have $\mathrm{dom}(\nabla h^*) = \nabla h(\mathrm{int}(K)) = \mathbb{R}^n$, where h^* denotes the conjugate function of h (see [2]). Thus, $\nabla h(x^k), \nabla h(x) \in \mathrm{int}(\mathrm{dom}(h^*))$ follows.

Since $D_{h^*}(\cdot, \nabla h(x))$ is coercive by Theorem 3.7 in [2], we can conclude

$$D_h(x, x^k) = D_{h^*}(\nabla h(x^k), \nabla h(x)) \to +\infty \;\text{ for } k \to \infty. \tag{24}$$

□

As the following theorem shows, the preceding results can be used to obtain convergence of $\{x^k\}$ to a solution of $VI(K, F)$ when the set K has a special (nonlinear) structure, which is related to the existence theory of Bregman-like functions. Note that for this central result we do not require any assumption implying the cut property; especially, no paramonotonicity is required.

Theorem 5.4. *Consider the situation of Lemma* 5.3*. Assume that* ∂K*, as described by Assumption* A*, does not contain any line segment*[2] *and that* (A.6), (A.7) *hold. Then* $\{x^k\} \to x^* \in SOL(K, F)$*.*

Proof. Let $\{x^{k_l}\} \to \overline{x}$ be a convergent subsequence with $\overline{x} \in \partial K$. It can be shown that in this situation for every $z \in K$, $z \neq \overline{x}$, it has to hold

$$D_h(z, x^{k_l}) \to +\infty, \quad l \to \infty. \tag{25}$$

For $z \in \mathrm{int}(K)$ this follows from Lemma 5.3; in consequence, the assertion remains to be shown for $z \in \partial K$ only.

Indeed, consider the used definition (4) of Bregman-like functions:

$$h(x) = \sum_{i=1}^{m} \varphi(g_i(x)) + \frac{\kappa}{2}||x||^2. \tag{26}$$

We can assume that $g_j(\overline{x}) = 0$ for some $j \in I_2$. From the convexity of $\varphi \circ g_i$ and $||\cdot||^2$ we obtain

$$D_h(z, x^{k_l}) \geq \varphi(g_j(z)) - \varphi(g_j(x^{k_l})) - \varphi'(g_j(x^{k_l}))\langle \nabla g_j(x^{k_l}),\, z - x^{k_l}\rangle, \tag{27}$$

since the remaining linearization terms in the definition of D_h are non-negative.

In view of (C.2) it follows that $\varphi'(g_j(x^{k_l})) \to +\infty$ for $l \to \infty$, whereas

$$\lim_{l\to\infty} \varphi(g_j(x^{k_l})) = 0 \quad \text{and} \quad \lim_{l\to\infty} \langle \nabla g_j(x^{k_l}),\, z - x^{k_l}\rangle = \langle \nabla g_j(\overline{x}),\, z - \overline{x}\rangle$$

[2]In view of Assumption A this holds when there are no affine constraints, i.e. $I_1 = \emptyset$.

are simple consequences of continuity. Thus, in view of (27), it will be sufficient to show $\langle \nabla g_j(\overline{x}),\, z-\overline{x}\rangle < 0$. Of course, we only have to consider the case $z \in \partial K$, since for $z \in \operatorname{int}(K)$ the assertion $D_h(z, x^{k_l}) \to +\infty$ already follows from Lemma 5.3.

For $z \in \partial K$ we can conclude that $v := \frac{1}{2}(z+\overline{x}) \in \operatorname{int}(K)$, since ∂K does not contain any line segment. Thus, $g_j(v) < 0$ and $g_j(\overline{x}) = 0$. In consequence,

$$0 > g_j(v) - g_j(\overline{x}) = \langle \nabla g_j(\overline{x}),\, v - \overline{x}\rangle = \frac{1}{2}\langle \nabla g_j(\overline{x}),\, z - \overline{x}\rangle. \tag{28}$$

Therefore, $\langle \nabla g_j(\overline{x}),\, z - \overline{x}\rangle < 0$, and in consequence, (25) holds.

Now let us turn to the conclusion of convergence of $\{x^k\}$. Remember that the proof presented here does not require the cut property.

Assume that $\overline{x}$ is not a solution. Then for any solution $\{D_h(x^*, x^{k_l})\} \to \infty$, which contradicts the convergence of $\{D_h(x^*, x^{k_l})\}$, shown in Theorem 3.2.

Thus, $\overline{x}$ has to be a solution and by Theorem 4.6, $\{x^k\} \to \overline{x}$. □

Remark 5.5. The assumptions on the special structure of the set K as well as the non-existence of solutions in $\operatorname{int}(K)$ seem to imply that $SOL(K, F)$ is a singleton.

Indeed, at least under reasonable assumptions, $SOL(K, F)$ is a convex set. If there were at least two solutions, thinking of the geometrical structure of K we directly see that there has to be a solution in $\operatorname{int}(K)$. But this has been excluded beginning with Lemma 5.2.

However, uniqueness of the solution in the discussed situation can also be reasoned without convexity of $SOL(K, F)$. The following results (derived from analogue results for a more general problem and a more general iteration scheme in [18]) establish a crucial relation between the structure of the set K and the solution set.

Lemma 5.6 (cf. Lemma 6.2 in [18]). *When properties of K, F and h guarantee the convergence of the sequences $\{x^k\}$ (to a solution) and $\{D_h(x^*, x^k)\}$, and there is more than one solution, then there is at least one solution x^* with $g_i(x^*) < 0$ for all $i \in I_2$.*

Proof. Assume that there are two solutions x^*, x^{**}, but $g_l(x^*) = g_j(x^{**}) = 0$ holds true for some $j, l \in I_2$. Without loss of generality we may assume $\{x^k\} \to x^*$ (indeed, if $\{x^k\}$ would converge to a solution x^{***} fulfilling $g_i(x^{***}) < 0$ for all $i \in I_2$ the assertion directly follows). Then we have, due to the convexity of the functions $\varphi \circ g_i$,

$$D_h(x^{**}, x^k) \geq \varphi(g_l(x^{**})) - \varphi(g_l(x^k)) - \varphi'(g_l(x^k))\langle \nabla g_l(x^k),\, x^{**} - x^k\rangle. \tag{29}$$

Now due to (C.2) we have

$$\lim_{k\to\infty} \varphi'(g_l(x^k)) = +\infty, \tag{30}$$

and due to simple continuity arguments the limits

$$\varphi(0) = \lim_{k\to\infty} \varphi(g_l(x^k)) \quad \text{and} \quad \lim_{k\to\infty} \langle \nabla g_l(x^k),\, x^{**} - x^k\rangle = \langle \nabla g_l(x^*),\, x^{**} - x^*\rangle \tag{31}$$

exist. $M := \{y \in K : \exists j_0 \in I_2 : g_{j_0}(y) = 0\}$ contains x^*, x^{**} as elements, but since it does not contain any line segment by (A.2), we have $\frac{1}{2}(x^* + x^{**}) =: v \notin M$. Hence, especially $g_l(v) < 0$ and therefore, just using convexity of g_l,

$$0 > g_l(v) - g_l(x^*) \geq \langle \nabla g_l(x^*),\, v - x^* \rangle = \frac{1}{2}\langle \nabla g_l(x^*),\, x^{**} - x^* \rangle \tag{32}$$

Now, concatenating (29)–(32) we obtain $D_h(x^{**}, x^k) \to +\infty$ for $k \to \infty$, which obviously is a contradiction to the convergence of $\{D_h(x^{**}, x^k)\}$. □

One directly obtains two interesting (and important) consequences.

Corollary 5.7 (cf. Corollary 6.3 in [18]). *Suppose that the assumptions of Lemma* 5.6 *are fulfilled.*

1. *The sequence* $\{x^k\}$ *generated by the BPPA converges to a solution* x^* *with* $g_i(x^*) < 0$ *for all* $i \in I_2$.
2. *If there are no affine restrictions, i.e. if* $I_1 = \emptyset$, *there is at least one solution* $x^* \in SOL(K, F) \cap \operatorname{int}(K)$.

Proof. *1.* Assume $\{x^k\} \to x^*$ and that there is another solution x^{**}. Then convergence of $\{D_h(x^{**}, x^k)\}$ is known. On the other hand, if $g_l(x^*) = 0$ for some $l \in I_2$, then following the proof of Lemma 5.6, we have $\{D_h(x^{**}, x^k)\} \to \infty$, which is a contradiction. Hence $g_i(x^*) < 0$ for every $i \in I_2$.

2. According to Lemma 5.6 there is a solution x^* which fulfills $g_i(x^*) < 0$ for each $i \in I_2$, but in the absence of affine constraints, $g_i(x^*) < 0$ holds for every $i = 1, \ldots, m$, i.e. $x^* \in \operatorname{int}(K)$. □

Especially, assuming the non-existence of solutions in $\operatorname{int}(K)$, Corollary 5.7 reasons that uniqueness of the solution is not restrictive. If there is more than one solution, then there is one solution in the interior, and thus convergence is ensured by Theorem 5.1. In other words, considering separate situations permits to conclude convergence in the entire framework.

We shall remark that the extended convergence sensing property (B.5) is crucial for the above analysis. This extension of the standard Bregman property (B.5)(a) has been introduced in [5] and further investigated in [13, 14, 15] for the existence of Bregman-like functions. However, the methods discussed in these references explicitly require paramonotonicity of F.

The following result summarizes the preceding discussion.

Theorem 5.8. *Let the assumptions of Theorem* 3.2 *as well as* (A.6) *and* (A.7) *hold true*[3] *and suppose that* ∂K *does not contain any line segment.*[4] *Then the sequence* $\{x^k\}$ *generated by the BPPA converges to a solution.*

[3]Especially, F is just assumed to be pseudomonotone with respect to $SOL(K, F)$.
[4]cf. Assumption A. This is especially true if every describing constraint g_i is strictly convex, and therefore e.g. balls are covered.

Consequently, in this situation the cut property (and thus also paramonotonicity etc.) are no longer necessary for the convergence of the sequence $\{x^k\}$ (generated by the BPPA) to a solution of the given problem.

6. Concluding Remarks

We considered the classical Bregman Proximal Point Algorithm which transforms an ill-posed constrained variational inequality into well-posed unconstrained systems of equations.

In the discussion of the BPPA the assumption of paramonotonicity and related properties implying the cut property are very customary ones. Since these properties are not given in saddle-point-problems of the Lagrangian of a convex program, they should be considered as rather restrictive.

The present paper contains a new analysis for the BPPA which permits:

- to show under only slightly more restrictive assumptions that if the generated sequence of iterates is convergent, then its limit has to be a solution,
- to prove convexity of the solution set under similar assumptions,
- to show using weak hypotheses that if the the boundary of the feasible set K does not contain any line segment, the method generates a sequence of iterates that converges to a solution.

If $F : \mathbb{R}^n \to \mathbb{R}^n$ is pseudomonotone*, then we obtain a new and shorter proof that each cluster point is a solution of $VI(K, F)$. Indeed:

Lemma 6.1. *If $F : \mathbb{R}^n \to \mathbb{R}^n$ is pseudomonotone* and the assumptions of Theorem* 3.2 *hold, then the generated sequence $\{x^k\}$ converges to some $x^* \in SOL(K, F)$.*

Proof. Suppose (in view of Theorem 4.6) that there is a cluster point $\overline{x}$ which is not a solution. Then there is $z \in K$ such that

$$\langle F(\overline{x}),\, z - \overline{x}\rangle < 0. \tag{33}$$

From Theorem 3.2 we know that

$$\langle F(\overline{x}),\, \overline{x} - x^*\rangle = 0, \tag{34}$$

and since F is pseudomonotone, we have $\langle F(x^*),\, \overline{x} - x^*\rangle = 0$. Now pseudomonotonicity* yields the existence of $k > 0$ such that $F(\overline{x}) = kF(x^*)$. Summing up (33) and (34), we obtain the contradiction

$$k \cdot \langle F(x^*),\, z - x^*\rangle = \langle F(\overline{x}),\, z - \overline{x}\rangle + \langle F(\overline{x}),\, \overline{x} - x^*\rangle < 0. \tag{35}$$

□

Let us shortly justify why we think that the iterates should also converge to a solution if K has a more general structure. To do this, we assume the contrary, namely that $\{x^k\}$ does not convergence. In view of the above results, this would have the following consequences:

- In view of Theorems 4.3 and 5.1: There are at least two subsequences $\{x^{k_l}\} \to \overline{x}$, $\{x^{k_j}\} \to x^\infty$ with $x^\infty \neq \overline{x}$, and every cluster point belongs to ∂K.
- In view of Theorems 4.6 and 3.2: $\{D_h(\cdot, x^k)\}$ is divergent for every cluster point $\overline{x}$ of $\{x^k\}$, but convergent for any solution $x^* \in SOL(K, F)$.
- In view of Theorems 4.6 and 3.2: None of the cluster points is a solution. That is, for each cluster point $\overline{x}$ there is some $z = z(\overline{x}) \in K$ such that

$$\langle \overline{f},\, z - \overline{x} \rangle < 0 \tag{36}$$

for $\overline{f} \in F(\overline{x})$ as above (without loss of generality, by continuity one might assume $z \in \operatorname{int}(K)$ to obtain $D_h(z, x^k) \to \infty$ from Lemma 5.3). However, such a z cannot be a solution of $VI(K, F)$, since for every solution x^* it holds (following Theorem 3.2)

$$\langle \overline{f},\, x^* - \overline{x} \rangle = 0. \tag{37}$$

Remark 6.2. As shown above, convergence of $\{x^k\}$ is related to Bregman distances. Let us point this out in some thoughts.

If $\{x^{k_l}\} \to \overline{x}$ is a convergent subsequence, but $\overline{x}$ is not a solution, due to continuity we have $\langle f^{k_l},\, z - x^{k_l+1} \rangle < 0$ for l large enough. But in view of the iteration scheme this implies

$$\begin{aligned} 0 &< \langle \nabla h(x^{k_l+1}) - \nabla h(x^{k_l}),\, z - x^{k_l+1} \rangle \\ &= D_h(z, x^{k_l}) - D_h(z, x^{k_l+1}) - D_h(x^{k_l+1}, x^{k_l}) \end{aligned}$$

Thus, $D_h(z, x^{k_l}) > D_h(z, x^{k_l+1})$ for all l sufficiently large. Since without loss of generality $z \in \operatorname{int}(K)$, this has to fit together with the proven fact $\{D_h(z, x^k)\} \to \infty$.

At this place we can see that if the divergent sequence $\{D_h(z, x^k)\}$ was shown to be monotonically increasing, a contradiction would be obtained.

References

[1] A. Auslender, M. Teboulle, S. Ben-Tiba: A logarithmic-quadratic proximal method for variational inequalities, Comput. Optim. Appl. 12 (1999) 31–40.

[2] H. Bauschke, J. Borwein: Legendre functions and the method of random Bregman projections, J. Convex Analysis 4 (1997) 27–67.

[3] E. Blum, W. Oettli: From optimization and variational inequalities to equilibrium problems, Math. Stud. 63 (1994) 123–145.

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

[5] Y. Censor, A. Iusem, S. A. Zenios: An interior point method with Bregman functions for the variational inequality problem with paramonotone operators, Math. Program. 81 (1998) 373–400.

[6] J.-P. Crouzeix: Pseudo-monotone variational inequality problems: existence of solutions, Math. Program. 78 (1997) 305–314.

[7] J.-P. Crouzeix, D. L. Zhu, P. Marcotte: Conditions ensuring the applicability of cutting-plane methods for solving variational inequalities, Math. Program. 88 (2000) 521–539.

[8] P. da Silva, J. Eckstein, C. Humes: Rescaling and stepsize selection in proximal methods using separable generalized distances, SIAM J. Optim. 12 (2001) 238–261.

[9] F. Facchinei, J.-S. Pang: Finite-Dimensional Variational Inequalities and Complementarity Problems, Springer, New York (2003).

[10] N. Hadjisavvas, S. Schaible: On a generalization of paramonotone maps and its application to solving the stampacchia variational inequality, Optimization 55 (2006) 593–604.

[11] A. Iusem: On some properties of paramonotone operators, J. Convex Analysis 5 (1998) 269–278.

[12] A. Kaplan, R. Tichatschke: Interior proximal method for variational inequalities: case of non-paramonotone operators, Set-Valued Anal. 12 (2004) 357–382.

[13] A. Kaplan, R. Tichatschke: Bregman functions and proximal methods for variational problems with nonlinear constraints, Optimization 56(1–2) (2007) 253–265.

[14] A. Kaplan, R. Tichatschke: Interior proximal method for variational inequalities on non-polyhedral sets, Discuss. Math., Differ. Incl. Control Optim. 27(1) (2007) 71–93.

[15] A. Kaplan, R. Tichatschke: On the convergence of interior proximal methods for variational inequalities on non-polyhedral sets, Discuss. Math., Differ. Incl. Control Optim. 30(1) (2010) 51–59.

[16] S. Karamardian: Complementarity problems over cones with monotone and pseudomonotone maps, J. Optimization Theory Appl. 18 (1976) 445–454.

[17] B. Kien, J.-C. Yao, N. Yen: On the solution existence of pseudomonotone variational inequalities, J. Glob. Optim. 41 (2008) 135–145.

[18] N. Langenberg: Convergence analysis of an extended auxiliary problem principle with various stopping criteria, Optim. Methods Softw. 26(1) (2011) 127–154.

[19] N. Langenberg: Pseudomonotone operators and the Bregman Proximal Point Algorithm, J. Glob. Optim. 47(4) (2010) 537–555.

[20] R. T. Rockafellar, R. J.-B. Wets: Variational Analysis, Springer, Berlin (1997).

[21] M. Solodov, B. Svaiter: An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions, Math. Oper. Res. 25 (2000) 214–230.