J Glob Optim (2010) 47:537–555
DOI 10.1007/s10898-009-9470-7

Pseudomonotone operators and the Bregman Proximal Point Algorithm

Nils Langenberg

Received: 22 May 2009 / Accepted: 27 August 2009 / Published online: 10 October 2009

Abstract To permit the stable solution of ill-posed problems, the Proximal Point Algorithm (PPA) was introduced by Martinet (RIRO 4:154–159, 1970) and further developed by Rockafellar (SIAM J Control Optim 14:877–898, 1976). Later on, the usual proximal distance function was replaced by the more general class of Bregman(-like) functions and related distances; see e.g. Chen and Teboulle (SIAM J Optim 3:538–543, 1993), Eckstein (Math Program 83:113–123, 1998), Kaplan and Tichatschke (Optimization 56(1–2):253–265, 2007), and Solodov and Svaiter (Math Oper Res 25:214–230, 2000). An adequate use of such generalized non-quadratic distance kernels admits to obtain an interior-point-effect, that is, the auxiliary problems may be treated as unconstrained ones. In the above mentioned works and nearly all other works related with this topic it was assumed that the operator of the considered variational inequality is a maximal monotone and paramonotone operator. The approaches of El-Farouq (JOTA 109:311–326, 2001), and Schaible et al. (Taiwan J Math 10(2):497–513, 2006) only need pseudomonotonicity (in the sense of Karamardian in JOTA 18:445–454, 1976); however, they make use of other restrictive assumptions which on the one hand contradict the desired interior-point-effect and on the other hand imply uniqueness of the solution of the problem. The present work points to the discussion of the Bregman algorithm under significantly weaker assumptions, namely pseudomonotonicity [and an additional assumption much less restrictive than the ones used by El-Farouq and Schaible et al. We will be able to show that convergence results known from the monotone case still hold true; some of them will be sharpened or are even new. An interior-point-effect is obtained, and for the generated subproblems we allow inexact solutions by means of a unified use of a summable-error-criterion and an error criterion of fixed-relative-error-type (this combination is also new in the literature).

Research is supported by the Landesgraduiertenförderung Rheinland-Pfalz.

N. Langenberg (✉)
Department of Mathematics, University of Trier, 54286 Trier, Germany
e-mail: langenberg@uni-trier.de

Keywords Pseudomonotone operators · Variational inequalities · Bregman distances · Proximal Point algorithm · Interior-point-effect

Mathematics Subject Classification (2000) 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 possibly multi-valued operator $\mathcal{T} : \mathbb{R}^n \to 2^{\mathbb{R}^n}$, the problem $VI(K, \mathcal{T})$ is to

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

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

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

Assumption A

(A.1) Existence of solutions: $x^* \in K$ is a solution and $t^* \in \mathcal{T}(x^*)$ fulfills (1). Further, the domain qualification ri $(\mathrm{dom}\mathcal{T}) \cap K \neq \emptyset$ holds true.

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

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

where the functions $g_i : \mathbb{R}^n \to \mathbb{R}$ are affine for $i \in I_1$ and convex and continuously differentiable for $i \in I_2$. Further assume that the set

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

contains no line segments [15].

(A.3) K has nonempty interior, i.e. there is some $\tilde{x} \in K$ such that $g_j(\tilde{x}) < 0$ for all $j \in I_1 \cup I_2$ (a kind of Slater's CQ).

Sufficient conditions for the existence of a solution are discussed in [7,11,20] and some references therein, for example. Our central purposes are:

- With respect to the existing approaches [10,25] with proximal-like methods for pseudomonotone inequalities, we will make essential relaxations of the assumptions on the problems under consideration. Up to now it has been assumed that the gradient of the regularizing (Bregman) functional is Lipschitz continuous which contradicts to consider the auxiliary problems as unconstrained. We do not impose any assumption which implies that $SOL(K, \mathcal{T})$ is neccessarily a singleton. Also an inexact solution of the auxiliary problems is allowed.
- With respect to the existing literature for monotone problems, we will consider a unified framework which combines two different stopping criteria for the subproblems. We will show that all known results can be transferred to the new framework, some of them are even strengthened. Not only since we deal with pseudomonotone operators, some of the convergence results are also new for the monotone case.

Especially, our method admits to solve pseudoconvex constrained optimization problems by means of well-posed unconstrained convex optimization problems. The relaxation of monotonicity to pseudomonotonicity is also of interest in the discussion of Nash equilibrium problems. The following simple example illustrates that monotonicity here is quite restrictive, even if each players' cost function is strongly convex in his strategy variables.

Example 1 Consider the Nash game with two players, each has one strategy variable x_i and an objective function (which are to be minimized):

$$f_1(x_1, x_2) = x_1^2 + ax_1x_2 + cx_1, \quad f_2(x_1, x_2) = x_2^2 + bx_1x_2 + dx_2.$$

The operator of the variational inequality [11] is monotone iff its Jacobian

$$JF(x) = \begin{pmatrix} 2 & a \\ b & 2 \end{pmatrix} \quad \forall\, x$$

is positive semidefinite, which is the case only when $(a+b)^2 \leq 16$. Thus, even if each function f_i is strongly convex in x_i, monotonicity only appears in rather special cases. Of course, when there are more players and when each player has more than one strategy variable, monotonicity should intuitively be an even more restrictive assumption.

Since a main purpose of this paper is an essential relaxation of the monotonicity properties of the operator of the variational inequality and 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 [8] for a more detailed overview.

Definition 1 (*Notions of* (*pseudo*)*monotonicity*) In the situation $\emptyset \neq D \subset S \subset \mathbb{R}^n$, a mapping $\mathcal{F}: S \to 2^{\mathbb{R}^n}$ is said to be

- *monotone on D* if $\langle f^x - f^y, x - y\rangle \geq 0$ for all $x, y \in D$, $f^x \in \mathcal{F}(x)$ and $f^y \in \mathcal{F}(y)$, and *maximal monotone* if the graph of $\mathcal{F}$ 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 $f^x \in \mathcal{F}(y)$ and $f^y \in \mathcal{F}(x)$,
- *pseudomonotone on D* if for any $x, y \in D$ and any $f^x \in \mathcal{F}(x)$ and $f^y \in \mathcal{F}(y)$ 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 $f^x \in \mathcal{F}(x)$, $f^y \in \mathcal{F}(y)$ the following implication holds true:

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

Since it will be of interest in the sequel, we additionally define an operator $\mathcal{F}$ to be pseudomonotone with respect to the solution set $SOL(K, \mathcal{F})$ (of a given problem) if one has $\langle f^x, x - x^*\rangle \geq 0$ for all feasible points $x \in K$, $f^x \in \mathcal{F}(x)$ and any solution $x^* \in SOL(K, \mathcal{F})$, see for example [11].

It is obvious that paramonotonicity implies all the other properties, whereas monotonicity and pseudomonotonicity* just imply pseudomonotonicity. Now let us give examples of important classes of operators.

Lemma 1 (see [14]) *If $f: D \to \mathbb{R}$ is a proper convex lower semicontinuous (lsc) function, its subdifferential ∂f is a maximal monotone and paramonotone operator. It is single-valued if f is differentiable in each $x \in D$.*

Despite Lemma 1, paramonotonicity is a rather restrictive assumption:

Example 2 Consider the convex optimization program

$$\min f(x) \quad \text{s.t. } g(x) \le 0. \tag{3}$$

Defining the associated Lagrange function by $\mathcal{L}(x, y) := f(x) + \langle y, g(x)\rangle$, the operator for the associated saddle-point-problem is $[\partial_x \mathcal{L}, -\partial_y \mathcal{L}]$, and it is known that this operator is not paramonotone in general [14].

Thus, it is worth considering (para-)monotonicity in some more detail. The following lemmata give examples for operators with some weaker properties.

Lemma 2 (see [8, Proposition 3]) *Let $f : D \to \mathbb{R}$ be differentiable and $g : f(D) \to \mathbb{R}$ differentiable with $g'(z) > 0$ for all $z \in f(D)$ such that $g \circ f$ is a convex function on D. Then f is pseudoconvex and its gradient ∇f is pseudomonotone* with $k = 1$.*

Lemma 3 (see [8, Proposition 4]) *If $f : D \to \mathbb{R}$ is pseudoconvex and differentiable, then its gradient ∇f is pseudomonotone* on D.*

In the pseudoconvex case, unfortunately the convenient additional property yielded by paramonotonicity in the monotone case does not hold true any longer; as seen in Lemma 3, only a weaker property is guaranteed in this case. But as will be seen in the sequel, it is strong enough to guarantee convergence of the method under consideration.

The following example deals with the set-valued case.

Lemma 4 (see [12, Corollary 3.2]) *If $f : D \to \mathbb{R} \cup \{+\infty\}$ is pseudoconvex and locally Lipschitz continuous, then its Clarke subdifferential $\partial^\circ f$ is a set-valued pseudomonotone operator.*

The latter situation might be considered as a prime example for set-valued pseudomonotone operators; thus it deserves to be discussed in some detail.

Lemma 5 (see [6, Proposition 2.4.3] and the following Corollary) *Given a locally Lipschitz function f on a convex subset $D \subset \mathbb{R}^n$, it holds that each minimum $x^* \in D$ of f over D also solves the variational inequality problem $VI(D, \partial^\circ f)$, where again $\partial^\circ f$ denotes the Clarke subdifferential of f.*

If f is pseudoconvex and if additionally we could show for some $\overline{x} \in D$ that there is $\overline{f} \in \partial^\circ f(\overline{x})$ such that $\langle \overline{f}, x^* - \overline{x}\rangle = 0$ holds true for any minimum $x^* \in D$, then pseudoconvexity yields $f(\overline{x}) = f(x^*)$, i.e. $\overline{x}$ is a solution.

As described in [12], operators of the form $\mathcal{T} = \partial^\circ f$ where f is a pseudoconvex and locally Lipschitz function do not necessarily have to be maximal in the sense that there is no proper pseudomonotone extension of their graphs, which is a difference to the monotone case. But in the same article it is shown that there is at least one equivalent operator $\hat{\mathcal{T}}$ which has the latter property and that the solution sets of $VI(D, \mathcal{T})$ and $VI(D, \hat{\mathcal{T}})$ coincide.

The main focus of this paper is to establish the convergence of a rather general Bregman-function based proximal-like method for pseudomonotone operators which are additionally pseudomonotone*. The latter assumption is not very restrictive: In the class of pseudomonotone mappings, pseudomonotone* operators are —under very mild additional assumptions—exactly those which have the cutting plane property (see [13] for the proof): If $x^* \in SOL(K, \mathcal{T})$ and there is any $x^{**} \in K$ with some $t^{**} \in \mathcal{T}(x^{**})$ such that $\langle t^{**}, x^* - x^{**}\rangle \ge 0$ holds true, then $x^{**} \in SOL(K, \mathcal{T})$ as well.

The above notion of pseudomonotonicity, taken from Karamardian [19], should not be mixed up with pseudomonotonicity in the sense of Brézis [1]:

Definition 2 (*Pseudomonotonicity in the sense of Brézis*) In the above situation $\emptyset \neq D \subset S \subset \mathbb{R}^n$, a mapping $\mathcal{F} : S \to 2^{\mathbb{R}^n}$ is said to be *pseudomonotone on D* (in the sense of Brézis), when in the situation

$$\{x^k\} \subset D, \; x^k \to \overline{x}, \; f^k \in \mathcal{F}(x^k), \; \limsup_{k\to\infty} \langle f^k, \; x^k - \overline{x} \rangle \leq 0 \tag{4}$$

(for all $k \in \mathbb{N}$) it always holds true that

$$\forall\, y \in D \quad \exists f_y \in \mathcal{F}(\overline{x}) : \quad \langle f_y, \overline{x} - y \rangle \leq \liminf_{k\to\infty} \langle f^k, x^k - y \rangle. \tag{5}$$

As can be seen easily, the latter concept rather deals with continuity, cf. [2].

2 Bregman-like functions and the BPPA

2.1 Bregman-like functions

We start with the definition of Bregman-like functions and some remarks.

Definition 3 (*Bregman-like functions, see* [17])
Let $S \subset \mathbb{R}^n$ be a nonempty, open and convex 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) h is continuous and strictly convex on $\mathrm{cl}(S)$ and differentiable on S.
(B.2) The set $\mathcal{M}(x, \alpha) := \{y \in S : \quad D_h(x, y) \leq \alpha\}$ is bounded for all fixed $\alpha \in \mathbb{R}$ and $x \in \mathrm{cl}(S)$, where the Bregman distance is defined by

$$D_h(x, y) := h(x) - h(y) - \langle \nabla h(y), \; x - y \rangle, \tag{6}$$

when $x \in \mathrm{cl}(S)$, $y \in S$.
(B.3) 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.4) 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.5) $\nabla h(S) = \mathbb{R}^n$.

A Bregman-like function h is said to have the cone property, if additionally

(B.6) For each $x \in K$ there are constants $\alpha(x) > 0$, $c(x) \in \mathbb{R}$ such that for all $y \in \mathrm{int}(K)$ we have

$$D_h(x, y) + c(x) \; \geq \; \alpha(x) \cdot ||x - y||. \tag{7}$$

Further, it is assumed that $\alpha(x) \geq \underline{\alpha} > 0$ and $c(x) \leq \overline{c}$.

One readily recognizes that the difference between such Bregman-like functions and standard Bregman functions can be found in condition (B.3). In Solodov and Svaiter [26] it is shown that (B.4) follows from (B.1). Further, it is well-known that D_h is non-negative and $D_h(x, y) = 0$ if and only if $x = y$. (B.6) strengthens (B.2) and especially holds true whenever h is strongly convex (cf. [16]).

Now let us discuss the existence of a (zone-coercive) Bregman-like function with zone $\text{int}(K)$ when $K \subset \mathbb{R}^n$ admits a description by (A.2) and (A.3). Kaplan and Tichatschke [15,17,18] considered for fixed $\kappa > 0$

$$h(x) := \sum_{i=1}^{m} \phi(g_i(x)) + \frac{\kappa}{2} \cdot ||x||^2, \tag{8}$$

where g_i, $i = 1, \ldots, m$, are the constraints describing K as in (A.2). Constructing ϕ according to the following construction assignments, one gets a broad class of Bregman-like functions.

Lemma 6 (Construction of Bregman-like functions) (see [15]) *If* $\phi : \mathbb{R} \to \mathbb{R}$ *used in* (8) *is strictly convex, continuous and increasing with* $\text{dom}\, \phi = (-\infty, 0]$, *and* ϕ *is continuously differentiable on* $(-\infty, 0)$ *with* $t \cdot \phi'(t) \to 0$ *and* $\phi'(t) \to \infty$ *for* $t \uparrow 0$, *then the function* h, *defined by* (8), *is a strongly convex (with modulus* κ*) and zone-coercive Bregman-like function with zone* $\text{int}(K)$.

There are functions ϕ with the above properties, e.g. $\phi(t) = -(-t)^p$ with $p \in (0, 1)$ fixed. However, even if ϕ is chosen as above, the standard Bregman condition (B.3)a) is not always fulfilled (see Example 1 in [17]).

By the way we emphasize that sets described by (A.2), (A.3) define the state of the art in the theory of the existence of Bregman-like functions [15]. For example, if all non-affine constraints are strictly convex, the described set satisfies (A.2). Nevertheless, as the following example shows, there are convex sets that do not satisfy these conditions.

Example 3 Assume that $K \subset \mathbb{R}^n$ with $n \geq 3$ is of the form

$$K = \{x \in \mathbb{R}^n : \rho_1(x_1, \ldots, x_j) + \rho_2(x_{j+1}, \ldots, x_n) \leq 0\} \tag{9}$$

for some $1 \leq j < n$, where ρ_1 is a convex, non-affine function and ρ_2 is an affine function.

For the sake of simplicity we consider the case $j = 1$, $\rho_1(x_1) = x_1^2$ and $\rho_2(x_2, x_3) = x_2 + x_3$. That is, $K = \{x \in \mathbb{R}^3 : \; x_1^2 + x_2 + x_3 \leq 0\}$. So K is described by exactly one restriction which is neither affine nor strictly convex. But the boundary $\partial K = \{x \in \mathbb{R}^3 : x_1^2 + x_2 + x_3 = 0\}$, which equals the set M in the definition of (A.2) here, and it can be seen in Fig. 1 that this boundary contains line segments (it even contains entire lines).

Analogously examples in the more general setting (9) can be constructed.

In the following we assume that h is a strictly convex, zone-coercive Bregman-like function with zone $\text{int}(K)$; whenever it is strongly convex, κ will denote the modulus of strong convexity. Property (A.2) is required for the existence of a Bregman-like function, but not for the subsequent analysis itself.[1]

[1] Of course, the classical PPA is obtained by $h(x) = \frac{1}{2}||x||^2$ and does not require properties like (A.2) and (A.3), since no interior-point-effect can be expected.

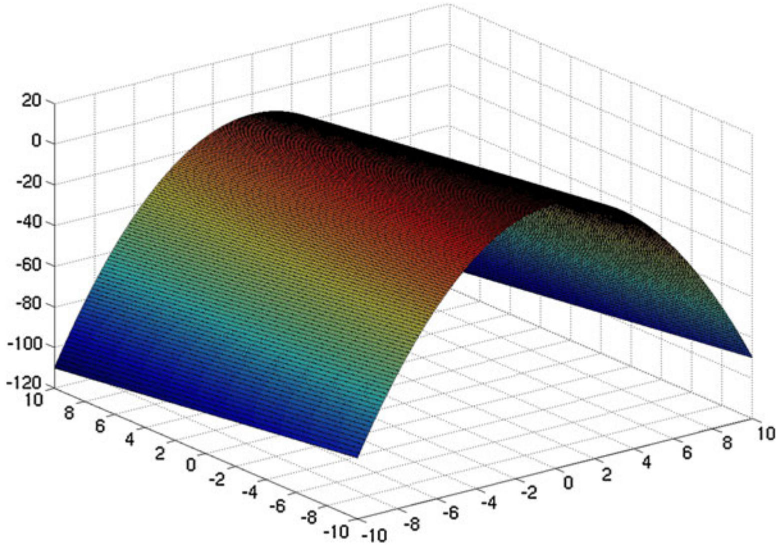

Fig. 1 The boundary of the set in example 3

2.2 The exact BPPA and well-definedness

Using the above concept of Bregman-like functions, we turn to the discussion of the Bregman Proximal Point Algorithm. For the sake of simplicity, we first introduce its exact form and turn to the discussion of some inexactness tolerance afterwards. Note that for $K = \mathbb{R}^n$ and $h(x) = \frac{1}{2}||x||^2$ we obtain the classical PPA with the convenient property that ∇h is Lipschitz continuous which fails to hold for zone-coercive Bregman-like functions.

Algorithm 1: BPPA for $VI(K, \mathcal{T})$ (exact version)

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

$$t^k + \chi_k \left(\nabla h(x^{k+1}) - \nabla h(x^k)\right) = t^k + \chi_k \nabla_1 D_h(x^{k+1}, x^k) = 0. \tag{10}$$

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

Here $\nabla_1 D_h$ denotes the gradient of D_h with respect to the first argument.

Let us care about well-definedness of Algorithm 1 (which will imply well-definedness of the methods mentioned and discussed in the sequel). In other words, we look for sufficient conditions for unique solvability of (10). In the maximal monotone case this is true, see e.g. Proposition 3 in [2].

The next sufficient condition is a significant relaxation since no monotonicity is required. First, remember that the considered algorithm should transform pseudomonotone problems into well-posed problems with strongly monotone operators. In other words, the operator $\mathcal{T}$

should admit such a transformation, i.e. there should be some constant $\underline{\chi}$ such that $\mathcal{T} + \underline{\chi}\mathcal{I}$ is strongly monotone, where here $\mathcal{I}$ denotes the identity mapping.

Definition 4 An operator $\mathcal{T}$ is said to be weakly monotone, if there exists some $L > 0$ such that $\langle t^x - t^y, x - y\rangle \geq -L \cdot ||x - y||^2$ for all $t^x \in \mathcal{T}(x)$, $t^y \in \mathcal{T}(y)$, i.e. such that $\mathcal{T} + L \cdot \mathcal{I}$ is monotone.

We see that if $\mathcal{T}$ is weakly monotone with modulus $L > 0$ and $\tilde{L} > L$ is given, then $\mathcal{T} + \tilde{L}\mathcal{I}$ is strongly monotone with modulus $\tilde{L} - L$. Also it is easy to see that if $\mathcal{T}$ is Lipschitz continuous with modulus L, then it is weakly monotone with the same modulus.

Now we are able to formulate a central theorem of well-definedness which seems to be the best result in literature, e.g. [10,25].

Theorem 1 *If $\mathcal{T}$ is a continuous and single-valued weakly monotone operator with modulus $L > 0$, ∇h is a strongly monotone operator with modulus $\kappa > 0$ and $\chi_k \geq \underline{\chi} > L/\kappa$ holds true, then (10) is uniquely solvable.*

In addition to results known up to now, we can also guarantee well-definedness for a class of set-valued operators.

Theorem 2 *If $\mathcal{T} = \partial^0 f$ is the Clarke subdifferential of a lower semicontinuous function f and weakly monotone with modulus $L > 0$, ∇h is a strongly monotone operator with modulus $\kappa > 0$ and $\chi_k \geq \underline{\chi} > L/\kappa$ holds true, then (10) is uniquely solvable.*

Proof We have that $\mathcal{T} + \chi_k \nabla h = \partial^0(f + \chi_k h)$ is the Clarke subdifferential of a proper, strongly convex and lower semicontinuous function. In this (convex) case the Clarke subdifferential equals the common convex subdifferential [6]. Therefore it is maximal monotone and strongly monotone; thus, a unique solution exists. □

In Counterexample 2.1 and Proposition 3.1 in [27] some pseudomonotone operator $\mathcal{T}$ is discussed. It is shown that there is no regularization parameter such that the regularized operator is at least pseudomonotone.

Consequently, the authors rightly pose the question why one should be interested in replacing one (maybe ill-posed) pseudomonotone by a sequence of (probably also ill-posed) problems where not even pseudomonotonicity is guaranteed any longer. However, some rather simple calculation shows that the operator in the cited example is not weakly monotone. In consequence, weak monotonicity is essential here.

To resume the above discussion we stress that the used assumption of weak monotonicity seems to be a very suitable one. So, for the sake of well-definedness only, we assume that $\mathcal{T}$ is weakly monotone. As will be seen below, this is not necessary to prove convergence to a solution.

2.3 Two inexact versions of the BPPA

Algorithm 1 is rather hard to execute for the following two reasons:

- No errors are permitted since equation (10) has to be solved exactly. To make it more tractable, some modification should be introduced.
- In the case of a set-valued operator $\mathcal{T}$ it might be challenging or even impossible to compute the set $\mathcal{T}(x)$ exactly for some/each $x \in K$. Thus, an outer approximation $\mathcal{T}(x) \subset \mathcal{T}^{\varepsilon}(x)$ is required (inspired by bundle methods for the computation of ε-subgradients).

Solodov and Svaiter [26] introduced a concept of inexactness that might be described as an error criterion of *fixed-relative-error-type*. Algorithm 2 illustrates their version of the BPPA.

Algorithm 2: BPPA for $VI(K, \mathcal{T})$ (Solodov and Svaiter)

1. Let some $x^0 \in \text{int}(K)$ be given. Choose $\chi_0 > 0$ and set $k := 0$.
2. If x^k solves the problem $VI(K, \mathcal{T}) \rightarrow$ STOP.
3. Find $x^{k+1} \in \text{int}(K)$, $y^k \in K$ and $t^k \in \mathcal{T}(y^k)$ such that:

$$t^k + \chi_k \left(\nabla h(x^{k+1}) - \nabla h(x^k)\right) = 0 \tag{11}$$

with, for some parameter $\sigma \in (0, 1)$,

$$D_h(y^k, x^{k+1}) \leq \sigma \cdot D_h(y^k, x^k). \tag{12}$$

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

Their idea of inexact computing consists in the option to choose $t^k \in \mathcal{T}(y^k)$ where y^k is some point near x^k (in the sense of (12)). Once y^k is computed, then x^{k+1} can be obtained easily whenever ∇h is explicitly invertible. According to [4], Bregman functions with this convenient property are available for polyhedra and balls, but despite this argumentation, Bregman-like functions of form (8) scarcely possess an explicitly invertible gradient [4].

Theorem 3 (see [26]) *Assume that $\mathcal{T}$ is a maximal monotone operator, a relative error $\sigma \in [0, 1)$ is chosen and h is a Bregman function with zone* int(K). *Let the regularization parameters $0 \leq \chi_k \leq \overline{\chi}$ be bounded. If in addition $\mathcal{T}$ is paramonotone or if there is at least one solution in* int(K)*, then the sequence $\{x^k\}$ generated by Algorithm 2 converges to some solution $x^* \in SOL(K, \mathcal{T})$.*

Clearly, although there are sufficient conditions for the existence of a solution in the interior (see e.g. [21]), it is rather hard or even impossible to check. Also, Solodov and Svaiter do not permit to make use of Bregman-like functions; since Bregman functions (in the proper sense) are only known for polyhedral sets, this is another restriction as well.

In Algorithm 3 the BPPA-version of Kaplan and Tichatschke, going back to Eckstein (see e.g. [9,17]), is formulated. They also allow an outer approximation of set-valued operators (namely the common ε-enlargement of maximal monotone operators).

Algorithm 3: BPPA for $VI(K, \mathcal{T})$ (Kaplan and Tichatschke)

1. Let some $x^0 \in \text{int}(K)$ be given. Choose $\chi_0 > 0$, $\varepsilon_k \geq 0$ and set $k := 0$.
2. If x^k solves the problem $VI(K, \mathcal{T}) \rightarrow$ STOP.
3. Find $x^{k+1} \in \text{int}(K)$ and $t^k \in \mathcal{T}^{\varepsilon_k}(x^{k+1})$ such that for any $x \in K$

$$\langle t^k + \chi_k \left(\nabla h(x^{k+1}) - \nabla h(x^k)\right), x - x^{k+1}\rangle \geq -\delta_k \cdot ||x - x^{k+1}||. \tag{13}$$

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

In this case, we have the following convergence result.

Theorem 4 (see [17]) *Assume that $\mathcal{T}$ is a maximal monotone operator, $\sum_{k=1}^{\infty} \max\{\delta_k, \varepsilon_k\} < \infty$ holds true and h is a Bregman-like function with zone* $\operatorname{int}(K)$. *Let the regularization parameters $0 \leq \chi_k \leq \overline{\chi}$ be bounded. If in addition $\mathcal{T}$ is the subdifferential of a proper convex lsc function or a paramonotone operator bounded on an open set containing K, then the sequence $\{x^k\}$ generated by Algorithm 3 converges to some $x^* \in SOL(K, \mathcal{T})$.*

Due to the required summability of δ_k the inexactness criterion is said to be of *summable-error-type*. This criterion can be interpreted as follows: If a pair (x^{k+1}, t^k) is found such that $||t^k + \chi_k \left(\nabla h(x^{k+1}) - \nabla h(x^k)\right)|| \leq \delta_k$, then x^{k+1} is a solution of (10).

3 The BPPA for pseudomonotone problems

We propose a generalization of both the inexact methods discussed above. The iteration scheme is illustrated in Algorithm 4.

Algorithm 4: Generalized BPPA for $VI(K, \mathcal{T})$

1. Let some $x^0 \in \operatorname{int}(K)$ be given. Choose $\chi_0 > 0$, $\delta_0 \geq 0$, set $k := 0$.
2. If x^k solves the problem $VI(K, \mathcal{T}) \rightarrow$ STOP.
3. Find the next iterate $x^{k+1} \in \operatorname{int}(K)$, a vector $y^k \in \operatorname{int}(K)$ and $t^k \in \mathcal{T}(y^k)$ such that:

$$\langle t^k + \chi_k \left(\nabla h(x^{k+1}) - \nabla h(x^k)\right), x - y^k\rangle \geq -\delta_k \cdot ||x - y^k|| \tag{14}$$

   holds for each $x \in K$ and

$$D_h(y^k, x^{k+1}) \leq \sigma \cdot D_h(y^k, x^k). \tag{15}$$

   for some parameter $\sigma \in [0, 1)$.
4. Choose $\chi_{k+1} > 0$, $\delta_{k+1} \geq 0$ and set $k := k + 1$ and go to Step 2.

Useful definitions of enlargements for pseudomonotone operators seem to be unknown in literature. One might define in analogy to the (maximal) monotone case the following one.

Definition 5 Let a mapping $\mathcal{F} : K \rightarrow 2^{\mathbb{R}^n}$ be given.

1. If $\mathcal{F}$ is a monotone operator, then $\mathcal{F}^\varepsilon$, defined by

$$\mathcal{F}^\varepsilon(x) := \{z \in \mathbb{R}^n : \langle z - f^y, x - y\rangle \geq -\varepsilon \quad \forall y \in \mathbb{R}^n, f^y \in \mathcal{F}(y)\}$$

   is called the ε-enlargement of $\mathcal{F}$ (see [3] for further details).
2. Analogously, if $\mathcal{F}$ is a pseudomonotone operator, then $\mathcal{F}^\varepsilon$, defined by

$$\mathcal{F}^\varepsilon(x) := \{z \in \mathbb{R}^n : \langle z, x - y\rangle \geq -\varepsilon \quad \forall y \in K : \exists f^y \in \mathcal{F}(y) : \ \langle f^y, x - y\rangle \geq 0\},$$

   is called the ε-enlargement of $\mathcal{F}$.

In the sequel, besides the summability of ε_k one would have to require the following additional assumption which is a direct consequence of the *Brønsted-Rockafellar-property* (cf. [3]) in the maximal monotone case:

(E.1) For every $(x_\varepsilon, t_\varepsilon) \in \text{gph}(\mathcal{T}^\varepsilon)$ there is a pair $(\tilde{x}_\varepsilon, \tilde{t}_\varepsilon) \in \text{gph}(\mathcal{T})$ (which might depend on ε) such that for $\varepsilon \to 0$ it holds

$$||t_\varepsilon - \tilde{t}_\varepsilon|| \to 0 \quad \text{and} \quad ||x_\varepsilon - \tilde{x}_\varepsilon|| \to 0. \tag{16}$$

Since such an approximation is unknown for pseudomonotone operators—in fact, it cannot even be guaranteed in the non-maximal monotone case—we continue without the use of $\mathcal{T}^{\varepsilon_k}$ to simplify the following argumentation.

Algorithm 4 covers both the algorithms in Sect. 2, if we assume that $y^k \in \text{int}(K)$ holds true (remember that in Algorithm 2 it was admitted to find an auxiliary point $y^k \in K$, but once found such a point, one might choose a new $y^k \in \text{int}(K)$ as a convex combination of the old $y^k \in K$ and the known $x^k \in \text{int}(K)$).

- Algorithm 2 by Solodov and Svaiter corresponds to the case $\delta_k = 0$ for each $k \in \mathbb{N}$. Then, since $y^k \in \text{int}(K)$,

$$\langle t^k + \chi_k \left(\nabla h(x^{k+1}) - \nabla h(x^k)\right), \; x - y^k\rangle \;\geq\; 0$$

  iff $t^k + \chi_k \left(\nabla h(x^{k+1}) - \nabla h(x^k)\right) = 0$, corresponding to Algorithm 2.
- Algorithm 3 by Kaplan and Tichatschke corresponds to the case $\sigma = 0$. Then $x^{k+1} = y^k$ in view of (15) and strict convexity of h.

We stress that the restriction of y^k to be an element of $\text{int}(K)$ was only required to show that Algorithm 2 is covered. In the sequel, y^k might also be an element of the boundary ∂K of K.

Let us begin with the convergence analysis of Algorithm 4. To do so, we need some additional assumptions on the occurring parameters.

Assumption A (continuation)

(A.4) We have $\sum_{k=0}^{\infty} \delta_k < \infty$ with $\delta_k \geq 0$ for each $k \in \mathbb{N}$ (summable-error-criterion), and for the parameter of the relative error we have $\sigma \in [0, 1)$.

(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).

For example, in case of weakly monotone operators $\underline{\chi}$ should be chosen such that the regularized operators are strongly monotone.

The following convergence lemma will be used in the sequel and extends a well-known result of Polyak (see Lemma 2.2.2 in [23]).

Lemma 7 *Assume that $\{a^k\}$, $\{b^k\}$, $\{c^k\}$, $\{d^k\}$ are sequences of non-negative numbers. If the relation*

$$a^{k+1} \leq (1 + b^k)a^k + c^k - d^k \tag{17}$$

holds true and $\sum_{k=0}^{\infty} \max\{b^k, c^k\} < \infty$, then the sequence $\{a^k\}$ is convergent and further $\sum_{k=0}^{\infty} d^k < \infty$ also holds true.

Proof Since $d^k \geq 0$, we have

$$a^{k+1} \leq (1 + b^k)a^k + c^k - d^k \leq (1 + b^k)a^k + c^k,$$

and thus an application of Polyak's Lemma 2.2.2 in [23] yields the convergence of the sequence $\{a^k\}$ to some $a \geq 0$. Especially, there is some $M > 0$ such that $||a^k|| \leq M$ for each $k \in \mathbb{N}$.

Rearranging (19) to $d^k \le c^k - a^{k+1} + a^k + b^k a^k$, and considering the sum of these inequalities, we have for any $r \in \mathbb{N}$

$$\sum_{k=1}^{r} d^k \le \sum_{k=1}^{r} c^k + \sum_{k=1}^{r} \left(a^k - a^{k+1}\right) + \sum_{k=1}^{r} b^k a^k \tag{18}$$

$$\le \sum_{k=1}^{r} c^k + a_1 - a_r + M \sum_{k=1}^{r} b^k, \tag{19}$$

and passing to the limit, we directly obtain the assertion due to the proven convergence of $\{a^k\}$ and the assumptions on $\{b^k\}$ as well as $\{c^k\}$. □

Since it will be a main part of the following discussion, remember that an operator $\mathcal{T}$ which is pseudomonotone with respect to the solution set of a given problem fulfills the condition $\langle t, x^* - z\rangle \le 0$ for every solution x^*, where $t \in \mathcal{T}(z)$.

For the following argumentation, we will also make use of the inequality

$$1 \le (1-r)^{-1} \le 1 + 2r \le 2, \tag{20}$$

which holds true for each $0 \le r \le \frac{1}{2}$. The proof is straightforward.

In the next theorem, we will prove several auxiliary results. We emphasize that it covers all and strengthens some of the corresponding statements known in the theory of proximal-like methods although the assumptions on the operator $\mathcal{T}$ are much less demanding.

Theorem 5 *Suppose that (A.1)–(A.5) hold true and that $x^* \in SOL(K, \mathcal{T})$ is an arbitrary solution. Assume that the operator $\mathcal{T}$ is pseudomonotone with respect to the solution set $SOL(K, \mathcal{T})$ and again $\{x^k\}$, $\{y^k\}$ denote the sequences generated by Algorithm 4. Then the following holds true:*

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

Proof The well-known three-point-formula (see [5, Lemma 3.1]) yields

$$\begin{aligned} D_h(x^*, x^k) - D_h(x^*, x^{k+1}) &= D_h(x^{k+1}, x^k) + \langle \nabla h(x^{k+1}) - \nabla h(x^k), x^* - x^{k+1}\rangle \\ &= D_h(x^{k+1}, x^k) + \langle \nabla h(x^{k+1}) - \nabla h(x^k),\ x^* - y^k\rangle \\ &\quad + \langle \nabla h(x^{k+1}) - \nabla h(x^k),\ y^k - x^{k+1}\rangle. \end{aligned} \tag{21}$$

Now let us investigate the terms occurring in (21). In view of the inexactness defined by (14), we have especially for $x = x^*$:

$$\langle \nabla h(x^{k+1}) - \nabla h(x^k), x^* - y^k\rangle \ge -\frac{1}{\chi_k}\langle t^k, x^* - y^k\rangle - \frac{\delta_k}{\chi_k}||x^* - y^k||, \tag{22}$$

and due to the three-point-formula, we have

$$\langle \nabla h(x^{k+1}) - \nabla h(x^k),\ y^k - x^{k+1}\rangle = D_h(y^k, x^k) - D_h(y^k, x^{k+1}) - D_h(x^{k+1}, x^k). \tag{23}$$

Concatenating (21)–(23), we obtain

$$D_h(x^*, x^k) - D_h(x^*, x^{k+1}) \geq D_h(y^k, x^k) - D_h(y^k, x^{k+1}) \\ -\frac{1}{\chi_k}(\langle t^k,\ x^* - y^k\rangle + \delta_k||x^* - y^k||). \tag{24}$$

Thus,

$$D_h\left(x^*, x^{k+1}\right) \leq D_h\left(x^*, x^k\right) + (\sigma - 1)D_h(y^k, x^k) \\ +\frac{1}{\chi_k}\langle t^k,\ x^* - y^k\rangle + \frac{\delta_k}{\chi_k}||x^* - y^k|| \tag{25}$$

It is clear that $(\sigma - 1)D_h(y^k, x^k) \leq 0$ since $\sigma < 1$ and that we also have $\frac{1}{\chi_k}\langle t^k, x^* - y^k\rangle \leq 0$ and only the term $\frac{\delta_k}{\chi_k}||x^* - y^k||$ remains to be investigated to apply Lemma 7.

We have with respect to (B.6):

$$\begin{aligned}
\frac{\delta_k}{\chi_k}||x^* - y^k|| &\leq \frac{\delta_k}{\chi_k}||x^* - x^{k+1}|| + \frac{\delta_k}{\chi_k}||x^{k+1} - y^k|| \\
&\leq \frac{\delta_k}{\chi_k}\left(\frac{1}{\alpha(y^k)}(D_h(y^k, x^{k+1}) + \overline{c}) + \frac{1}{\alpha(x^*)}(D_h(x^*, x^{k+1}) + \overline{c})\right) \\
&\leq \frac{\delta_k\sigma}{\underline{\alpha}\underline{\chi}}D_h(y^k, x^k) + \frac{\delta_k}{\chi_k\underline{\alpha}}\overline{c} + \frac{\delta_k}{\underline{\alpha}\underline{\chi}}D_h(x^*, x^{k+1}) + \frac{\delta_k}{\chi_k\underline{\alpha}}\overline{c} \\
&= \frac{\delta_k\sigma}{\underline{\alpha}\underline{\chi}}D_h(y^k, x^k) + \frac{\delta_k}{\underline{\alpha}\underline{\chi}}D_h(x^*, x^{k+1}) + \frac{2\overline{c}}{\underline{\alpha}\chi_k}\delta_k.
\end{aligned} \tag{26}$$

Now (25) turns to

$$\begin{aligned}
D_h(x^*, x^{k+1}) &\leq D_h\left(x^*, x^k\right) + (\sigma - 1)D_h(y^k, x^k) \\
&\quad +\frac{1}{\chi_k}\langle t^k,\ x^* - y^k\rangle + \frac{\delta_k}{\chi_k}||x^* - y^k|| \\
&\leq D_h\left(x^*, x^k\right) + (\sigma - 1)D_h(y^k, x^k) + \frac{1}{\chi_k}\langle t^k,\ x^* - y^k\rangle \\
&\quad +\frac{\delta_k\sigma}{\underline{\alpha}\underline{\chi}}D_h(y^k, x^k) + \frac{\delta_k}{\underline{\alpha}\underline{\chi}}D_h(x^*, x^{k+1}) + \frac{2\overline{c}}{\underline{\alpha}\chi_k}\delta_k,
\end{aligned}$$

and after the rearrangement of some terms,

$$\left(1 - \frac{\delta_k}{\underline{\alpha}\underline{\chi}}\right)D_h(x^*, x^{k+1}) \leq D_h(x^*, x^k) + \frac{2\overline{c}}{\underline{\alpha}\chi_k}\delta_k + \frac{1}{\chi_k}\langle t^k,\ x^* - y^k\rangle \\ +\underbrace{\left(\frac{\delta_k\sigma}{\underline{\alpha}\underline{\chi}} + \sigma - 1\right)}_{=:\theta_k}D_h(y^k, x^k). \tag{27}$$

Due to (A.4) it is known that there is some $k_0 \in \mathbb{N}$ such that

$$\delta_k \leq \underline{\alpha}\underline{\chi}\left(\frac{1}{\sigma} - 1\right), \tag{28}$$

which is easily shown to be equivalent to $\theta_k \leq 0$. Thus, in the following we can (and will) assume that $\theta_k \leq 0$ holds true.

Now we apply (20) to obtain for any $k \in \mathbb{N}$:

$$\left(1 - \frac{\delta_k}{\underline{\alpha}\underline{\chi}}\right)^{-1} \leq 1 + 2\frac{\delta_k}{\underline{\alpha}\underline{\chi}}, \tag{29}$$

from which we deduce

$$\begin{aligned} D_h(x^*, x^{k+1}) \leq & \left(1 + 2\frac{\delta_k}{\underline{\alpha}\underline{\chi}}\right) D_h(x^*, x^k) + \left(1 + 2\frac{\delta_k}{\underline{\alpha}\underline{\chi}}\right)\frac{2\overline{c}}{\underline{\alpha}\chi_k}\delta_k \\ & + \left(1 + 2\frac{\delta_k}{\underline{\alpha}\underline{\chi}}\right)\frac{1}{\chi_k}\langle t^k,\ x^* - y^k\rangle \\ & + \left(1 + 2\frac{\delta_k}{\underline{\alpha}\underline{\chi}}\right)\left(\frac{\delta_k\sigma}{\underline{\alpha}\underline{\chi}} + \sigma - 1\right) D_h(y^k, x^k), \end{aligned} \tag{30}$$

At this place, Lemma 7 is applicable with

- $a^k := D_h(x^*, x^k)$
- $b^k := 2\frac{\delta_k}{\underline{\alpha}\underline{\chi}}$
- $c^k := \left(1 + 2\frac{\delta_k}{\underline{\alpha}\underline{\chi}}\right)\frac{2\overline{c}}{\underline{\alpha}\chi_k}\delta_k$
- $d^k := -\left(1 + 2\frac{\delta_k}{\underline{\alpha}\underline{\chi}}\right)\frac{1}{\chi_k}\langle t^k,\ x^* - y^k\rangle - \left(1 + 2\frac{\delta_k}{\underline{\alpha}\underline{\chi}}\right)\left(\frac{\delta_k\sigma}{\underline{\alpha}\underline{\chi}} + \sigma - 1\right) D_h(y^k, x^k) \geq 0$

and yields the convergence of $\{a_k\} = \{D_h(x^*, x^k)\}$ (Assertion 1) and that further Assertions 3 and 5 hold true, i.e.

$$\sum_{k=1}^{\infty} D_h(y^k, x^k) < \infty \tag{31}$$

and

$$0 \leq \sum_{k=1}^{\infty} -\frac{1}{\chi_k}\langle t^k, x^* - y^k\rangle < \infty. \tag{32}$$

The boundedness of $\{x^k\}$ (Assertion 2) is a consequence the convergence of $\{D_h(x^*, x^k)\}$ and the Bregman axiom (B.2). Assertion 4 (boundedness of the sequence $\{y^k\}$) can be deduced as shown in [21].

Finally, Assertion 6 is a direct consequence of the well-known four-point-formula (see [26, Corollary 2.6]) for Bregman(-like) functions and the proven Assertions 1 and 3:

$$\begin{aligned} \langle \nabla h(x^{k+1}) - \nabla h(x^k), x^* - y^k\rangle = & D_h(x^*, x^k) - D_h(x^*, x^{k+1}) \\ & - D_h(y^k, x^k) + D_h(y^k, x^{k+1}). \end{aligned}$$

□

The above result deserves to be commented in some detail.

Remark 1 1. The extension of Polyak's Lemma allows to preserve some results that are well-known for the case that $\mathcal{T}$ is a maximal monotone operator. Here, only pseudomonotonicity with respect to the solution set (in the sense of Karamardian) is assumed.

2. Nevertheless, some of the above results are even new for the maximal monotone case, at least they are genuinely stronger as known results, for example the convergence of $\sum_{k=1}^{\infty} D_h(y^k, x^k)$ which, concerning the version using a stopping criterion of summable-error-type only, corresponds to the convergence of $\sum_{k=1}^{\infty} D_h(x^{k+1}, x^k)$ has not been proven up to now. An analogue statement for $\sum_{k=1}^{\infty} \langle t^k,\ x^* - x^{k+1}\rangle$ holds true.
3. Although we deal with weaker hypotheses here, the above proof is somewhat easier than some arguments in preceding works, see for example Chapter 4 in [26].
4. For the above proof we did not any longer require that $y^k \in \mathrm{int}(K)$.

Corollary 1 *As a direct consequence of Theorem 5, we obtain:*

1. *The sequences $\{x^k\}$ and $\{y^k\}$ both have at least one cluster point, and each cluster point belongs to K.*
2. *Each cluster point of $\{x^k\}$ also is a cluster point of $\{y^k\}$ and vice versa, i.e. the cluster sets of both sequences coincide.*
3. *If $\{x^{k_l}\} \to \overline{x}$ denote a convergent subsequence and a cluster point, respectively, then we also have $\{x^{k_l+1}\} \to \overline{x}$.*

Proof 1. Existence of cluster points follows from boundedness of the sequences. Each cluster point belongs to K since K is closed.

2. Let $\{x^{k_l}\} \to \overline{x}$. Since especially $D_h(y^{k_l}, x^{k_l}) \to 0$, it follows from (B.4) that also $\{y^{k_l}\} \to \overline{x}$. If the convergence of some subsequence $\{y^{k_l}\}$ is known, the same argument holds.
3. Analogously, from 2. it follows $\{y^{k_l}\} \to \overline{x}$ and due to the iteration scheme, it holds $D_h(y^{k_l}, x^{k_l+1}) \le \sigma D_h(y^{k_l}, x^{k_l}) \to 0$, using (B.4) we deduce $\{x^{k_l+1}\} \to \overline{x}$.

□

Our next task will be to show that each cluster point of $\{x^k\}$ (and thus also of $\{y^k\}$) is a solution of $VI(K, \mathcal{T})$. From now on, $x^{k_l} \to \overline{x}$ denote a convergent subsequence and a cluster point, respectively. By Corollary 1 we know that $y^{k_l} \to \overline{x}$ and also $x^{k_l+1} \to \overline{x}$ and by Theorem 5 it is clear that

$$\langle t^{k_l}, x^* - y^{k_l}\rangle \to 0 \tag{33}$$

has to hold.

Some additional hypothesis is necessary. In the sequel, suppose that besides the assumptions of Theorem 5 one of the following holds true:
Assumption A (continuation)[2]

(A.6) $\mathcal{T} : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is a locally bounded operator and $\mathrm{gph}(\mathcal{T})$ is closed.
(A.7) $\mathcal{T} : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is pseudomonotone in the sense of Brézis.
(A.8) $SOL(K, \mathcal{T}) \cap \mathrm{int}(K) \neq \emptyset$ and $\mathrm{gph}(\mathcal{T})$ is closed.

First, we care about (A.6) and (A.7), where the following result is useful.

[2] It is easy to see that, for example, single-valued and continuous operators have the required additional properties (A.6) and (A.7). Also the possibly set-valued Clarke sub- differential of a locally Lipschitz function is covered, see e.g. [11, Proposition 7.1.4].

Lemma 8 *When the assumptions of Theorem 5 as well as (A.6) or (A.7) hold true, then we have that* $\{t^{k_l}\} \to \bar{t}$ *for some* $\bar{t} \in \mathcal{T}(\bar{x})$ *with the property that* $\langle \bar{t}, x^* - \bar{x} \rangle = 0$.

Proof Let us discuss (A.6) first. Since $\mathcal{T}$ is locally bounded around $\bar{x}$, the sequence $\{t^{k_l}\}$ is bounded because of $\{y^{k_l}\} \to \bar{x}$. So we can assume without loss of generality that $\{t^{k_l}\} \to \bar{t}$ holds true for some $\bar{t}$. Since $\mathrm{gph}(\mathcal{T})$ is closed, $\bar{t} \in \mathcal{T}(\bar{x})$ is valid. The equation $\langle \bar{t}, x^* - \bar{x} \rangle = 0$ now is a direct consequence of (33).

Now turn to (A.7). In view of (33) and Definition 2 the existence of some $\bar{t} \in \mathcal{T}(\bar{x})$ with $\langle \bar{t}, x^* - \bar{x} \rangle \geq 0$ is guaranteed. That indeed $\langle \bar{t}, x^* - \bar{x} \rangle = 0$ holds true is a direct consequence of the fact that $\mathcal{T}$ is pseudomonotone with respect to the solution set. □

Now we are able to prove the following important result.

Theorem 6 [3] *Assume that the assumptions of Lemma 8 hold true and that* $\mathcal{T}$ *either is pseudomonotone** *or the Clarke subdifferential of a pseudoconvex function* f. *Then each cluster point* $\{x^k\}$ *is a solution of the considered problem.*

Proof Passing to the limit $l \to \infty$ in the iteration scheme (14), we obtained in Lemma 8

$$\langle \bar{t}, x^* - \bar{x} \rangle = 0, \tag{34}$$

and due to the pseudomonotonicity of $\mathcal{T}$ we also have $\langle t^*, \bar{x} - x^* \rangle = 0$.

Now relation (2) in the definition of pseudomonotone* operators implies the existence of $k > 0$ such that $k \cdot t^* \in \mathcal{T}(\bar{x})$, and it follows for each $x \in K$:

$$\langle kt^*, x - \bar{x} \rangle = k \cdot \langle t^*, x^* - \bar{x} \rangle + k \cdot \langle t^*, x - x^* \rangle \geq 0.$$

In the case that $\mathcal{T}$ is the generalized gradient of some function f, from (34) and the pseudoconvexity of f we infer that $f(x^*) \geq f(\bar{x})$ which directly implies the assertion. Therefore, $\bar{x} \in SOL(K, \mathcal{T})$ in both cases. □

To conclude this discussion, we can make a comparison to the existing literature dealing with Bregman-based interior point methods. Besides minor improvements, we could replace monotonicity by pseudomonotonicity, maximality by the closedness of the graph and finally paramonotonicity by pseudomonotonicity*. Further we should mention that when working with the above notion of enlargements one does not have to make explicit use of the known *Brønsted-Rockafellar-property* of enlargements of maximal monotone operators, a weakened property is sufficient.

Thus, the conditions imposed above are all covered by the usual ones in the maximal monotone case. Note that, under suitable assumptions, the Clarke subdifferential has the properties used above, namely a closed graph and local boundedness (see Propositions 2.1.2 and 2.1.5 in [6]).

For the sake of completeness, we attach the following result. It generalizes some statements in [21,26] to the pseudomonotone case. We emphasize that if there is a solution in the interior of the feasible set K, neither boundedness nor pseudomonotonicity in the sense of Brézis have to be required.

Lemma 9 *If the assumptions of Theorem 5 as well as (A.8) hold true, each cluster point of the generated sequences* $\{x^k\}$, $\{y^k\}$ *is a solution as well.*

[3] If $\mathcal{T}$ is a maximal monotone operator, (local) boundedness of $\mathcal{T}$ as well as pseudomonotonicity in the sense of Brézis do not have to be required anymore. Indeed, Solodov and Svaiter proved several auxiliary results which enable to omit this condition (since the case of maximal monotone operators is not our central topic, we refer to [26]).

Proof Let again $\{x^{k_l}\} \to \overline{x}$ denote a convergent subsequence. In view of Corollary 1 we directly obtain $\{y^{k_l}\} \to \overline{x}$ as well as $\{x^{k_l+1}\} \to \overline{x}$.

Now for some $x^{**} \in SOL(K, \mathcal{T}) \cap \text{int}(K)$ 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* (see e.g. [17]), we can conclude that $\overline{x}$ and therefore each cluster point of the generated sequences has to belong to $\text{int}(K)$.

Therefore, $\nabla h(\overline{x})$ exists and 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 (14). In view of $x - y^{k_l} \to x - \overline{x}$ and $\overline{x} \in \text{int}(K)$ it follows $t^{k_l} \to 0$.

Closedness of the graph of $\mathcal{T}$ implies $0 \in \mathcal{T}(\overline{x})$, i.e. $\overline{x} \in SOL(K, \mathcal{T})$. □

Recall that if $\mathcal{T}$ is pseudomonotone and continuous, then $SOL(K, \mathcal{T})$ is a convex set—if in this case there is more than one solution and the boundary of K does not contain any line segment (cf. (A.2)), then one solution has to belong to $\text{int}(K)$ (see also [21] for the existence of a solution in $\text{int}(K)$).

The following theorem contains the conclusion of convergence of the entire sequence. Note that in the latter proof only the convergence of $\{D_h(x^*, x^k)\}$ for each solution x^* and some of the Bregman axioms, but no property of $\mathcal{T}$, is required. The proof is rather standard, see e.g. [21].

Theorem 7 *If each cluster point of $\{x^k\}$ solves $VI(K, \mathcal{T})$, then the sequence $\{x^k\}$ generated by Algorithm 4 converges to some $x^* \in SOL(K, \mathcal{T})$. Clearly, in this case also $\{y^k\}$ converges to x^*.*

4 Concluding remarks

We considered the well-known Bregman Proximal Point Algorithm (BPPA) for solving variational inequalities and established the convergence of the latter method for problems with only pseudomonotone* operators which is a much weaker hypothesis than the customary use of paramonotonicity.

In view of the examples considered above, from now on also constrained optimization problems with pseudoconvex functions might be solved by solving several (unconstrained!) well-posed convex optimization problems, at least in the case of weakly monotone operators, when the regularization parameters χ_k are chosen appropriately. But also a significantly larger class of Nash equilibrium problems might be treated in this way. However, a sufficient condition on the cost functions f_i to guarantee monotonicity or pseudomonotonicity of the operator of the underlying variational inequality seems to be unknown.

Further, we unified two different inexactness tolerances customary used in literature. Surprisingly, also the relaxed error tolerance admitted from now on does not require stronger assumptions neither it allows to prove weaker convergence results only. In fact, the results proven in the present article are at least as strong as they existed in literature when only one of the error criteria is used and the operator of the considered variational inequality is paramonotone. Clearly, our results can be transferred to the discussion of one stopping criterion only by setting $\sigma = 0$ or $\delta_k = 0$, respectively.

Related to the papers of El-Farouq [10] and Schaible et al. [25] our approach admits to omit the assumptions of a Lipschitz-continuous gradient of h (this contradicts the zone-coerciveness) and the assumption of $\mathcal{T}$ being strongly pseudomonotone (which implies uniqueness of solutions of the considered problem). Thus, also in comparison with this approach the method presented in the present paper is applicable to a much broader class of problems and it is far more tractable since inexact solutions of the auxiliary problems are permitted.

Nevertheless, if, as for example in the original Proximal Point Algorithm, ∇h is Lipschitz continuous, then it is easily proved that each cluster point of the generated sequences is a solution (indeed, just apply Theorem 5, the iteration scheme and Lipschitz continuity of ∇h), and the results in the present paper are genuinely stronger in this case as well, for example since only pseudomonotonicity with respect to the solution set has to be required.

Acknowledgements The author is grateful to the reviewer for the valuable and helpful comments.

References

1. Brézis, H.: Équations et inéquations non linéaires dans les espaces vectoriels en dualité. Ann. Inst. Fourier **18**, 115–175 (1968)
2. Burachik, R., Iusem, A.: A generalized Proximal Point algorithm for the variational inequality problem in Hilbert space. SIAM J. Optim. **8**, 197–216 (1998)
3. Burachik, R., Svaiter, B.: ε-enlargement of maximal monotone operators in Banach spaces. Set-Valued Anal. **7**, 117–132 (1999)
4. Censor, Y., Iusem, A., Zenios, S.A.: An interior point method with Bregman functions for the variational inequality problem with paramonotone operators. Math. Program. **81**, 373–400 (1998)
5. Chen, G., Teboulle, M.: Convergence analysis of a proximal-like minimization algorithm using Bregman functions. SIAM J. Optim. **3**, 538–543 (1993)
6. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
7. Crouzeix, J.-P.: Pseudo-monotone variational inequality problems: existence of solutions. Math. Program. **78**, 305–314 (1997)
8. Crouzeix, J.-P., Zhu, D.L., Marcotte, P.: Conditions ensuring the applicability of cutting-plane methods for solving variational inequalities. JOTA **87**, 457–471 (1995)
9. Eckstein, J.: Approximate iterations in Bregman-function-based proximal algorithms. Math. Program. **83**, 113–123 (1998)
10. El-Farouq, N.: Pseudomonotone variational inequalities: convergence of proximal methods. JOTA **109**, 311–326 (2001)
11. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, New York (2003)
12. Hadjisavvas, N.: Continuity and maximality properties of pseudomonotone operators. J. Convex Anal. **10**, 459–469 (2003)
13. Hadjisavvas, N., Schaible, S.: On a generalization of paramonotone maps and its application to solving the Stampacchia variational inequality. Optimization **55**, 593–604 (2006)
14. Iusem, A.: On some properties of paramonotone operators. J. Conv. Anal. **5**, 269–278 (1998)
15. Kaplan, A., Tichatschke, R.: On the convergence of interior proximal methods for variational inequalities on non-polyhedral sets. Discuss. Math. (2009, to appear)
16. Kaplan, A., Tichatschke, R.: On inexact generalized proximal methods with a weakened error tolerance criterion. Optimization **53**, 3–17 (2004)
17. Kaplan, A., Tichatschke, R.: Bregman functions and proximal methods for variational problems with nonlinear constraints. Optimization **56**(1–2), 253–265 (2007)
18. Kaplan, A., Tichatschke, R.: Interior proximal method for variational inequalities on non-polyhedral sets. Discuss. Math. **27**(1), 71–93 (2007)
19. Karamardian, S.: Complementarity problems over cones with monotone and pseudomonotone maps. JOTA **18**, 445–454 (1976)
20. Kien, B., Yao, J.-C., Yen, N.: On the solution existence of pseudomonotone variational inequalities. J. Glob. Optim. **41**, 135–145 (2008)
21. Langenberg, N.: Convergence analysis of an extended auxiliary problem principle with various stopping criteria. Optim. Methods Softw. (2009, accepted)

22. Martinet, B.: Régularisation d'inéquations variationelles par approximations successives. RIRO **4**, 154–159 (1970)
23. Polyak, B.T.: Introduction to Optimization. Optimization Software, Inc. Publ. Division, New York (1987)
24. Rockafellar, R.T.: Monotone operators and the Proximal Point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
25. Schaible, S., Yao, J., Zeng, L.: A proximal method for pseudomonotone type variational-like inequalities. Taiwan. J. Math. **10**(2), 497–513 (2006)
26. Solodov, M., Svaiter, B.: An inexact hybrid generalized proximal point algorithm and some new results on the theory of Bregman functions. Math. Oper. Res. **25**, 214–230 (2000)
27. Tam, N., Yao, J., Yen, N.: Solution methods for pseudomonotone variational inequalities. JOTA **138**, 253–273 (2008)