Convergence analysis of an extended Auxiliary Problem Principle for solving variational inequalities

Nils Langenberg

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

Abstract

We study the *Extended Proximal Auxiliary Problem Principle*-method (EPAPP) by Kaplan and Tichatschke [17, 20] for solving variational inequalities whose operator is the sum of a maximal monotone and a continuous operator.

As in comparable methods using Bregman distances the authors required that the operator of the considered variational inequality (here called main operator) is paramonotone (see [11] for definition and properties of paramonotone operators).

The main purpose of this paper is to establish the convergence of the EPAPP-method without use of paramonotonicity. A sort of error summability criterion is used to allow inexact solutions of the auxiliary problems, and we also admit an outer approximation of the set-valued component of the operator. Due to the use of Bregman-like functions to construct the symmetric components of the auxiliary operators an *interior point effect* is provided, that is, – with a certain precaution – the auxiliary problems can be treated as unconstrained ones.

Key words: variational inequalities, auxiliary problem principle, Bregman distances, multi-valued and maximal monotone operators, non-paramonotone operators

AMS Subject Classifications: 47H05, 47J20, 65J20, 65K10, 90C25

1 Introduction

Cohen [5] originally introduced the *Auxiliary Problem Principle* (APP) for optimization problems aiming for a unification of convergence analysis of optimization algorithms like (sub-) gradient and decomposition algorithms. Later on the APP-method was extended to a more general class of variational inequalities [6]. The utility of APP and APP-based theory is wide-spread, for example, new convergence results for the Lions-Mercier splitting algorithm [22] could be found on this way.

The APP provides another useful advantage. Namely, on this way it is possible to solve variational inequalities with non-potential operators (that is, variational inequalities that do not arise from optimization problems) by means of solving a sequence of optimization problems (see [6]).

Martinet [23] introduced the classical *Proximal-Point-Algorithm* (PPA) for solving ill-posed optimization problems. This method has been extended to variational inequalities by Rockafellar [24]. Up to now, there have been created numerous variants and extensions of this method, even using non-quadratic distances, in particular Bregman distances (see e.g. [3, 4, 8, 9]) in the Bregman Proximal Point Algorithm (BPPA).

A connection of the concepts of BPPA and APP was studied by Kaplan and Tichatschke (cf. [17, 20]) in the framework of an *Extended Proximal APP* (EPAPP). In these and other papers on Bregman-based methods (e.g. [3, 7, 9, 10, 14, 16, 19, 25]) paramonotonicity of the main operator was supposed.

Kaplan and Tichatschke [18] also modified the concept of Bregman functions which allowed to construct Bregman-function-based methods with an interior point effect also for problems with nonlinear constraints.

This paper is organized as follows: Section 2 contains the statement of problem and the state of the art in the theory of Bregman-like functions, whereas Section 3 treats the EPAPP-method and its convergence analysis without the paramonotonicity assumption. Finally we draw some conclusions in Section 4.

2 Statement of Problem & Bregman-like functions

Let $K \subset \mathbb{R}^n$ be a closed convex set, $\mathcal{F}: K \to \mathbb{R}^n$ a continuous operator with certain monotonicity properties and $\mathcal{Q}: \mathbb{R}^n \to 2^{\mathbb{R}^n}$ be a maximal monotone and multi-valued operator. Consider the variational inequality

$$VI(K, \mathcal{F}, \mathcal{Q}) \qquad \begin{array}{l} \text{find } x^* \in K \text{ and } q^* \in \mathcal{Q}(x^*): \\ \langle \mathcal{F}(x^*) + q^*,\ x - x^* \rangle \ \geq\ 0 \quad \forall\, x \in K, \end{array} \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the canonical inner product in $\mathbb{R}^n$. The solution set of this problem $VI(K, \mathcal{F}, \mathcal{Q})$ will be denoted by $SOL(K, \mathcal{F}, \mathcal{Q})$.

Splitting the main operator is of interest, for instance, in mathematical physics (e.g. for Bingham problems or problems with friction), where an appropriate splitting allows efficient approximations of $\mathcal{Q}$ (cf. [15]).

Throughout this article $x^* \in SOL(K, \mathcal{F}, \mathcal{Q})$ is an arbitrary (but fixed) solution and q^* denotes an element of $\mathcal{Q}(x^*)$ such that (1) is fulfilled. Further, we will make the following general assumptions.

Assumption A
The problem under consideration generally has the following properties:

(A.1) $SOL(K, \mathcal{F}, \mathcal{Q}) \neq \emptyset$.

(A.2) $\mathcal{F}$ is continuous on K.

(A.3) $\mathcal{Q}$ is a maximal monotone operator and it holds $\mathrm{ri}(\mathrm{dom}\mathcal{Q} \cap K) \neq \emptyset$.

(A.4) 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 $i \in I_2$. Furthermore assume that the set

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

contains no line segments.

(A.5) Slater's constraint qualification holds:

$$\exists \, x \in K : \quad g_i(x) < 0 \quad \forall \, i \in I_1 \cup I_2.$$

While the necessity of (A.1) and (A.5) is obvious, (A.2) and (A.3) represent some extension of COHEN's APP which is obtained by setting $\mathcal{F} := \mathcal{F} + \mathcal{Q}$, where often the single-valuedness of $\mathcal{F} + \mathcal{Q}$ is supposed additionally.

The assumption (A.4) represents the state of the art in the theory of Bregman-like functions. It seems to be quite theoretically, but it is for example fulfilled, if each g_i $(i \in I_2)$ is strictly convex or if the elementwise maximum of the functions g_i $(i \in I_2)$ is a strictly convex function (for related discussions see [18, 19]). However, the example $K := \{x \in \mathbb{R}^2 : \ -x_1 + x_2^2 \leq 0\}$ shows that not even one of the constraints has to be strictly convex.

Geometrically, if one assumes $I_1 = \emptyset$ then M describes the boundary ∂K, so (A.4) requires a certain curvature of the boundary of K. More generally, this assumption demands that if there is some line segment contained in ∂K, this can be traced back to one of the affine constraints.

Let us discuss the concept of Bregman-like functions in some detail.

Definition 1. ***(Bregman-like functions)***
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 $\mathcal{M}(x,\alpha) := \{y \in S: \quad D_h(x,y) \le \alpha\}$ is bounded for all fixed $\alpha \in \mathbb{R}$ and $x \in \mathrm{cl}(S)$, where the Bregman distance is defined by*

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

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 the following holds:

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

One readily recognizes that the difference between such Bregman-like functions and standard Bregman functions can be found in condition (B.5). Solodov and Svaiter [25] showed that (B.6) is a consequence of (B.2) and (B.3). 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 consider a (zone-coercive) Bregman-like function h with zone $\mathrm{int}(K)$ when $K \subset \mathbb{R}^n$ admits a description by (A.4) and (A.5). Kaplan and Tichatschke [13, 18, 19] considered the function

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

for fixed $\kappa > 0$, where g_i are the constraints describing K.

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

Lemma 1. ***(Construction of Bregman-like functions)*** *(see [13])*
Let ϕ in (2) be constructed with the following properties:

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

(C.2) *ϕ is continuously differentiable on $(-\infty, 0)$.*

(C.3) *It holds $t \cdot \phi'(t) \to 0$ for $t \uparrow 0$.*

(C.4) *It holds $\phi'(t) \to \infty$ for $t \uparrow 0$.*

Then the function h, defined by (2), is a strongly convex (with modulus κ) and zone-coercive Bregman-like function with zone $\operatorname{int}(K)$.

Note that there are functions ϕ satisfying (C.1)-(C.4), 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 [18]).

In the following we assume that h is a strongly convex (with modulus κ), zone-coercive Bregman-like function with zone $\operatorname{int}(K)$, where $K \subset \mathbb{R}^n$ admits a description like in (A.4) and (A.5).

3 The EPAPP-method and its convergence

The EPAPP-method by Kaplan and Tichatschke [20] is illustrated in Algorithm 1.

Algorithm 1: EPAPP-algorithm

---

1. Let a start-iterate $x^1 \in \operatorname{int}(K)$ be given. Choose scalars $\delta_1 \geq 0$, $\chi_1 > 0$, $\varepsilon_1 \geq 0$, an auxiliary operator $\mathcal{L}_1$ and an operator $\mathcal{Q}^1$ with $\mathcal{Q} \subset \mathcal{Q}^1 \subset \mathcal{Q}_{\varepsilon_1}$ (for the notation see (A.6) in the sequel). Set $k := 1$.
2. If x^k solves the problem $\mathrm{VI}(\mathcal{F}, \mathcal{Q}, K) \to$ STOP.
3. Calculate $x^{k+1} \in K$, $q^{k+1} \in \mathcal{Q}^k(x^{k+1})$ by solving (P_δ^k).
4. Choose $\delta_{k+1} \geq 0$, $\chi_{k+1} > 0$, $\varepsilon_{k+1} \geq 0$, $\mathcal{L}_{k+1}$ and $\mathcal{Q}^{k+1}$.
   Set $k := k + 1$ and go to step 2.

The core of this method can be found in step 3 (solving the problems (P_δ^k) which will be referred to as auxiliary problems), all the other steps serve for updates and optimality testing.

The auxiliary problems (P_δ^k) in the third step are given by

$$(\boldsymbol{P_\delta^k}) \qquad \text{find } x^{k+1} \in K,\ q^{k+1} \in \mathcal{Q}^k(x^{k+1}) \text{ such that :}$$

$$\langle \mathcal{F}(x^k) + q^{k+1} + \mathcal{L}_k(x^{k+1}) - \mathcal{L}_k(x^k) + \chi_k\big(\nabla h(x^{k+1}) - \nabla h(x^k)\big),\ x - x^{k+1}\rangle$$

$$\geq \quad -\delta_k \cdot ||x - x^{k+1}|| \qquad \forall\, x \in K.$$

See e.g. [20] for a short overview of some special cases (regularized Newton method, projection methods etc.) of the EPAPP method; we will come back to them later on.

Remark 1. *Sometimes the auxiliary problems are formulated slightly different. Consider the following one:*

$$(\boldsymbol{P_e^k}) \quad \textit{find } x^{k+1} \in K,\ q^{k+1} \in \mathcal{Q}^k(x^{k+1}),\ e^{k+1} \in \mathbb{R}^n :$$
$$\mathcal{F}(x^k) + q^{k+1} + \mathcal{L}_k(x^{k+1}) - \mathcal{L}_k(x^k) + \chi_k \nabla_1 D_h(x^{k+1}, x^k) = e^{k+1},$$

where $||e^{k+1}|| \leq \delta_k$ *and* ∇_1 *denotes the partial gradient with respect to the first argument. Then it is easy to see that each solution* (x^{k+1}, q^{k+1}) *of* (P_e^k) *also is a solution of* (P_δ^k) *and there is no problem if we continue with the discussion of* (P_δ^k)*.*

In the sequel we use the following additional assumptions on operators and variables introduced by Kaplan and Tichatschke in [20]. They also assume the paramonotonicity of the main operator $\mathcal{F} + \mathcal{Q}$. This paramonotonicity assumption for the involved operator is standard for Bregman-function-based proximal methods. It is avoided in logarithmic-quadratic proximal methods, which are applicable however only to linearly constrained problems [1, 12].

Assumption A (continuation)

(A.6) $\mathcal{Q}^k$ is an outer approximation of $\mathcal{Q}$ satisfying $\mathcal{Q} \subset \mathcal{Q}^k \subset \mathcal{Q}_{\varepsilon_k}$ where $\mathcal{Q}_{\varepsilon_k}$ stands for the ε_k-enlargement of $\mathcal{Q}$ (cf. [2] for properties of $\mathcal{Q}_{\varepsilon_k}$).

(A.7) Given a family $\{\mathcal{L}_y\}_{y \in K}$ with each $\mathcal{L}_y : K \to \mathbb{R}^n$ a single-valued, continuous and monotone operator, the operators $\mathcal{F} - \mathcal{L}_y$ admit the existence of $\gamma > 0$ with

$$\langle \mathcal{F}(y) - \mathcal{L}_y(y) + q^* + \mathcal{L}_y(x^*), y - x^*\rangle \geq \gamma ||\mathcal{F}(y) - \mathcal{L}_y(y) - \mathcal{F}(x^*) + \mathcal{L}_y(x^*)||^2$$

for all $y \in \operatorname{int}(K) \cap \operatorname{dom} \mathcal{Q}$ and $x^* \in SOL(K, \mathcal{F}, \mathcal{Q})$ arbitrary but fixed.

(A.8) For any convergent sequence $\{y^k\} \subset \mathrm{int}(K) \cap \mathrm{dom}\mathcal{Q}$ it holds

$$\mathcal{L}_{y^k}(y^{k+1}) - \mathcal{L}_{y^k}(y^k) \to 0 \quad \text{as } k \to \infty.$$

(A.9) It holds $\sum_{k=1}^{\infty} \max\{\delta_k, \varepsilon_k\} < \infty$ (remember $\delta_k, \varepsilon_k \geq 0$).

(A.10) There are $\underline{\chi}, \overline{\chi}$ such that $(4\gamma\kappa)^{-1} \leq \underline{\chi} \leq \chi_k \leq \overline{\chi} < \infty$ and further

$$\sum_{k=1}^{\infty} \max\{0, \chi_k - \chi_{k+1}\} \;<\; \infty.$$

As indicated by (A.8), discussing the auxiliary operators $\mathcal{L}_y$ one should think of a parametrization $\mathcal{L}_k := \mathcal{L}_y|_{y=x^k}$, where x^k are the iterates generated by the method under consideration.

Let us briefly discuss some of these assumptions. (A.6) and (A.9) allow an inexact solution and the use of auxiliary operators in the auxiliary problems and permit a quite simple implementation. Instead of (A.8) it is frequently assumed that the operators $\mathcal{L}_y$ are Lipschitz continuous with a Lipschitz constant independent of the choice of y. Easy to see that property (A.8) used here is much weaker than this Lipschitz assumption.

Concerning (A.7) we state that this is fulfilled if the operators $\mathcal{F} - \mathcal{L}_y$ are co-coercive (have the Dunn property, see [26] for definition and elementary properties of co-coercive operators).

Example 1. *(see [16])*
Consider the example with $n = 1,\; K = [-2, 2],\; \mathcal{L}_y \equiv 0 \;\forall y,$

$$\mathcal{F}(x) \;=\; \begin{cases} x^2, & x \geq -1, \\ x + 2, & x < -1 \end{cases}$$

and $\mathcal{Q}(x) = x + 4$ *in which (A.7) is fulfilled, although the operators* $\mathcal{F} - \mathcal{L}_y$ *are not monotone (so they can a fortiori not possess the Dunn property). Note that here* $\mathcal{F}$ *is not even pseudomonotone (in the sense of Karamardian, see [21] for the definition).*

In addition to these comments in [16] we observe that (A.7) also admits that not even the sum of $\mathcal{F}$ *and a strongly monotone operator is pseudomonotone (consider* $\mathcal{F} + \mathcal{I}$ *in the above example,* $\mathcal{I}$ *stands for the identity mapping).*

Before discussing the convergence of the method we should take care of well-definedness. We just give the result briefly.

Theorem 1. *(see [20])*
Suppose that (A.1)-(A.3) hold, that the operators $\mathcal{L}_k$ are continuous and monotone on K and that the control parameters fulfill $\chi_k > 0$ and $\delta_k, \varepsilon_k \geq 0$ for each k. Then the problem $(\mathcal{P}_\delta^k)$ with $\mathcal{Q}^k := \mathcal{Q}$ and $\delta_k = 0$ has a unique solution and the generated sequence $\{x^k\}$ belongs to $\mathrm{int}(K)$.

It is a direct consequence that the considered auxiliary problems (P_δ^k) with the original $\mathcal{Q}^k$ and $\delta_k > 0$ are solvable in $\mathrm{int}(K)$ (clearly, one cannot expect uniqueness of their solutions). Thus, the method under consideration is well-defined.

Now let us answer the question of convergence. Kaplan and Tichatschke proposed, e.g. in [17], to consider the following function, serving as a Lyapunov-like function. Let $\Gamma : \quad SOL(K, \mathcal{F}, \mathcal{Q}) \times \mathrm{int}(K) \times [\underline{\chi}, \overline{\chi}] \rightarrow \mathbb{R}$ be defined by

$$\Gamma(x^*, x, \chi) \;:=\; D_h(x^*, x) + \frac{1}{\chi}\langle \mathcal{F}(x^*) + q^*,\; x - x^*\rangle.$$

Proposition 1. *(see [17])*
If the assumptions (A.1)-(A.3), (A.7), (A.9), (A.10) are valid and if h is a zone-coercive, strongly convex Bregman-like function with zone $\mathrm{int}(K)$, *the following statements for the sequence $\{x^k\}$ (generated by EPAPP) hold:*

(1) The sequence $\{\Gamma(x^*, x^k, \chi_k)\}$ is convergent for each $x^* \in SOL(K, \mathcal{F}, \mathcal{Q})$.

(2) The generated sequence $\{x^k\}$ is bounded.

(3) $D_h(x^{k+1}, x^k) \rightarrow 0$ for $k \rightarrow \infty$.

(4) $||x^{k+1} - x^k|| \rightarrow 0$ for $k \rightarrow \infty$.

Now we are going to prove the convergence of the EPAPP-method without using paramonotonicity. Doing so, we will consider two cases in this article.

Case 1: $\mathcal{Q}$ is the subdifferential ∂f of a proper convex, lower semicontinuous (lsc) function f and $\mathcal{Q}_{\varepsilon_k}$ here is the ε_k-subdifferential. Assume that (A.7) also holds for the family $\{\mathcal{L}_y\}$ with $\mathcal{L}_y \equiv 0$ for all $y \in \mathrm{int}(K)$:

$$\langle \mathcal{F}(y) + q^*,\; y - x^*\rangle \;\geq\; \gamma \cdot ||\mathcal{F}(y) - \mathcal{F}(x^*)||^2. \tag{3}$$

Case 2: The operator $\mathcal{F} + \mathcal{Q}$ satisfies a weakened Dunn property, meaning that for each $x^* \in SOL(K, \mathcal{F}, \mathcal{Q})$ there is an $\alpha(x^*) > 0$ such that

$$\langle \mathcal{F}(y) + q - \mathcal{F}(x^*) - q^*,\; y - x^*\rangle \;\geq\; \alpha(x^*) \cdot ||\mathcal{F}(y) + q - \mathcal{F}(x^*) - q^*||^2 \tag{4}$$

for all $y \in \mathrm{int}(K)$ and $q \in \mathcal{Q}(y)$.

Remark 2. *It is worth noting that $\mathcal{Q}(x^*)$ has to be a singleton in Case 2. More generally we can proof the following:*

If an operator $\mathcal{Q}: K \to 2^{\mathbb{R}^n}$ is co-coercive with modulus $c > 0$, then $\mathcal{Q}$ is single-valued. Indeed, the co-coercivity assumption means that for all $x, y \in K$ and all $q^x \in \mathcal{Q}(x)$, $q^y \in \mathcal{Q}(y)$ the following holds:

$$\langle q^x - q^y,\ x - y \rangle \ \geq \ c \cdot ||q^x - q^y||^2. \tag{5}$$

Now letting $x = y$ in (5) and assuming $q^x \neq q^y$ yields a contradiction.

$\diamond$

Remark 3. *Sometimes (e.g. in [8, 20]) it is assumed that if $q^k \in \mathcal{Q}(y^k)$ when $y^k \to x \in K$, then $\{q^k\}$ is a bounded sequence. In Case 2, this is implied by (4). Indeed, using the Cauchy-Schwarz inequality, we deduce*

$$\begin{array}{rl} & \alpha(x^*) \cdot ||\mathcal{F}(y^k) + q^k - \mathcal{F}(x^*) - q^*||^2 \\ \leq & \langle \mathcal{F}(y^k) + q^k - \mathcal{F}(x^*) - q^*,\ y^k - x^* \rangle \\ \leq & ||y^k - x^*|| \cdot ||\mathcal{F}(y^k) + q^k - \mathcal{F}(x^*) - q^*||. \end{array}$$

Division by $\alpha(x^*) \cdot ||\mathcal{F}(y^k) + q^k - \mathcal{F}(x^*) - q^*||$ and finally using the estimate $||a|| - ||b|| \leq ||a - b||$ yield the inequality

$$||q^k - q^*|| \ \leq \ ||\mathcal{F}(y^k) - \mathcal{F}(x^*)|| + \frac{1}{\alpha(x^*)}||y^k - x^*||. \tag{6}$$

Using the convergence of $\{y^k\}$ and continuity of $\mathcal{F}$ we obtain the assertion.

$\diamond$

We observe that the operator in Example 1 also fulfills the hypothesis of Case 2 (with $2\alpha(x^*) \leq 1$). Thus, even if the sum of $\mathcal{F}$ and a strongly monotone operator is not pseudomonotone, this assumption can be fulfilled, too.

Turning to the convergence analysis, let us begin with some statements that do not require any of the additional assumptions in one of the cases. They will be used in the proof of convergence later.

By Proposition 1 we know that $\{x^k\}$ is a bounded sequence, so there has to be at least one cluster point $\overline{x}$, which belongs to K, because K is closed. For the rest of this article, denote $\{x^{k_l}\}$ a subsequence with

$$x^{k_l} \to \overline{x}. \tag{7}$$

It is known that the sequence $\{\chi_k\}$ is convergent (see Remark 1 in [17]). Using this fact and the well-known three-point-formula ([4], Lemma 3.1) we obtain

$$\langle \nabla h(x^{k_l+1}) - \nabla h(x^{k_l}),\ x^* - x^{k_l+1} \rangle \ \to \ 0, \quad l \to \infty. \tag{8}$$

Now let us turn to the discussion of Case 1. Clearly, $\mathcal{Q}$ then is also paramonotone, but since $\mathcal{F}$ does not have to be monotone, $\mathcal{F}+\mathcal{Q}$ also does not have to be monotone, a fortiori not paramonotone (see Example 1 and related explanations).

In the first case we will make use of the following result.

Lemma 2. *(see [20])*
Assume that $\mathcal{Q}$ is the subdifferential ∂f of a proper convex, lsc function f; (A.2) is fulfilled and $x^* \in SOL(K,\mathcal{F},\mathcal{Q})$ and $\overline{x} \in K$. If additionally (A.7) is valid or $\mathcal{F}$ is monotone then the following statements are equivalent:

1. *It holds*
$$\langle \mathcal{F}(\overline{x}),\ \overline{x}-x^*\rangle + f(\overline{x}) - f(x^*) \ \leq\ 0.$$
2. *q^* – which belongs to $\mathcal{Q}(x^*)$ – also is an element of $\mathcal{Q}(\overline{x})$ and it holds*
$$\langle \mathcal{F}(\overline{x}) + q^*,\ \overline{x}-x^*\rangle \ \leq\ 0.$$

We begin to prove the convergence of the method for Case 1 and Case 2. We will start with a lemma concerning the cluster points of the generated sequence $\{x^k\}$.

Lemma 3. *Assume that the case-specific hypotheses of Case 1 or 2 are fulfilled, respectively. Then each cluster point of the sequence $\{x^k\}$, generated by EPAPP, belongs to $SOL(K,\mathcal{F},\mathcal{Q})$.*

Proof. Consider Case 1.

By the iteration scheme $(P_\delta^{k_l})$, the monotonicity of each $\mathcal{L}_k$ and the definition of $\mathcal{Q}^k$ we have for the subsequence $\{x^{k_l}\}$:

$$\begin{aligned}
& -\chi_{k_l}\cdot\langle \nabla h(x^{k_l+1}) - \nabla h(x^{k_l}),\ x^*-x^{k_l+1}\rangle - \delta_{k_l}\cdot ||x^*-x^{k_l+1}|| && (9)\\
\leq\ & \langle \mathcal{F}(x^{k_l}) + q^{k_l+1} + \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^*-x^{k_l+1}\rangle \\
\leq\ & \langle \mathcal{F}(x^{k_l}) + q^* + \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^*-x^{k_l}\rangle && (10)\\
& +\ \langle \mathcal{F}(x^{k_l}) + q^* + \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^{k_l}-x^{k_l+1}\rangle + \varepsilon_{k_l}. && (11)
\end{aligned}$$

Denote $T := \langle \mathcal{F}(x^{k_l}) + q^* + \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^*-x^{k_l}\rangle$ in (10). Due to (A.8), (8) and the continuity of $\mathcal{F}$ one easily recognizes that (9) and (11) converge to zero for $l\to\infty$. So let us discuss what happens with (10), i.e.

with T; remember that all $\mathcal{L}_k$ are monotone.

$$\begin{aligned}
T &= \langle \mathcal{F}(x^{k_l}) + q^* + \mathcal{L}_{k_l}(x^*) - \mathcal{L}_{k_l}(x^{k_l}),\ x^* - x^{k_l}\rangle \\
&\quad + \langle \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^*),\ x^* - x^{k_l}\rangle \\
&= \langle \mathcal{F}(x^{k_l}) + q^* + \mathcal{L}_{k_l}(x^*) - \mathcal{L}_{k_l}(x^{k_l}),\ x^* - x^{k_l}\rangle \\
&\quad + \langle \mathcal{L}_{k_l}(x^{k_l}) - \mathcal{L}_{k_l}(x^*),\ x^* - x^{k_l}\rangle \\
&\quad + \langle \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^* - x^{k_l}\rangle \\
&\leq -\gamma \cdot ||\mathcal{F}(x^{k_l}) - \mathcal{F}(x^*) + \mathcal{L}_{k_l}(x^*) - \mathcal{L}_{k_l}(x^{k_l})||^2 \qquad (12) \\
&\quad + \langle \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^* - x^{k_l}\rangle. \qquad (13)
\end{aligned}$$

Since (9), (11) and (13) tend to zero for $l \to \infty$, we can state that the non-positive term (12) must tend to zero as well, that means

$$\mathcal{F}(x^{k_l}) - \mathcal{F}(x^*) + \mathcal{L}_{k_l}(x^*) - \mathcal{L}_{k_l}(x^{k_l}) \to 0, \quad l \to \infty. \qquad (14)$$

On the other hand, due to (A.8) and the special assumption of Case 1 it obviously holds

$$\begin{aligned}
T &= \langle \mathcal{F}(x^{k_l}) + q^*,\ x^* - x^{k_l}\rangle \ + \ \langle \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^* - x^{k_l}\rangle \qquad (15) \\
&\leq -\gamma \cdot ||\mathcal{F}(x^{k_l}) - \mathcal{F}(x^*)||^2 \ + \ \langle \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^* - x^{k_l}\rangle.
\end{aligned}$$

Passing to the limit in (15), we obtain that

$$\mathcal{F}(x^{k_l}) \ \to \ \mathcal{F}(\overline{x}) \ = \ \mathcal{F}(x^*). \qquad (16)$$

On the other hand, because of $\mathcal{Q} = \partial f$ the subgradient inequality holds:

$$\langle q^{k_l+1},\ x^* - x^{k_l+1}\rangle \ \leq \ f(x^*) - f(x^{k_l+1}) + \varepsilon_{k_l+1}.$$

Inserting the last inequality into the iteration scheme $(\mathcal{P}_\delta^{k_l})$, and passing to the limit $l \to \infty$, we get

$$\langle \mathcal{F}(\overline{x}),\ x^* - \overline{x}\rangle + f(x^*) - f(\overline{x}) \ \geq \ 0.$$

Now $q^* \in \mathcal{Q}(\overline{x})$ and $\langle \mathcal{F}(\overline{x}) + q^*,\ \overline{x} - x^*\rangle \ \leq \ 0$ are a direct consequence of Lemma 2. So there is $\overline{q} \in \mathcal{Q}(\overline{x})$ (namely $\overline{q} := q^*$) with

$$\langle \mathcal{F}(\overline{x}) + \overline{q},\ x^* - \overline{x}\rangle \ \geq \ 0. \qquad (17)$$

Combining (17) and (16) yields

$$\begin{aligned}
\langle \mathcal{F}(\overline{x}) + \overline{q},\ x - \overline{x}\rangle &= \langle \mathcal{F}(\overline{x}) + \overline{q},\ x - x^*\rangle + \langle \mathcal{F}(\overline{x}) + \overline{q},\ x^* - \overline{x}\rangle \\
&\geq \langle \mathcal{F}(\overline{x}) + \overline{q},\ x - x^*\rangle \\
&= \langle \mathcal{F}(x^*) + q^*,\ x - x^*\rangle \\
&\geq 0,
\end{aligned}$$

and thus $\overline{x} \in SOL(K, \mathcal{F}, \mathcal{Q})$ is proven for Case 1.

Now consider Case 2.

Owing to the iteration scheme $(\mathcal{P}_\delta^{k_l})$ there is $q^{k_l+1} \in \mathcal{Q}^{k_l}(x^{k_l+1})$ with

$$\begin{array}{rl} & \langle \mathcal{F}(x^{k_l}) + q^{k_l+1} + \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^* - x^{k_l+1}\rangle \\ \geq & -\delta_{k_l} \cdot ||x^* - x^{k_l+1}|| - \chi_{k_l} \cdot \langle \nabla h(x^{k_l+1}) - \nabla h(x^{k_l}),\ x^* - x^{k_l+1}\rangle. \end{array}$$

Because of $x^* \in SOL(K, \mathcal{F}, \mathcal{Q})$ there is $q^* \in \mathcal{Q}(x^*)$ with

$$\langle \mathcal{F}(x^*) + q^*,\ x^* - x^{k_l+1}\rangle \ \leq\ 0.$$

Combining the last two inequalities leads to

$$\begin{array}{rl} & \langle \mathcal{F}(x^{k_l}) - \mathcal{F}(x^*) + q^{k_l+1} - q^* + \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^* - x^{k_l+1}\rangle \\ \geq & -\delta_{k_l} \cdot ||x^* - x^{k_l+1}|| - \chi_{k_l} \cdot \langle \nabla h(x^{k_l+1}) - \nabla h(x^{k_l}),\ x^* - x^{k_l+1}\rangle, \end{array}$$

and continuing

$$\begin{array}{rl} & \delta_{k_l} \cdot ||x^* - x^{k_l+1}|| \ +\ \langle \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^* - x^{k_l+1}\rangle \\ & \quad + \ \chi_{k_l} \cdot \langle \nabla h(x^{k_l+1}) - \nabla h(x^{k_l}),\ x^* - x^{k_l+1}\rangle \\ \geq & \langle \mathcal{F}(x^{k_l}) - \mathcal{F}(x^*) + q^{k_l+1} - q^*,\ x^{k_l+1} - x^*\rangle \\ = & \langle \mathcal{F}(x^{k_l+1}) - \mathcal{F}(x^*) + q^{k_l+1} - q^*,\ x^{k_l+1} - x^*\rangle \\ & \quad + \ \langle \mathcal{F}(x^{k_l}) - \mathcal{F}(x^{k_l+1}),\ x^{k_l+1} - x^*\rangle. \end{array} \tag{18}$$

To apply the special assumption of the second case one has to pay attention since $q^{k_l+1} \in \mathcal{Q}^{k_l}(x^{k_l+1})$ does not have to belong to $\mathcal{Q}(x^{k_l+1})$. But owing to the *Brønsted-Rockafellar-property* (cf. [2]) of the $\varepsilon-$enlargement we know that for each l there is $\tilde{x}^{k_l+1}$ and some $\tilde{q}^{k_l+1} \in \mathcal{Q}(\tilde{x}^{k_l+1})$ such that

$$||\tilde{x}^{k_l+1} - x^{k_l+1}|| \ \leq\ \sqrt{\varepsilon_{k_l+1}} \quad \text{and} \quad ||\tilde{q}^{k_l+1} - q^{k_l+1}|| \ \leq \sqrt{\varepsilon_{k_l+1}}. \tag{19}$$

The convergence $\{\tilde{x}^{k_l}\} \to \overline{x}$ is obvious. Denote $f^{k_l} := \mathcal{F}(\tilde{x}^{k_l}) - \mathcal{F}(x^{k_l})$; since $\mathcal{F}$ is continuous, we know that $f^{k_l} \to 0$ for $l \to \infty$. Analogeously, by (19) we know that $\{\tilde{q}^{k_l} - q^{k_l}\} \to 0$ for $l \to \infty$, too.

Now (18) turns to

$$\begin{array}{rcl} T^{k_l} & := & \langle \chi_{k_l} \cdot \big(\nabla h(x^{k_l+1}) - \nabla h(x^{k_l})\big) + \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^* - x^{k_l+1}\rangle \\ & & \quad + \ \delta_{k_l} \cdot ||x^* - x^{k_l+1}|| - \langle \mathcal{F}(x^{k_l}) - \mathcal{F}(x^{k_l+1}),\ x^{k_l+1} - x^*\rangle \\ & \geq & \langle \mathcal{F}(x^{k_l+1}) - \mathcal{F}(x^*) + q^{k_l+1} - q^*,\ x^{k_l+1} - x^*\rangle \\ & = & \langle \mathcal{F}(\tilde{x}^{k_l+1}) - \mathcal{F}(x^*) + \tilde{q}^{k_l+1} - q^*,\ \tilde{x}^{k_l+1} - x^*\rangle \\ & & \quad + \ \langle \mathcal{F}(\tilde{x}^{k_l+1}) - \mathcal{F}(x^*) + \tilde{q}^{k_l+1} - q^*,\ x^{k_l+1} - \tilde{x}^{k_l+1}\rangle \\ & & \quad + \ \langle \mathcal{F}(x^{k_l+1}) - \mathcal{F}(\tilde{x}^{k_l+1}),\ x^{k_l+1} - x^*\rangle \\ & & \quad + \ \langle q^{k_l+1} - \tilde{q}^{k_l+1},\ x^{k_l+1} - x^*\rangle. \end{array}$$

Using the weakened Dunn property (4) we deduce

$$
\begin{aligned}
T^{k_l} \;\geq\; & \alpha(x^*) \cdot ||\mathcal{F}(\tilde{x}^{k_l+1}) - \mathcal{F}(x^*) + \tilde{q}^{k_l+1} - q^*||^2 && (20)\\
& + \langle \tilde{q}^{k_l+1} - q^*,\ x^{k_l+1} - \tilde{x}^{k_l+1} \rangle && (21)\\
& + \langle \mathcal{F}(\tilde{x}^{k_l+1}) - \mathcal{F}(x^*),\ x^{k_l+1} - \tilde{x}^{k_l+1} \rangle && (22)\\
& + \langle \mathcal{F}(x^{k_l+1}) - \mathcal{F}(\tilde{x}^{k_l+1}),\ x^{k_l+1} - x^* \rangle && (23)\\
& + \langle q^{k_l+1} - \tilde{q}^{k_l+1},\ x^{k_l+1} - x^* \rangle. && (24)
\end{aligned}
$$

Passing to the limit $l \to \infty$ we see that $T^{k_l} \to 0$ (remember (8), (A.8), Proposition 1 and the continuity of $\mathcal{F}$), (21) - (24) also vanish (for (21) remember Remark 3). Therefore, for $l \to \infty$

$$
\mathcal{F}(\tilde{x}^{k_l+1}) - \mathcal{F}(x^*) + \tilde{q}^{k_l+1} - q^* \;\to\; 0 \tag{25}
$$

has to be since all other terms tend to zero. With respect to the continuity of $\mathcal{F}$ the convergence $\mathcal{F}(\tilde{x}^{k_l+1}) \to \mathcal{F}(\overline{x})$ is obvious, and that is why the sequence $\{q^{k_l}\}$ has to be convergent, say $\{q^{k_l}\} \to \overline{q}$.

Since also $\{\tilde{q}^{k_l}\} \to \overline{q}$ is valid, maximal monotonicity of $\mathcal{Q}$ implies $\overline{q} \in \mathcal{Q}(\overline{x})$, and using (25) we conclude that

$$
\mathcal{F}(x^*) + q^* \;=\; \mathcal{F}(\overline{x}) + \overline{q}, \tag{26}
$$

where q^* is defined as above. Using this fact, $x^* \in SOL(K, \mathcal{F}, \mathcal{Q})$ and the iteration scheme, $\overline{x} \in SOL(K, \mathcal{F}, \mathcal{Q})$ is derived straightforward:

$$
\begin{aligned}
\langle \mathcal{F}(\overline{x}) + \overline{q},\ x - \overline{x} \rangle \;&=\; \langle \mathcal{F}(\overline{x}) + \overline{q},\ x^* - \overline{x} \rangle + \langle \mathcal{F}(\overline{x}) + \overline{q},\ x - x^* \rangle\\
&\geq\; \langle \mathcal{F}(\overline{x}) + \overline{q},\ x^* - \overline{x} \rangle\\
&=\; \lim_{l\to\infty} \langle \mathcal{F}(x^{k_l}) + q^{k_l+1},\ x^* - x^{k_l+1} \rangle\\
&\geq\; -\lim_{l\to\infty} \langle \mathcal{L}_{k_l}(x^{k_l+1}) - \mathcal{L}_{k_l}(x^{k_l}),\ x^* - x^{k_l+1} \rangle\\
&\quad\; + \lim_{l\to\infty} \delta_{k_l} \cdot ||x^* - x^{k_l+1}||\\
&\quad\; - \lim_{l\to\infty} \chi_{k_l} \cdot \langle \nabla h(x^{k_l+1}) - \nabla h(x^{k_l}),\ x^* - x^{k_l+1} \rangle\\
&=\; 0.
\end{aligned}
$$

Thus, $\overline{x} \in SOL(K, \mathcal{F}, \mathcal{Q})$ is also proven for the second case. □

Before concluding the convergence of the sequence $\{x^k\}$ we will discuss the necessity of symmetry of $\mathcal{Q}$. Let us specify the consequences of the common use of (A.7) and (3). Although the latter assumption does not deal with the $\mathcal{L}_y$ directly, it permits via (14) and (16) to conclude

$$
\lim_{l\to\infty} \mathcal{L}_{k_l}(x^*) - \mathcal{L}_{k_l}(x^{k_l}) \;=\; 0. \tag{27}
$$

So we are looking for a condition on the $\mathcal{L}_y$ which permit to conclude $x^* = \overline{x}$ in this situation.

Lemma 4. *Assume that (3) holds and further for all sequences $y^k \to y$ with $y^k \in \operatorname{int}(K)$ and $z^k \to z$ with $z^k \in K$ the continuity property*

$$\mathcal{L}_{y^k}(z^k) \to \mathcal{L}_y(z) \tag{28}$$

is valid. Additionally assume that each $\mathcal{L}_y$ is an injective mapping. Then each cluster point of the generated sequence belongs to the solution set, too.

Proof. In the situation above we obtain $\mathcal{L}_{\overline{x}}(x^*) = \mathcal{L}_{\overline{x}}(\overline{x})$ by using the continuity hypothesis. But since $\mathcal{L}_{\overline{x}}$ is injective, we obtain $\overline{x} = x^*$ and therefore, $\overline{x} \in SOL(K, \mathcal{F}, \mathcal{Q})$ is valid also for non-symmetric operators $\mathcal{Q}$. □

Up to now we have shown that the generated sequence has at least one cluster point and each of its cluster points belongs to the solution set. Now we deduce the convergence of the entire sequence.

Theorem 2. *The sequence $\{x^k\}$ generated by the method under consideration converges to a solution of $VI(K, \mathcal{F}, \mathcal{Q})$.*

Proof. Since each cluster point $\overline{x}$ belongs to $SOL(K, \mathcal{F}, \mathcal{Q})$ by Lemma 3 (and Lemma 4, respectively), it holds due to Proposition 1

$$D_h(\overline{x}, x^k) \to \overline{D}, \quad k \to \infty,$$

for some $\overline{D} \geq 0$. Now 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 $x^{k_l} \to \overline{x}$ we obtain from (B.5)a)

$$D_h(\overline{x}, x^{k_l}) \to 0, \quad l \to \infty,$$

i.e. $\overline{D} = 0$ has to be valid. Now $D_h(\overline{x}, x^k) \to 0$ for $k \to \infty$ and (B.6) yield the assertion

$$x^k \to \overline{x}, \quad k \to \infty.$$

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

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

and suppose $\overline{x} \neq x'$. Then (B.5)b) yields

$$D_h(\overline{x}, x^{k_m}) \to \infty,$$

but that is a contradiction to the convergence of the latter sequence (remember that $\overline{x} \in SOL(K, \mathcal{F}, \mathcal{Q})$ and then use Proposition 1). Therefore, there can only be one cluster point, and for a bounded sequence this is equivalent to its convergence.

□

Remark 4. *The continuity property discussed in Lemma 4 is fulfilled for many auxiliary operators. Indeed, if e.g.* $\mathcal{L}_k(x) := \nabla\mathcal{F}(x^k)x$ *(correspondence to a Newton method), then (28) is fulfilled; for projection methods,* $\mathcal{L}_k(x) := Ax$*, this holds true, too.*

Further, also the injectivity property of the $\mathcal{L}_y$ *is fulfilled for commonly used auxiliary operators. Indeed, if we consider some projection methods,* $\mathcal{L}_k(x) := Ax$ *with a positive definite matrix* A*, then – due to the regularity of* A *– (27) is equivalent to* $\overline{x} = x^*$*.*

On the other hand, discussing the Newton method $\mathcal{L}_k(x) := \nabla\mathcal{F}(x^k)x$*, we analogeously obtain* $\overline{x} - x^* \in \ker(\nabla\mathcal{F}(\overline{x}))$*; if the latter Jacobian is regular, this means* $\overline{x} = x^*$*.*

So, considering commonly used auxiliary operators with a certain full-rank-property it is no longer of interest whether $\mathcal{Q}$ *is a subdifferential or not, since the proof of* $\overline{x} = x^*$ *does not need this argument when considering concrete families* $\{\mathcal{L}_y\}$ *with such a property. In other words, only using (A.1) - (A.10) and (3) the EPAPP-method always converges to a solution if an auxiliary operator with some – quite natural – properties is chosen.*

$\diamond$

Finally, we can make the following statement concerning the behaviour of the EPAPP-method whenever $SOL(K, \mathcal{F}, \mathcal{Q})$ is not a singleton.

Corollary 1. *(see [20])*
Suppose that the above mentioned conditions are fulfilled. If $SOL(K, \mathcal{F}, \mathcal{Q})$ *contains more than just one element, the sequence* $\{x^k\}$ *converges to an element* $\overline{x}$ *with* $g_i(\overline{x}) < 0$ *for all* $i \in I_2$*.*

Note that in case of non-uniqueness of solutions such an $\overline{x}$ always has to exist. Otherwise one would obtain a contradiction applying the referred proof to each $z \in SOL(K, \mathcal{F}, \mathcal{Q})$.

4 Concluding Remarks

Up to now, convergence results of Bregman function based methods have always been created under the hypothesis of paramonotonicity of some operator and some other, more or less, theoretical assumptions. On the one hand paramonotonicity allowed one to avoid co-coercivity-like assumptions, but, on the other hand, the latter ones are much easier to check. Although co-coerciveness implies paramonotonicity in general, Example 1 shows that weakened co-coerciveness does not even imply monotonicity.

Zhu and Marcotte [26] started their investigations with some co-coercivity assumptions, and presented a method which requires a (standard) Bregman function with a Lipschitz-continuous gradient. This implicates a lack of applicability (there are not many sets for which such Bregman functions are known) and, even worse, the Lipschitz-continuity of the gradient of Bregman functions contradicts their zone-coercivity. But the latter one is the reason for the mentioned *interior-point-effect*, i.e. the effect that the auxiliary problems can – with a certain precaution – be treated as unconstrained problems.

The method presented here (including the used assumptions) is a unification of these two research fields in some way. It admits

- non-paramonotone operators with some special properties,
- a multi-valued operator $\mathcal{Q}$,
- a successive (outer) approximation of the multi-valued operator $\mathcal{Q}$ by means of the ε-enlargement,
- treating the auxiliary problems as unconstrained ones,
- an inexact solution of the auxiliary problems under the criterion of summability of the errors.

To finish, let us shortly classify the discussed cases. In the absence of auxiliary operators, Case 1 only requires the assumptions made by Kaplan and Tichatschke except the missing paramonotonicity. The assumption $\mathcal{Q} = \partial f$ can be avoided by the choice of adequate auxiliary operators (see Lemma 4) with some quite natural properties.

Case 2 presents a convergence analysis based on a co-coercivity-like assumption which is much more general than the classical assumption of co-coercivity. In the case of the latter one we would discuss single-valued operators only (see Remark 2). As demonstrated in Example 1, the weakened hypothesis even admits a set-valued operator $\mathcal{F}$ whose sum with some strongly monotone operator is not even pseudomonotone.

References

[1] A. Auslender, M. Teboulle, and S. Ben-Tiba. A logarithmic-quadratic proximal method for variational inequalities. *Computational Optimization and Applications*, 12:31–40, 1999.

[2] R. Burachik and B. Svaiter. ε-enlargement of maximal monotone operators in Banach spaces. *Set-Valued Analysis*, 7:117–132, 1999.

[3] Y. Censor, A. Iusem, and S. A. Zenios. An interior point method with Bregman functions for the variational inequality problem with paramonotone operators. *Math. Programming*, 81:373–400, 1998.

[4] G. Chen and M. Teboulle. Convergence analysis of a proximal-like minimization algorithm using Bregman functions. *SIAM J. Optim.*, 3:538–543, 1993.

[5] G. Cohen. Auxiliary problem principle and decomposition of optimization problems. *JOTA*, 32:277–305, 1980.

[6] G. Cohen. Auxiliary problem principle extended to variational inequalities. *JOTA*, 59:325–333, 1988.

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

[8] J. Eckstein. Nonlinear proximal point algorithms using Bregman functions, with application to convex programming. *Math. Oper. Res.*, 18:202–226, 1993.

[9] J. Eckstein. Approximate iterations in Bregman-function-based proximal algorithms. *Math. Programming*, 83:113–123, 1998.

[10] A. Iusem. On some properties of generalized proximal point methods for variational inequalities. *JOTA*, 96:337–362, 1998.

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

[12] C. Jager. Auxiliary problem principle using logarithmic-quadratic functions. *Forschungsbericht*, 07-01:pp. 23, 2007. Mathematik/Informatik, Universität Trier.

[13] A. Kaplan and R. Tichatschke. On the convergence of interior proximal methods for variational inequalities on non-polyhedral sets (submitted).

[14] A. Kaplan and R. Tichatschke. Convergence analysis of non-quadratic proximal methods for variational inequalities in Hilbert spaces. *J. of Global Optimization*, 22:119–136, 2002.

[15] A. Kaplan and R. Tichatschke. Proximal-based regularization methods and a successive approximation of variational inequalities in Hilbert spaces. *Control and Cybernetics*, 31:521–544, 2002.

[16] A. Kaplan and R. Tichatschke. Extended auxiliary problem principle to variational inequalities with multi-valued operators. *Optimization*, 53:223–252, 2004.

[17] A. Kaplan and R. Tichatschke. Extended auxiliary problem principle using Bregman distances. *Optimization*, 53:603–623, 2004.

[18] A. Kaplan and R. Tichatschke. Bregman functions and proximal methods for variational problems with nonlinear constraints. *Optimization*, 56, No 1-2:253–265, 2007.

[19] A. Kaplan and R. Tichatschke. Interior proximal method for variational inequalities on non-polyhedral sets. *Discussiones Mathematicae*, 27, 1:71–93, 2007.

[20] A. Kaplan and R. Tichatschke. Bregman functions and auxiliary problem principle. *Optimization Methods and Software*, 23, 1:95–107, 2008.

[21] S. Karamardian. Complementarity problems over cones with monotone and pseudomonotone maps. *JOTA*, 18:445–454, 1976.

[22] P.-L. Lions and B. Mercier. Splitting algorithm for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16:964–979, 1979.

[23] B. Martinet. Régularisation d'inéquations variationelles par approximations successives. *RIRO*, 4:154–159, 1970.

[24] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14:877–898, 1976.

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

[26] D. L. Zhu and P. Marcotte. Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities. *SIAM J. Optimization*, 6:714–726, 1996.