Korpelevich's method for variational inequality problem in Banach spaces

Alfredo N. Iusem * and Mostafa Nasri †

July 21, 2008

Abstract

We propose a variant of Korpelevich's method for solving variational inequality problems with operators in Banach spaces. A full convergence analysis of the method is presented under reasonable assumptions on the problem data.

Keywords: Bregman function, Bregman projection, Korpelevich's method, Variational inequality problem.

AMS Classification: 49J40, 49M07, 90C25.

1 Introduction

Assume that B is a reflexive Banach space with norm $\|\cdot\|$, B^* is the topological dual of B with norm $\|\cdot\|_*$, and the symbol $\langle\cdot,\cdot\rangle$ indicates the duality coupling in $B^* \times B$, defined by $\langle \phi, x\rangle = \phi(x)$ for all $x \in B$ and all $\phi \in B^*$. The underlying problem, called variational inequality problem and denoted by $\mathrm{VIP}(T, C)$ from now on, consists of finding an $x^* \in C$ such that

$$\langle T(x^*), x - x^*\rangle \geq 0 \quad \forall x \in C,$$

where C is a nonempty closed convex subset of B and $T : B \to B^*$ is an operator. The set of solutions of $\mathrm{VIP}(T, C)$ will be denoted by $S(T, C)$.

Variational inequality problems arise in a wide variety of application areas (se. e.g. [18]). They encompass as particular cases convex optimization problems, linear and monotone complementarity problems, equilibrium problems, etc.

*Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ, CEP 22460-320, Brazil, e-mail: iusp@impa.br

†Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Rio de Janeiro, RJ, CEP 22460-320, Brazil, e-mail: mostafa@impa.br

In this paper, we will extend Korpelevich's method to infinite dimensional Banach spaces, and thus we start with an introduction to its well known finite dimensional formulation, i.e., we assume that $B = \mathbb{R}^n$. In this setting, there are several iterative methods for solving VIP(T, C). The simplest one is the natural extension of the projected gradient method for optimization problems, substituting the operator T for the gradient, so that we generate a sequenced $\{x^k\} \subset \mathbb{R}^n$ through:

$$x^{k+1} = P_C(x^k - \alpha_k T(x^k)), \tag{1}$$

where α_k is some positive real number and P_C, is the orthogonal projection onto C. This method converges under quite strong hypotheses. If T is Lipschitz continuous and strongly monotone, i.e.

$$\|T(x) - T(y)\| \leq L \|x - y\| \quad \forall\, x, y \in \mathbb{R}^n,$$

and

$$\langle T(x) - T(y), x - y\rangle \geq \sigma \|x - y\|^2 \quad \forall\, x, y \in \mathbb{R}^n,$$

where $L > 0$ and $\sigma > 0$ are the Lipschitz and strong monotonicity constants respectively, then the sequence generated by (1) converges to a solution of VIP(T, C) (provided that the problem has solutions) if the stepsizes α_k are taken as $\alpha_k = \alpha \in (0, 2\sigma/L^2)$ for all k (see e.g., [6], [10]). If we relax the strong monotonicity assumption to plain monotonicity, i.,e.

$$\langle T(x) - T(y), x - y\rangle \geq 0 \quad \forall\, x, y \in \mathbb{R}^n,$$

then the situation becomes more complicated, and we may get a divergent sequence independently of the choice of the stepsizes α_k. The typical example consists of taking $B = C = \mathbb{R}^2$ and T a rotation with a $\pi/2$ angle, which is certainly monotone and Lipschitz continuous. The unique solution of VIP(T, C) is the origin, but (1) gives rise to a sequence satisfying $\|x^{k+1}\| > \|x^k\|$ for all k. In order to deal with this situation, Korpelevich suggested in [19] an algorithm of the form:

$$y^k = P_C(x^k - \alpha_k T(x^k)), \tag{2}$$

$$x^{k+1} = P_C(x^k - \alpha_k T(y^k)). \tag{3}$$

In order to clarify the geometric motivation behind this procedure, consider VIP(T, C) with a monotone T. Let $H_k = \{x \in \mathbb{R}^n : \langle T(y^k), x - y^k\rangle = 0\}$. If $x^* \in S(T, C)$, then the monotonicity of T guarantees that (3) is a step from x^k in the direction of its orthogonal projection onto the hyperplane H_k separating x^k from any solution of VIP(T, C), so that for α_k small enough, x^{k+1} will be closer than x^k to any solution of VIP(T, C). This property, called Fejér monotonicity of $\{x^k\}$ with respect to the solution set of VIP(T, C), is the basis of the convergence analysis. In fact, if T is Lipschitz continuous with constant L and VIP(T, C) has solutions, then the sequence generated by (2)–(3) converges to a solution of VIP(T, C) provided that $\alpha_k = \alpha \in (0, 1/L)$ (see, [19]).

In the absence of Lipschitz continuity of T, it is natural to emulate once again the projected gradient method for optimization, and search for an appropriate stepsize in an inner loop. This is achieved by taking first an arbitrary stepsize β_k, compute $z^k = P_C(x^k - \beta_k T(x^k))$ and then try vectors of the form $y(\alpha) = \alpha v^k + (1-\alpha)x^k$ with $\alpha \in (0,1]$, until a value of α is found such that $y^k = y(\alpha)$ satisfies

$$\langle T(y^k), x^k - P_C(z^k)\rangle \geq \frac{\delta}{\beta_k}\left\|x^k - P_C(z^k)\right\|^2, \tag{4}$$

where $\delta \in (0,1)$ is some constant. Then the orthogonal projection of x^k onto the hyperplane $H_k = \{x \in \mathbb{R}^n : \langle T(y^k), x - y^k\rangle = 0\}$ is computed, and finally x^{k+1} is defined as orthogonal projection of $P_{H_k}(x^k)$ onto C. We remark that along the search for the proper α the right hand side of (4) is kept constant, and that, through T is evaluated at several points in the segment between z^k and x^k, no orthogonal projection onto C are required during the inner loop, and we have only two projections onto C per iteration, in the computation of z^k and x^{k+1}, exactly as in the original method (2)–(3).

The above backtracking procedure for determining th right α is sometimes called an Armijo-type search (see [1]). It has been analyzed for VIP(T,C) in [16]. Other variants of Korpelevich method can be found in [17], [21], and other methods for the problem appear in [5], [11], [13], [23] and [24] for the case in which T is point-to-point, as in this paper. Extensions of Korpelevich method to the of point-to-set case can be found in [15] and [4]. All these references deal with finite dimensional spaces.

In this paper, we are interested in infinite dimensional Banach spaces, for which direct methods for VIP(T,C) are much scarcer. A descent method was proposed in [27], and a projection method, which works in reflexive Banach spaces, is analyzed in [2], [12]. We proceed to describe the latter. Let $J : B \to B^*$ be the normalized duality mapping (i.e., the subdifferential of $g(x) = \frac{1}{2}\|x\|^2$; see [9]), which can also be defined as

$$J(x) = \{x^* \in B^* : \langle x^*, x\rangle = \|x^*\|_* \|x\|, \|x^*\|_* = \|x\|\}.$$

Given $x^k \in B$, x^{k+1} is calculated as the Bregman projection with respect to g of the point $J^{-1}\left[J\left(x^k - \lambda_k T(x^k)\right)\right]$ onto C, where $\{\lambda_k\} \subset \mathbb{R}_{++}$ is an exogenous bounded sequence (see Definition 2.6 below for the formal definition of Bregman projection). Formally, the method has the form

$$x^{k+1} = \Pi_C^g\left\{J^{-1}\left[J\left(x^k - \lambda_k T(x^k)\right)\right]\right\}, \tag{5}$$

where Π_C^g is the Bregman projection onto C with respect to g. The convergence result for this method is as follows.

Theorem 1.1. *Suppose that B is uniformly convex and uniformly smooth and that*

i) T is uniformly monotone, that is, $\langle T(x) - T(y), x - y\rangle \geq \psi(\|T(x) - T(y)\|_*)$, where $\psi(t)$ is a continuous strictly increasing function for all $t \geq 0$ with $\psi(0) = 0$, *ii) T has ϕ-arbitrary growth, that is, $\|T(y)\|_* \leq \phi(\|y - z\|)$ for all $y \in C$ and $\{z\} = S(T, C)$, where ϕ is a continuous nondecreasing function with $\phi(0) \geq 0$,*

iii) $\{\lambda_k\}$ is a positive nonincreasing sequence that satisfies $\lim_{k\to\infty} \lambda_k = 0$ and $\sum_{k=0}^{\infty} \lambda_k = \infty$.

Then the sequence $\{x^k\}$ generated by (5) converges strongly to a unique point $z \in S(T, C)$.

Proof. See [2]. □

Another result for this method, establishing weak convergence, rather than strong, can be found in [12]. It reads as follows:

Theorem 1.2. *Let B be a uniformly smooth Banach space, also 2-uniformly convex with constant $1/\gamma$, whose duality mapping J is weakly sequentially continuous. Assume that* VIP(T, C) *satisfies:*

i) there exists a real positive number α such that for all $x, y \in C$, it holds that

$$\langle T(x) - T(y), x - y\rangle \geq \alpha \|T(x) - T(y)\|_*^2,$$

ii) for all $y \in C$ and all $u \in S(T, C)$, it holds that

$$\|T(y)\|_* \leq \|T(y) - T(u)\|_*.$$

If $S(T, C) \neq \emptyset$, $\{\lambda_k\} \subset [\hat{\beta}, \tilde{\beta}]$, with $0 < \hat{\beta} < \tilde{\beta} < (\gamma^2\alpha)/2$, and x^0 belongs to C, then the sequence $\{x^k\}$ generated by (5) is weakly convergent to the point $z \in S(T, C)$ characterized as $z = \lim_{k\to+\infty} \Pi_{S(T,C)}(x^k)$.

Proof. See Theorem 3.1 of [12]. □

Related convergence results for Cesaro averages of sequences related to $\{x^k\}$ can be found in Theorem 4.2 of [3]. We will see later on that the convergence properties of our algorithm hold under assumptions quite weaker than those demanded by Theorems 1.1 and 1.2 (see Theorem 4.8 below).

The outline of this paper is as follows. In Section 2 we present some theoretical tools needed in the sequel. In Section 3 we state our algorithm formally. In Section 4 we establish the convergence properties of the algorithm.

2 Preliminaries

Definition 2.1. *Consider an operator $T : B \to B^*$.*

i) T is said to be monotone if for all $x, y \in B$, it holds that

$$\langle T(x) - T(y), x - y\rangle \geq 0.$$

ii) T is said to be pseudomonotone if for all $x, y \in B$, it holds that

$$\langle T(y), x-y\rangle \geq 0 \Rightarrow \langle T(x), x-y\rangle \geq 0.$$

iii) T is said to be hemicontinuous on a subset C of B if for all $x, y \in C$, the mapping $h : [0,1] \to B^*$ defined as $h(t) = T(tx + (1-t)y)$ is continuous with respect to the weak* topology of B^*.

iv) T is said to be uniformly continuous on a subset E of B if for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $x, y \in E$, it hold that

$$\|x-y\| < \delta \Rightarrow \|T(x) - T(y)\|_* < \epsilon.$$

We will prove that the sequence generated by our algorithm is an asymptotically solving sequence (see Definition 4.4) for VIP(T, C) when T is uniformly continuous on bounded subsets of C, $S(T, C) \neq \emptyset$, and VIP(T, C) satisfies property **A**, stated below.

A: For some $x^* \in S(T, C)$, it holds that

$$\langle T(y), y - x^*\rangle \geq 0 \quad \forall y \in C. \tag{6}$$

It is worthwhile mentioning that the problem of finding an $x^* \in C$ such that (6) is satisfied, is known as *Minty variational inequality problem.* Some existence results for this problem has been presented in [20]. We also mention that assumption **A** has been already used for solving VIP(T, C) in finite dimensional spaces (see, e.g., [23]). It is not difficult to prove that pseudomonotonicity implies property **A**, while the converse is not true, as illustrated by the following example.

Example 2.2. *Consider $T : \mathbb{R} \to \mathbb{R}$ defined as $T(x) = \cos(x)$ with $C = [0, \frac{\pi}{2}]$.*

We have that $S(T, C) = \{0, \frac{\pi}{2}\}$. VIP$(T, C)$ satisfies the property **A**, because for $x^* = 0$ the statement in (6) holds. But if we take $x = 0$ and $y = \frac{\pi}{2}$ in Definition 2.1(ii), we conclude that T is not pseudomonotone.

The next lemma will be useful for proving that all weak cluster points of the sequence generated by our algorithm solves $S(T, C)$.

Lemma 2.3. *Consider* VIP(T, C). *If $T : C \to B^*$ is monotone and hemicontinuous on C, then*

$$S(T, C) = \{x \in C : \langle T(y), y - x\rangle \geq 0 \quad \forall y \in C\}.$$

Proof. See [25]. □

Next we state some properties of Bregman projections which will be used in the remainder of this paper, taken from [7]. We consider an auxiliary function $g : B \to \mathbb{R}$, which is strictly convex, lower semicontinuous, and Gâteaux differentiable. We will denote the family of such functions as $\mathcal{F}$. The Gâteaux derivative of g will be denoted by g'.

Definition 2.4. *Let $g : B \to \mathbb{R}$ be a convex and Gâteaux differentiable function.*

i) The Bregman distance with respect to g is the function $D_g : B \times B \to \mathbb{R}$ defined as $D_g(x, y) = g(x) - g(y) - \langle g'(y), x - y\rangle$.

ii) The modulus of total convexity of g is the function $\nu_g : B \times [0, +\infty) \to [0, +\infty)$ defined as $\nu_g(x, t) = \inf\{D_g(y, x) : y \in B, \|y - x\| = t\}$.

iii) g is said to be a totally convex function at $x \in B$ if $\nu_g(x, t) > 0$ for all $t > 0$.

iv) g is said to be a totally convex function if $\nu_g(x, t) > 0$ for all $t > 0$ and all $x \in B$.

v) g is said to be a uniformly totally convex function on $E \subset B$ if $\inf_{x\in\tilde{E}} \nu_g(x, t) > 0$ for all $t > 0$ and all bounded subsets $\tilde{E} \subset E$.

We will present next some additional conditions on g, which are needed in the convergence analysis of our algorithm.

H1: The level sets of $D_g(x, \cdot)$ are bounded for all $x \in B$.

H2: $\inf_{x\in C} \nu_g(x, t) > 0$ for all bounded set $C \subset B$ and all $t > 0$.

H3: g' is uniformly continuous on bounded subsets of B.

H4: g' is onto, i.e., for all $y \in B^*$, there exists $x \in B$ such that $g'(x) = y$.

H5: $(g')^{-1}$ is continuous.

H6: If $\{y^k\}$ and $\{z^k\}$ are sequences in C which converge weakly to y and z, respectively and $y \neq z$, then

$$\liminf_{j\to\infty} \left|\langle g'(y^k) - g'(z^k), y - z\rangle\right| > 0.$$

These properties were identified in [14]. We make a few remarks on them. H2 is known to hold when g is lower semicontinuous and uniformly convex on bounded sets (see [8]). It has been proved in page 75 of [7], that sequential weak-to-weak* continuity of g' ensures H6. Existence of $(g')^{-1}$ will be a consequence of H4 for any $g \in \mathcal{F}$. We mention that for the case of strictly convex and smooth B and $g(x) = \|x\|^r$, we have an explicit formula for $(g')^{-1}$, in terms of ϕ', where $\phi(\cdot) = \frac{1}{s}\|\cdot\|_*^r$ with $\frac{1}{s} + \frac{1}{r} = 1$, namely $(g')^{-1} = r^{1-s}\phi'$.

It is important to check that functions satisfying these properties are available in a wide class of Banach spaces. The prototypical example is $g(x) = \frac{1}{2}\|x\|^2$, in which case g' is the duality operator, and the identity operator in the case of Hilbert space. It is convenient to deal with a general g rather than just the square of the norm because in Banach spaces this function lacks the privileged status it enjoys in Hilbert spaces. In the spaces $\mathcal{L}^p$ and ℓ_p, for instance, the function

$g(x) = \frac{1}{p}\|x\|^p$ leads to simpler calculations than the square of the norm. It has been shown in [14] that the function $g(x) = r\|x\|^s$, works satisfactorily in any reflexive, uniformly smooth and uniformly convex Banach space, for any $r > 0$, $s > 1$. We have the following result.

Proposition 2.5.

i) If B is a uniformly smooth and uniformly convex Banach space, then $g(x) = r\|x\|^s$ satisfies H1–H4 *for all $r > 0$ and all $s > 1$.*

ii) If B is a Hilbert space, then $g(x) = \frac{1}{2}\|x\|^2$ satisfies H6*. The same holds for $g(x) = \frac{1}{p}\|x\|^p$ when $B = \ell_p$ $(1 < p < \infty)$.*

Proof. See Proposition 2 of [14]. □

We remark that the only problematic property is H6, in the sense that the only example we have of a nonhilbertian Banach space for which we know functions satisfying it is ℓ_p with $1 < p < \infty$. As we will see in Section 4, most of our convergence results demand only H1–H5.

Now we present some properties of Bregman projection in Banach spaces. A full discussion about this issue can be found in [7].

Definition 2.6. *Assume that B is a Banach space. Let $g \in \mathcal{F}$ be a totally convex function on B satisfying* H1*. The Bregman projection of $x \in B$ onto C, denoted by $\Pi_C^g(x)$, is defined as unique solution of the following minimization problem.*

$$\Pi_C^g(x) = \operatorname*{argmin}_{y \in C} D_g(y, x).$$

It is worthwhile mentioning that $D_g(x, y) = \frac{1}{2}\|x - y\|^2$ whenever $g(x) = \frac{1}{2}\|x\|^2$ and B is a Hilbert space. The next proposition lists some properties of Bregman projections.

Proposition 2.7. *Assume that B is a Banach space. Let $g \in \mathcal{F}$ be a totally convex function on B satisfying* H1*. In this situation, the following two statements are true.*

i) The operator $\Pi_C^g : B \to C$ is well defined.

ii) $\bar{x} = \Pi_C^g(x)$ if and only if $g'(x) - g'(\bar{x}) \in N_C(\bar{x})$, or equivalently, $\bar{x} \in C$ and

$$\langle g'(x) - g'(\bar{x}), z - \bar{x}\rangle \leq 0 \quad \forall z \in C.$$

Proof. See page 70 of [7]. □

We will utilize the following properties in our convergence analysis.

Proposition 2.8. *Assume that $g : B \to \mathbb{R}$ is convex and Gâteaux differentiable. For any $x, y, z \in B$, it holds that*

$$D_g(y,z) + D_g(z,x) - D_g(y,x) = \langle g'(z) - g'(x), z - y \rangle. \tag{7}$$

Proof. See 1.3.9 of [7]. □

Proposition 2.9. *Let $g \in \mathcal{F}$ be a totally convex function on B satisfying* H1. *Then for all $0 \neq v \in B^*$, $\tilde{y} \in B$, $x \in H^+$ and $\bar{x} \in H^-$, it holds that $D_g(\bar{x}, x) \geq D_g(\bar{x}, z) + D_g(z, x)$ where z is the unique minimizer of $D_g(\cdot, x)$ on H where $H = \{y \in B : \langle v, y - \tilde{y} \rangle = 0\}$, $H^+ = \{y \in B : \langle v, y - \tilde{y} \rangle \geq 0\}$, $H^- = \{y \in B : \langle v, y - \tilde{y} \rangle \leq 0\}$.*

Proof. See Lemma 1 of [14]. □

Proposition 2.10. *Assume that $g \in \mathcal{F}$ satisfies* H2. *Let $\{x^k\}, \{y^k\} \subset B$ be two sequences such that at least one of them is bounded. If $\lim_{k\to\infty} D_g(y^k, x^k) = 0$, then $\lim_{k\to\infty} \left\| x^k - y^k \right\| = 0$.*

Proof. See Proposition 5 of [14]. □

Proposition 2.11. *Let $g : B \to \mathbb{R}$ be totally convex and Fréchet differentiable on B. If* H1 *holds, then $\Pi_C^g : B \to C$, the Bregman projection operator, is norm-to-norm continuous on B.*

Proof. See Proposition 4.3 of [22], where this result is proved under weaker assumptions. □

3 Statement of the algorithm

Now we present the formal statement of the algorithm. It requires three exogenous parameters: $\delta \in (0,1)$, $\hat{\beta}$ and $\tilde{\beta}$ with $0 < \hat{\beta} \leq \tilde{\beta}$, and an exogenous sequence $\{\beta_k\} \subset [\hat{\beta}, \tilde{\beta}]$.

Korpelevich's method for VIP(T, C):

1. Initialization:

$$x^0 \in C. \tag{8}$$

2. Iterative step: Given x^k, define

$$z^k = (g')^{-1}[g'(x^k) - \beta_k T(x^k)]. \tag{9}$$

If $x^k = \Pi_C^g(z^k)$ stop. Otherwise, let

$$\ell(k) = \min\{\ell \geq 0 : \langle T(y^{\ell(k)}), x^k - \Pi_C^g(z^k) \rangle \geq \frac{\delta}{\beta_k} D_g(\Pi_C^g(z^k), x^k), \tag{10}$$

where

$$y^{\ell(k)} = 2^{-\ell}\Pi^g_C(z^k) + (1-2^{-\ell})x^k. \tag{11}$$

We put

$$\alpha_k = 2^{-\ell(k)}, \tag{12}$$

$$y^k = \alpha_k \Pi^g_C(z^k) + (1-\alpha_k)x^k, \tag{13}$$

$$w^k = \Pi^g_{H_k}(x^k), \tag{14}$$

where

$$H_k = \{y \in B : \langle T(y^k), y - y^k\rangle = 0\}.$$

$$x^{k+1} = \Pi^g_C(w^k). \tag{15}$$

4 Convergence analysis

We start by establishing that Korpelevich's method for $\mathrm{VIP}(T,C)$ is well defined, and proving some elementary properties.

Proposition 4.1. *Suppose that Algorithm* (8)–(15) *stops after j steps. If $g \in \mathcal{F}$ is totally convex on B and satisfies* H1 *and* H4*, then x^k generated by the algorithm is a solution of* $\mathrm{VIP}(T,C)$*.*

Proof. Assume that $x^k = \Pi^g_C(z^k)$. Using (9), we have $g'(z^k) = g'(x^k) - \beta_k T(x^k)$. Proposition 2.7(ii) entails that

$$\langle g'(z^k) - g'(x^k), z - x^k\rangle = \langle g'(z^k) - g'(\Pi^g_C(z^k)), z - \Pi^g_C(z^k)\rangle \le 0 \quad \forall z \in C,$$

which in turns implies

$$\beta_k \langle T(x^k), z - x^k\rangle \ge 0 \quad \forall z \in C.$$

Since $\beta_k > 0$, we conclude that $x^k \in S(T,C)$. □

Proposition 4.2. *Assume that T is continuous on C. If $g \in \mathcal{F}$ is totally convex and it satisfies* H1*, then the following statements hold for Algorithm* (8)–(15)*.*

i) $x^k \in C \quad \forall k \ge 0$.

ii) If H4 *is satisfied, then $\ell(k)$ is well defined (i.e. the Armijo-type search for α_k is finite).*

iii) If the Algorithm does not stop at iteration k, then $\langle T(y^k), x^k - y^k\rangle > 0$.

Proof. i) It follows from (8) and (15).

ii) Assume by contradiction that

$$\langle T(y^{\ell(k)}), x^k - \Pi_C^g(z^k)\rangle < \frac{\delta}{\beta_k} D_g(\Pi_C^g(z^k), x^k) \quad \forall \ell \geq 0. \tag{16}$$

Since T is continuous and $y^{\ell(k)} \to x^k$ as $\ell \to \infty$, we get, multiplying both sides of (16) by β_k,

$$\beta_k \langle T(x^k), x^k - \Pi_C^g(z^k)\rangle \leq \delta D_g(\Pi_C^g(z^k), x^k),$$

or equivalently,

$$\langle g'(x^k) - g'(z^k), x^k - \Pi_C^g(z^k)\rangle \leq \delta D_g(\Pi_C^g(z^k), x^k), \tag{17}$$

using H4 and (9). Applying (7) to the left side of (17), we obtain

$$D_g(\Pi_C^g(z^k), x^k) + D_g(x^k, z^k) - D_g(\Pi_C^g(z^k), z^k) \leq \delta D_g(\Pi_C^g(z^k), x^k). \tag{18}$$

Since g is strictly convex, Definition 2.4(i) and the stopping criterion imply that

$$D_g(\Pi_C^g(z^k), x^k) > 0.$$

Therefore, using (18) and the fact that $\delta \in (0, 1)$, we get

$$D_g(x^k, z^k) < D_g(\Pi_C^g(z^k), z^k),$$

which contradicts Definition 2.6, because $x^k \in C$.

iii) Combining statements (10)–(13), we get

$$\langle T(y^k), x^k - y^k\rangle = \alpha_k \langle T(y^k), x^k - \Pi_C^g(z^k)\rangle \geq \frac{\delta \alpha_k}{\beta_k} D_g(\Pi_C^g(z^k), x^k) > 0,$$

in view of the stopping criterion. □

The next proposition establishes the Fejér monotonicity property of the sequence $\{x^k\}$ generated by the algorithm with respect to $S(T, C)$.

Proposition 4.3. *Assume that T is continuous on C, that* VIP(T, C) *satisfies property* **A***, and that g satisfies* H1–H5*. Let $\{x^k\}, \{y^k\}, \{z^k\}$ be the sequences generated by Algorithm* (8)–(15)*. If the algorithm does not have finite termination, then*

i) the sequence $\{D_g(x^*, x^k)\}$ is nonincreasing (and henceforth convergent) for any $x^* \in S(T, C)$ satisfying (6)*.*

ii) The sequence $\{x^k\}$ is bounded, therefore it has weak cluster points.

iii) $\lim_{k\to\infty} \|w^k - x^k\| = 0$.

iv) The sequence $\{z^k\}$ is bounded.

v) $\lim_{k\to\infty}\langle T(y^k), x^k - y^k\rangle = 0$.

Proof. i) For each k, define $H_k^- = \{x \in B : \langle T(y^k), x-y^k\rangle \le 0\}$, $H_k = \{x \in B : \langle T(y^k), x-y^k\rangle = 0\}$, and $H_k^+ = \{x \in B : \langle T(y^k), x - y^k\rangle \ge 0\}$ where $\{y^k\}$ is the sequence generated by (13). Take $x^* \in S(T, C)$ satisfying (6), so that $x^* \in H_k^-$ for all k. On the other hand, by Proposition 4.2(iii), $x^k \in H_k^+$ and $x^k \notin H_k^-$. Therefore, Proposition 2.9 implies that

$$D_g(x^*, x^k) \ge D_g(x^*, w^k) + D_g(w^k, x^k). \tag{19}$$

By (7), Proposition 2.7(ii), and the fact that $x^{k+1} = \Pi_C^g(w^k)$, we have that

$$D_g(x^*, x^{k+1}) + D_g(x^{k+1}, w^k) - D_g(x^*, w^k) = \langle g'(x^{k+1}) - g'(w^k), x^{k+1} - x^*\rangle \le 0,$$

which implies

$$D_g(x^*, w^k) \ge D_g(x^*, x^{k+1}) + D_g(x^{k+1}, w^k). \tag{20}$$

By combining (19) and (20), we get

$$D_g(x^*, x^k) \ge D_g(x^*, x^{k+1}) + D_g(x^{k+1}, w^k) + D_g(w^k, x^k). \tag{21}$$

Since $D_g(x^{k+1}, w^k), D_g(w^k, x^k) \ge 0$, we get the result from (21).

ii) Take any $x^* \in S(T, C)$ satisfying (6). Then the result follows from H1 and the reflexivity of B.

iii) Taking limits in (21) and using (i), we obtain $\lim_{j\to\infty} D_g(w^k, x^k) = 0$, which in turns implies, using Proposition 2.10 and (ii), $\lim_{j\to\infty} \left\|w^k - x^k\right\| = 0$.

iv) Note that T, g', and $(g')^{-1}$ are continuous functions by our assumptions. On the other hand, $\{x^k\}$ is bounded by (ii). Henceforth $\{z^k\}$ is bounded, because $\{\beta_k\}$ is bounded.

v) We have that

$$0 = \langle T(y^k), w^k - y^k\rangle = \langle T(y^k), w^k - x^k\rangle + \langle T(y^k), x^k - y^k\rangle \quad \forall k,$$

since $w^k = \Pi_{H_k}^g(x^k)$ belongs to H_k, by (14) and the definition of Bregman projection. Hence,

$$\left|\langle T(y^k), x^k - y^k\rangle\right| = \left|\langle T(y^k), x^k - w^k\rangle\right| \le \left\|T(y^k)\right\|_* \left\|x^k - w^k\right\| \quad \forall k. \tag{22}$$

We remind that assumption H3 implies Fréchet differentiability of g (see Proposition 4.8 of [26]). So using Proposition 2.11, boundedness of the sequence $\{x^k\}$ and $\{z^k\}$ established in (ii) and (iv), and the fact that $\{\alpha_k\} \subset [0, 1]$, we conclude that $\{y^k\}$ is bounded, which in turns implies boundedness of the sequence $\{\left\|T(y^k)\right\|_*\}$. Now, taking limits in (22) and invoking (iii), we complete the proof of (v). □

We need now to the following concept.

Definition 4.4. *We say that* $\{x^k\}$ *is an asymptotically solving sequence for* $\mathrm{VIP}(T, C)$ *if* $0 \leq \liminf_{k\to\infty}\langle T(x^k), z - x^k\rangle$ *for each* $z \in C$.

Proposition 4.5. *Assume that* T *is continuous on* C*, that* $\mathrm{VIP}(T, C)$ *satisfies the property* **A** *and that* g *satisfies* H1–H5. *Let* $\{x^k\}$ *and* $\{z^k\}$ *be the sequences generated by Algorithm* (8)–(15). *Suppose that* $\{x^{i_k}\}$ *is a subsequence of* $\{x^k\}$ *satisfying* $\lim_{k\to\infty} D_g(\Pi_C^g(z^{i_k}), x^{i_k}) = 0$. *Then* $\{x^{i_k}\}$ *is an asymptotically solving sequence for* $\mathrm{VIP}(T, C)$.

Proof. Note that $\{x^k\}$ is bounded by Proposition 4.3(ii), and H2 is satisfied by assumption. Thus, Proposition 2.10 implies that

$$\lim_{k\to\infty} \left\|x^{i_k} - \Pi_C^g(z^{i_k})\right\| = 0. \tag{23}$$

Now we apply Proposition 2.7(ii) to obtain

$$\langle g'(z^{i_k}) - g'(\Pi_C^g(z^{i_k})), z - \Pi_C^g(z^{i_k})\rangle \leq 0 \quad \forall z \in C,$$

or equivalently, in view of (9),

$$\frac{1}{\beta_{i_k}}\langle g'(x^{i_k}) - g'(\Pi_C^g(z^{i_k})), z - \Pi_C^g(z^{i_k})\rangle \leq \langle T(x^{i_k}), z - \Pi_C^g(z^{i_k})\rangle \quad \forall z \in C,$$

which is equivalent to

$$\frac{1}{\beta_{i_k}}\langle g'(x^{i_k}) - g'(\Pi_C^g(z^{i_k})), z - \Pi_C^g(z^{i_k})\rangle + \langle T(x^{i_k}), \Pi_C^g(z^{i_k}) - x^{i_k}\rangle \leq \langle T(x^{i_k}), z - x^{i_k}\rangle \quad \forall z \in C. \tag{24}$$

Now fix $z \in C$, and let $k \to \infty$ in (24). Using H3, (23), continuity of T, the fact that $\{\beta_k\} \subset [\hat{\beta}, \tilde{\beta}]$, and the boundedness of the sequences $\{T(x^{i_k})\}$, $\{\Pi_C^g(z^{i_k})\}$ (which follow from Propositions 2.11, 4.3(iv)), we obtain

$$0 \leq \liminf_{k\to\infty}\langle T(x^{i_k}), z - x^{i_k}\rangle.$$

□

Proposition 4.6. *Assume that* T *is uniformly continuous on bounded subsets of* C*, that* $\mathrm{VIP}(T, C)$ *satisfies property* **A***, and that* g *satisfies* H1–H5. *If a subsequence* $\{\alpha_{i_k}\}$ *of the sequence* $\{\alpha_k\}$ *defined in* (12) *converges to* 0 *then* $\{x^{i_k}\}$ *is an asymptotically solving sequence for* $\mathrm{VIP}(T, C)$.

Proof. To prove this assertion, we use Proposition 4.5. Thus, we must show that

$$\lim_{k\to\infty} D_g(\Pi_C^g(z^{i_k}), x^{i_k}) = 0.$$

By contradiction, and without loss of generality, let us assume that $\lim_{k\to\infty} D_g(\Pi_C^g(z^{i_k}), x^{i_k}) = \eta > 0$. Define

$$\bar{y}^k = 2\alpha_{i_k}\Pi_C^g(z^{i_k}) + (1 - 2\alpha_{i_k})x^{i_k}$$

or equivalently

$$\bar{y}^k - x^{i_k} = 2\alpha_{i_k}[\Pi_C^g(z^{i_k}) - x^{i_k}]. \tag{25}$$

Since $\{\Pi_C^g(z^{i_k}) - x^{i_k}\}$ is bounded and $\lim_{k\to\infty} \alpha_{i_k} = 0$, it follows from (25) that

$$\lim_{k\to 0} \left\| \bar{y}^k - x^{i_k} \right\| = 0. \tag{26}$$

From (10) and definition of $\bar{y}^k$ we get

$$\langle T(\bar{y}^k), x^{i_k} - \Pi_C^g(z^{i_k})\rangle < \frac{\delta}{\beta_{i_k}} D_g(\Pi_C^g(z^{i_k}), x^{i_k})$$

for all k. Since T is uniformly continuous on bounded subsets of C and $\delta \in (0,1)$, using (26) we can find $N \in \mathbb{N}$ such that

$$\langle \beta_{i_k} T(x^{i_k}), x^{i_k} - \Pi_C^g(z^{i_k})\rangle < D_g(\Pi_C^g(z^{i_k}), x^{i_k}) \quad \forall\, k \geq N,$$

which implies, using (9),

$$\langle g'(x^{i_k}) - g'(z^{i_k}), x^{i_k} - \Pi_C^g(z^{i_k})\rangle < D_g(\Pi_C^g(z^{i_k}), x^{i_k}) \quad \forall\, k \geq N.$$

Proposition 2.8 implies that

$$D_g(\Pi_C^g(z^{i_k}), x^{i_k}) + D_g(x^{i_k}, z^{i_k}) - D_g(\Pi_C^g(z^{i_k}), z^{i_k}) < D_g(\Pi_C^g(z^{i_k}), x^{i_k}) \quad \forall\, k \geq N,$$

which is equivalent to $D_g(x^{i_k}, z^{i_k}) < D_g(\Pi_C^g(z^{i_k}), z^{i_k})$, contradicting Definition 2.6 and the fact that $x^k \in C$. □

Corollary 4.7. *Assume that T is uniformly continuous on bounded subsets of C, that* VIP(T,C) *satisfies property* **A** *and that g satisfies* H1–H5*. Then the sequence $\{x^k\}$ generated by Algorithm* (8)–(15) *is an asymptotically solving sequence for* VIP(T,C)*.*

Proof. First assume that there exists a subsequence $\{\alpha_{i_k}\}$ of $\{\alpha_k\}$ which converges to 0. In this case, we obtain $0 \leq \liminf_{k\to\infty}\langle T(x^{i_k}), z - x^{i_k}\rangle$ from Proposition 4.6. Now assume that $\{\alpha_{i_k}\}$ is any subsequence of $\{\alpha_k\}$ bounded away from zero (say $\alpha_{i_k} \geq \bar{\alpha} > 0$. It follows from (10) and (13) that

$$\langle T(y^{i_k}), x^{i_k} - y^{i_k}\rangle \geq \frac{\delta \alpha_{i_k}}{\beta_{i_k}} D_g(\Pi_C^g(z^{i_k}), x^{i_k}). \tag{27}$$

Taking limits in (27) as $k \to \infty$, and taking into account Proposition 4.3(v), we get

$$\lim_{k\to\infty} D_g(\Pi_C^g(z^{i_k}), x^{i_k}) = 0,$$

which in turns implies $0 \leq \liminf_{k\to\infty}\langle T(x^{i_k}), z - x^{i_k}\rangle$, using Proposition 4.5. □

Now we can state and prove our main convergence result.

Theorem 4.8. *Assume that T is monotone and uniformly continuous on bounded subsets of C and that g satisfies* H1–H5. *Let $\{x^k\}$ be the sequence generated by* (8)–(15). *Then*

i) $\liminf_{k\to\infty}\langle T(z), z-x^k\rangle \geq 0$ *for all $z \in C$.*

ii) $\{x^k\}$ *has weak cluster points and all of them solve* VIP(T,C).

iii) If VIP(T,C) *has a unique solution or* H6 *is satisfied, then the whole sequence $\{x^k\}$ is weakly convergent to some solution of* VIP(T,C).

Proof. i) Note that monotonicity of T implies property **A**. Take an arbitrary $z \in C$. By monotonicity of T we have that

$$\langle T(z), z-x^k\rangle \geq \langle T(x^k), z-x^k\rangle \quad \forall k. \tag{28}$$

Taking lim inf on both sides of statement (28) as $k\to\infty$, we get

$$\liminf_{k\to\infty}\langle T(z), z-x^k\rangle \geq \liminf_{k\to\infty}\langle T(x^k), z-x^k\rangle \geq 0,$$

where the rightmost inequality follows from Definition 4.4 and Corollary 4.7.

(ii) Note that $\{x^k\}$ has at least one weak cluster point by reflexivity of B and Proposition 4.3(ii). Thus, let $\bar{x}$ be any cluster point of $\{x^k\}$ and $\{x^{i_k}\}$ a subsequence of $\{x^k\}$ such that $\lim_{k\to\infty} x^{i_k} = \bar{x}$. In view of (i),

$$\langle T(z), z-\bar{x}\rangle = \lim_{k\to\infty}\langle T(z), z-x^{i_k}\rangle \geq 0,$$

for each $z \in C$. On the other hand, norm-to-norm continuity of T on C gives norm-to-weak* continuity of T on C, and hence T is hemicontinuous on C. We conclude that (ii) holds using Lemma 2.3.

(iii) If VIP(T,C) has a unique solution, then the result follows from (ii). Otherwise, assume that $\hat{x} \in C$ is another weak cluster point of $\{x^k\}$ solving VIP(T,C), and let $\{x^{\ell_k}\}$ be a subsequence of $\{x^k\}$ such that $\lim_{k\to\infty} x^{\ell_k} = \hat{x}$. By (ii), both $\bar{x}$ and $\hat{x}$ solve VIP(T,C). By Proposition 4.3(i), both $D_g(\bar{x}, x^k)$ and $D_g(\hat{x}, x^k)$ converge, say to $\eta \geq 0$ and $\mu \geq 0$, respectively. Using the definition of D_g, we have that

$$\langle g'(x^{\ell_k}) - g'(x^{i_k}), \bar{x}-\hat{x}\rangle = D_g(\bar{x}, x^{i_k}) - D_g(\bar{x}, x^{\ell_k}) + D_g(\hat{x}, x^{\ell_k}) - D_g(\hat{x}, x^{i_k}).$$

Therefore

$$\left|\langle g'(x^{\ell_k}) - g'(x^{i_k}), \bar{x}-\hat{x}\rangle\right| \leq \left|D_g(\bar{x}, x^{i_k}) - D_g(\bar{x}, x^{\ell_k})\right| + \left|D_g(\hat{x}, x^{\ell_k}) - D_g(\hat{x}, x^{i_k})\right|. \tag{29}$$

Taking limits in (29) with $k\to\infty$, we get

$$\liminf_{k\to\infty}\left|\langle g'(x^{\ell_k}) - g'(x^{i_k}), \bar{x}-\hat{x}\rangle\right| \leq |\eta-\eta| + |\mu-\mu| = 0,$$

which contradicts H6. As a result, $\tilde{x} = \hat{x}$. □

Acknowledgment

Research for this paper by the first author was partially supported by CNPq grant N^o 301280-86. The second author acknowledges his scholarship for his doctoral studies, granted jointly by CNPq and TWAS.

References

[1] Armijo, L. Minimization of functions having continuous partial derivatives. *Pacific Journal of Mathematics* **16** (1966) 1–3.

[2] Alber, Y.I. Metric and generalized projection operators in Banach spaces: properties and applications. Theory and applications of nonlinear operators of accretive and monotone type. *Lecture Notes in Pure and Applied Mathematics* **178** (1996) 15–50.

[3] Alber, Y.I. On average convergence of the iterative projection methods. *Taiwanese Journal of Mathematics* **6** (2002) 323–341.

[4] Bao, T.Q., Khanh, P.Q. A projection-type algorithm for pseudomonotone nonlipschitzian multivalued variational inequalities. *Nonconvex Optimization and Its Applications* **77** (2005) 113–129.

[5] Bao, T.Q., Khanh, P.Q. Some algorithms for solving mixed variational inequalities. *Acta Mathematica Vietnamica* **31** (2006) 83–103.

[6] Bertsekas, D.P., Tsitsiklis, J.N. *Parallel and Distributed Computation: Numerical Methods.* Prentice Hall, New Jersey (1989).

[7] Butnariu, D., Iusem, A.N. *Totally convex functions for fixed points computation and infinite dimensional optimization.* Kluwer, Dordrecht (2000).

[8] Butnariu, D., Resmerita, E. Bregman distances, totally convex functions and a method for solving operator equations in Banach spaces. *Abstract and Applied Analysis* (2006) Art. ID 84919.

[9] Cioranescu, I. *Geometry of banach spaces, duality mappings, and nonlinear problems.* Kluwer, Dordrecht (1990).

[10] Fang, S.-C. An iterative method for generalized complementarity problems. *IEEE Transactions on Automatic Control* **25** (1980) 1225–1227.

[11] He, B.S. A new method for a class of variational inequalities. *Mathematical Programming* **66** (1994) 137–144.

[12] Iiduka, H., Takahashi, W. Weak convergence of a projection algorithm for variational inequalities in a Banach space. *Journal of Mathematical Analysis and Applications* **339** (2008) 668–679.

[13] Iusem, A.N. An iterative algorithm for the variational inequality problem. *Computational and Applied Mathematics* **13** (1994) 103–114.

[14] Iusem, A.N., Gárciga Otero, R. Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces. *Numerical Functional Analysis and Optimization* **22** (2001) 609–640.

[15] Iusem, A.N., Lucambio Pérez, L.R. An extragradient-type algorithm for non-smooth variational inequalities. *Optimization* **48** (2000) 309–332.

[16] Iusem, A.N., Svaiter, B.F. A variant of Korpelevich's method for variational inequalities with a new search strategy. *Optimization* **42** (1997) 309–321.

[17] Khobotov, E.N. Modifications of the extragradient method for solving variational inequalities and certain optimization problems. *USSR Computational Mathematics and Mathematical Physics* **27** (1987) 120–127.

[18] Kinderlehrer, D., Stampacchia, G. *An Introduction to Variational Inequalities and Their Applications*. Academic Press, New York

[19] Korpelevich, G.M. The extragradient method for finding saddle points and other problems. *Ekonomika i Matematcheskie Metody* **12** (1976) 747–756.

[20] Lin, L.J., Yang, M.F., Ansari, Q.H., Kassay, G. Existence results for Stampacchia and Minty type implicit variational inequalities with multivalued maps. *Nonlinear Analysis. Theory, Methods and Applications* **61** (2005) 1–19.

[21] Marcotte, P. Application of Khobotov's algorithm to variational inequalities and network equilibrium problems. *Information Systems and Operational Research* **29** (1991) 258–270.

[22] Resmerita, E. On total convexity, Bregman projections and stability in Banach spaces. *Journal of Convex Analysis* **11** (2004) 1–16.

[23] Solodov, M.V., Svaiter, B.F. A new projection method for variational inequality problems. *SIAM Journal on Control and Optimization* **37** (1999) 765–776.

[24] Solodov, M.V., Tseng, P. Modified projection-type methods for monotone variational inequalities. *SIAM Journal on Control and Optimization* **34** (1996) 1814–1830.

[25] Takahashi, W. *Nonlinear Functional Analysis.* Yokohama Publishers, Yokohama (2000).

[26] Zeidler, E. *Nonlinear Functional Analysis and its Applications I*. Springer, New York (1985).

[27] Zhu, D.L., Marcotte, P. Convergence properties of feasible descent methods for solving variational inequalities in Banach spaces. *Computational Optimization and Applications* **10** (1998) 35–49.