A Strongly Convergent Direct Method for Monotone Variational Inequalities in Hilbert Spaces

J. Y. Bello Cruz* A. N. Iusem†

October 16, 2008

Abstract

We introduce a two-step direct method, like Korpelevich's, for solving monotone variational inequalities. The advantage of our method over that one is that ours converges strongly in Hilbert spaces, while only weak convergence has been proved for Korpelevich's algorithm. Our method also has the following desirable property: the sequence converges to the solution of the problem which lies closest to the initial iterate.

Keywords: Monotone variational inequalities, Korpelevich's method, Maximal monotone operators, Projection method, Armijo-type search, Strong convergence.
Mathematical Subject Classification (2008): 90C47, 49J35.

1 Introduction

Let C be a nonempty, closed and convex subset of a real Hilbert space $\mathcal{H}$ and $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ a point-to-set operator. The variational inequality problem for T and C, denoted VIP(T, C), is the following:

find $x^* \in C$ such that there exists $u^* \in T(x^*)$ satisfying

$$\langle u^*, x - x^* \rangle \geq 0 \qquad \forall\, x \in C.$$

We denote the solution set of this problem by $S(T, C)$.

The variational inequality problem was first introduced by P. Hartman and G. Stampacchia [6] in 1966. An excellent survey of methods for finite dimensional variational inequality problems ($\mathcal{H} = \mathbb{R}^n$) can be found in [4].

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

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

Here, we are interested in direct methods for solving VIP(T, C). The basic idea consists of extending the projected gradient method for constrained optimization, i.e., for the problem of minimizing $f(x)$ subject to $x \in C$. This problem is a particular case of VIP(T, C) taking $T = \nabla f$. This procedure is given by the following iterative scheme:

$$x^0 \in C, \tag{1}$$

$$x^{k+1} = P_C(x^k - \alpha_k \nabla f(x^k)), \tag{2}$$

with $\alpha_k > 0$ for all k. The coefficients α_k are called stepsizes and $P_C : \mathcal{H} \to C$ is the orthogonal projection onto C, i.e. $P_C(x) = \underset{y \in C}{\operatorname{argmin}} \|x - y\|$.

An immediate extension of the method (1)–(2) to VIP(T, C) for the case in which T is point-to-point, is the iterative procedure given by

$$x^0 \in C, \tag{3}$$

$$x^{k+1} = P_C(x^k - \alpha_k T(x^k)). \tag{4}$$

Convergence results for this method require some monotonicity properties of T. We introduce next several possible options.

Definition 1. *Consider $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ and $W \subset \mathcal{H}$ convex. T is said to be:*

i) monotone on W if $\langle u - v, x - y \rangle \geq 0$ for all x, $y \in W$ and all $u \in T(x)$, $v \in T(y)$,

ii) paramonotone on W if it is monotone in W, and whenever $\langle u - v, x - y \rangle = 0$ with $x, y \in W$, $u \in T(x)$, $v \in T(y)$ it holds that $u \in T(y)$ and $v \in T(x)$,

iii) uniformly monotone on W if $\langle u - v, x - y \rangle \geq \psi(\|x - y\|)$ for all x, $y \in W$ and all $u \in T(x)$, $v \in T(y)$, where $\psi : \mathbb{R}_+ \to \mathbb{R}$ is an increasing function, with $\psi(0) = 0$,

iv) strongly monotone on W if $\langle u - v, x - y \rangle \geq \omega \|x - y\|^2$ for some $\omega > 0$ and for all x, $y \in W$ and all $u \in T(x)$, $v \in T(y)$.

It follows from Definition 1 that the following implications hold: (iv) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i). The reverse assertions are not true in general.

It has been proved in [5] that when T is strongly monotone and Lipschitz continuous, i.e. there exists $L > 0$ such that $\|T(x) - T(y)\| \leq L\|x - y\|$ for all $x, y \in \mathbb{R}^n$, then the scheme (3)-(4) converges to the unique solution of VIP(T, C), provided that $\alpha_k \in \left(\epsilon, \frac{2\omega}{L^2}\right)$ for all k and for some $\epsilon > 0$.

These results were later improved for the case in which T is point-to-set, establishing convergence under weaker assumptions on T, for suitable selections of the stepsizes α_k. In this case, the iterative procedure is given by

$$x^{k+1} = P_C(x^k - \alpha_k u^k), \tag{5}$$

with $u^k \in T(x^k)$. The case of uniformly monotone operators is analyzed in [1] and the case of paramonotone ones in [2].

We remark that there is no chance to relax the assumption on T to plain monotonicity. For example, consider $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $T(x) = Ax$, with $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. T is monotone and the unique solution of VIP(T, C) is $x^* = 0$. However, it is easy to check that $\|x^k - \alpha_k T(x^k)\| > \|x^k\|$ for all $x^k \neq 0$ and all $\alpha_k > 0$, and therefore the sequence generated by (5) moves away from the solution, independently of the choice of the stepsize α_k.

In order to overcome this difficulty, the following iteration, called extragradient method, was proposed by G. M. Korpelevich in [11] for the finite dimensional case:

$$y^k = P_C\left(x^k - \beta T(x^k)\right), \tag{6}$$

$$x^{k+1} = P_C\left(x^k - \beta T(y^k)\right), \tag{7}$$

where $\beta > 0$ is a fixed number. Assuming that T is monotone and Lipschitz continuous with constant L, and that $\beta \in \left(\epsilon, \frac{1}{L}\right)$ for some $\epsilon > 0$, Korpelevich showed that the sequence $\{x^k\}$ generated by (6)-(7) converges to some point in $S(T, C)$. When T is not Lipschitz, or it is Lipschitz but the constant L is not known, the fixed stepsize β in the first step must be replaced by a stepsize computed through an Armijo-type search, as in the following method, present in [10]. The algorithm requires the following exogenous parameters: $\delta \in (0, 1)$, $\hat{\beta}$, $\tilde{\beta}$ satisfying $0 < \hat{\beta} \leq \tilde{\beta}$, and a sequence $\{\beta_k\} \subseteq [\hat{\beta}, \tilde{\beta}]$.

Initialization step. Take

$$x^0 \in C.$$

Iterative step. Given x^k define

$$z^k := x^k - \beta_k T(x^k).$$

If $x^k = P_C(z^k)$ stop. Otherwise take

$$j(k) := \min\left\{ j \geq 0 : \left\langle T(2^{-j} P_C(z^k) + (1 - 2^{-j})x^k), x^k - P_C(z^k) \right\rangle \geq \frac{\delta}{\beta_k} \|x^k - P_C(z^k)\|^2 \right\},$$

$$\alpha_k := 2^{-j(k)},$$

$$y^k := \alpha_k P_C(z_k) + (1 - \alpha_k)x^k,$$

$$H_k := \left\{ z \in \mathcal{H} : \langle z - y^k, T(y^k) \rangle \leq 0 \right\},$$

$$x^{k+1} := P_C(P_{H_k}(x^k)).$$

This strategy for determining the stepsizes guarantees convergence under the only assumptions of monotonicity and continuity of T and existence of solutions of VIP(T, C), without assuming

Lipschitz continuity of T. Also, this algorithm demands only two projections onto C per iteration, unlike other variants, e.g. [7], with projections onto C inside the inner loop for the search.

This algorithm can be extended to an infinite dimensional Banach space, achieving weak convergence under mild assumptions, see [9].

We will introduce a new Korpelevich-type algorithm with strong convergence in Hilbert spaces. It is related to the method by M. Solodov and B. Svaiter in [13], where a similar modification is performed upon the proximal point method for solving $\mathrm{VIP}(T, C)$, with the same goal, namely upgrading weak convergence to strong one. Strong convergence is forced by combining Korpelevich-type iterations with simple projection steps onto the intersection of C and two halfspaces, containing $S(T, C)$.

Additionally, our algorithm has the distinctive feature that the limit of the generated sequence is the closest solution of the problem to the initial iterate x^0. This property is useful in many specific applications, e.g. in image reconstruction. We emphasize that this feature is of interest also in finite dimension, differently from on the strong versus weak convergence issue.

We mention that the method in [13], as all proximal point algorithm in general, requires in each step the solution of a rather hard subproblem, while our method inherits from Korpelevich's an explicit nature, without subproblems to be solved, up to the projection onto the intersection of C with two half-spaces. The presence of the half-spaces does not entail any significant additional cost over the computation of the projection onto C itself. The computational cost of this projection is negligible as compared to the cost of a proximal step, for instance, and thus both Korpelevich's method and ours can be considered as direct methods for $\mathrm{VIP}(T, C)$.

We impose two additional conditions on T, besides maximal monotonicity: T must be point-to-point and uniformly continuous on bounded sets. We comment now on these assumptions. Uniform continuous on bounded sets holds automatically in finite dimension, due to the continuity of point-to-point maximal monotone operators (e.g. Theorem 4.6.3 in [3]). We also mention that it is required in the analysis of [9] for proving weak convergence of Korpelevich's method in infinite dimensional spaces. In connection with the possibility of considering point-to-set, rather than point-to-point operators, we mention that a variant of Korpelevich method for point-to-set maximal monotone operators was proposed in [8], but with the following serious limitation: in principle, we should replace $T(x^k)$ by some $u^k \in T(x^k)$ everywhere in the algorithm, but, due to the lack of inner continuity of T, an arbitrary $u^k \in T(x^k)$ does not work; u^k must satisfy some additional conditions, which are not adequate for computational implementation. In particular, the method cannot be applied to cases in which T is given by an "oracle", which provides just one $u \in T(x)$ for each x. This is a rather frequent situation for point-to-set operators. Thus, we have opted to present our strongly convergent method only for the point-to-point setting.

2 Preliminary results

In this section, we present some definitions and results that are needed for the convergence analysis of the proposed methods. First, we state two well known facts on orthogonal projections.

Lemma 1. *Let K be any nonempty closed and convex set in $\mathcal{H}$ and P_K the orthogonal projection onto K. For all $x, y \in \mathcal{H}$ and all $z \in K$, the following properties hold:*

i) $\langle z - y, z - P_K(y) \rangle \geq \|z - P_K(y)\|^2$.

ii) $\langle x - P_K(x), z - P_K(x) \rangle \leq 0$.

Proof. See Lemma 1 in [13]. □

Proposition 1. *Let $T : \mathcal{H} \to \mathcal{H}$ be a point-to-point monotone operator. If $P_K(x - \beta T(x)) = x$ for some $\beta > 0$ then $x \in S(T, K)$.*

Proof. See Proposition 2 in [10]. □

We recall now the definition of maximal monotone operators.

Definition 2. *Let $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ be a monotone operator. T is maximal monotone if $T = T'$ for all monotone $T' : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ such that $G(T) \subseteq G(T')$, where $G(T) := \{(x, u) \in \mathcal{H} \times \mathcal{H} \, : \, u \in T(x)\}$.*

We also need the following results on maximal monotone and operators.

Lemma 2. *Let $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ be a maximal monotone operator. Then*

i) $G(T)$ *is closed.*

ii) $S(T, C)$ *is closed and convex for all closed and convex* $C \subset \mathcal{H}$.

Proof. i) See Proposition 4.2.1(ii) of [3].

ii) Closedness of $S(T, C)$ follows easily from (i). Convexity of $S(T, C)$ is elementary.

□

3 An extragradient method with strong convergence

In this section, we introduce a new iterative method for solving VIP(T, C), which generates a sequence strongly convergent to some point belonging to $S(T, C)$, differently from Korpelevich's method, for which only weak convergence has been established. Of course, weak and strong convergence are only distinguishable in the infinite-dimensional setting. We assume in this section that T is maximal monotone, point-to-point, and uniformly continuous on bounded sets.

The algorithm requires the following exogenous parameters: $\delta \in (0,1)$, $\hat{\beta}$, $\tilde{\beta}$ satisfying $0 < \hat{\beta} \leq \tilde{\beta}$, and a sequence $\{\beta_k\} \subseteq [\hat{\beta}, \tilde{\beta}]$. It is defined as follows:

Algorithm A
Initialization step. Take

$$x^0 \in C.$$

Iterative step. Given x^k define

$$z^k := x^k - \beta_k T(x^k). \tag{8}$$

If $x^k = P_C(z^k)$ stop. Otherwise let,

$$j(k) := \min\left\{ j \geq 0 : \langle T(2^{-j} P_C(z^k) + (1 - 2^{-j})x^k), x^k - P_C(z^k)\rangle \geq \frac{\delta}{\beta_k} \|x^k - P_C(z^k)\|^2 \right\}, \tag{9}$$

$$\alpha_k := 2^{-j(k)}, \tag{10}$$

$$y^k := \alpha_k P_C(z_k) + (1 - \alpha_k)x^k. \tag{11}$$

Define

$$H_k := \left\{ z \in \mathcal{H} \,:\, \langle z - y^k, T(y^k)\rangle \leq 0 \right\},$$

$$W_k := \left\{ z \in \mathcal{H} \,:\, \langle z - x^k, x^0 - x^k\rangle \leq 0 \right\},$$

$$x^{k+1} := P_{H_k \cap W_k \cap C}(x^0). \tag{12}$$

3.1 Convergence analysis of Algorithm A

First, we establish that Algorithm A is well defined.

Proposition 2. *i) $x^k \in C$ for all $k \geq 0$.*

ii) $j(k)$ is well defined.

iii) $y^k \in C$ for all $k \geq 0$.

Proof. i) Follows from (12).

ii) Assume by contradiction that the minimum in (9) is not achieved. In this case, for all $\alpha > 0$, it holds that

$$\langle T(y^k(\alpha)), x^k - P_C(z^k)\rangle < \frac{\delta}{\beta_k}\|x^k - P_C(z^k)\|^2, \tag{13}$$

where $y^k(\alpha) = \alpha P_C(z^k) + (1 - \alpha)x^k$. Note that

$$\|x^k - P_C(z^k)\|^2 \leq \langle x^k - z^k, x^k - P_C(z^k)\rangle = \beta_k \langle T(x^k), x^k - P_C(z^k)\rangle \leq \delta\|x^k - P_C(z^k)\|^2,$$

using Lemma 1(i) in the first inequality, (8) in the equality and (13) in the second inequality, after taking limits with $\alpha \to \infty$, in view of the continuity of T. Since $\|x^k - P_C(z^k)\| > 0$ by the stopping criterion and $\delta \in (0,1)$, we arrive at a contradiction.

iii) Follows from (11), taking into account that $\alpha_k \in [0,1]$ for all $k \geq 0$ by (9)-(10).

□

Next, we establish some properties of Algorithm A.

Proposition 3. *For all k,*

$$\|x^{k+1} - x^0\|^2 \geq \|x^k - x^0\|^2 + \|x^{k+1} - x^k\|^2 \tag{14}$$

and

$$\|x^{k+1} - x^k\| \geq \alpha_k \frac{\delta}{\tilde{\beta}} \frac{\|x^k - P_C(z^k)\|^2}{\|T(y^k)\|}. \tag{15}$$

Proof. Since $x^{k+1} \in W_k$,

$$0 \geq \langle x^{k+1} - x^k, x^0 - x^k \rangle = \frac{1}{2}\left(\|x^{k+1} - x^k\|^2 - \|x^{k+1} - x^0\|^2 + \|x^k - x^0\|^2\right),$$

which implies (14).

Now, using that $\|x^k - P_{H_k}(x^k)\| \leq \|z - x^k\|$ for all $z \in H_k$, since $x^{k+1} \in H_k$, we have that $\|x^{k+1} - x^k\| \geq \|x^k - P_{H_k}(x^k)\|$. Since $P_{H_k}(x^k) = x^k - \langle T(y^k), x^k - y^k \rangle \dfrac{T(y^k)}{\|T(y^k)\|^2}$, we obtain that

$$\begin{aligned} \|x^{k+1} - x^k\| &\geq \|x^k - P_{H_k}(x^k)\| = \frac{\langle T(y^k), x^k - y^k \rangle}{\|T(y^k)\|} = \alpha_k \frac{\langle T(y^k), x^k - P_C(z^k) \rangle}{\|T(y^k)\|} \\ &\geq \delta \frac{\alpha_k}{\beta_k} \frac{\|x^k - P_C(z^k)\|^2}{\|T(y^k)\|} \geq \alpha_k \frac{\delta}{\tilde{\beta}} \frac{\|x^k - P_C(z^k)\|^2}{\|T(y^k)\|}, \end{aligned}$$

using (8)-(11) in the second inequality and the fact that $\beta_k \leq \tilde{\beta}$ for all k in the third one. □

Next we prove optimality of the weak cluster points of $\{x^k\}$.

Theorem 1. *Suppose that Algorithm A generates an infinite sequence $\{x^k\}$. Then either $\{x^k\}$ is bounded and each of its weak cluster points belongs to $S(T,C) \neq \emptyset$, or $S(T,C) = \emptyset$ and $\lim_{k\to\infty} \|x^k\| = \infty$.*

Proof. If $\{x^k\}$ is bounded, we obtain from (14) that the sequence $\{\|x^k - x^0\|\}$ is nondecreasing and bounded, hence convergent. By (14) again, $0 \leq \|x^{k+1} - x^k\|^2 \leq \|x^{k+1} - x^0\|^2 - \|x^k - x^0\|^2$, and we conclude that

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0. \tag{16}$$

Since $\{x^k\}$ is bounded, there exists a subsequence $\{x^{i_k}\}$ of $\{x^k\}$ that converges weakly to some x^*. It follows from (15) and (16) that

$$\lim_{k\to\infty} \alpha_k \frac{\|x^k - P_C(z^k)\|^2}{\|T(y^k)\|} = 0.$$

The sequence $\{P_C(z^k)\}$ is bounded, using boundedness of $\{x^k\}$ and (8), and (9)-(11) imply that $\{y^k\}$ is bounded. It follows from the uniform continuity of T that $\{T(y^k)\}$ is also bounded. Thus,

$$\lim_{k\to\infty} \alpha_k \left\| x^k - P_C(x^k) \right\| = 0. \tag{17}$$

We consider now two cases.

Case 1. Suppose that $\{\alpha_k\}$ does not converge to 0, i.e. there exists a subsequence $\{\alpha_{i_k}\}$ of $\{\alpha_k\}$ and some $\alpha > 0$ such that $\alpha_{i_k} \geq \alpha$ for all k. In this case, we define $w^k := P_C(z^k)$ and it follows from (17) that

$$\lim_{k\to\infty} \|x^{i_k} - w^{i_k}\| = 0. \tag{18}$$

Since T is uniformly continuous, we have

$$\lim_{k\to\infty} \|T(x^{i_k}) - T(w^{i_k})\| = 0. \tag{19}$$

Let x^* be a weak cluster point of $\{x^{i_k}\}$. By (18), it is also a weak cluster point of $\{w^{i_k}\}$. Without loss of generality, we assume that $\{x^{i_k}\}$ and $\{w^{i_k}\}$ converge weakly to x^*. Let $N_C(x)$ be the normal cone to C at $x \in C$, i.e., $N_C(x) = \{z \in \mathcal{H} : \langle x - y, z\rangle \geq 0 \quad \forall y \in C\}$. Define

$$\hat{T}(x) := T(x) + N_C(x). \tag{20}$$

It is known that $\hat{T}$, as given in (20), is maximal monotone and that $0 \in \hat{T}(x)$ if and only if $x \in S(T, C)$; see [12].

In order to prove that $x^* \in S(T, C)$, take $(x, u) \in G(\hat{T})$, so that $x \in C$ and $u \in \hat{T}(x) = T(x) + N_C(x)$, implying that $u - T(x) \in N_C(x)$. So, we have

$$\langle x - y, u - T(x)\rangle \geq 0 \quad \forall y \in C. \tag{21}$$

On the other hand, since $w^k = P_C(x^k - \beta_k T(x^k))$ and $x \in C$, it follows from Lemma 1(ii), with $K = C$ and $x = x^k - \beta_k T(x^k)$, that

$$\langle x - w^k, x^k - \beta_k T(x^k) - w^k\rangle \leq 0 \quad \forall\, x \in C \text{ and } k \geq 0. \tag{22}$$

Since β_k is positive for all k, we get from (22)

$$\left\langle x - w^k, \frac{x^k - w^k}{\beta_k} - T(x^k) \right\rangle \leq 0 \quad \forall\, x \in C \text{ and } k \geq 0. \tag{23}$$

Thus,

$$
\begin{aligned}
\langle x-w^k,u\rangle &\geq \langle x-w^k,T(x)\rangle \geq \langle x-w^k,T(x)\rangle + \left\langle x-w^k,\frac{x^k-w^k}{\beta_k}-T(x^k)\right\rangle \\
&= \langle x-w^k,T(x)-T(w^k)\rangle + \langle x-w^k,T(w^k)-T(x^k)\rangle + \left\langle x-w^k,\frac{x^k-w^k}{\beta_k}\right\rangle \\
&\geq \langle x-w^k,T(w^k)-T(x^k)\rangle + \left\langle x-w^k,\frac{x^k-w^k}{\beta_k}\right\rangle \\
&\geq -\|x-w^k\|\left(\|T(w^k)-T(x^k)\| + \frac{1}{\beta_k}\|w^k-x^k\|\right) \\
&\geq -\|x-w^k\|\left(\|T(w^k)-T(x^k)\| + \frac{1}{\hat{\beta}}\|w^k-x^k\|\right), \qquad (24)
\end{aligned}
$$

using (21) with $y=w^k$ because $w^k\in C$ in the first inequality, (23) in the second inequality, the monotonicity of T in the third one, Cauchy-Schwartz inequality in the fourth one and the fact that $\beta_k\geq\hat{\beta}>0$ for all k in the last one.

Now, using (18) and (19), we get that the subsequences $\{w^{i_k}-x^{i_k}\}$ and $\{T(w^{i_k})-T(x^{i_k})\}$ strongly converge to zero. Then, we can take limits with $k\to\infty$ in (24) over the subsequence with superindices $\{i_k\}$ and, using that $\{w^{i_k}\}$ converges weakly to x^*, we obtain that

$$\langle x-x^*,u\rangle\geq 0 \qquad \forall\,(x,u)\in G(\hat{T}). \tag{25}$$

Since $\hat{T}$ is maximal monotone, it follows from (25) that $(x^*,0)\in G(\hat{T})$ i.e. $0\in\hat{T}(x^*)=T(x^*)+N_C(x^*)$ and hence $x^*\in S(T,C)$.

Case 2. Suppose that $\lim_{k\to\infty}\alpha_k=0$. Taking

$$\hat{y}^k=2\alpha_k P_C(z^k)+(1-2\alpha_k)x^k, \tag{26}$$

it follows from the definition of $j(k)$ in (9) that

$$\langle T(\hat{y}^k),x^k-P_C(z^k)\rangle<\frac{\delta}{\beta_k}\|x^k-P_C(z^k)\|^2. \tag{27}$$

Note that $\hat{y}^k-x^k=2\alpha_k(P_C(x^k)-x^k)$ by (26). Since, as discussed above, $\{P_C(z^k)\}$ and $\{x^k\}$ are bounded, it follows from the assumption of this case that $\lim_{k\to\infty}\|\hat{y}^k-x^k\|=0$. Thus, we get from (27)

$$
\begin{aligned}
\frac{\delta}{\beta_k}\|x^k-P_C(z^k)\|^2 &> \langle T(\hat{y}^k),x^k-P_C(z^k)\rangle \\
&= \langle T(\hat{y}^k)-T(x^k),x^k-P_C(z^k)\rangle + \langle T(x^k),x^k-P_C(z^k)\rangle \\
&= \langle T(\hat{y}^k)-T(x^k),x^k-P_C(z^k)\rangle + \frac{1}{\beta_k}\langle x^k-z^k,x^k-P_C(z^k)\rangle \\
&\geq -\|T(\hat{y}^k)-T(x^k)\|\|x^k-P_C(x^k)\| + \frac{1}{\beta_k}\|x^k-P_C(x^k)\|^2,
\end{aligned}
$$

using (8) in the second equality, and Cauchy-Schwartz inequality and Lemma 1(i) in the second inequality. Now, an elementary rearrangement yields

$$\|T(\hat{y}^k) - T(x^k)\|\|x^k - P_C(x^k)\| \geq \frac{(1-\delta)}{\beta_k}\|x^k - P_C(x^k)\|^2 \geq \frac{(1-\delta)}{\tilde{\beta}}\|x^k - P_C(x^k)\|^2, \tag{28}$$

using the fact that $\beta_k \leq \tilde{\beta}$ for all k in the second inequality. Since $\|x^k - P_C(x^k)\| > 0$ for all k because $x^k \notin S(T,C)$ for all k in view of Proposition 1, it follows from (28) that

$$\|T(\hat{y}^k) - T(x^k)\| \geq \frac{(1-\delta)}{\tilde{\beta}}\|x^k - P_C(x^k)\| \geq 0. \tag{29}$$

Since $\lim_{k\to\infty} \|\hat{y}^k - x^k\| = 0$ and T is uniformly continuous on bounded sets, we obtain

$$\lim_{k\to\infty} \|T(\hat{y}^k) - T(x^k)\| = 0. \tag{30}$$

Taking limits with $k \to \infty$ in (29) and using (30), we get $0 \geq \lim_{k\to\infty} \|x^k - P_C(x^k)\| \geq 0$. Therefore, $\lim_{k\to\infty} \|x^k - P_C(x^k)\| = \lim_{k\to\infty} \|x^k - w^k\| = 0$. From here on, we can proceed as in the previous case, from (18) on, taking the whole sequences $\{x^k\}$, $\{w^k\}$ instead of $\{x^{i_k}\}$, $\{w^{i_k}\}$, in order to complete the proof of the first assertion.

Suppose now that $S(T,C) = \emptyset$. Using the preceding assertion in this proposition, we obtain that $\{x^k\}$ is unbounded. Since the sequence $\{\|x^k - x^0\|\}$ is nondecreasing by (14), it follows that $\lim_{k\to\infty} \|x^k - x^0\| = \infty$ and so $\lim_{k\to\infty} \|x^k\| = \infty$. □

We assume from now on that $S(T,C)$ is nonempty. Define

$$Y := \{x \in \mathcal{H} \,:\, \langle z - x, x^0 - x\rangle \leq 0 \quad \forall z \in S(T,C)\,\}. \tag{31}$$

Next we show that the generated sequence $\{x^k\}$ is contained in Y.

Proposition 4. *If $x^k \in Y$ then*

i) $S(T,C) \subseteq H_k \cap W_k \cap C$,

ii) x^{k+1} is well defined and $x^{k+1} \in Y$.

Proof. i) Note that

$$\begin{aligned}\langle T(y^k), x^* - y^k\rangle &= \langle T(y^k) - T(x^*), x^* - y^k\rangle + \langle T(x^*), x^* - y^k\rangle \\ &\leq \langle T(x^*), x^* - y^k\rangle \leq 0,\end{aligned} \tag{32}$$

for any $x^* \in S(T,C)$, using the monotonicity of T in the first inequality, and the definition of $S(T,C)$ together with Proposition 2(iii) in the second inequality. It follows from (32) that $S(T,C) \subseteq H_k$.

Since $x^k \in Y$, we have that $\langle x^* - x^k, x^0 - x^k\rangle \leq 0$ for all $x^* \in S(T,C)$. By definition of W_k, we obtain that $S(T,C) \subseteq W_k$. We conclude that $S(T,C) \subseteq H_k \cap W_k \cap C$.

ii) Since $S(T,C) \subseteq H_k \cap W_k \cap C$ and $S(T,C)$ is nonempty, it follows that $H_k \cap W_k \cap C$ is nonempty. Thus the next iterate x^{k+1} is well defined, in view of (12). By Lemma 1(ii), we have that

$$\langle z - x^{k+1}, x^0 - x^{k+1}\rangle \leq 0 \quad \forall z \in H_k \cap W_k \cap C. \tag{33}$$

Since $S(T,C) \subseteq H_k \cap W_k \cap C$ for all k, (33) holds for all $z \in S(T,C)$, and so $x^{k+1} \in Y$ by (31).

□

Corollary 1. *Algorithm A is well defined and generates infinite sequences $\{x^k\}$, $\{y^k\}$ and $\{u^k\}$ such that $\{x^k\} \subset Y$ and $S(T,C) \subseteq H_k \cap W_k \cap C$ for all k.*

Proof. It is enough to observe that $x^0 \in Y$ and apply inductively Proposition 4.

□

Corollary 2. *The sequence $\{x^k\}$ generated by Algorithm A is bounded and each of its weak cluster points belong to $S(T,C)$.*

Proof. If the solution set is nonempty, in view of (12) we have that $\|x^{k+1} - x^0\| \leq \|z - x^0\|$ for all $z \in H_k \cap W_k \cap C$. Since $S(T,C) \subseteq H_k \cap W_k \cap C$ by Corollary 1, it follows that $\|x^{k+1} - x^0\| \leq \|x^* - x^0\|$ for all $x^* \in S(T,C)$. Thus, $\{x^k\}$ is bounded, and by Theorem 1, all its weak cluster points belong to $S(T,C)$. □

Finally, we can now state and prove our main result.

Theorem 2. *Assume that $S(T,C) \neq \emptyset$ and let $\{x^k\}$ be a sequence generated by Algorithm A. Define $x^* = P_{S(T,C)}(x^0)$. Then $\{x^k\}$ converges strongly to x^*.*

Proof. Note that x^*, the orthogonal projection of x^0 onto $S(T,C)$, exists because the solution set $S(T,C)$ is nonempty by assumption, and closed and convex by Lemma 2(ii). By the definition of x^{k+1}, we have that

$$\|x^{k+1} - x^0\| \leq \|z - x^0\| \qquad \forall\, z \in H_k \cap W_k \cap C. \tag{34}$$

Since $x^* \in S(T,C) \subseteq H_k \cap W_k \cap C$ for all k, it follows from (34) that

$$\|x^k - x^0\| \leq \|x^* - x^0\| \tag{35}$$

for all k. By Corollary 2, $\{x^k\}$ is bounded and each of its weak cluster points belongs to $S(T,C)$. Let $\{x^{i_k}\}$ be any weakly convergent subsequence of $\{x^k\}$, and let $\hat{x} \in S(T,C)$ be its weak limit. Observe that

$$\begin{aligned}
\|x^{i_k} - x^*\|^2 &= \|x^{i_k} - x^0 - (x^* - x^0)\|^2 \\
&= \|x^{i_k} - x^0\|^2 + \|x^* - x^0\|^2 - 2\langle x^{i_k} - x^0, x^* - x^0\rangle \\
&\leq 2\|x^* - x^0\|^2 - 2\langle x^{i_k} - x^0, x^* - x^0\rangle,
\end{aligned}$$

where the inequality follows from (35). By the weak convergence of $\{x^{i_k}\}$ to $\hat{x}$, we obtain

$$\limsup_{k\to\infty} \|x^{i_k} - x^*\|^2 \leq 2(\|x^* - x^0\|^2 - \langle \hat{x} - x^0, x^* - x^0 \rangle). \tag{36}$$

Applying Lemma 1(ii) with $K = S(T,C)$, $x = x^0$ and $z = \hat{x} \in S(T,C)$, and taking into account that x^* is the projection of x^0 onto $S(T,C)$, we have that

$$\langle x^0 - x^*, \hat{x} - x^* \rangle \leq 0. \tag{37}$$

Now, using (37) we have

$$\begin{aligned} 0 &\geq -\langle \hat{x} - x^*, x^* - x^0 \rangle = -\langle \hat{x} - x^0, x^* - x^0 \rangle - \langle x^0 - x^*, x^* - x^0 \rangle \\ &\geq -\langle \hat{x} - x^0, x^* - x^0 \rangle + \|x^* - x^0\|^2. \end{aligned}$$

It follows that

$$\langle \hat{x} - x^0, x^* - x^0 \rangle \geq \|x^* - x^0\|^2. \tag{38}$$

Combining (38) with (36), we conclude that $\{x^{i_k}\}$ converges strongly to x^*. Thus, we have shown that every weakly convergent subsequence of $\{x^k\}$ converges strongly to x^*. Hence, the whole sequence $\{x^k\}$ converges strongly to $x^* \in S(T,C)$. □

Acknowledgment

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

References

[1] Alber, Ya.I. Recurrence relations and variational inequalities. *Soviet Mathematicheskie Doklady* **27** (1983) 511–517.

[2] Bello Cruz, J.Y., Iusem, A.N. Convergence of direct methods for paramonotone variational inequalities (to be published).

[3] Burachik, R.S., Iusem, A.N. *Set-Valued Mappings and Enlargements of Monotone Operators*. Springer, Berlin (2007).

[4] Facchinei, F., Pang, J.S. *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer, Berlin (2003).

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

[6] Hartman, P. and Stampacchia, G. On some non-linear elliptic differential-functional equations. *Acta Mathematica* **115** (1966) 271-310.

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

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

[9] Iusem, A.N., Nasri, M. Korpelevich's method for variational inequality problems in Banach spaces (to be published).

[10] 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; Addendum *Optimization* **43** (1998) 85.

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

[12] Rockafellar, R.T. On the maximality of sums of nonlinear monotone operators. *Transactions of the American Mathematical Society* **149** (1970) 75-88.

[13] Solodov, M.V., Svaiter, B.F. Forcing strong convergence of proximal point iterations in a Hilbert space. *Mathematical Programming* **87** (2000) 189-202.