Full convergence of an approximate projections method for nonsmooth variational inequalities

J. Y. Bello Cruz* A. N. Iusem†

February 2, 2010

Abstract

We analyze an explicit method for solving nonsmooth variational inequality problems, establishing convergence of the whole sequence, under paramonotonicity of the operator. Previous results on similar methods required much more demanding assumptions, like coerciveness of the operator.

Keywords: Monotone variational inequalities, Maximal monotone operators, Projection method.
Mathematical Subject Classification (2008): 90C47, 49J35.

1 Introduction

Let C be a nonempty, closed and convex subset of $\mathbb{R}^n$ and $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ a point-to-set operator. The variational inequality problem for T and C, denoted $\mathrm{VIP}(T, C)$, 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 [12] in 1966. An excellent survey of methods for finite dimensional variational inequality problems can be found in [9].

Here, we are interested in direct methods for solving $\mathrm{VIP}(T, C)$. They are called direct because the solution of subproblems at each iteration is not required. Iterate x^{k+1} is computed using only

*Universidade Federal de Goiás, IME, CEP 74001-970 GO, Goiânia, 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

information on the previous point x^k and easy computations. 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 $P_C : \mathbb{R}^n \to C$ is the orthogonal projection onto C, i.e. $P_C(x) = \operatorname*{argmin}_{y \in C} \|x - y\|$.

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

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

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

where $u^k \in T(x^k)$, and the positive sequence α_k satisfies some conditions.

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

Definition 1. *Consider $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ and $W \subset \mathbb{R}^n$ 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) strictly monotone on W if $\langle u - v, x - y\rangle > 0$ for all x, $y \in W$ such that $x \neq y$, and all $u \in T(x)$, $v \in T(y)$,

iv) 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 $\psi(0) = 0$,

v) 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: (v) $\Rightarrow$ (iv) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i). The reverse assertions are not true in general.

Convergence of the scheme (3)-(4), is established in [1] assuming uniform monotonicity of T, and in [3] assuming paramonotonicity of T.

We remark that there is no chance to relax the assumption on T to plain monotonicity, to case one-step iteration. 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 $\mathrm{VIP}(T, C)$ 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 (4) moves away from the solution, independently of the choice of the stepsize α_k.

Thus, the scheme (3)-(4) fails to converges for arbitrary monotone operators. In such a case, an available option is Korpelevich's method and its variants, which perform a double-step iteration of the form:

$$y^k = P_C(x^k - \alpha_k T(x^k)) \tag{5}$$

$$x^{k+1} = P_C(x^k - \gamma_k T(y^k)), \tag{6}$$

where y^k is an auxiliary point. See, e.g. [13],[16],[17],[18].

In this paper we will deal with one-step iterations, and thus we keep the paramonotonicity assumption. We comment next on this assumption.

The notion of paramonotonicity, which is in-between monotonicity and strict monotonicity, was introduced in [5], and many of its properties were established in [8] and [14]. Among them, we mention the following:

i) If T is the subdifferential of a convex function, then T is paramonotone; see Proposition 2.2 in [14].

ii) If $T : \mathbb{R}^n \to \mathbb{R}^n$ is monotone and differentiable, and $J_T(x)$ denotes the Jacobian matrix of T at x, then T is paramonotone if and only if $\mathrm{Rank}(J_T(x) + J_T(x)^t) = \mathrm{Rank}(J_T(x))$ for all x; see Proposition 4.2 in [14].

It follows that affine operators of the form $T(x) = Ax + b$ are paramonotone when A is positive semidefinite (not necessarily symmetric), and $\mathrm{Rank}(A + A^t) = \mathrm{Rank}(A)$. This situation includes cases of nonsymmetric and singular matrices, where $S(T, \mathbb{R}^n)$ can be a subspace, differently from the case of strictly or strongly monotone operators, for which $S(T, C)$ is always a singleton, when nonempty. Of course, this can happen also for nonlinear operators.

1.1 Relaxed projection methods

The method given by (3)-(4) is fully direct only in a few specific instances, namely when P_C is given by an explicit formula (e.g. when C is a halfspace, or a ball, or a subspace). When C is a general closed convex set, however, one has to solve the problem $\min\{\|x - (x^k - \alpha_k T(x^k))\| : x \in C\}$, in order to compute the projection onto C.

One option for avoiding this difficulty consists of replacing at iteration k P_C by P_{C_k}, where C_k is a halfspace containing the given set C and not x^k. For variational inequality problems, this approach was introduced by M. Fukushima in [11].

Observe that projections onto halfspaces are easily computable. We consider the case in which C is of the form

$$C = \{z \in \mathbb{R}^n \,:\, g(z) \leq 0\}, \tag{7}$$

where $g : \mathbb{R}^n \to \mathbb{R}$ is a convex function. The differentiability of g is not assumed and the representation (7) is therefore rather general, because any system of inequalities $g_j(x) \leq 0$ with $j \in J$, where all the g_j's are convex, may be represented as in (7) with $g(x) = \sup\{g_j(x) \,:\, j \in J\}$.

An explicit method for solving VIP(T, C) was studied in [3], using the following relaxed iteration:

$$x^{k+1} = P_{C_k}\left(x^k - \frac{\beta_k}{\eta_k}u^k\right), \tag{8}$$

where $\eta_k = \max\{1, \|u^k\|\}$, β_k is an exogenous stepsize satisfying

$$\sum_{k=0}^{\infty} \beta_k = \infty. \tag{9}$$

$$\sum_{k=0}^{\infty} \beta_k^2 < \infty, \tag{10}$$

and C_k is defined as

$$C_k := \{z \in \mathbb{R}^n \,:\, g(x^k) + \langle v^k, z - x^k\rangle \leq 0\},$$

with $v^k \in \partial g(x^k)$, where $\partial g(x^k)$ is the subdifferential of g at x^k.

It was proved in [3] that the sequence generated by (8) is bounded, the difference between consecutive iterates converges to zero, and all its cluster points belong to $S(T, C)$, under a quite demanding assumption, besides paramonotonicity; namely T must satisfy the following coerciveness condition:

(Q) There exist $z \in C$ and a bounded set $D \subseteq \mathbb{R}^n$ such that $\langle u, x - z\rangle \geq 0$ for all $x \notin D$ and for all $u \in T(x)$.

In this paper we will analyze a new algorithm relaxing the hypotheses in [3]. We do not need any coerciveness condition. Also, we obtain convergence of the whole sequence to some the solution of VIP(T, C), assuming only existence of solutions.

We describe next our method. We construct the main sequence $\{x^k\}$ as follows: we perform a finite inner loop starting at the current iterate x^k, consisting of projections onto suitable hyperplanes containing C, until a point $\tilde{y}^k$ is obtained, whose distance to C is smaller than a certain multiple of the current exogenous steplength β_k. After this inner loop, a step is taken from $\tilde{y}^k$ in the opposite direction to $\tilde{u}^k \in T(\tilde{y}^k)$ with an exogenous steplength related to β_k, and the resulting point is projected onto another auxiliary hyperplane containing C, thus obtaining the next main iterate x^{k+1}. The inner loop of projections onto hyperplanes hence substitutes for the exact projection onto C, demanded in the exact algorithm given by (3)-(4).

A related inner loop has been proposed in [4], combined with a two-step strategy like in (5)-(6), for solving point-to-point monotone variational inequality problems, thus relaxing the paramonotonicity assumption. We emphasize that the method proposed in this paper allows for point-to-set operators.

There are just a few options for the case in which T is point-to-set and just monotone. Two of them can be found in [2] and [15], but these methods are not easily implementable, and cannot be considered direct or explicit methods.

2 Preliminary results

In this section, we present some definitions and results that are needed for the convergence analysis of the proposed method.

Definition 2. *Let S be a nonempty subset of $\mathbb{R}^n$. A sequence $\{x^k\}$ in $\mathbb{R}^n$ is said to be quasi-Fejér convergent to S if and only if for all $x \in S$ there exist $k_0 \geq 0$ and a sequence $\{\delta_k\} \subset \mathbb{R}_+$ such that $\sum_{k=0}^{\infty} \delta_k < \infty$ and $\|x^{k+1} - x\|^2 \leq \|x^k - x\|^2 + \delta_k$ for all $k \geq k_0$.*

Proposition 1. *If $\{x^k\}$ is quasi-Fejér convergent to S then:*

i) $\{x^k\}$ is bounded,

ii) if a cluster point x^* of $\{x^k\}$ belongs to S, then the whole sequence $\{x^k\}$ converges to x^*.

Proof. See Theorem 1 in [6]. □

It is convenient to introduce the following notation: let $g : \mathbb{R}^n \to \mathbb{R}$ be a convex function, and X a nonempty, compact and convex subset of $\mathbb{R}^n$. Given a point $x \in X$ and $v \in \partial g(x)$, the solution of the problem

$$\min\{\|z - x\| \ : \ g(x) + \langle v, z - x \rangle \leq 0 \ , \ z \in X\}$$

is denoted by $\tilde{z}(x, v)$. Let $C = \{z \in \mathbb{R}^n \ : \ g(z) \leq 0\}$.

Lemma 1. *There exists $\tilde{\alpha} \in [0, 1)$ such that $\operatorname{dist}(\tilde{z}(x, v), C) \leq \tilde{\alpha} \operatorname{dist}(x, C)$ for all $x \in X \setminus C$ and for all $v \in \partial g(x)$, where $\operatorname{dist}(x, C) = \min_{y \in C} \|x - y\|$.*

Proof. See Lemma 4 in [10]. □

Now, we state two well known facts on orthogonal projections.

Lemma 2. *Let K be any nonempty closed and convex set in $\mathbb{R}^n$ and P_K the orthogonal projection onto K. For all $x, y \in \mathbb{R}^n$ and all $z \in K$, the following properties hold:*

i) $\|P_K(x) - P_K(y)\|^2 \leq \|x-y\|^2 - \|(P_K(x) - x) - (P_K(y) - y)\|^2$.

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

Proof. See Lemma 4.1 in [15]. □

We recall now the definition of maximal monotone operators.

Definition 3. *Let $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ be a monotone operator. T is maximal monotone if $T = T'$ for all monotone $T' : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ such that $G(T) \subseteq G(T')$, where $G(T) := \{(x, u) \in \mathbb{R}^n \times \mathbb{R}^n : u \in T(x)\}$.*

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

Lemma 3. *Let $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ be a maximal monotone operator. Then*

i) T is locally bounded at any point in the interior of its domain.

ii) $G(T)$ is closed.

iii) T is bounded on bounded subsets of the interior of its domain.

Proof. i) See Theorem 4.6.1(ii) of [7].

ii) See Proposition 4.2.1(ii) of [7].

iii) It follows easily from (i).

□

Proposition 2. *Let T be a paramonotone operator in C. Take $x \in S(T, C)$ and $x^* \in C$. If there exists $u^* \in T(x^*)$ such that $\langle u^*, x^* - x\rangle = 0$ then x^* is also solution of VIP(T, C).*

Proof. See Proposition 13 in [8]. □

Lemma 4. *Let T be a maximal monotone and paramonotone operator. Let $\{(x^k, u^k)\} \subset G(T)$ be a bounded sequence such that all cluster points of $\{x^k\}$ belong to C. For each $x \in S(T, C)$ define $\gamma_k(x) := \langle u^k, x^k - x\rangle$. If for some $x \in S(T, C)$ there exists a subsequence $\{\gamma_{j_k}(x)\}$ of $\{\gamma_k(x)\}$ such that $\lim_{k\to\infty} \gamma_{j_k}(x) \leq 0$, then there exists a cluster point of $\{x^{j_k}\}$ belonging to $S(T, C)$.*

Proof. See Lemma 6 of [3]. $\square$

The next lemma provides a computable upper bound for the distance from a point to the feasible set C.

Lemma 5. *Let $g : \mathbb{R}^n \to \mathbb{R}$ be a convex function and $C := \{z \in \mathbb{R}^n : g(z) \le 0\}$. Assume that there exists $w \in C$ such that $g(w) < 0$. Then, for all x such that $g(x) > 0$, we have*

$$\operatorname{dist}(x, C) \le \frac{\|x - w\|}{g(x) - g(w)}\, g(x)\,.$$

Proof. Take $x_\lambda := \lambda y + (1 - \lambda)x$ with $\lambda := \frac{g(x)}{g(x)-g(y)}$. Note that $\lambda \in (0, 1)$. Then

$$g(x_\lambda) = g(\lambda y + (1 - \lambda)x) \le \lambda g(y) + (1 - \lambda)g(x) = g(x) - \lambda(g(x) - g(y)) = 0.$$

Thus, $x_\lambda \in C$ and

$$\operatorname{dist}(x, C) \le \|x - x_\lambda\| = \|x - (\lambda y + (1 - \lambda)x)\| = \lambda\|x - y\| = \frac{g(x)}{g(x) - g(y)}\|x - y\|.$$

$\square$

3 A relaxed projection algorithm

We introduce an algorithm which replaces projections onto the feasible set by easily computable projections onto suitable hyperplanes. We assume that the operator T is point-to-set, maximal monotone and paramonotone.

We assume also that C is of the form given in (7), which we repeat here:

$$C = \{z \in \mathcal{H} : g(z) \le 0\}, \tag{11}$$

where $g : \mathcal{H} \to \mathbb{R}$ is a convex function, and that a Slater point is available, i.e. we will explicitly use a point w such that $g(w) < 0$.

The algorithm presented here has higher computational demands than the algorithm introduced in [3], basically the inner loop of projections onto separating hyperplanes, but as a compensation we obtain better convergence results, namely convergence of the whole sequence, and we do not assume any coercivity condition.

3.1 Statement of Algorithm

Consider an exogenous sequence $\{\beta_k\} \subseteq \mathbb{R}_{++}$ satisfying

$$\sum_{k=0}^{\infty} \beta_k = \infty, \tag{12}$$

$$\sum_{k=0}^{\infty} \beta_k^2 < \infty. \tag{13}$$

The algorithm is defined as follows.

Algorithm A

Initialization step: Take

$$x^0 \in \mathbb{R}^n.$$

Iterative step: Given x^k, if $g(x^k) \leq 0$ then take $\tilde{y}^k := x^k$. Else, perform the following inner loop, generating points $y^{k,0}$, $y^{k,1}$, ... Take $y^{k,0} = x^k$, $\theta > 0$ and choose $v^{k,0} \in \partial g(y^{k,0})$. For $j = 0, 1, \dots,$ let

$$C_{k,j} := \{z \in \mathbb{R}^n \, : \, g(y^{k,j}) + \langle v^{k,j}, z - y^{k,j} \rangle \leq 0\}, \tag{14}$$

with $v^{k,j} \in \partial g(y^{k,j})$. Define

$$y^{k,j+1} := P_{C_{k,j}}(y^{k,j}). \tag{15}$$

Stop the inner loop when $j = j(k)$, defined as

$$j(k) := \min \left\{ j \geq 0 : \frac{g(y^{k,j}) \, \|y^{k,j} - w\|}{g(y^{k,j}) - g(w)} \leq \theta \beta_k \right\}. \tag{16}$$

Let

$$\tilde{y}^k = y^{k,j(k)}. \tag{17}$$

Take $\tilde{u}^k \in T(\tilde{y}^k)$ and defined $\eta_k := \max\{1, \|\tilde{u}^k\|\}$ and $C_k := C_{k,j(k)}$. Compute

$$x^{k+1} = P_{C_k} \left(\tilde{y}^k - \frac{\beta_k}{\eta_k} \tilde{u}^k \right), \tag{18}$$

with β_k satisfying (12)-(13).

If $x^{k+1} = \tilde{y}^k$ then stop.

Unlike other projection methods, Algorithm A generates a sequence $\{x^k\}$ which is not necessarily contained in the set C. As will be shown in the next subsection, the generated sequence is asymptotically feasible and, in fact, converges to some solution of VIP(T, C).

Algorithm A can be easily implemented, because $P_{C_{k,j}}$ and P_{C_k} have explicit formulae, which we present next.

Proposition 3. *Define* $C_x := \{z \in \mathbb{R}^n : g(x) + \langle v, z - x\rangle \le 0\}$ *with* $v \in \partial g(x)$. *Then for any* $y \in \mathbb{R}^n$,

$$P_{C_x}(y) = \begin{cases} y - \dfrac{g(x) + \langle v, y - x\rangle}{\|v\|^2} v & \text{if} \quad g(x) + \langle x, y - x\rangle > 0 \\ y & \text{if} \quad g(x) + \langle v, y - x\rangle \le 0. \end{cases}$$

Proof. See Proposition 3.1 in [19]. □

It follows from Proposition 3, (14), and (18) that

$$y^{k,j+1} = P_{C_{k,j}}(y^{k,j}) = y^{k,j} - \frac{1}{\|v^{k,j}\|^2} \max\left\{0, g(y^{k,j})\right\} v^{k,j},$$

and

$$x^{k+1} = P_{C_k}\left(\tilde{y}^k - \frac{\beta_k}{\eta_k}\tilde{u}^k\right) = \tilde{y}^k - \frac{\beta_k}{\eta_k}\tilde{u}^k - \frac{1}{\|v^k\|^2} \max\left\{0, g(\tilde{y}^k) - \frac{\beta_k}{\eta_k}\langle \tilde{u}^k, \tilde{v}^k\rangle\right\} \tilde{v}^k,$$

so that Algorithm A can be considered as a fully direct method for VIP(T, C).

The iteration formulae of the algorithm become more explicit in the smooth case, i.e. when C is of the form $C = \{z \in \mathbb{R}^n : g_i(z) \le 0,\ 1 \le i \le m\}$ where the g_i's are convex and differentiable. The set C can be rewritten in our notation with $g(x) = \max_{1\le i\le m}\{g_i(x)\}$. In this situation, the well known formula for the subdifferential of the maximum of convex functions allows us to take

$$v^{k,j} = \nabla g_{\ell(k,j)}(y^{k,j}), \qquad \text{with} \quad \ell(k,j) \in \operatorname{argmax}_{0\le i\le m}\{g_i(y^{k,j})\},$$

so that the hyperplane onto which each inner-loop iterate is projected is the first order approximation of the most violated constraint at that iterate.

Observe that $\partial g(x) \neq \emptyset$ for all $x \in \mathbb{R}^n$, because we assume that g is convex and $\mathrm{dom}(g) = \mathbb{R}^n$.

3.2 Convergence analysis of Algorithm A

Before establishing convergence of Algorithm A, we need to ascertain the validity of the stopping criterion and the fact that the algorithm is well defined.

Proposition 4. *Take* C, $C_{k,j}$, $\tilde{y}^k$ *and* x^k *defined by Algorithm A. Then,*

i) $C \subseteq C_{k,j}$ *for all* k *and for all* j.

ii) If $x^{k+1} = \tilde{y}^k$ *for some* k, *then* $\tilde{y}^k \in S(T, C)$.

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

Proof. i) It follows from (14) and the definition of subgradient.

ii) Suppose that $x^{k+1} = \tilde{y}^k$. Then , since $x^{k+1} \in C_k$, we have $g(\tilde{y}^k)+\langle \tilde{v}^k, x^{k+1}-\tilde{y}^k\rangle = g(x^k) \leq 0$, i.e. $\tilde{y}^k \in C$. Moreover, since x^{k+1} is given by (18), using Lemma 2(ii) with $x = \tilde{y}^k - \frac{\beta_k}{\eta_k}\tilde{u}^k$ and $K = C_k$, we obtain

$$\left\langle x^{k+1} - \left(\tilde{y}^k - \frac{\beta_k}{\eta_k}\tilde{u}^k\right), z - x^{k+1}\right\rangle \geq 0 \quad \forall z \in C_k. \tag{19}$$

Taking $x^{k+1} = \tilde{y}^k$ in (19) and taking into account the facts that $\beta_k > 0$, $\eta_k \geq 1$ for all k, and $C \subseteq C_k$, we get $\langle \tilde{u}^k, z - \tilde{y}^k\rangle \geq 0$ for all $z \in C$. Since $\tilde{u}^k \in \tilde{u}^k$, we conclude that $\tilde{y}^k \in S(T, C)$.

iii) Assume by contradiction that $\dfrac{g(y^{k,j})\,\|y^{k,j}-w\|}{g(y^{k,j})-g(w)} > \theta\beta_k$ for all j. Thus, we get an infinite sequence $\{y^{k,j}\}_{j=0}^{\infty}$ such that

$$\lim_{j\to\infty} \frac{g(y^{k,j})\,\|y^{k,j}-w\|}{g(y^{k,j})-g(w)} \geq \theta\beta_k > 0. \tag{20}$$

Taking into account the inner loop in j given in (16)-(15), i.e. $y^{k,j+1} = P_{C_{k,j}}(y^{k,j})$ for each k, we obtain, for each $x \in C$,

$$\begin{aligned} \|y^{k,j+1}-x\|^2 &= \|P_{C_{k,j}}(y^{k,j}) - P_{C_{k,j}}(x)\|^2 \leq \|y^{k,j}-x\|^2 - \|y^{k,j+1}-y^{k,j}\|^2 \\ &\leq \|y^{k,j}-x\|^2, \end{aligned} \tag{21}$$

using Lemma 2(i) with $x = y^{k,j}$, $y = x$ and $K = C_{k,j}$. Thus, $\{y^{k,j}\}_{j=0}^{\infty}$ is quasi-Fejér convergent to C, and hence it is bounded by Proposition 1(i). It follows that $\tau := \dfrac{1}{-g(w)} \sup\limits_{0\leq j\leq\infty} \|y^{k,j} - w\|$ is finite and also,

$$g(y^{k,j}) > 0 \qquad \forall\, j. \tag{22}$$

Using (21), we get

$$\lim_{j\to\infty} \|y^{k,j+1}-y^{k,j}\| = 0. \tag{23}$$

Since $y^{k,j+1}$ belongs to $C_{k,j}$, we have from (14) that

$$g(y^{k,j}) \leq \langle v^{k,j}, y^{k,j}-y^{k,j+1}\rangle \leq \|v^{k,j}\|\|y^{k,j}-y^{k,j+1}\|, \tag{24}$$

using Cauchy-Schwartz inequality in the last inequality.

Since $\{y^{k,j}\}_{j=0}^{\infty}$ is bounded and the subdifferential of g is bounded on bounded sets, we obtain that $\{\|v^{k,j}\|\}_{j=0}^{\infty}$ is bounded. In view of (23) and (24),

$$\lim_{j\to\infty} g(y^{k,j}) \leq 0. \tag{25}$$

It follows from (22) and (25) that

$$\begin{aligned}
\lim_{j\to\infty} \frac{g(y^{k,j})\,\|y^{k,j}-w\|}{g(y^{k,j})-g(w)} &\leq \lim_{j\to\infty} \frac{g(y^{k,j})\,\|y^{k,j}-w\|}{-g(w)} \\
&\leq \frac{1}{-g(w)} \sup_{0\leq j\leq\infty} \|y^{k,j}-w\| \lim_{j\to\infty} g(y^{k,j}) \\
&= \tau \lim_{j\to\infty} g(y^{k,j}) \leq 0,
\end{aligned}$$

contradicting (20). It follows that $j(k)$ is well defined.

□

We continue by proving the quasi-Fejér properties of the sequences $\{x^k\}$ and $\{\tilde{y}^k\}$ generated by Algorithm A.

Proposition 5. *If $S(T,C)$ is nonempty, then $\{\tilde{y}^k\}$ and $\{x^k\}$ are quasi-Fejér convergent to $S(T,C)$.*

Proof. Observe that $\eta_k \geq \|\tilde{u}^k\|$ and $\eta_k \geq 1$ for all k by the definition of η_k. Then, for all k,

$$\frac{1}{\eta_k} \leq 1 \tag{26}$$

and

$$\frac{\|\tilde{u}^k\|}{\eta_k} \leq 1. \tag{27}$$

Take $\bar{x} \in S(T,C)$. Thus, there exists $\bar{u} \in T(\bar{x})$ such that

$$\langle \bar{u}, x-\bar{x}\rangle \geq 0 \quad \forall\, x \in C. \tag{28}$$

First note that,

$$\begin{aligned}
\|\tilde{y}^k-\bar{x}\| &= \|y^{k,j(k)}-\bar{x}\| = \|P_{C_{k,j(k)-1}}(y^{k,j(k)-1}) - P_{C_{k,j(k)-1}}(\bar{x})\| \\
&\leq \|y^{k,j(k)-1}-\bar{x}\| = \|P_{C_{k,j(k)-2}}(y^{k,j(k)-2}) - P_{C_{k,j(k)-2}}(\bar{x})\| \\
&\leq \|y^{k,j(k)-2}-\bar{x}\| \leq \cdots \leq \|y^{k,0}-\bar{x}\| = \|x^k-\bar{x}\|,
\end{aligned} \tag{29}$$

using Lemma 2(i) and (15).

Let $\tilde{\theta} = 1 + \theta \|\bar{u}\| \geq 1 + \theta \dfrac{\|\bar{u}\|}{\eta_k}$, by (26). Then

$$
\begin{aligned}
\|\tilde{y}^{k+1} - \bar{x}\|^2 &\leq \|x^{k+1} - \bar{x}\|^2 = \left\| P_{C_k}\left(\tilde{y}^k - \frac{\beta_k}{\eta_k}\tilde{u}^k\right) - P_{C_k}(\bar{x}) \right\|^2 \\
&\leq \left\| \tilde{y}^k - \frac{\beta_k}{\eta_k}\tilde{u}^k - \bar{x} \right\|^2 \\
&= \|\tilde{y}^k - \bar{x}\|^2 + \frac{\|\tilde{u}^k\|^2}{\eta_k^2}\beta_k^2 - 2\frac{\beta_k}{\eta_k} \langle \tilde{u}^k, \tilde{y}^k - \bar{x} \rangle \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 - 2\frac{\beta_k}{\eta_k} \langle \bar{u}, \tilde{y}^k - \bar{x} \rangle \\
&= \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 - 2\frac{\beta_k}{\eta_k} \Big(\langle \bar{u}, \tilde{y}^k - P_C(\tilde{y}^k) \rangle + \langle \bar{u}, P_C(\tilde{y}^k) - \bar{x} \rangle \Big) \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 + 2\frac{\beta_k}{\eta_k} \langle \bar{u}, P_C(\tilde{y}^k) - \tilde{y}^k \rangle \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 + \frac{\beta_k}{\eta_k} \|\bar{u}\| \|P_C(\tilde{y}^k) - \tilde{y}^k\| \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 + \frac{\beta_k}{\eta_k} \|\bar{u}\| \operatorname{dist}(\tilde{y}^k, C) \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 + \frac{\beta_k}{\eta_k} \|\bar{u}\| \frac{g(y^{k,j}) \|y^{k,j} - w\|}{g(y^{k,j}) - g(w)} \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 \left(1 + \theta \frac{\|\bar{u}\|}{\eta_k}\right) \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \tilde{\theta}\beta_k^2 \leq \|x^k - \bar{x}\|^2 + \tilde{\theta}\beta_k^2, \qquad (30)
\end{aligned}
$$

using (29) in the first inequality, Lemma 2(i) in the second one, the monotonicity of T and (27) in the third one, the definition of $S(T,C)$ in the fourth one, Cauchy-Schwartz inequality in the fifth one, Lemma 5 and the definition of $j(k)$ in the sixth and seventh one, and (29) in the last one.

Using Definition 2, (30) and (13), we conclude that the sequences $\{\tilde{y}^k\}$ and $\{x^k\}$ are quasi-Fejér convergent to $S(T,C)$. □

Next we establish some important convergence properties of Algorithm A.

Proposition 6. *Let $\{x^k\}$, $\{\tilde{y}^k\}$ and $\{\tilde{u}^k\}$ be the sequences generated by Algorithm A. Then,*

i) $\{x^k\}$, $\{\tilde{y}^k\}$ *and* $\{\tilde{u}^k\}$ *are bounded.*

ii) $\lim_{k\to\infty} \operatorname{dist}(x^k, C) = \lim_{k\to\infty} \operatorname{dist}(\tilde{y}^k, C) = 0$.

iii) $\lim_{k\to\infty} \|x^{k+1} - \tilde{y}^k\| = 0$.

iv) All cluster points of $\{x^k\}$ *and* $\{\tilde{y}^k\}$ *belong to* C.

Proof. i) For $\{x^k\}$ and $\{\tilde{y}^k\}$ use Proposition 5(i) and Proposition 1(i). For $\{\tilde{u}^k\}$, use boundedness of $\{\tilde{y}^k\}$ and Lemma 3(iii).

ii) We have that $\operatorname{dist}(\tilde{y}^k, C) \le \dfrac{g(y^{k,j})\,\|y^{k,j} - w\|}{g(y^{k,j}) - g(w)} \le \theta\beta_k$ by definition of $j(k)$, using Lemma 5. In view of (10), $\lim_{k\to\infty} \operatorname{dist}(\tilde{y}^k, C) = 0$.

For all k we have that

$$\|x^{k+1} - P_{C_k}(\tilde{y}^k)\| = \left\| P_{C_k}\left(\tilde{y}^k - \frac{\beta_k}{\eta_k}\tilde{u}^k\right) - P_{C_k}(\tilde{y}^k)\right\| \le \beta_k, \tag{31}$$

using (18) and Lemma 2(i) in the first inequality.

We may apply Lemma 1 because $\{\tilde{y}^k\}$ is bounded by (i). It follows that there exists a compact set containing $\{\tilde{y}^k\}$, and we can conclude that there exists $\tilde{\mu} \in [0, 1)$ such that

$$\operatorname{dist}(\tilde{z}(x, v), C) \le \tilde{\mu}\ \operatorname{dist}(x, C) \tag{32}$$

for all $x \in X \setminus C$ and all $v \in \partial g(x)$.

In view of the definition of $\tilde{z}(x, v)$,

$$\tilde{z}(\tilde{y}^k, \tilde{v}^k) = P_{C_k}(\tilde{y}^k).$$

Therefore, it follows from (32) that

$$\operatorname{dist}(P_{C_k}(\tilde{y}^k), C) = \operatorname{dist}(\tilde{z}(\tilde{y}^k, \tilde{v}^k), C) \le \tilde{\mu}\ \operatorname{dist}(\tilde{y}^k, C), \tag{33}$$

for all k such that $\tilde{y}^k \notin C$. If $\tilde{y}^k \in C$, (33) holds trivially because $C \subseteq C_k$ by Proposition 4(i). Observe that

$$\operatorname{dist}(x^{k+1}, C) \le \|x^{k+1} - P_{C_k}(\tilde{y}^k)\| + \operatorname{dist}(P_{C_k}(\tilde{y}^k), C) \le \beta_k + \tilde{\mu}\ \operatorname{dist}(\tilde{y}^k, C) \le \beta_k + \tilde{\mu}\,\theta\beta_k,$$

using (31) and (33) in the second inequality. Therefore, we obtain $\lim\limits_{k\to\infty} \operatorname{dist}(x^k, C) = 0$, establishing (ii).

iii) Using (31), we get

$$\|x^{k+1} - \tilde{y}^k\| \le \|x^{k+1} - P_{C_k}(\tilde{y}^k)\| + \|P_{C_k}(\tilde{y}^k) - \tilde{y}^k\| \le \beta_k + \operatorname{dist}(\tilde{y}^k, C). \tag{34}$$

Since $\lim_{k\to\infty} \beta_k = 0$ by (10), it follows from (ii) and (34) that $\lim_{k\to\infty} \|x^{k+1} - \tilde{y}^k\| = 0$.

iv) Follows from (ii).

$\square$

Paramonotonicity of T is used for the first time in this section in the following theorem.

Theorem 1. *Assume that T is paramonotone. If $S(T,C) \neq \emptyset$ then the sequence $\{x^k\}$ generated by Algorithm A converges to some solution of VIP(T,C).*

Proof. Assume that $S(T,C) \neq \emptyset$. Let $\{x^k\}$, $\{\tilde{y}^k\}$ and $\{\tilde{u}^k\}$ be the sequences generated by Algorithm A. Define $\gamma_k : S(T,C) \to \mathbb{R}$ as

$$\gamma_k(x) := \langle \tilde{u}^k, \tilde{y}^k - x \rangle. \tag{35}$$

Note that

$$\begin{aligned} \|x^{k+1} - x\|^2 &= \left\| P_{C_k}\left(\tilde{y}^k - \frac{\beta_k}{\eta_k}\tilde{u}^k\right) - P_{C_k}(x)\right\|^2 \le \left\| \left(\tilde{y}^k - \frac{\beta_k}{\eta_k}\tilde{u}^k\right) - x\right\|^2 \\ &= \|\tilde{y}^k - x\|^2 + \frac{\beta_k^2}{\eta_k^2}\|\tilde{u}^k\|^2 - 2\frac{\beta_k}{\eta_k}\ \langle \tilde{u}^k, \tilde{y}^k - x\rangle \\ &\le \|\tilde{y}^k - x\|^2 - \beta_k\left(2\frac{\gamma_k(x)}{\eta_k} - \beta_k\right) \\ &\le \|x^k - x\|^2 - \beta_k\left(2\frac{\gamma_k(x)}{\eta_k} - \beta_k\right). \end{aligned} \tag{36}$$

We prove first that $\{\tilde{y}^k\}$ has a cluster point which belongs to $S(T,C)$. Since $\{(\tilde{y}^k, \tilde{u}^k)\}$ is bounded by Proposition 6(i), it suffices to prove that $\{\gamma_k\}$ has a nonpositive cluster point. Assume that this is not true, and fix some $\bar{x} \in S(T,C)$. Clearly $\{\gamma_k(\bar{x})\}$ must be bounded away from zero for large k, i.e. there exist $\bar{k}$ and $\rho > 0$ such that $\gamma_k(\bar{x}) \ge \rho$ for all $k \ge \bar{k}$. Since $\{\tilde{u}^k\}$ is bounded, there exists $\theta > 1$ such that $\|\tilde{u}^k\| \le \theta$ for all k. Therefore

$$\eta_k = \max\{1, \|\tilde{u}^k\|\} \le \max\{1, \theta\} = \theta$$

for all k. In view of Lemma 4 and Proposition 6(ii), we can find $\bar{\rho} > 0$ such that $\dfrac{\gamma_k(\bar{x})}{\eta_k} \ge \dfrac{\gamma_k(\bar{x})}{\theta} > \bar{\rho}$ and hence, in view of (36), we obtain

$$\|x^{k+1} - \bar{x}\|^2 \le \|x^k - \bar{x}\|^2 - \beta_k(2\bar{\rho} - \beta_k) \tag{37}$$

for all $k \ge \bar{k}$. Since $\lim_{k\to\infty} \beta_k = 0$ by (10), there exists $k' \ge \bar{k}$ such that $\beta_k \le \bar{\rho}$ for all $k \ge \bar{k}$. So, we get from (37), for all $k \ge k'$,

$$\bar{\rho}\beta_k \le \|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2. \tag{38}$$

Summing (38) with k between k' and m, we obtain:

$$\bar{\rho}\sum_{k=k'}^{m}\beta_k \leq \sum_{k=k'}^{m}\left(\|x^k-\bar{x}\|^2-\|x^{k+1}-\bar{x}\|^2\right) \leq \|x^{k'}-\bar{x}\|^2-\|x^{m+1}-\bar{x}\|^2 \leq \|x^{k'}-\bar{x}\|^2. \tag{39}$$

Taking limits in (39) with $m\to\infty$, we contradict the assumption that $\sum_{k=0}^{\infty}\beta_k=\infty$. Thus, there exists a cluster point of $\{\tilde{y}^k\}$ belonging to $S(T,C)$. It follows from Proposition 5 and Proposition 1(ii) that $\{\tilde{y}^k\}$ is convergent to some point in $S(T,C)$. Using Proposition 6(iii), we obtain that $\{x^k\}$ converges to some point onto $S(T,C)$. □

Remark 1. We have included the assumption that a Slater point w is available, only for obtaining a fully explicit algorithm for a quite general convex set C. In fact, such assumption can be replaced by a rather weaker one, namely:

H) There exists an easily computable and continuous $\tilde{g}:\mathcal{H}\to\mathbb{R}$ such that $\operatorname{dist}(x,C)\leq\tilde{g}(x)$ for all $x\in\mathcal{H}$, and $\tilde{g}(x)=0$ if and only if $g(x)=0$.

Assuming (H), we can replace the left hand side of the inequality in (16) by $\tilde{g}(y^{k,j})$, and all our convergence results are preserved; in fact only the proof of Proposition 4(iii) has to modified.

Assuming existence of a Slater point w allows us to give an explicit formula for $\tilde{g}$, namely

$$\tilde{g}(x)=\begin{cases}\dfrac{g(x)}{g(x)-g(w)}\,\|x-w\| & \text{if } x\notin C\\[2ex] 0 & \text{if } x\in C,\end{cases}$$

but there are examples of sets C for which no Slater point is available, while (H) holds, including instances in which $int(C)=\emptyset$.

Acknowledgment

Research for this paper by 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] Bao, T.Q., Khanh, P.Q. A projection-type algorithm for pseudomonotone nonlipschitzian multi-valued variational inequalities. *Nonconvex Optimization and Its Applications* **77** (2005) 113-129.

[3] Bello Cruz, J.Y., Iusem, A.N. Convergence of direct methods for paramonotone variational inequalities. (to be published in *Computational Optimization and Applications*).

[4] Bello Cruz, J.Y., Iusem, A.N. An explicit algorithm for monotone variational inequalities (to be published).

[5] Bruck, R.E. An iterative solution of a variational inequality for certain monotone operators in a Hilbert space. *Bulletin of the American Mathematical Society* **81** (1975) 890-892.

[6] Burachik, R., Graña Drummond, L.M., Iusem, A.N., Svaiter, B.F. Full convergence of the steepest descent method with inexact line searches. *Optimization* **32** (1995) 137-146.

[7] Burachik, R.S., Iusem, A.N. *Set-valued mappings and enlargements of monotone operators*. Springer, Berlin (2008).

[8] Censor, Y., Iusem, A.N., Zenios, S.A. An interior point method with Bregman functions for the variational inequality problem with paramonotone operators. *Mathematical Programming* **81** (1998) 373-400.

[9] Facchinei, F., Pang, J.S. *Finite-dimensional variational inequalities and complementarity problems*. Springer, Berlin (2003).

[10] Fukushima, M. An outer approximation algorithm for solving general convex programs. *Operations Research* **31** (1983) 101-113.

[11] Fukushima, M. A Relaxed projection for variational inequalities. *Mathematical Programming* **35** (1986) 58-70.

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

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

[14] Iusem, A.N. On some properties of paramonotone operators. *Journal of Convex Analysis* **5** (1998) 269-278.

[15] Iusem, A.N., Lucambio Pérez, L.R. An extragradient-type method 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; Addendum *Optimization* **43** (1998) 85.

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

[19] Qu, B., Xiu, N. A new halfspace-relaxation projection method for the split feasibility problem. *Linear Algebra and its Applications* **428** (2008) 1218-1229.