An explicit algorithm for monotone variational inequalities

J. Y. Bello Cruz* A. N. Iusem†

September 30, 2010

Abstract

We introduce a fully explicit method for solving monotone variational inequalities in Hilbert spaces, where orthogonal projections onto the feasible set are replaced by projections onto suitable hyperplanes. We prove weak convergence of the whole generated sequence to a solution of the problem, under the only assumptions of continuity and monotonicity of the operator and existence of solutions.

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

1 Introduction

Let $\mathcal{H}$ be a Hilbert space, C be a nonempty, closed and convex subset of $\mathcal{H}$ and $T:\mathcal{H}\to\mathcal{H}$ a point-to-point operator. The variational inequality problem for T and C, denoted VIP(T,C), is the problem of finding a point $x^*\in C$ satisfying

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

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. Variational inequalities have a wide range of applications. Several of them are described in [22]. An excellent survey of methods for finite dimensional variational inequality problems can be found in [9].

Here, we are interested in explicit methods for solving VIP(T,C), that is to say, methods whose iterations are given by closed formulae, without demanding solution of subproblems. 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:

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

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

*Instituto de Matemática e Estatística, Universidade Federal de Goiás, Campus Samambaia, 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

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. $P_C(x) = \mathrm{argmin}_{y \in C} \|x - y\|$. See [2], [3] and [13] for convergence properties of this method for the case in which f is convex, which are related to results in this paper.

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

$$x^{k+1} = P_C\left(x^k - \alpha_k T(x^k)\right) , \tag{3}$$

where the sequence α_k satisfies some conditions.

Convergence results for this method require some monotonicity properties of T. Next, we introduce 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) pseudomonotone on W if for all x, $y \in W$ it holds that $\langle T(y), x - y\rangle \geq 0$ implies $\langle T(x), x - y\rangle \geq 0$,

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

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

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

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

It follows from Definition 1 that the following implications hold: (v) $\Rightarrow$ (iv) $\Rightarrow$ (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i). The reverse assertions are

Convergence results for the scheme (3) have been established in [1] for the case of uniformly monotone operators, and in [5] for the case of paramonotone ones.

We remark that for the method given by (3) there is no chance of relaxing the assumption on T to plain monotonicity. For example, consider $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $T(x) = Ax$, with $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and $C = \mathbb{R}^2$. 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 (3) moves away from the solution, independently of the choice of the stepsize α_k.

To overcome this weakness of the method defined by (3), Korpelevich proposed in [23] a modification of the method, called the extragradient algorithm. It generates iterates using the formulae:

$$y^k = P_C(x^k - \beta T(x^k)), \tag{4}$$

$$x^{k+1} = P_C(x^k - \beta T(y^k)), \tag{5}$$

where $\beta > 0$ is a fixed number. The difference with (3) is that T is evaluated twice and the projection is computed twice at each iteration, but the benefit is significant, because to the resulting algorithm is applicable to the whole class of monotone variational inequalities. However, in order to establish convergence, Korpelevich assumed that T is Lipschitz continuous and that an estimate of the Lipschitz constant (called L) is available. It has been proved in [23] that the extragradient method is globally convergent if T is monotone and Lipschitz continuous on C and $\beta \in \left(0, \frac{1}{L}\right)$.

When T is not Lipschitz continuous, or it is Lipschitz but the constant L is not known, the fixed parameter β must be replaced by stepsizes computed through an Armijo-type search, as in the following method, presented in [16] (see also [18] for another related approach).

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

$$x^0 \in C. \tag{6}$$

Iterative step. Given x^k define

$$z^k = P_C(x^k - \beta_k T(x^k)). \tag{7}$$

If $x^k = z^k$ stop. Otherwise take

$$j(k) = \min\left\{ j \geq 0 : \langle T(2^{-j}z^k + (1-2^{-j})x^k), x^k - z^k \rangle \geq \frac{\delta}{\beta_k} \|x^k - z^k\|^2 \right\}, \tag{8}$$

$$y^k = 2^{-j(k)}z^k + (1 - 2^{-j(k)})x^k, \tag{9}$$

$$\tilde{x}^k = x^k - \frac{\langle T(y^k), x^k - y^k \rangle}{\|T(y^k)\|^2} T(y^k), \tag{10}$$

$$x^{k+1} = P_C(\tilde{x}^k). \tag{11}$$

This strategy for determining the stepsizes guarantees convergence under the only assumptions of monotonicity and continuity of T and existence of solutions of $\mathrm{VIP}(T, C)$, that is to say, without assuming Lipschitz continuity of T. Also, this algorithm demands only two projections onto C per iteration, unlike other variants, e.g. [14] and [24], with projections onto C inside the inner loop for the search of the stepsize.

Another approach for solving variational inequality problems can be found in [19], where the following general scheme is proposed for solving $\mathrm{VIP}(T, C)$ in finite dimensional spaces:

Basic scheme. Take $x^0 \in C$ and a sequence $\{\gamma_k\} \subseteq [\eta, 2 - \eta]$ with $\eta > 0$.

Step 1. Given x^k, find $y^k \in C$ such that

$$\langle T(y^k), x^k - y^k \rangle > 0. \tag{12}$$

Step 2. Set

$$\tilde{x}^k = x^k - \gamma_k \frac{\langle T(y^k), x^k - y^k \rangle}{\|T(y^k)\|^2} T(y^k). \tag{13}$$

Step 3. Find $x^{k+1} \in C$ such that

$$\|x^{k+1} - y\| \leq \|\tilde{x}^k - y\| \qquad \forall y \in C. \tag{14}$$

It has been proved in [19] that the sequence $\{x^k\}$ so defined converges to a solution of $\mathrm{VIP}(T, C)$ assuming that T is locally Lipschitz continuous and pseudomonotone, and that $\mathrm{VIP}(T, C)$ has solutions.

1.1 Relaxed projection methods

The method given by (3) 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 $\min\{\|x - (x^k - \alpha_k u^k)\| : x \in C\}$, in order to compute the projection onto C. The same drawback affects the algorithms given by (4)-(5) and (6)-(11), which demand in fact two orthogonal projections per iteration.

A similar limitation applies to the method in [19], though the orthogonal projections do not appear as such in the formulation of the algorithm. We discuss next the extent to which vectors y^k and x^{k+1} satisfying (12)-(14) can be determined through an explicit procedure.

It is clear that the hyperplane separating x^k from the solution set, implicit in (12), can be found by taking

$$y^k = \alpha_k x^k + (1 - \alpha_k) P_C(x^k - T(x^k)), \tag{15}$$

as in Korpelevich's method but, again, the computation of P_C cannot be achieved explicitly. Other procedures for finding an appropriate y^k are discussed in [19], [20] and [21], where several specific variants of the scheme given by (12)-(14) are presented, but none of them is explicit.

The situation is different regarding the computation of x^{k+1}. In [19] an inner loop is introduced in order to find an x^{k+1} satisfying (14), instead of the projection onto C. A finite sequence $\{z^{k,i}\}$ is generated, starting from $\tilde{x}^k$, by successive reflections through hyperplanes separating $\{z^{k,i}\}$ from C. Assuming a Slater condition on C a point in C satisfying (14) is found, after a finite number of such reflections. In conclusion, x^{k+1} can be explicitly determined, but y^k cannot, and hence Konnov's methods are not fully explicit.

One option for avoiding the use of orthogonal projections onto C, 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. Observe that projections onto halfspaces are easily computable. This concept of relaxed projection was first proposed by M. Fukushima in [11] for convex optimization.

We consider the case in which C is of the form

$$C = \{x \in \mathcal{H} : g(x) \leq 0\}, \tag{16}$$

where $g : \mathcal{H} \to \mathbb{R}$ is a convex function, satisfying Slater's condition, i.e. there exists a point w such that $g(w) < 0$. The differentiability of g is not assumed and the representation (16) 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 (16) with $g(x) = \sup\{g_j(x) : j \in J\}$.

Fukushima introduced in [10] a method for solving VIP(T, C) in a finite dimensional space, i.e. $\mathcal{H} = \mathbb{R}^n$, using the following relaxed iteration:

$$x^{k+1} = P_{C_k}\left(x^k - \beta_k \frac{T(x^k)}{\|T(x^k)\|}\right), \tag{17}$$

where β_k is an exogenous stepsize satisfying

$$\sum_{k=0}^{\infty} \beta_k = \infty \qquad \text{and} \qquad \sum_{k=0}^{\infty} \beta_k^2 < \infty, \tag{18}$$

and C_k is defined as

$$C_k := \{x \in \mathbb{R}^n : g(x^k) + \langle v^k, x - x^k \rangle \leq 0\}, \tag{19}$$

with $v^k \in \partial g(x^k)$, where $\partial g(x^k)$ is the subdifferential of g at x^k, i.e. $\partial g(x^k) = \{v : g(x) \geq g(x^k) + \langle v, x - x^k \rangle\}$.

In [10] convergence of $\{x^k\}$ to a point in $S(T, C)$ was proved, assuming that T is strongly monotone and that it satisfies the following coerciveness condition:

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

A related relaxed iteration has been proposed in [5], using the iteration:

$$x^{k+1} = P_{C_k}\left(x^k - \frac{\beta_k}{\eta_k} T(x^k)\right), \tag{20}$$

where $\eta_k = \max\{1, \|T(x^k)\|\}$, β_k satisfy (18), and C_k is given by (19).

It was proved that the sequence generated by (20) is bounded, the difference between consecutive iterates converges to zero, and all its cluster points belong to $S(T, C)$. These results were established under quite demanding assumptions: T must be paramonotone and it must satisfy the following coerciveness condition:

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

1.2 The new method

In this paper we will analyze a new algorithm, relaxing the hypotheses in [5] in two directions: we assume plain monotonicity of T instead of paramonotonicity, and we do not need any coerciveness condition. Additionally, we obtain convergence results stronger than those in [5]; namely we get weak convergence of the whole sequence to some solution of VIP(T, C), assuming only existence of solutions, and all results hold in a Hilbert space (of course, in finite dimensional case we get strong, rather than weak, convergence).

The main advantage over Korpelevich's method (4)-(5) and its variants (e.g. (6)-(11)) is that it replaces orthogonal projections onto C, which in general are not easily computable, by projections onto hyperplanes, which have simple closed formulae. Thus, the method is indeed fully explicit.

Next we describe our method and compare it with (6)-(11). In (6)-(11) a step is taken from the current iterate x^k in the direction of $-T(x^k)$, resulting in an auxiliary point z^k. A line search is then performed in the segment between x^k and $P_C(z^k)$, resulting in the point y^k. Then, a step with a specified steplength is taken from x^k in the direction of $-T(y^k)$, and the next iterate is obtained by projecting the resulting point onto C.

In our method, we construct simultaneously two sequences, the main sequence $\{x^k\}$ and the auxiliary sequence $\{\tilde{y}^k\}$. A step is taken from $\tilde{y}^{k-1}$ in the direction of $-T(\tilde{y}^{k-1})$ with an exogenous steplength, and the resulting point is projected onto an auxiliary hyperplane containing C. This projection step is repeated in a finite inner loop, changing the auxiliary hyperplanes, until a point $\tilde{y}^k$ is obtained, whose distance to C is smaller than a certain multiple of the current exogenous steplength.

After this inner loop, the next main iterate x^{k+1} is a convex combination with exogenous coefficients of $\tilde{y}^k$ and x^k. The inner loop of projections onto hyperplanes hence it substitutes the exact projection onto C, demanded in (4)-(5) and (6)-(11).

Our inner loop has some similarities with the inner loop in Konnov's method [19], described above. Ours ends up with a point $\tilde{y}^k$ satisfying $\|\tilde{y}^k - y\| \leq \|z^k - y\|$ for all $y \in C$, similar to (14). The difference is that we use projections, rather then reflections, onto separating hyperplanes, so that our sequence will in general be infeasible, while the sequences $\{y^k\}$, $\{x^k\}$ in Konnov's general procedure are always feasible. Generally speaking, the generated sequence approaches the solution from the outside of the feasible set in our method, and from the inside in Konnov's one.

In connection with the method in [5], the algorithm in this paper works under weaker assumptions on T, but it demands continuity of the operator. Thus, it cannot be used for point-to-set operators T, which are admissible in the convergence analysis in [5]. Extensions of Korpelevich's method to the case of point-to-set operators can be found in [15] and [4].

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

2 Preliminary results

In this section, we present some definitions and results that are needed for the convergence analysis of the proposed method. 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) $\|P_K(x) - P_K(y)\|^2 \le \|x-y\|^2 - \|(P_K(x)-x) - (P_K(y)-y)\|^2$.

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

Proof. See Lemma 1 in [26]. □

Next we deal with the so called quasi-Fejér convergence and its properties.

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

This definition originates in [8] and has been further elaborated in [17].

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

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

ii) $\{\|x^k - x\|\}$ *converges for all* $x \in S$,

iii) if all weak cluster point of $\{x^k\}$ *belong to* S, *then the sequence* $\{x^k\}$ *is weakly convergent.*

Proof. See Proposition 1 in [2]. □

For $S \subseteq \mathcal{H}$, define $\operatorname{dist}(x, S) := \inf_{z\in S} \|z - x\|$. It is clear that if S is a closed and convex set then $\operatorname{dist}(x, S) = \min_{z\in S} \|z - x\| = \|P_S(x) - x\|$ where $P_S(x) = \operatorname{argmin}_{z\in S} \|x - z\|$. Next, we establish two properties of quasi-Fejér convergent sequences.

Lemma 2. *If a sequence $\{x^k\}$ is quasi-Fejér convergent to a closed and convex set S, then*

i) the sequence $\{\operatorname{dist}(x^k, S)\}$ *is convergent,*

ii) the sequence $\{P_S(x^k)\}$ *is strongly convergent.*

Proof. i) The sequence $\{\text{dist}(x^k, S)\}$ is bounded, since $0 \leq \text{dist}(x^k, S) \leq \|x^k - x\|$ for all $x \in S$, and $\{\|x^k - x\|\}$ converges for all $x \in S$, by Proposition 1(ii).

Assume that $\{\text{dist}(x^k, S)\}$ has two cluster points, say λ and μ, with $\lambda < \mu$. It follows that $\{\text{dist}(x^k, S)^2\}$ has λ^2 and μ^2 as cluster points. Take $\nu = (\mu^2 - \lambda^2)/3$, and subsequences $\{\text{dist}(x^{j(k)}, S)^2\}$ and $\{\text{dist}(x^{\ell(k)}, S)^2\}$ of $\{\text{dist}(x^k, S)^2\}$ such that $\lim_{k\to\infty}\{\text{dist}(x^{j(k)}, S)^2\} = \lambda^2$, $\lim_{k\to\infty}\{\text{dist}(x^{\ell(k)}, S)^2\} = \mu^2$. For each k take j_k, ℓ_k such that $k < \ell_k < j_k$, with $\text{dist}(x^{j_k}, S)^2 < \lambda^2 + \nu$, $\text{dist}(x^{\ell_k}, S)^2 > \mu^2 - \nu$. Defining $w = P_S(x^{j_k})$, we get

$$\begin{aligned} 0 &< \nu = 3\nu - 2\nu = \mu^2 - \lambda^2 - 2\nu = (\mu^2 - \nu) - (\lambda^2 + \nu) < \text{dist}(x^{\ell_k}, S)^2 - \text{dist}(x^{j_k}, S)^2 \\ &= \text{dist}(x^{\ell_k}, S)^2 - \|x^{j_k} - w\|^2 \leq \|x^{\ell_k} - w\|^2 - \|x^{j_k} - w\|^2 \\ &= \sum_{j=\ell_k-1}^{j_k} (\|x^{j+1} - w\|^2 - \|x^j - w\|^2) \leq \sum_{j=\ell_k-1}^{j_k} \delta_j \leq \sum_{j=k}^{\infty} \delta_j. \end{aligned}$$

Thus, $\nu < \sum_{j=k}^{\infty} \delta_j$ for all k, contradicting the fact that $\sum_{j=0}^{\infty} \delta_j < \infty$. Hence, $\nu = 0$, i.e. $\lambda^2 = \mu^2$ implying $\lambda = \mu$. It follows that all cluster points of $\{\text{dist}(x^k, S)\}$ coincide, i.e. the sequence $\{\text{dist}(x^k, S)\}$ converges.

ii) We will now prove that $\{u^k\} := \{P_S(x^k)\}$ is a Cauchy sequence, hence strongly convergent. Using Lemma 1(i) with $K = S$, $x = x^q$ and $y = u^p$, we obtain

$$\|u^q - u^p\|^2 = \|P_S(x^q) - P_S(u^p)\|^2 \leq \|x^q - u^p\|^2 - \|x^q - u^q\|^2. \tag{21}$$

Since $\{x^k\}$ is quasi-Fejér convergent to S and $p < q$, we get from (21) that

$$\begin{aligned} \|u^q - u^p\|^2 &\leq \|x^p - u^p\|^2 - \|x^q - u^q\|^2 + \sum_{j=p}^{q-1} \delta_j \\ &\leq \text{dist}(x^p, S)^2 - \text{dist}(x^q, S)^2 + \sum_{j=p}^{\infty} \delta_j. \end{aligned} \tag{22}$$

By (i), $\{\text{dist}(x^k, S)\}$ converges, and using the fact $\sum_{j=0}^{\infty} \delta_j < \infty$, we obtain from (22) that $\{u^k\}$ is a Cauchy sequence.

□

Now we recall the definition of maximal monotone operators.

Definition 3. *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 monotone variational inequalities.

Proposition 2. *Let $T : \mathcal{H} \to \mathcal{H}$ be a monotone and continuous operator and C a closed and convex set. Then $S(T, C)$ if nonempty, is closed and convex.*

Proof. See e.g.Lemma 2.4(ii) of [6]. □

The next lemma will be useful for proving that all weak cluster points of the sequence generated by our algorithm belong to $S(T, C) = \{x \in C \, : \, \langle T(x), y - x\rangle \geq 0 \, , \, \forall y \in C\}$.

Lemma 3. *Consider VIP(T, C). If $T : \mathcal{H} \to \mathcal{H}$ is continuous and monotone, then*

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

Proof. By the monotonicity of T, we have $\langle T(x), y - x\rangle \leq \langle T(y), y - x\rangle$ for all x, $y \in C$. Thus, it is clear that $S(T, C) \subseteq \{x \in C\, :\, \langle T(y), y - x\rangle \geq 0\,,\, \forall\, y \in C\}$.

Conversely, assume that $x \in \{x \in C\, :\, \langle T(y), y - x\rangle \geq 0\;,\, \forall\, y \in C\}$. Take $y(\alpha) = (1-\alpha)x + \alpha y$, $y \in C$ with $\alpha \in (0, 1)$. It is clear that $y(\alpha) \in C$ and therefore

$$\begin{aligned} 0 \leq \langle T(y(\alpha)), y(\alpha) - x\rangle &= \langle T((1-\alpha)x + \alpha y), (1-\alpha)x + \alpha y - x\rangle \\ &= \alpha\langle T((1-\alpha)x + \alpha y), y - x\rangle. \end{aligned}$$

Dividing by $\alpha > 0$, we get

$$0 \leq \langle T((1-\alpha)x + \alpha y), y - x\rangle, \tag{23}$$

for all $\alpha \in (0, 1)$. Making $\alpha \to 0$ and using the continuity of T, we obtain from (23) that $\langle T(x), y - x\rangle \geq 0$, for all $y \in C$, i.e. $x \in S(T, C)$. □

The next lemma provides a computable upper bound on the distance between a point and the feasible set C.

Lemma 4. *Let $g : \mathcal{H} \to \mathbb{R}$ be a convex function and $C := \{x \in \mathcal{H}\, :\, g(x) \leq 0\}$. Assume that there exists $w \in C$ such that $g(w) < 0$. Then, for all y such that $g(y) > 0$, we have*

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

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

$$g(w_\lambda) = g(\lambda w + (1-\lambda)y) \leq \lambda g(w) + (1-\lambda)g(y) = g(y) - \lambda(g(y) - g(w)) = 0.$$

Thus, $w_\lambda \in C$ and

$$\operatorname{dist}(y, C) \leq \|y - w_\lambda\| = \|y - (\lambda w + (1-\lambda)y)\| = \lambda\|y - w\| = \frac{g(y)}{g(y) - g(w)}\|y - w\|.$$

□

Finally, we will need the following elementary result on sequence averages.

Proposition 3. *Let $\{u^k\} \subset \mathcal{H}$ be a sequence strongly convergent to $\tilde{u}$. Take nonnegative real numbers $\zeta_{k,j}$ $(k \geq 0,\ 0 \leq j \leq k)$ such that $\lim_{k\to\infty} \zeta_{k,j} = 0$ for all j and $\sum_{j=0}^{k} \zeta_{k,j} = 1$ for all k. Define*

$$w^k := \sum_{j=0}^{k} \zeta_{k,j} u^j\,.$$

Then, $\{w^k\}$ also converges strongly to $\tilde{u}$.

Proof. Elementary.

□

3 A relaxed projection algorithm

In this section, we introduce an algorithm which replaces projections onto C by easily computable projections onto suitable hyperplanes. We assume that that T is point-to-point and maximal monotone and C of the form given in (16), which we repeat here:

$$C = \{x \in \mathcal{H} \,:\, g(x) \leq 0\}, \tag{24}$$

where $g : \mathcal{H} \to \mathbb{R}$ is a convex function.

We need the following boundedness assumptions on ∂g and T.

(R) ∂g is bounded on bounded sets.

(S) T is bounded on bounded sets.

In finite dimensional spaces, these two assumptions are always satisfied because T and ∂g are maximal monotone operators (see e.g. Theorem 4.6.1(ii) of [7]).

We also assume that a Slater point is available, i.e. we will explicitly use a point w such that $g(w) < 0$.

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

$$\sum_{k=0}^{\infty} \beta_k = \infty, \tag{25}$$

$$\sum_{k=0}^{\infty} \beta_k^2 < \infty. \tag{26}$$

The algorithm is defined as follows.

Algorithm A

Initialization step: Fix an exogenous constant $\theta > 0$ and define

$$x^0 := 0 \qquad \text{and} \qquad z^0 \in \mathcal{H}.$$

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

$$C_{k,j} := \{z \in \mathcal{H} \,:\, g(y^{k,j}) + \langle v^{k,j}, z - y^{k,j} \rangle \leq 0\}, \tag{27}$$

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

$$y^{k,j+1} := P_{C_{k,j}}(y^{k,j}). \tag{28}$$

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

$$j(k) := \min \left\{ j \geq 0 : \frac{g(y^{k,j}) \, \|y^{k,j} - w\|}{g(y^{k,j}) - g(w)} \leq \theta \beta_k \right\}. \tag{29}$$

Set

$$\tilde{y}^k := y^{k,j(k)}. \tag{30}$$

Choose $\tilde{v}^k \in \partial g(\tilde{y}^k)$ and let

$$C_k := C_{k,j(k)} = \{z \in \mathcal{H} \, : \, g(\tilde{y}^k) + \langle \tilde{v}^k, z - \tilde{y}^k \rangle \leq 0\}\,. \tag{31}$$

Define $\eta_k := \max\{1, \|T(\tilde{y}^k)\|\}$. Take

$$z^{k+1} := P_{C_k}\left(\tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k)\right). \tag{32}$$

If $z^{k+1} = \tilde{y}^k$, stop. Otherwise, define

$$\sigma_k := \sum_{j=0}^{k} \frac{\beta_j}{\eta_j} = \sigma_{k-1} + \frac{\beta_k}{\eta_k}, \tag{33}$$

$$x^{k+1} := \left(1 - \frac{\beta_k}{\eta_k \sigma_k}\right) x^k + \frac{\beta_k}{\eta_k \sigma_k} \tilde{y}^k. \tag{34}$$

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 $\mathrm{VIP}(T, C)$.

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

Proposition 4. *Define $C_x := \{z \in \mathcal{H} \, : \, g(x) + \langle v, z - x \rangle \leq 0\}$ with $v \in \partial g(x)$. Then for any $y \in \mathcal{H}$,*

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

Proof. See Proposition 3.1 in [25]. □

It follows from Proposition 4, (27), (28), (31) and (32) that

$$\begin{aligned} y^{k,j+1} &= P_{C_{k,j}}(y^{k,j}) = y^{k,j} - \frac{1}{\|v^{k,j}\|^2} \max\left\{0, g(y^{k,j})\right\} v^{k,j}, \\ z^{k+1} &= P_{C_k}\left(\tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k)\right) = \tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k) - \frac{1}{\|\tilde{v}^k\|^2} \max\left\{0, g(\tilde{y}^k) - \frac{\beta_k}{\eta_k} \langle T(\tilde{y}^k), \tilde{v}^k \rangle\right\} \tilde{v}^k, \end{aligned}$$

so that Algorithm A can be considered as a fully explicit method for $\mathrm{VIP}(T, C)$.

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

$$\begin{aligned} v^{k,j} &= \nabla g_{\ell(k,j)}(y^{k,j}), \qquad \text{with} \quad \ell(k,j) \in \arg\max_{0 \leq i \leq m} \{g_i(y^{k,j})\} \\ v^k &= \nabla g_{\ell(k)}(\tilde{y}^k), \qquad \text{with} \quad \ell(k) \in \arg\max_{0 \leq i \leq m} \{g_i(\tilde{y}^k)\}, \end{aligned}$$

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.

We observe that the inner loop in (27)-(28), starts with the point z^k and ends with a point $\tilde{y}^k$ close to C, in fact $\mathrm{dist}(\tilde{y}^k, C) \leq \theta\beta_k$, and $\lim_{k\to\infty} \beta_k = 0$ by (26). It might seem that this inner loop can be replaced by any finite procedure leading to an approximation of $P_C(z^k)$, say a point $\tilde{y}^k$ such that $\|\tilde{y}^k - P_C(z^k)\|$ is small enough. This not the case. In the first place, depending on the location of the intermediate hyperplanes $C_{k,j}$, the sequence $\{y^{k,j}\}_j$ may approach points in C far from $P_C(z^k)$; in fact the computational cost of our inner loop is lower than the computation of an inexact orthogonal projection of z^k onto C. On the other hand, it is not the case that any point z close enough to $P_C(z^k)$ will do the job. The crucial relation for convergence of our method is $\|\tilde{y}^k - x\| \leq \|z^k - x\|$ for all $x \in C$ (see also (14)), which may fail if we replace $\tilde{y}^k$ by points z arbitrarily closed to $P_C(z^k)$.

4 Convergence analysis of Algorithm A

For convergence of our method, we assume that T is monotone and continuous. Observe that $\partial g(x) \neq \emptyset$ for all $x \in \mathcal{H}$, because we assume that g is convex and $\mathrm{dom}(g) = \mathcal{H}$.
First we establish that Algorithm A is well defined.

Proposition 5. *Take C, $C_{k,j}$, C_k, $\tilde{y}^k$, z^k and x^k defined by* (16), (27), (31), (30), (32) *and* (34) *respectively. Then,*

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

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

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

Proof. i) It follows from (27), (31) and the definition of the subdifferential.

ii) Suppose that $z^{k+1} = \tilde{y}^k$. Then, since $z^{k+1} \in C_k$, we have $g(\tilde{y}^k) + \langle \tilde{v}^k, z^{k+1} - \tilde{y}^k \rangle = g(\tilde{y}^k) \leq 0$, i.e. $\tilde{y}^k \in C$. Moreover, since z^{k+1} is given by (32), using Lemma 1(ii) with $x = \tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k)$ and $K = C_k$, we obtain

$$\left\langle z^{k+1} - \left(\tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k) \right), z - z^{k+1} \right\rangle \geq 0 \quad \forall z \in C_k. \tag{35}$$

Taking $z^{k+1} = \tilde{y}^k$ in (35) and using the facts that $\beta_k > 0$, and $C \subseteq C_k$ for all k, we get $\langle T(\tilde{y}^k), z - \tilde{y}^k \rangle \geq 0$ for all $z \in C$. 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 obtain an infinite sequence $\{y^{k,j}\}_{j=0}^{\infty}$, such that

$$\liminf_{j\to\infty} \frac{g(y^{k,j})\, \|y^{k,j} - w\|}{g(y^{k,j}) - g(w)} \geq \theta\beta_k > 0. \tag{36}$$

Taking into account the inner loop in j given in (30) i.e. $y^{k,j+1} = P_{C_{k,j}}(y^{k,j})$ for each k, we obtain, for each $x \in C$,

$$\|y^{k,j+1} - x\|^2 = \|P_{C_{k,j}}(y^{k,j}) - P_{C_{k,j}}(x)\|^2 \leq \|y^{k,j} - x\|^2 - \|y^{k,j+1} - y^{k,j}\|^2 \leq \|y^{k,j} - x\|^2, \tag{37}$$

using Lemma 1(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 $\tau := \frac{1}{-g(w)} \sup_{0 \le j \le \infty} \|y^{k,j} - w\|$ is finite and also,

$$g(y^{k,j}) > 0 \qquad \text{for all} \quad j. \tag{38}$$

Using (37), we get

$$\lim_{j\to\infty} \|y^{k,j+1} - y^{k,j}\| = \lim_{j\to\infty} \|P_{C_{k,j}}(y^{k,j}) - y^{k,j}\| = 0. \tag{39}$$

Since $y^{k,j+1}$ belongs to $C_{k,j}$, we have from (27) that

$$g(y^{k,j}) \le \langle v^{k,j}, y^{k,j} - y^{k,j+1}\rangle \le \|v^{k,j}\|\|y^{k,j} - y^{k,j+1}\|, \tag{40}$$

using Cauchy-Schwartz in the last inequality.

Since $\{y^{k,j}\}_{j=0}^{\infty}$ is bounded and the subdifferential of g is bounded on bounded sets by assumption (R), we obtain that $\{\|v^{k,j}\|\}_{j=0}^{\infty}$ is bounded. In view of (39) and (40),

$$\liminf_{j\to\infty} g(y^{k,j}) \le 0. \tag{41}$$

It follows from (38) and (41) that

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

contradicting (36). Therefore, it follows that $j(k)$ is well defined. □

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

Proposition 6. *If $S(T, C)$ is nonempty, then $\{\tilde{y}^k\}$ and $\{z^k\}$ are quasi-Fejér convergent to $S(T, C)$.*

Proof. Observe that $\eta_k \ge \|T(\tilde{y}^k)\|$ and $\eta_k \ge 1$ for all k by the definition of η_k. Then, for all k,

$$\frac{1}{\eta_k} \le 1 \tag{42}$$

and

$$\frac{\|T(\tilde{y}^k)\|}{\eta_k} \le 1. \tag{43}$$

Take $\bar{x} \in S(T, C)$. Then,

$$\begin{aligned}
\|\tilde{y}^k - \bar{x}\|^2 &= \|y^{k,j(k)} - \bar{x}\|^2 = \|P_{C_{k,j(k)-1}}(y^{k,j(k)-1}) - P_{C_{k,j(k)-1}}(\bar{x})\|^2 \le \|y^{k,j(k)-1} - \bar{x}\|^2 \\
&= \|P_{C_{k,j(k)-2}}(y^{k,j(k)-2}) - P_{C_{k,j(k)-2}}(\bar{x})\|^2 \le \|y^{k,j(k)-2} - \bar{x}\|^2 \le \cdots \le \|z^k - \bar{x}\|^2.
\end{aligned} \tag{44}$$

Let $\tilde{\theta} = 1 + \theta \|T(\bar{x})\| \geq 1 + \theta \dfrac{\|T(\bar{x})\|}{\eta_k}$, by (42). Then

$$\begin{aligned}
\|\tilde{y}^{k+1} - \bar{x}\|^2 &\leq \|z^{k+1} - \bar{x}\|^2 = \left\| P_{C_k}\left(\tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k)\right) - P_{C_k}(\bar{x}) \right\|^2 \leq \left\| \tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k) - \bar{x} \right\|^2 \\
&= \|\tilde{y}^k - \bar{x}\|^2 + \frac{\|T(\tilde{y}^k)\|^2}{\eta_k^2}\beta_k^2 - 2\frac{\beta_k}{\eta_k}\,\langle T(\tilde{y}^k), \tilde{y}^k - \bar{x}\rangle \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 - 2\,\frac{\beta_k}{\eta_k}\langle T(\bar{x}), \tilde{y}^k - \bar{x}\rangle \\
&= \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 - 2\,\frac{\beta_k}{\eta_k}\left(\langle T(\bar{x}), \tilde{y}^k - P_C(\tilde{y}^k)\rangle + \langle T(\bar{x}), P_C(\tilde{y}^k) - \bar{x}\rangle\right) \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 + 2\,\frac{\beta_k}{\eta_k}\langle T(\bar{x}), P_C(\tilde{y}^k) - \tilde{y}^k\rangle \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 + \frac{\beta_k}{\eta_k}\|T(\bar{x})\|\|P_C(\tilde{y}^k) - \tilde{y}^k\| \leq \|\tilde{y}^k - \bar{x}\|^2 + \tilde{\theta}\beta_k^2 \\
&\leq \|z^k - \bar{x}\|^2 + \tilde{\theta}\beta_k^2, \qquad (45)
\end{aligned}$$

using (44) in the first inequality, Lemma 1(i) in the second one, the monotonicity of T and (43) in the third one, the definition of $S(T, C)$ in the fourth one, Cauchy-Schwartz inequality in the fifth one, Lemma 4 and the definition of $j(k)$ in the sixth one, and (44) in the last one.

Using Definition 2, (45) and (26), we conclude that the sequences $\{\tilde{y}^k\}$ and $\{z^k\}$ are quasi-Fejér convergent to $S(T, C)$. □

Proposition 7. *Let $\{z^k\}$, $\{\tilde{y}^k\}$ and $\{x^k\}$ be the sequences generated by Algorithm A. Assume that $S(T, C)$ is nonempty. Then,*

i) $x^{k+1} = \dfrac{1}{\sigma_k}\displaystyle\sum_{j=0}^{k} \frac{\beta_j}{\eta_j}\tilde{y}^j$,

ii) $\{\tilde{y}^k\}$, $\{x^k\}$ *and* $\{T(\tilde{y}^k)\}$ *are bounded,*

iii) $\lim_{k\to\infty} \operatorname{dist}(x^k, C) = 0$,

iv) all weak cluster points of $\{x^k\}$ *belong to* C.

Proof. i) In view of (33), (34) can be rewritten as

$$\sigma_k x^{k+1} = \left(\sigma_k - \frac{\beta_k}{\eta_k}\right) x^k + \frac{\beta_k}{\eta_k}\tilde{y}^k = \sigma_{k-1}x^k + \frac{\beta_k}{\eta_k}\tilde{y}^k = \sigma_{k-2}x^{k-1} + \frac{\beta_{k-1}}{\eta_{k-1}}\tilde{y}^{k-1} + \frac{\beta_k}{\eta_k}\tilde{y}^k = \cdots =$$

$$\sigma_0 x^1 + \sum_{j=0}^{k}\frac{\beta_j}{\eta_j}\tilde{y}^j = \sum_{j=0}^{k}\frac{\beta_j}{\eta_j}\tilde{y}^j,$$

using the fact that $\sigma_0 = 0$, which follows from (33), and we obtain the result after dividing both sides by σ_k.

ii) For $\{\tilde{y}^k\}$ use Proposition 6 and Proposition 1(i). For $\{T(\tilde{y}^k)\}$, use boundedness of $\{\tilde{y}^k\}$ and assumption (S). For $\{x^k\}$, use boundedness of $\{\tilde{y}^k\}$ and (i).

iii) It follows from Lemma 4 and (29)-(30) that

$$\operatorname{dist}(\tilde{y}^k, C) \leq \theta\beta_k. \tag{46}$$

Define

$$\tilde{x}^{k+1} := \frac{1}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}P_C(\tilde{y}^j). \tag{47}$$

Since $\frac{1}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j} = 1$ by (33), we get from the convexity of C that $\tilde{x}^{k+1} \in C$. Let

$$\tilde{\beta} := \sum_{j=0}^{\infty}\beta_j^2. \tag{48}$$

Note that $\tilde{\beta}$ is finite by (26). Then

$$\begin{aligned}\operatorname{dist}(x^{k+1}, C) &\leq \|x^{k+1} - \tilde{x}^{k+1}\| = \left\|\frac{1}{\sigma_k}\left(\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}(\tilde{y}^j - P_C(\tilde{y}^j))\right)\right\| \leq \frac{1}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}\|\tilde{y}^j - P_C(\tilde{y}^j)\| \\ &= \frac{1}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}\operatorname{dist}(\tilde{y}^j, C) \leq \frac{\theta}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j^2}{\eta_j} \leq \frac{\theta}{\sigma_k}\sum_{j=0}^{k}\beta_j^2 \leq \theta\frac{\tilde{\beta}}{\sigma_k},\end{aligned} \tag{49}$$

using the fact that $\tilde{x}^{k+1}$ belongs to C in the first inequality, (i) and (47) in the first equality, convexity of $\|\cdot\|$ in the second inequality, (46) in the third one, (42) in the fourth one and (48) in the last one.

Take $\gamma > 1$ such that $\|T(\tilde{y}^k)\| \leq \gamma$ for all k. Existence of γ follows from (i). Thus,

$$\lim_{k\to\infty}\sigma_k = \lim_{k\to\infty}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j} \geq \lim_{k\to\infty}\frac{1}{\gamma}\sum_{j=0}^{k}\beta_j = \infty, \tag{50}$$

using that $\eta_j = \max\{1, \|T(\tilde{y}^j)\|\} \leq \max\{1, \gamma\} \leq \gamma$ for all j in the first inequality and (25) in the last equality. Thus, taking limits in (49), we get $\lim_{k\to\infty}\operatorname{dist}(x^k, C) = 0$, establishing (iii).

iv) Follows from (iii). □

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

Theorem 1. *If $S(T, C) \neq \emptyset$ then all weak cluster points of the sequence $\{x^k\}$ generated by Algorithm A solve VIP(T, C).*

Proof. For any $x \in C$ we have

$$\begin{aligned}\|z^{j+1} - x\|^2 &= \left\|P_{C_j}\left(\tilde{y}^j - \frac{\beta_j}{\eta_j}T(\tilde{y}^j)\right) - P_{C_j}(x)\right\|^2 \leq \left\|\left(\tilde{y}^j - \frac{\beta_j}{\eta_j}T(\tilde{y}^j)\right) - x\right\|^2 \\ &= \|\tilde{y}^j - x\|^2 + \frac{\|T(\tilde{y}^j)\|^2}{\eta_j^2}\beta_j^2 - 2\frac{\beta_j}{\eta_j}\left\langle T(\tilde{y}^j), \tilde{y}^j - x\right\rangle \\ &\leq \|z^j - x\|^2 + \beta_j^2 + 2\frac{\beta_j}{\eta_j}\langle T(x), x - \tilde{y}^j\rangle,\end{aligned} \tag{51}$$

using Lemma 1(i) in the first inequality, and the monotonicity of T and (43) in the last inequality. Summing (51) from 0 to $k-1$ and dividing by σ_{k-1}, we obtain from Proposition 7(ii)

$$\frac{(\|z^k - x\|^2 - \|z^0 - x\|^2)}{\sigma_{k-1}} \quad \leq \quad \frac{\sum_{j=0}^{k-1} \beta_j^2}{\sigma_{k-1}} + \langle T(x), x - x^k \rangle. \tag{52}$$

Let $\hat{x}$ be a weak cluster point of $\{x^k\}$. Existence of $\hat{x}$ is guaranteed by Proposition 7(i). Note that $\hat{x} \in C$ by Proposition 7(iv).

By (25), (26), (50) and boundedness of $\{z^k\}$, taking limits in (52) with $k \to \infty$, we obtain that $\langle T(x), x - \hat{x} \rangle \geq 0$ for all $x \in C$. Using Lemma 3, we get that $\hat{x} \in S(T, C)$. Therefore, all weak cluster points of $\{x^k\}$ solve VIP(T, C). □

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

Theorem 2. *Define $x^* = \lim_{k\to\infty} P_{S(T,C)}(\tilde{y}^k)$. Then either $S(T, C) \neq \emptyset$ and $\{x^k\}$ converges weakly to x^*, or $S(T, C) = \emptyset$ and $\lim_{k\to\infty} \|x^k\| = \infty$.*

Proof. Assume that $S(T, C) \neq \emptyset$ and define $u^k = P_{S(T,C)}(\tilde{y}^k)$. Note that u^k, the orthogonal projection of $\tilde{y}^k$ onto $S(T, C)$, exists since the solution set $S(T, C)$ is nonempty by assumption, and closed and convex by Proposition 2(v). By Proposition 6, $\{\tilde{y}^k\}$ is quasi-Fejér convergent to $S(T, C)$. Therefore, it follows from Lemma 2(ii) that $\{P_{S(T,C)}(\tilde{y}^k)\}$ is strongly convergent. Let

$$x^* = \lim_{k\to\infty} P_{S(T,C)}(\tilde{y}^k) = \lim_{k\to\infty} u^k. \tag{53}$$

By Proposition 7(i) and Theorem 1, $\{x^k\}$ is bounded and each of its weak cluster points belong 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. It suffices to show that $\hat{x} = x^*$.

By Lemma 1(ii) we have that $\langle \hat{x} - u^j, \tilde{y}^j - u^j \rangle \leq 0$ for all j. Let $\xi = \sup_{0\leq j\leq\infty} \|\tilde{y}^j - u^j\|$. By Proposition 7(i), $\xi < \infty$. Then,

$$\langle \hat{x} - x^*, \tilde{y}^j - u^j \rangle \leq \langle u^j - x^*, \tilde{y}^j - u^j \rangle \leq \xi \|u^j - x^*\| \qquad \forall j. \tag{54}$$

Multiplying (54) by $\dfrac{\beta_j}{\eta_j \sigma_{k-1}}$ and summing from 0 to $k-1$, we get from Proposition 7(ii)

$$\left\langle \hat{x} - x^*, x^k - \frac{1}{\sigma_{k-1}} \sum_{j=0}^{k-1} \frac{\beta_j}{\eta_j} u^j \right\rangle \leq \frac{\xi}{\sigma_{k-1}} \sum_{j=0}^{k-1} \frac{\beta_j}{\eta_j} \|u^j - x^*\|. \tag{55}$$

Define $\zeta_{k,j} := \frac{1}{\sigma_k} \frac{\beta_j}{\eta_j}$ $(k \geq 0,\ 0 \leq j \leq k)$. In view of (50), $\lim_{k\to\infty} \zeta_{k,j} = 0$ for all j. By (33), $\sum_{j=0}^{k} \zeta_{k,j} = 1$ for all k. Then, using (53) and Proposition 3 with $w^k = \sum_{j=0}^{k} \zeta_{k,j} u^j = \frac{1}{\sigma_k} \sum_{j=0}^{k} \frac{\beta_j}{\eta_j} u^j$, we obtain that

$$\lim_{k\to\infty} \frac{1}{\sigma_{k-1}} \sum_{j=0}^{k-1} \frac{\beta_j}{\eta_j} u^j = x^*, \tag{56}$$

and hence

$$\lim_{k\to\infty} \frac{1}{\sigma_{k-1}} \sum_{j=0}^{k-1} \frac{\beta_j}{\eta_j} \|u^j - x^*\| = 0, \tag{57}$$

using the fact that $\frac{1}{\sigma_{k-1}}\sum_{j=0}^{k-1}\frac{\beta_j}{\eta_j}=1$.

From (56) and (57), since $\lim_{k\to\infty} x^{i_k}=\hat{x}$, taking limits in (55) with $k\to\infty$ along the subsequence with subindices $\{i_k\}$, we conclude that $\langle \hat{x}-x^*, \hat{x}-x^*\rangle \le 0$, implying that $\hat{x}=x^*$.

If $S(T,C)=\emptyset$ then by Theorem 1 no subsequence of $\{x^k\}$ can be bounded, and hence $\lim_{k\to\infty}\|x^k\|=\infty$. $\square$

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 $\mathrm{dist}(x,C)\le \tilde{g}(x)$ 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 (29) by $\tilde{g}(y^{k,j})$, and all our convergence results are preserved; in fact only the proof of Proposition 5(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 C has empty interior.

Acknowledgements

The first author was partially supported by PROCAD UFG/UnB/IMPA. The second author was partially supported by CNPq grant No. 301280-86.

References

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

[2] Alber, Ya.I., Iusem, A.N., Solodov, M.V. On the projected subgradient method for nonsmooth convex optimization in a Hilbert space. *Mathematical Programming* **81** (1998) 23-37.

[3] Alber, Ya.I., Iusem, A.N. Extension of subgradient techniques for nonsmooth optimization in Banach spaces. *Set-Valued Analysis* **9** (2001) 315-335.

[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] Bello Cruz, J.Y., Iusem, A.N. Convergence of direct methods for paramonotone variational inequalities. *Computational Optimization and Application* **46** (2010) 247-263.

[6] Bello Cruz, J.Y., Iusem, A.N. A strongly convergent direct method for monotone variational inequalities in Hilbert spaces. *Numerical Functional Analysis and Optimization* **30** (2009) 23-36.

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

[8] Ermoliev, Yu.M. On the method of generalized stochastic gradients and quasi-Fejér sequences. *Cybernetics* **5** (1969) 208-220.

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

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

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

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

[13] Iusem, A.N. On the convergence properties of the projected gradient method for convex optimization. *Computational and Applied Mathematics* **22** (2003) 37-52.

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

[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] Iusem, A.N., Svaiter, B.F., Teboulle, M. Entropy-like proximal methods in convex programming. *Mathematics of Operations Research* **19** (1994) 790-814.

[18] Khobotov, E.N. Modifications of the extragradient method for solving variational inequalities and certain optimization problems. *USSR Computational Mathematics and Mathematical Physics* **27**,5 (1987) 120-127.

[19] Konnov, I.V. A class of combined iterative methods for solving variational inequalities. *Journal of Optimization Theory and Applications* **94** (1997) 677-693.

[20] Konnov, I.V. A combined relaxation method for variational inequalities with nonlinear constraints. *Mathematical Programming* **80** (1998) 239-252.

[21] Konnov, I.V. *Combined relaxation methods for variational inequalities* . Springer-Velarg, Berlin (2001).

[22] Kinderlehrer, D., Stampacchia, G. *An Introduction to Variational Inequalities and Their Applications*. Academic Press, New York (1980).

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

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

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

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