An Explicit Algorithm for Monotone Variational Inequalities

J. Y. Bello Cruz* A. N. Iusem†

February 19, 2009

Abstract

We introduce a fully explicit method for solving monotone variational inequalities in Hilbert spaces, where orthogonal projections onto C 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 $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ 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 [10] in 1966. Variational inequalities have a wide range of applications. Several of them are described in [18]. An excellent survey of methods for finite dimensional variational inequality problems can be found in [9].

*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 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:

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

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

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 [12] 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

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

where $u^k \in T(x^k)$, and 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) 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 $\psi(0) = 0$.

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

Convergence results for scheme (3) have been established in [1] for the case of uniformly monotone operators, and in [6] 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 $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. 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 [19] 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 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 [19] 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, 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 [15] (see also [17] 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

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

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

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

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

$$x^{k+1} := P_C\left(x^k - \frac{\langle T(y^k), x^k - y^k\rangle}{\|T(y^k)\|^2}\, T(y^k)\right). \tag{10}$$

This strategy for determining the stepsizes guarantees convergence under the only assumptions of monotonicity and continuity of T and existence of solutions of $\mathrm{VIP}(T, C)$, without assuming Lipschitz continuity of T. Also, this algorithm demands only two projections onto C per iteration, unlike other variants, e.g. [13] and [20], with projections onto C inside the inner loop for the search of the stepsize. Other algorithms for $\mathrm{VIP}(T, C)$, less directly related to Korpelevich's method, can be found in [11], [22] and [24].

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)-(10), which demand in fact two orthogonal projections per iteration.

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

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

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 (11) is therefore rather general, because any system of inequalities $g_j(x)\le 0$ with $j\in J$, where all the g_j's are convex, may be represented as in (11) with $g(x)=\sup\{g_j(x) \,:\, j\in J\}$.

We studied in [6] a method for solving $\mathrm{VIP}(T,C)$ for the case in which T is point-to-set in a finite dimensional space, i.e. $\mathcal{H}=\mathbb{R}^n$, using the following relaxed iteration:

$$x^{k+1}=P_{C_k}\left(x^k-\frac{\beta_k}{\eta_k}u^k\right), \tag{12}$$

where $u^k\in T(x^k)$, $\eta_k=\max\{1,\|u^k\|\}$, β_k is an exogenous stepsize satisfying $\sum_{k=0}^{\infty}\beta_k=\infty$, $\sum_{k=0}^{\infty}\beta_k^2<\infty$, and C_k is defined as

$$C_k:=\{z\in\mathcal{H} \,:\, g(x^k)+\langle v^k,z-x^k\rangle\le 0\},$$

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

We proved that the sequence generated by (12) 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 $D\subseteq\mathcal{H}$ such that $\langle u,x-z\rangle\ge 0$ for all $x\notin D$ and for all $u\in T(x)$.

In this paper we will analyze a new algorithm for the case in which T is a point-to-point operator, relaxing the hypotheses in [6] in two directions: we assume plain monotonicity of T instead of paramonotonicity, and we don't need any coerciveness condition. Additionally, we obtain convergence results stronger them those in [6]; namely we get weak convergence of the whole sequence to some solution of $\mathrm{VIP}(T, C)$, 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)-(10)) 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.

We describe next our method and compare it with (6)-(10). In (6)-(10) 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 a 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 substitutes for the exact projection onto C, demanded in (4)-(5) and (6)-(10).

In connection with the method in [6], 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 [6]. Extensions of Korpelevich method to the of point-to-set case can be found in [14] 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 state our algorithm formally. 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 [23]. □

We next 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 $k_0 \geq 0$ 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 $k \geq k_0$.*

This definition originates in [8] and has been further elaborated in [16].

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 $\{\operatorname{dist}(x^k, S)\}$ is bounded, because $0 \leq \operatorname{dist}(x^k, S) \leq \|x^k - x\|$ for all $x \in S$, and $\{\|x^k - x\|\}$ converges for all $x \in S$, by Lemma 1(ii).

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

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

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 $\{\operatorname{dist}(x^k, S)\}$ coincide, i.e. that the sequence $\{\operatorname{dist}(x^k, S)\}$ converges.

ii) We will 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 get

$$\|u^q - u^p\|^2 = \|P_S(x^q) - P_S(u^p)\|^2 \leq \|x^q - u^p\|^2 - \|x^q - u^q\|^2. \tag{13}$$

Since $\{x^k\}$ is quasi-Fejér convergent to S and $p < q$, we get from (13) that

$$\begin{aligned} \|u^q - u^p\|^2 &\leq \|x^p - u^p\|^2 - \|x^q - u^q\|^2 + \sum_{j=p}^{q} \delta_j \\ &\leq \operatorname{dist}(x^p, S)^2 - \operatorname{dist}(x^q, S)^2 + \sum_{j=p}^{\infty} \delta_j. \end{aligned} \tag{14}$$

By (i), $\{\operatorname{dist}(x^k, S)\}$ converges, and using the fact $\sum_{j=0}^{\infty} \delta_j < \infty$, we obtain from (14) that $\{u^k\}$ is a Cauchy sequence.

□

We recall now 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 maximal monotone operators and monotone variational inequalities.

Lemma 3. *Let $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ be a maximal monotone operator and C a closed and convex set. Then*

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

ii) $G(T)$ is closed.

iii) If $\mathcal{H}$ is finite dimensional then T is bounded on bounded subsets of the interior of its domain.

iv) If T is point-to-point then T is continuous.

v) $S(T, C)$, if nonempty, is closed and convex.

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) using a compactness argument.

iv) See Theorem 4.6.3 of [7].

v) See Lemma 2.4(ii) of [5].

$\square$

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 4. *Consider VIP(T,C). If $T : \mathcal{H} \to \mathcal{H}$ is maximal monotone and point-to-point, then*

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

Proof. By Lemma 3(iv), if T is point-to-point and maximal monotone then T is continuous. 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

$$0 \leq \langle T((1-\alpha)x + \alpha y), y - x\rangle, \tag{15}$$

for all $\alpha \in (0,1)$. Making $\alpha \to 0$ and using the continuity of T, we obtain from (15) that $\langle T(x), y - x\rangle \geq 0$, for all $y \in C$, i.e. $x \in S(T,C)$. $\square$

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

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

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

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

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

Thus, $x_\lambda \in C$ and

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

$\square$

We will also need the following elementary result on sequence averages.

Proposition 2. *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* $\lim \zeta_{k,j} = 0$ *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 C is of the form given in (11) and satisfies Slater's condition, and that T is point-to-point and maximal monotone. 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 in view of Lemma 3(iii), because T and ∂g are maximal monotone operators. 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 $\{\beta_k\} \subseteq \mathbb{R}_{++}$ satisfying

$$\sum_{k=0}^{\infty} \beta_k = \infty, \tag{16}$$

$$\sum_{k=0}^{\infty} \beta_k^2 < \infty. \tag{17}$$

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 $g(z^k) \le 0$ 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 $j = 0, 1, \dots$, let

$$C_{k,j} := \{z \in \mathcal{H} \,:\, g(y^{k,j}) + \langle v^{k,j}, z - y^{k,j}\rangle \le 0\}, \tag{18}$$

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

$$y^{k,j+1} := P_{C_{k,j}}(y^{k,j}). \tag{19}$$

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

$$j(k) := \min\left\{ j \ge 0 : \frac{g(y^{k,j})\, \|y^{k,j} - w\|}{g(y^{k,j}) - g(w)} \le \theta\beta_k \right\}. \tag{20}$$

Let

$$\tilde{y}^k := y^{k,j(k)}. \tag{21}$$

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

$$C_k := C_{k,j(k)} = \{z \in \mathcal{H} \,:\, g(\tilde{y}^k) + \langle \tilde{v}^k, z - \tilde{y}^k\rangle \le 0\}\,. \tag{22}$$

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

$$z^{k+1} := P_{C_k}\left(\tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k)\right). \tag{23}$$

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

$$\sigma_k := \sum_{j=0}^{k} \frac{\beta_j}{\eta_j} = \sigma_{k-1} + \frac{\beta_k}{\eta_k}, \tag{24}$$

$$x^{k+1} := \left(1 - \frac{\beta_k}{\eta_k \sigma_k}\right) x^k + \frac{\beta_k}{\eta_k \sigma_k} \tilde{y}^k. \tag{25}$$

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

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

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

Proof. See Proposition 3.1 in [21]. □

It follows from Proposition 3, (18), (19), (22) and (23) that

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

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

The iteration formulae of the algorithm become more explicit in the smooth case, i.e. when C is of the form $C = \{x \in \mathcal{H} : g_i(x) \le 0,\ 1 \le i \le m\}$ where the g_i's are convex and Gateaux 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

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

$$v^k = \nabla g_{\ell(k)}(\tilde{y}^k), \qquad \text{with} \quad \ell(k) \in \arg\max_{0\le i\le m}\{g_i(\tilde{y}^k)\},$$

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.

4 Convergence analysis of Algorithm A

For convergence of our method, we assume that T is point-to-point and maximal monotone, and hence continuous by Lemma 3(iv). Observe that $\partial g(x) \ne \emptyset$ 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 4. *Take C, $C_{k,j}$, C_k, $\tilde{y}^k$, z^k and x^k defined by* (11), (18), (22), (21), (23) *and* (25) *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 (18), (22) and the definition of 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 (23), using Lemma 1(ii) with $x = \tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k)$ and $K = C_k$, we obtain

$$\left\langle z^{k+1} - \left(\tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k) \right), z - z^{k+1} \right\rangle \geq 0 \quad \forall z \in C_k. \tag{26}$$

Taking $z^{k+1} = \tilde{y}^k$ in (26) 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 get an infinite sequence $\{y^{k,j}\}_{j=0}^{\infty}$ such that

$$\liminf_{j \to \infty} \frac{g(y^{k,j})\, \|y^{k,j} - w\|}{g(y^{k,j}) - g(w)} \geq \theta \beta_k > 0. \tag{27}$$

Taking into account the inner loop in j given in (21) i.e. $y^{k,j+1} = P_{C_{k,j}}(y^{k,j})$ for each k, we obtain, for each $x \in C$,

$$\|y^{k,j+1} - x\|^2 = \|P_{C_{k,j}}(y^{k,j}) - P_{C_{k,j}}(x)\|^2 \leq \|y^{k,j} - x\|^2 - \|y^{k,j+1} - y^{k,j}\|^2 \leq \|y^{k,j} - x\|^2, \tag{28}$$

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 $\tau := \dfrac{1}{-g(w)} \sup\limits_{0 \leq j \leq \infty} \|y^{k,j} - w\|$ is finite and also,

$$g(y^{k,j}) > 0 \qquad \text{for all} \quad j. \tag{29}$$

Using (28), we get

$$\lim_{j \to \infty} \|y^{k,j+1} - y^{k,j}\| = \lim_{j \to \infty} \|P_{C_{k,j}}(y^{k,j}) - y^{k,j}\| = 0. \tag{30}$$

Since $y^{k,j+1}$ belongs to $C_{k,j}$, we have from (18) that

$$g(y^{k,j}) \leq \langle v^{k,j}, y^{k,j} - y^{k,j+1} \rangle \leq \|v^{k,j}\| \|y^{k,j} - y^{k,j+1}\|, \tag{31}$$

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 from Lemma 3(iii) that $\{\|v^{k,j}\|\}_{j=0}^{\infty}$ is bounded. In view of (30) and (31),

$$\liminf_{j \to \infty} g(y^{k,j}) \leq 0. \tag{32}$$

It follows from (29) and (32) that

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

contradicting (27). 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 5. *If $S(T,C)$ is nonempty, then $\{\tilde{y}^k\}$ and $\{z^k\}$ are quasi-Fejér convergent to $S(T,C)$.*

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

$$\frac{1}{\eta_k} \leq 1 \tag{33}$$

and

$$\frac{\|T(\tilde{y}^k)\|}{\eta_k} \leq 1. \tag{34}$$

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

$$\begin{aligned}
\|\tilde{y}^k - \bar{x}\| &= \|y^{k,j(k)} - \bar{x}\| = \|P_{C_{k,j(k)-1}}(y^{k,j(k)-1}) - P_{C_{k,j(k)-1}}(\bar{x})\| \leq \|y^{k,j(k)-1} - \bar{x}\| \\
&= \|P_{C_{k,j(k)-2}}(y^{k,j(k)-2}) - P_{C_{k,j(k)-2}}(\bar{x})\| \leq \|y^{k,j(k)-2} - \bar{x}\| \leq \cdots \leq \|z^k - \bar{x}\|.
\end{aligned} \tag{35}$$

Let $\tilde{\theta} = 1 + \theta\, T(\bar{x}) \geq 1 + \theta \dfrac{T(\bar{x})}{\eta_k}$, by (33). Then

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

$$
\begin{aligned}
&= \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 - 2\,\frac{\beta_k}{\eta_k}\Big(\langle T(\bar{x}), \tilde{y}^k - P_C(\tilde{y}^k)\rangle + \langle T(\bar{x}), P_C(\tilde{y}^k) - \bar{x}\rangle\Big) \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 + 2\,\frac{\beta_k}{\eta_k}\langle T(\bar{x}), P_C(\tilde{y}^k) - \tilde{y}^k\rangle \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 + \frac{\beta_k}{\eta_k}\|T(\bar{x})\|\|P_C(\tilde{y}^k) - \tilde{y}^k\| \leq \|\tilde{y}^k - \bar{x}\|^2 + \tilde{\theta}\beta_k^2 \\
&\leq \|z^k - \bar{x}\|^2 + \tilde{\theta}\beta_k^2, \qquad (36)
\end{aligned}
$$

using (35) in the first inequality, Lemma 1(i) in the second one, the monotonicity of T and (34) 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 one, and (35) in the last one.

Using Definition 2, (36) and (17), we conclude that the sequences $\{\tilde{y}^k\}$ and $\{z^k\}$ are quasi-Fejér convergent to $S(T,C)$. □

Proposition 6. *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) $\{\tilde{y}^k\}$, $\{x^k\}$ *and* $\{T(\tilde{y}^k)\}$ *are bounded,*

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

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

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

Proof. i) For $\{\tilde{y}^k\}$ use Proposition 5 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 (25).

ii) Apply (25) recursively.

iii) It follows from Lemma 5 and (20)-(21) that

$$\mathrm{dist}(\tilde{y}^k, C) \leq \theta\beta_k. \qquad (37)$$

Define

$$\tilde{x}^{k+1} := \frac{1}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}P_C(\tilde{y}^j). \qquad (38)$$

Since $\frac{1}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}=1$ by (24), we get from the convexity of C that $\tilde{x}^{k+1}\in C$. Let

$$\tilde{\beta}:=\sum_{j=0}^{\infty}\beta_j^2. \tag{39}$$

Note that $\tilde{\beta}$ is finite by (17). Then

$$\begin{aligned}\operatorname{dist}(x^{k+1},C) &\le \|x^{k+1}-\tilde{x}^{k+1}\| = \left\|\frac{1}{\sigma_k}\left(\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}(\tilde{y}^j-P_C(\tilde{y}^j))\right)\right\| \le \frac{1}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}\|\tilde{y}^j-P_C(\tilde{y}^j)\|\\ &= \frac{1}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}\operatorname{dist}(\tilde{y}^j,C)\le\frac{\theta}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j^2}{\eta_j}\le\frac{\theta}{\sigma_k}\sum_{j=0}^{k}\beta_j^2\le\theta\frac{\tilde{\beta}}{\sigma_k},\end{aligned} \tag{40}$$

using the fact that $\tilde{x}^{k+1}$ belongs to C in the first inequality, (ii) and (38) in the first equality, convexity of $\|\cdot\|$ in the second inequality, (37) in the third one, (33) in the fourth one and (39) in the last one.

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

$$\lim_{k\to\infty}\sigma_k=\lim_{k\to\infty}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}\ge\lim_{k\to\infty}\frac{1}{\gamma}\sum_{j=0}^{k}\beta_j=\infty, \tag{41}$$

using that $\eta_j=\max\{1,\|T(\tilde{y}^j)\|\}\le\max\{1,\gamma\}\le\gamma$ for all j in the first inequality and (16) in the last equality. Thus, taking limits in (40), 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

$$\begin{aligned}\|z^{j+1}-x\|^2 &= \left\|P_{C_j}\left(\tilde{y}^j-\frac{\beta_j}{\eta_j}T(\tilde{y}^j)\right)-P_{C_j}(x)\right\|^2\le\left\|\left(\tilde{y}^j-\frac{\beta_j}{\eta_j}T(\tilde{y}^j)\right)-x\right\|^2\\ &= \|\tilde{y}^j-x\|^2+\frac{\|T(\tilde{y}^j)\|^2}{\eta_j^2}\beta_j^2-2\frac{\beta_j}{\eta_j}\langle T(\tilde{y}^j),\tilde{y}^j-x\rangle\\ &\le \|z^j-x\|^2+\beta_j^2+2\frac{\beta_j}{\eta_j}\langle T(x),x-\tilde{y}^j\rangle,\end{aligned} \tag{42}$$

using Lemma 1(i) in the first inequality, and the monotonicity of T and (34) in the last inequality. Summing (42) from 0 to $k-1$ and dividing by σ_{k-1}, we obtain from Proposition 6(ii)

$$\frac{(\|z^k - x\|^2 - \|z^0 - x\|^2)}{\sigma_{k-1}} \quad \leq \quad \frac{\sum_{j=0}^{k-1} \beta_j^2}{\sigma_{k-1}} + \langle T(x), x - x^k \rangle. \tag{43}$$

Let $\hat{x}$ be a weak cluster point of $\{x^k\}$. Existence of $\hat{x}$ is guaranteed by Proposition 6(i). Note that $\hat{x} \in C$ by Proposition 6(iv).

By (16), (17), (41) and boundedness of $\{z^k\}$, taking limits in (43) we obtain that $\langle T(x), x - \hat{x} \rangle \geq 0$ for all $x \in C$. Using Lemma 4 with $k \to \infty$ 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 because the solution set $S(T, C)$ is nonempty by assumption, and closed and convex by Lemma 3(v). By Proposition 5, $\{\tilde{y}^k\}$ 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

$$x^* = \lim_{k\to\infty} P_{S(T,C)}(\tilde{y}^k) = \lim_{k\to\infty} u^k. \tag{44}$$

By Proposition 6(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 6(i), $\xi < \infty$. Then,

$$\langle \hat{x} - x^*, \tilde{y}^j - u^j \rangle \leq \langle u^j - x^*, \tilde{y}^j - u^j \rangle \leq \xi \, \|u^j - x^*\| \qquad \forall j. \tag{45}$$

Multiplying (45) by $\dfrac{\beta_j}{\eta_j \sigma_{k-1}}$ and summing from 0 to $k-1$, we get from Proposition 6(ii)

$$\left\langle \hat{x} - x^*, x^k - \frac{1}{\sigma_{k-1}} \sum_{j=0}^{k-1} \frac{\beta_j}{\eta_j} u^j \right\rangle \leq \frac{\xi}{\sigma_{k-1}} \sum_{j=0}^{k-1} \frac{\beta_j}{\eta_j} \|u^j - x^*\|. \tag{46}$$

Define $\zeta_{k,j} := \frac{1}{\sigma_k} \frac{\beta_j}{\eta_j}$ $(k \geq 0,\ 0 \leq j \leq k)$. In view of (41), $\lim_{k\to\infty} \zeta_{k,j} = 0$ for all j. By (24), $\sum_{j=0}^{k} \zeta_{k,j} = 1$ for all k. Then, using (44) and Proposition 2 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

$$\lim_{k\to\infty} \frac{1}{\sigma_{k-1}} \sum_{j=0}^{k-1} \frac{\beta_j}{\eta_j} u^j = x^*, \tag{47}$$

and hence

$$\lim_{k\to\infty} \frac{1}{\sigma_{k-1}} \sum_{j=0}^{k-1} \frac{\beta_j}{\eta_j} \|u^j - x^*\| = 0, \tag{48}$$

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

¿From (47) and (48), since $\lim_{k\to\infty} x^{i_k} = \hat{x}$, taking limits in (46) with $k \to \infty$ along the subsequence with subindices $\{i_k\}$, we conclude that $\langle \hat{x} - x^*, \hat{x} - x^* \rangle \leq 0$, implying that $\hat{x} = x^*$.

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

□

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 grant N^o 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. A strongly convergent direct method for monotone variational inequalities in Hilbert spaces. *Numerical Functional Analysis and Optimization* **30** (2009) 23-36.

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

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

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

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

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

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

[15] 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.

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

[17] 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.

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

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

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

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

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

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

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