Convergence of Direct Methods for Paramonotone Variational Inequalities

J. Y. Bello Cruz* A. N. Iusem†

October 16, 2008

Abstract

We analyze one-step direct methods for variational inequality problems, establishing convergence under paramonotonicity of the operator. Previous results on the method required much more demanding assumptions, like strong or uniform monotonicity, implying uniqueness of solution, which is not the case for our approach.

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 [11] in 1966. An excellent survey of methods for finite dimensional variational inequality problems can be found in [7].

*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

1.1 Direct methods for VIP(T,C)

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

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

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

with $\alpha_k > 0$ for all k. The coefficients α_k are called stepsizes and $P_C : \mathbb{R}^n \to C$ 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-point, is the iterative procedure given by

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

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

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

Definition 1. *Consider $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ 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.

It has been proved in [8] that when T is strongly monotone and Lipschitz continuous, i.e. there exists $L > 0$ such that $\|T(x) - T(y)\| \leq L\|x - y\|$ for all $x, y \in \mathbb{R}^n$, then the scheme (3)-(4) converges to the unique solution of VIP(T, C), provided that $\alpha_k \in \left(\epsilon, \frac{2\omega}{L^2}\right)$ for all k and for some $\epsilon > 0$.

Y. Alber extended this method in three directions: he considers point-to-set operators, works in a general Hilbert space, and demands uniform monotonicity of T instead of strong monotonicity. Under these assumptions he proved that the iterative procedure given by

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

where $u^k \in T(x^k)$, and the sequence α_k satisfies some conditions related to ψ, is strongly convergent to a solution of $\mathrm{VIP}(T,C)$, see [1].

These results are somewhat undesirable for several reasons. The hypotheses of strong or uniform monotonicity are too demanding, since they imply uniqueness of the solution of $\mathrm{VIP}(T,C)$. Even if they hold, it may happen that the constants ω, L or the function ψ are not known "a priori", and even when they are known they may lead to too small estimates of the stepsize α_k, entailing a slow convergence of the method.

We remark that there is no chance to relax the assumption on T to plain monotonicity. For example, consider $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $T(x) = Ax$, with $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. T is monotone and the unique solution of $\mathrm{VIP}(T,C)$ 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 (5) moves away from the solution, independently of the choice of the stepsize α_k.

Thus, we cannot proceed much further unless we impose a condition stronger than monotonicity of T. In this paper we assume paramonotonicity of T. The notion of paramonotonicity, which is in-between monotonicity and strict monotonicity, was introduced in [6], and many of its properties were established in [12]. 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 [12].

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 [12].

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, in which case $S(T,\mathbb{R}^n)$ 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.

In Section 3, we consider Algorithm (5). Assuming that T is a paramonotone operator, we obtain that the sequence generated by (5) is globally convergent to some point on $S(T,C)$, if $S(T,C)$ is nonempty and the stepsizes $\{\alpha_k\}$ satisfy: $\alpha_k = \frac{\beta_k}{\eta_k}$ where $\eta_k = \max\{1, \|u^k\|\}$, and

$$\sum_{k=0}^{\infty} \beta_k = \infty, \tag{6}$$

and

$$\sum_{k=0}^{\infty} \beta_k^2 < \infty. \tag{7}$$

This selection rule has been considered several times for similar methods, see [14], [3] and [2]. Methods of this type, like (5), are called direct, because they do not require the solution of subproblems at each iteration, and it is easy to compute x^{k+1} using only the previous point x^k.

1.2 Relaxed projection methods

The method given by (5) 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. Observe that projections onto halfspaces are easily computable.

We consider the case in which C is of the form

$$C = \{z \in \mathbb{R}^n \, : \, g(z) \le 0\}, \tag{8}$$

where $g : \mathbb{R}^n \to \mathbb{R}$ is a convex function, satisfying Slater's condition, i.e. there exists a point $\hat{x}$ such that $g(\hat{x}) < 0$. The differentiability of g is not assumed and the representation (8) 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 (8) with $g(x) = \sup\{g_j(x) \, : \, j \in J\}$.

M. Fukushima introduced in [10] a method for solving VIP(T, C), using the following relaxed iteration:

$$x^{k+1} = P_{C_k}\left(x^k - \beta_k \frac{T(x^k)}{\|T(x^k)\|}\right), \tag{9}$$

where β_k is an exogenous stepsize satisfying (6)-(7) and C_k is defined as

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

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

He proved convergence of $\{x^k\}$ to a point in $S(T, C)$, under quite demanding assumptions: T must be strongly monotone and it must satisfy the following coerciveness condition:

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

In this paper we will analyze the algorithm given by (9), except that T is a point-to-set paramonotone operator, and with the following condition, instead of (P):

(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 $u \in T(x)$.

It is clear that condition (Q) is weaker than (P). The following two conditions are sufficient for establishing (Q):

i) T is monotone and there exists $x^* \in C$ such that $0 \in T(x^*)$, (we can take $z = x^*$ in condition (Q)).

ii) T is uniformly monotone and ψ satisfies $\lim_{t\to\infty} \frac{\psi(t)}{t} = \infty$. Indeed, we have, in view of Definition 1(iv), $\langle u, x - z \rangle \geq \langle v, x - z \rangle + \psi(\|x - z\|) \geq \|x - z\| \left(\frac{\psi(\|x-z\|)}{\|x-z\|} - \|v\| \right)$ for all (x, u), $(z, v) \in G(T)$, so that (Q) holds for any $z \in C$, taking as D a large enough ball centered at z.

It is easy to check that if T is Lipschitz continuous and uniformly monotone with $\lim_{t\to\infty} \frac{\psi(t)}{t} = \infty$, then property (Q) holds. As an example of an operator satisfying (Q) but not (P), take $T(x) = x - P_L(x)$ where $L \subset \mathbb{R}^n$ is a subspace, and C such that $C \cap L \neq \emptyset$. T is paramonotone, because $T(x) = \nabla f(x)$ with $f(x) = (\text{dist}(x, C))^2$, and satisfies (Q) because the points in $C \cap L$ are zeroes of T. It can be easily shown that T does not satisfy (P).

In Section 4, we analyze the method given by (9), relaxing the hypotheses in [10] in three directions: T can be point-to-set, we assume paramonotonicity of T instead of strong monotonicity, and use (Q) instead of (P). Under these conditions, we prove that the sequence generated by (9) is bounded, the difference between consecutive iterates converges to zero, and all its cluster points belong to $S(T, C)$.

2 Preliminary results

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

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 [4]. □

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

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

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

Lemma 1. *There exists* $\kappa \in [0, 1)$ *such that* $\operatorname{dist}(\tilde{z}(x, v), C) \le \kappa \operatorname{dist}(x, C)$ *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 [9]. □

Lemma 2. *Take* $\{\xi_k\}, \{\nu_k\} \subset \mathbb{R}_+$ *and* $\mu \in [0, 1)$*. If the inequalities*

$$\xi_{k+1} \le \mu \xi_k + \nu_k, \quad k \in \mathbb{N}$$

hold and $\lim_{k\to\infty} \nu_k = 0$*, then* $\lim_{k\to\infty} \xi_k = 0$.

Proof. See Lemma 2 in [10]. □

Lemma 3. *Take* $K \subset \mathbb{R}^n$ *closed and* $z \notin K$*. Let* $\{z^k\} \subseteq \mathbb{R}^n$ *be such that* $\lim_{k\to\infty} \|z^{k+1} - z^k\| = 0$ *and both* z *and some point in* K *are cluster points of* $\{z^k\}$*. Then there exist* $\zeta > 0$ *and a subsequence* $\{z^{j_k}\}$ *of* $\{z^k\}$ *such that*

$$\operatorname{dist}(z^{j_k+1}, K) > \operatorname{dist}(z^{j_k}, K) \tag{10}$$

and

$$\operatorname{dist}(z^{j_k}, K) > \zeta. \tag{11}$$

Proof. Let $\zeta = \frac{1}{3} \operatorname{dist}(z, K) > 0$ and define

$$U := \{x \in \mathbb{R}^n \; : \; \operatorname{dist}(x, K) \le 2\,\zeta\} \,. \tag{12}$$

Clearly, there exists a subsequence $\{z^{j_k}\}$ of $\{z^k\}$ such that $z^{j_k} \in U$, $z^{j_k+1} \notin U$. Otherwise either $\{z^k\}$ eventually remains out of U, in which case $\operatorname{dist}(z^k, K) > 2\,\zeta$ for large k and then $\{z^k\}$ cannot have a cluster point belonging to K, or it eventually remains in U, in which case all its cluster points, including z, belong to U, but $\operatorname{dist}(z, K) = 3\,\zeta$, contradicting the definition of U given in (12). Thus, $\operatorname{dist}(z^{j_k+1}, K) > 2\,\zeta \ge \operatorname{dist}(z^{j_k}, K)$ by definition of U, so that (10) holds.

Since $\lim_{k\to\infty} \|z^{k+1} - z^k\| = 0$ there exists $\tilde{k} \ge 0$ such that $\|z^{j_k+1} - z^{j_k}\| < \zeta$ for all $k \ge \tilde{k}$, so that

$$\operatorname{dist}(z^{j_k}, K) \ge \operatorname{dist}(z^{j_k+1}, K) - \|z^{j_k+1} - z^{j_k}\| > 2\,\zeta - \zeta = \zeta$$

for all $k \ge \tilde{k}$, and hence $\{z^{j_k}\}_{k \ge \tilde{k}}$ satisfies (10) and (11). □

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

Lemma 4. *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) $\|P_K(x) - P_K(y)\| \leq \|x - y\|$.

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

Proof. See Lemma 4.1 in [13]. □

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 5. *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 [5].

ii) See Proposition 4.2.1(ii) of [5].

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 [6]. □

Lemma 6. *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. Suppose that there exist $x \in S(T,C)$ and a subsequence $\{\gamma_{j_k}(x)\}$ of $\{\gamma_k(x)\}$ such that $\lim_{k\to\infty} \gamma_{j_k}(x) \leq 0$. Let (x^*,u^*) be a cluster point of the bounded subsequence $\{(x^{j_k},u^{j_k})\}$. Since T is maximal monotone, $u^* \in T(x^*)$ by Lemma 5(ii). Without loss of generality we assume that $\lim_{k\to\infty}(x^{j_k},u^{j_k}) = (x^*,u^*)$. Therefore, $\lim_{k\to\infty} \gamma_{j_k}(x) = \lim_{k\to\infty}\langle u^{j_k}, x^{j_k} - x\rangle = \langle u^*, x^* - x\rangle \leq 0$. Since $x \in S(T,C)$, there exists $u \in T(x)$ such that $\langle u, x^* - x\rangle \geq 0$, and using the monotonicity of T we obtain

$$0 \geq \lim_{k\to\infty} \gamma_{j_k}(x) = \langle u^*, x^* - x\rangle \geq \langle u, x^* - x\rangle \geq 0. \tag{13}$$

It follows from (13) that $\langle u^*, x^* - x\rangle = 0$, and we conclude from Proposition 2 that $x^* \in S(T,C)$. $\square$

3 A direct method for VIP(T,C) with paramonotone operators

In this section we assume that T is maximal monotone and paramonotone.

Our algorithm requires an exogenous sequence $\{\beta_k\} \subset \mathbb{R}_{++}$ of stepsizes satisfying (6)-(7). The algorithm is defined as:

Algorithm 1
Initialization step:

$$x^0 \in C.$$

Iterative step: Given x^k, if $0 \in T(x^k)$ then stop. Otherwise, take $u^k \in T(x^k)$, $u^k \neq 0$, $\eta_k := \max\{1, \|u^k\|\}$ and

$$x^{k+1} = P_C\left(x^k - \frac{\beta_k}{\eta_k}u^k\right). \tag{14}$$

3.1 Convergence analysis of Algorithm 1

We start by proving the quasi-Fejér properties of the sequence $\{x^k\}$ generated by Algorithm 1.

Proposition 3. *If Algorithm 1 generates an infinite sequence $\{x^k\}$ and $S(T,C)$ is nonempty, then:*

i) $\{x^k\}$ is quasi-Fejér convergent to $S(T,C)$.

ii) If a cluster point of $\{x^k\}$ belongs to $S(T,C)$ then $\{x^k\}$ converges to a point in $S(T,C)$.

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

$$\langle \bar{u}, x - \bar{x}\rangle \geq 0 \quad \forall\, x \in C. \tag{15}$$

Then,

$$
\begin{aligned}
\|x^{k+1}-\bar{x}\|^2 &= \left\|P_C\left(x^k-\frac{\beta_k}{\eta_k}u^k\right)-P_C(\bar{x})\right\|^2 \le \left\|\left(x^k-\frac{\beta_k}{\eta_k}u^k\right)-\bar{x}\right\|^2 \\
&= \|x^k-\bar{x}\|^2+\beta_k^2-2\frac{\beta_k}{\eta_k}\,\langle u^k, x^k-\bar{x}\rangle \\
&= \|x^k-\bar{x}\|^2+\beta_k^2-2\frac{\beta_k}{\eta_k}\Big(\langle u^k-\bar{u}, x^k-\bar{x}\rangle+\langle \bar{u}, x^k-\bar{x}\rangle\Big) \\
&\le \|x^k-\bar{x}\|^2+\beta_k^2-\,2\,\frac{\beta_k}{\eta_k}\langle \bar{u}, x^k-\bar{x}\rangle \le \|x^k-\bar{x}\|^2+\beta_k^2,
\end{aligned}
$$

using (14), the fact that $\bar{x} \in S(T, C)$ and Lemma 4(i) in the first inequality, the monotonicity of T in the second inequality and (15) in the third inequality.

Using Definition 2 we conclude, in view of the fact that β_k satisfies (7), that the sequence $\{x^k\}$ is quasi-Fejér convergent to $S(T, C)$.

ii) Follows from (i) and Proposition 1(ii).

□

Corollary 1. *The sequences $\{x^k\}$, $\{u^k\}$ generated by Algorithm 1 are bounded.*

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

Paramonotonicity of T is used for the first time in the following theorem, which is our main convergence result on Algorithm 1.

Theorem 1. *Assume that T is maximal monotone and paramonotone. If $S(T, C)$ is nonempty then either Algorithm 1 stops at some iteration k, in which case $x^k \in S(T, C)$, or it generates an infinite sequence which converges to some $x^* \in S(T, C)$.*

Proof. If the algorithm stops at iteration k, i.e. $0 \in T(x^k)$, the result follows from the definition of $S(T, C)$. Therefore, we assume that the sequence $\{x^k\}$ is infinite. By Corollary 1, $\{x^k\}$ has cluster points. We claim that there exists a cluster point of $\{x^k\}$ belonging to $S(T, C)$. Otherwise, in view of Lemma 6, for each $\bar{x} \in S(T, C)$ there exists $\bar{k} \ge 0$ and $\rho > 0$ such that $\gamma_k(\bar{x}) = \langle u^k, x^k - \bar{x}\rangle \ge \rho$ for all $k \ge \bar{k}$. Fix some $\bar{x} \in S(T, C)$, nonempty by hypothesis, and consider the corresponding ρ, $\bar{k}$. Since $\{u^k\}$ is bounded by Corollary 1, there exists $\theta > 1$ such that $\|u^k\| \le \theta$ for all k. Therefore

$$\eta_k = \max\{1, \|u^k\|\} \le \max\{1, \theta\} = \theta \quad \forall k, \tag{16}$$

and since $\lim_{k\to\infty} \beta_k = 0$, we can find $\bar{k} \ge 0$ such that

$$\beta_k \le \frac{\rho}{\theta} \quad \text{and} \quad \langle u^k, x^k - \bar{x}\rangle \ge \rho \qquad \forall k \ge \bar{k}. \tag{17}$$

Thus, for all $k \geq \bar{k}$,

$$\begin{aligned}
\|x^{k+1} - \bar{x}\|^2 &= \left\| P_C\left(x^k - \frac{\beta_k}{\eta_k} u^k\right) - P_C(\bar{x}) \right\|^2 \leq \left\| \left(x^k - \frac{\beta_k}{\eta_k} u^k\right) - \bar{x} \right\|^2 \\
&= \|x^k - \bar{x}\|^2 + \beta_k^2 - 2\frac{\beta_k}{\eta_k} \langle u^k, x^k - \bar{x} \rangle \\
&\leq \|x^k - \bar{x}\|^2 - 2\beta_k \frac{\rho}{\theta} + \beta_k^2 = \|x^k - \bar{x}\|^2 - \beta_k \left(2\frac{\rho}{\theta} - \beta_k\right) \\
&\leq \|x^k - \bar{x}\|^2 - \beta_k \frac{\rho}{\theta}, \qquad (18)
\end{aligned}$$

using Lemma 4(i) in the first inequality, with $K = C$, $x = x^k - \frac{\beta_k}{\eta_k} u^k$ and $y = \bar{x}$, (16) in the second one and (17) in the third one. It follows from (18) that

$$\frac{\rho}{\theta} \beta_k \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2. \qquad (19)$$

Summing (19) with k between $\bar{k}$ and m,

$$\frac{\rho}{\theta} \sum_{k=\bar{k}}^{m} \beta_k \leq \sum_{k=\bar{k}}^{m} \left(\|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2 \right) = \|x^{\bar{k}} - \bar{x}\|^2 - \|x^{m+1} - \bar{x}\|^2 \leq \|x^{\bar{k}} - \bar{x}\|^2. \qquad (20)$$

Taking limits in (20) with $m \to \infty$, we contradict the assumption $\sum_{k=0}^{\infty} \beta_k = \infty$. Thus, there exists a cluster point of $\{x^k\}$ belonging to $S(T, C)$. In view of Proposition 3(ii), $\{x^k\}$ converges to a point in $S(T, C)$. □

4 A relaxed projection algorithm

In this section, we introduce an algorithm which eliminates the projection onto C. We assume that C is of the form given in (8). The algorithm is defined as follows.

Algorithm 2
Initialization step:

$$x^0 \in C.$$

Iterative step: Given x^k, if $0 \in T(x^k)$ then stop. Otherwise, take $u^k \in T(x^k)$, $u^k \neq 0$, choose $\eta_k := \max\{1, \|u^k\|\}$, $v^k \in \partial g(x^k)$ and let

$$C_k := \{z \in \mathbb{R}^n \ : \ g(x^k) + \langle v^k, z - x^k \rangle \leq 0\}. \qquad (21)$$

Compute

$$x^{k+1} = P_{C_k}\left(x^k - \frac{\beta_k}{\eta_k} u^k\right), \qquad (22)$$

with β_k satisfying (6)-(7). If $x^{k+1} = x^k$ then stop.

Unlike other projection methods, Algorithm 2 generates a sequence $\{x^k\}$ which is not necessarily contained in the set C. Note that Algorithm 2 can be easily implemented, because P_{C_k} has the following explicit formula.

Proposition 4. *For any $y \in \mathbb{R}^n$,*

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

Proof. See Proposition 3.1 in [15]. □

It follows from Proposition 4 that

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

so that Algorithm 2 can be considered as a fully direct method for VIP(T, C).

4.1 Convergence analysis of Algorithm 2

For convergence of our method, we assume that T is maximal monotone, paramonotone and satisfies Condition (Q), with we repeat here:

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

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

Proposition 5. *Take C, C_k and x^k defined by (8), (21) and (22) respectively. Then*

i) $C \subseteq C_k$ for all k.

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

Proof. i) It follows from (21) and the definition of subgradient.

ii) Suppose that $x^{k+1} = x^k$. Then , since $x^{k+1} \in C_k$, we have $g(x^k) + \langle v^k, x^{k+1} - x^k \rangle = g(x^k) \leq 0$, i.e. $x^k \in C$. Moreover, since x^{k+1} is given by (22), using Lemma 4(ii) with $x = x^k - \frac{\beta_k}{\eta_k}u^k$ and $K = C_k$, we obtain

$$\left\langle x^{k+1} - \left(x^k - \frac{\beta_k}{\eta_k}u^k\right), z - x^{k+1} \right\rangle \geq 0 \quad \forall z \in C_k. \tag{23}$$

Taking $x^{k+1} = x^k$ in (23) and taking into account the facts that $\beta_k > 0$, $\eta_k \geq 1$ for all k, and $C \subseteq C_k$, we get $\langle u^k, z - x^k \rangle \geq 0$ for all $z \in C$. Since $u^k \in T(x^k)$, we conclude that $x^k \in S(T, C)$.

□

Lemma 7. *Take $\hat{z} \in C$ and D satisfying (Q), let $\{x^k\}$ be the sequence generated by Algorithm 2 and choose $\lambda > 0$ such that $\|x^0 - \hat{z}\| \leq \lambda$ and $D \subseteq B(\hat{z}, \lambda)$. Then,*

i) if $x^k \in D$ then $\|x^{k+1} - \hat{z}\|^2 \leq \lambda^2 + \beta_k^2 + 2\beta_k\lambda$,

ii) if $x^k \notin D$ then $\|x^{k+1} - \hat{z}\|^2 \leq \|x^k - \hat{z}\|^2 + \beta_k^2$.

Proof. Since $\hat{z} \in C$, we get from Proposition 5(i) that $\hat{z} \in C_k$ for all k, i.e. $\hat{z} = P_{C_k}(\hat{z})$. Then, in view of (22) and Lemma 4(i) with $K = C_k$, $x = x^k - \frac{\beta_k}{\eta_k} u^k$ and $y = \hat{z}$, we obtain

$$\begin{aligned} \|x^{k+1} - \hat{z}\|^2 &= \left\| P_{C_k}\left(x^k - \frac{\beta_k}{\eta_k} u^k\right) - P_{C_k}(\hat{z}) \right\|^2 \leq \left\| \left(x^k - \frac{\beta_k}{\eta_k} u^k\right) - \hat{z} \right\|^2 \\ &= \|x^k - \hat{z}\|^2 + \frac{\beta_k^2}{\eta_k^2}\|u^k\|^2 - 2\frac{\beta_k}{\eta_k} \langle u^k, x^k - \hat{z}\rangle \\ &\leq \|x^k - \hat{z}\|^2 + \beta_k^2 - 2\frac{\beta_k}{\eta_k} \langle u^k, x^k - \hat{z}\rangle. \end{aligned} \tag{24}$$

Thus,

i) if $x^k \in D$, applying Cauchy-Schwartz inequality in (24) and the fact that $D \subseteq B(\hat{z}, \lambda)$, we obtain that

$$\|x^{k+1} - z\|^2 \leq \|x^k - \hat{z}\|^2 + \beta_k^2 + 2\frac{\beta_k}{\eta_k} \|u^k\|\|x^k - \hat{z}\| \leq \lambda^2 + \beta_k^2 + 2\beta_k\lambda,$$

ii) if $x^k \notin D$, it follows from (Q) that $\langle u^k, x^k - \hat{z}\rangle \geq 0$ and we get from (24) that

$$\|x^{k+1} - z\|^2 \leq \|x^k - \hat{z}\|^2 + \beta_k^2.$$

□

Proposition 6. *Let $\{x^k\}$, $\{u^k\}$ be the sequences generated by Algorithm 2. Then,*

i) $\{x^k\}$ and $\{u^k\}$ are bounded.

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

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

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

Proof. i) Take $\hat{z}$ and D satisfying (Q), $\lambda > 0$ such that $\|x^0 - \hat{z}\| \leq \lambda$ and $D \subseteq B(\hat{z}, \lambda)$, and $\bar{\beta} > 0$ such that $\beta_k \leq \bar{\beta}$ for all k ($\bar{\beta}$ exists by (7)). Let $\sigma = \sum_{j=0}^{\infty} \beta_j^2$. Define $\bar{\lambda} := \sqrt{\lambda^2 + 2\bar{\beta}\lambda + \sigma}$. We claim that $\{x^k\} \subseteq B(\hat{z}, \bar{\lambda})$. We consider two cases.

If $x^k \in B(\hat{z}, \lambda)$, we have $x^k \in B(\hat{z}, \bar{\lambda})$ because $\bar{\lambda} > \lambda$. If $x^k \notin B(\hat{z}, \lambda)$, let $\ell(k) = \max\left\{\ell < k : x^\ell \in B(\hat{z}, \lambda)\right\}$. $\ell(k)$ is well defined because $\|x^0 - \hat{z}\| \leq \lambda$, so that $x^0 \in B(\hat{z}, \lambda)$. By Lemma 7(i),

$$\|x^{\ell(k)+1} - \hat{z}\|^2 \leq \lambda^2 + \beta_{\ell(k)}^2 + 2\beta_{\ell(k)}\lambda \leq \lambda^2 + 2\bar{\beta}\lambda + \beta_{l(k)}^2. \tag{25}$$

Iterating the inequality in Lemma 7(ii), since $x^j \notin D$ for j between $\ell(k) + 1$ and k,

$$\|x^k - \hat{z}\|^2 \leq \|x^{\ell(k)+1} - \hat{z}\|^2 + \sum_{j=\ell(k)+1}^{k-1} \beta_j^2. \tag{26}$$

Combining (25) and (26)

$$\|x^k - \hat{z}\|^2 \leq \lambda^2 + 2\bar{\beta}\lambda + \sum_{j=\ell(k)}^{k-1} \beta_j^2 \leq \lambda^2 + 2\bar{\beta}\lambda + \sum_{j=0}^{\infty} \beta_j^2 = \lambda^2 + 2\bar{\beta}\lambda + \sigma = \bar{\lambda}^2.$$

Thus, $x^k \in B(\hat{z}, \bar{\lambda})$ and hence $\{x^k\}$ is bounded. For $\{u^k\}$ use boundedness of $\{x^k\}$ and Lemma 5(iii).

ii) For all k we have that

$$\|x^{k+1} - P_{C_k}(x^k)\| = \left\| P_{C_k}\left(x^k - \frac{\beta_k}{\eta_k} u^k\right) - P_{C_k}(x^k) \right\| \leq \frac{\beta_k}{\eta_k}\|u^k\| \leq \beta_k, \tag{27}$$

using (22) and Lemma 4(i) in the first inequality, and the fact that $\eta_k \geq \|u^k\|$ for all k in the second one.

We apply Lemma 1 with $X = B(\hat{z}, \bar{\lambda})$ and conclude that there exists $\tilde{\mu} \in [0, 1)$ such that

$$\text{dist}(\tilde{z}(x, v), C) \leq \tilde{\mu} \ \text{dist}(x, C) \tag{28}$$

for all $x \in B(\hat{z}, \bar{\lambda}) \setminus C$ and all $v \in \partial g(x)$.

By (i), $\{x^k\} \subseteq B(\hat{z}, \bar{\lambda})$, and we obtain, using the definition of $\tilde{z}(x, v)$, that $\tilde{z}(x^k, v^k) = P_{C_k}(x^k)$. Therefore, it follows from (28) that

$$\text{dist}(P_{C_k}(x^k), C) = \text{dist}(\tilde{z}(x^k, v^k), C) \leq \tilde{\mu} \ \text{dist}(x^k, C), \tag{29}$$

for all k such that $x^k \notin C$. If $x^k \in C$, (29) holds trivially because $C \subseteq C_k$ by Proposition 5(i). Observe that

$$\text{dist}(x^{k+1}, C) \leq \|x^{k+1} - P_{C_k}(x^k)\| + \text{dist}(P_{C_k}(x^k), C) \leq \beta_k + \tilde{\mu}\ \text{dist}(x^k, C),$$

using (27) and (29) in the second inequality. Therefore, using Lemma 2 with $\nu_k = \beta_k$ and $\xi_k = \text{dist}(x^k, C)$, we obtain $\lim\limits_{k\to\infty} \text{dist}(x^k, C) = 0$, establishing (ii).

iii) Using (27), we get

$$\|x^{k+1} - x^k\| \leq \|x^{k+1} - P_{C_k}(x^k)\| + \|P_{C_k}(x^k) - x^k\| \leq \beta_k + \text{dist}(x^k, C). \tag{30}$$

Since $\lim_{k\to\infty} \beta_k = 0$ by (7), it follows from (ii) and (30) that $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$.

iv) Follows from (ii).

□

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

Theorem 2. *If T is paramonotone and $S(T, C) \neq \emptyset$ then all cluster points of the sequence $\{x^k\}$ generated by Algorithm 2 solve VIP(T, C).*

Proof. Let $\{x^k\}$, $\{u^k\}$ be the sequences generated by Algorithm 2. Define $\gamma_k : S(T, C) \to \mathbb{R}$ as

$$\gamma_k(x) := \langle u^k, x^k - x \rangle. \tag{31}$$

Note that

$$\begin{aligned} \|x^{k+1} - x\|^2 &= \left\| P_{C_k}\left(x^k - \frac{\beta_k}{\eta_k} u^k\right) - P_{C_k}(x) \right\|^2 \leq \left\| \left(x^k - \frac{\beta_k}{\eta_k} u^k\right) - x \right\|^2 \\ &= \|x^k - x\|^2 + \frac{\beta_k^2}{\eta_k^2}\|u^k\|^2 - 2\frac{\beta_k}{\eta_k}\ \langle u^k, x^k - x \rangle \\ &\leq \|x^k - x\|^2 - \beta_k \left(2\frac{\gamma_k(x)}{\eta_k} - \beta_k\right). \end{aligned} \tag{32}$$

We prove first that $\{x^k\}$ has cluster points in $S(T, C)$. Since $\{(x^k, u^k)\}$ is bounded by Proposition 6(i), it suffices to prove that $\{\gamma_k\}$ has a nonpositive cluster point for some $x \in S(T, C)$. 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 $\gamma_k(\bar{x}) \geq \rho$ for all $k \geq \bar{k}$. Since $\{u^k\}$ is bounded, there exists $\theta > 1$ such that $\|u^k\| \leq \theta$ for all k. Therefore

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

for all k. Thus, we can find $\bar{\rho} > 0$ such that $\frac{\gamma_k(\bar{x})}{\eta_k} \geq \frac{\gamma_k(\bar{x})}{\theta} > \bar{\rho}$ and hence, in view of (32), we obtain

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \beta_k(2\bar{\rho} - \beta_k) \tag{33}$$

for all $k \geq \bar{k}$. Since $\lim_{k\to\infty} \beta_k = 0$ by (7), there exists $k' \geq \bar{k}$ such that $\beta_k \leq \bar{\rho}$ for all $k \geq \bar{k}$. So, we get from (33), for all $k \geq k'$,

$$\bar{\rho}\beta_k \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2. \tag{34}$$

Summing (34) with k between k' and m, we obtain:

$$\bar{\rho}\sum_{k=k'}^{m} \beta_k \leq \sum_{k=k'}^{m} \left(\|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2\right) \leq \|x^{k'} - \bar{x}\|^2 - \|x^{m+1} - \bar{x}\|^2 \leq \|x^{k'} - \bar{x}\|^2. \tag{35}$$

Taking limits in (35) with $m \to \infty$, we contradict the assumption that $\sum_{k=0}^{\infty} \beta_k = \infty$. Thus, there exists a cluster point of $\{x^k\}$ belonging to $S(T, C)$.

Finally, we prove that all cluster points of $\{x^k\}$ belong to $S(T, C)$. Suppose that $\{x^k\}$ has a cluster point $z \notin S(T, C)$. Since $S(T, C)$ is closed by Lemma 5(iii), we invoke Lemma 3 to obtain a subsequence $\{x^{j_k}\}$ of $\{x^k\}$ and a real number $\zeta > 0$ such that

$$\operatorname{dist}(x^{j_k}, S(T, C)) > \zeta, \tag{36}$$

and

$$\operatorname{dist}(x^{j_k+1}, S(T, C)) > \operatorname{dist}(x^{j_k}, S(T, C)). \tag{37}$$

Take $\gamma_k(x)$ as defined by (31). Note that $\{\gamma_k(x)\}$ is bounded by Proposition 6(ii) and define $\gamma : S(T, C) \to \mathbb{R}$ as $\gamma(x) := \liminf_{k\to\infty} \gamma_{j_k}(x)$. We claim that $\gamma(x) > 0$ for all $x \in S(T, C)$. Otherwise, by Lemma 6 $\{x^{j_k}\}$ has a cluster point in $S(T, C)$, in contradiction with (36). We claim now that γ is continuous in $S(T, C)$. Take x, $x' \in S(T, C)$. Note that $\gamma_{j_k}(x) = \langle u^{j_k}, x^{j_k} - x\rangle = \langle u^{j_k}, x^{j_k} - x'\rangle + \langle u^{j_k}, x' - x\rangle \leq \gamma_{j_k}(x') + \theta\|x - x'\|$. Thus, $\gamma(x) \leq \gamma(x') + \theta\|x - x'\|$, where θ is a upper bound of $\{\|u^k\|\}$. Reversing the role of x, x', we obtain $|\gamma(x) - \gamma(x')| \leq \theta\|x - x'\|$ establishing the claim.

Let V be the set of cluster points of $\{x^k\}$. We have shown above that $V \cap S(T, C) \neq \emptyset$. Since $\{x^k\}$ is bounded, V is compact and so is $V \cap S(T, C)$. It follows that γ attains its minimum on $V \cap S(T, C)$ at some x^*, so that $\gamma(x) \geq \gamma(x^*) > 0$ for all $x \in V \cap S(T, C)$, using the claim above.

Take $\hat{k}$ such that

$$\gamma_{j_k}(x) \geq \frac{\gamma(x)}{2}, \tag{38}$$

and

$$\beta_{j_k} < \frac{\gamma(x^*)}{\theta}, \tag{39}$$

for all $k \geq \hat{k}$. In view of (32), we get, for all $x \in V \cap S(T,C)$ and all $k \geq \hat{k}$,

$$\|x^{j_k+1} - x\|^2 \leq \|x^{j_k} - x\|^2 - \beta_{j_k}\left(2\frac{\gamma_{j_k}(x)}{\eta_{j_k}} - \beta_{j_k}\right) \leq \|x^{j_k} - x\|^2 - \beta_{j_k}\left(\frac{\gamma(x^*)}{\theta} - \beta_{j_k}\right) < \|x^{j_k} - x\|^2,$$

using (38) in the second inequality and (39) in the third one. It follows that $\mathrm{dist}(x^{j_k+1}, V \cap S(T,C)) \leq \mathrm{dist}(x^{j_k}, V \cap S(T,C))$ for all $k \geq \hat{k}$, in contradiction with (37). The contradiction arises from assuming that $\{x^k\}$ has clusters points out of $S(T,C)$, and therefore all cluster points of $\{x^k\}$ solve VIP(T,C). □

We summarize the convergence sequence properties of Algorithm 2 in the following corollary.

Corollary 2. *If T is paramonotone and $S(T,C) \neq \emptyset$, then the sequence $\{x^k\}$ generated by Algorithm 2 is bounded,* $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$ *and all cluster points of $\{x^k\}$ belong to $S(T,C)$. If VIP(T,C) has a unique solution then the whole sequence $\{x^k\}$ converges to it.*

Proof. It follows from Proposition 6(i), Proposition 6(iii) and Theorem 2. □

Remark 1. Note that we have convergence of the whole sequence under any hypothesis on T ensuring uniqueness of solutions of VIP(T,C), like e.g. strict monotonicity. This is much weaker than strong monotonicity, as demanded in [10] for obtaining a similar result.

Acknowledgment

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

References

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

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

[3] 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.

[4] 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.

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

[6] 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.

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

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

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

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

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

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

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

[14] Polyak, B.T. *Introduction to Optimization*. Optimization Software, New York (1987).

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