impa INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA

DIRECT METHODS FOR MONOTONE VARIATIONAL INEQUALITIES

Doctoral Thesis by

José Yunier Bello Cruz

Supervised by

Prof. Alfredo Noel Iusem

IMPA - Instituto Nacional de Matemática Pura e Aplicada
Rio de Janeiro, Brazil
July 2009

Dedicated to:

My children

Maria Elena Bello Rodríguez

José Alejandro Bello Rodríguez

Acknowledgments

My foremost thanks go to my thesis adviser Dr. Alfredo Noel Iusem. Without him, this thesis would not have been possible. I thank him for his patience and encouragement that carried me on through difficult times, and for his insights and suggestions that helped to shape my research skills. His valuable feedback contributed greatly to this thesis.

I am grateful to my former advisers Dr. Héctor Pijeira and Dr. Luis Ramiro Piñeiro, who introduced and helped me to start my graduate student life in Mathematics.

I thank the other members of my thesis committee: Dr. Roberto Andreani, Dr. Rolando Gárciga Otero, Dr. Luis Mauricio Graña Drummond, Dr. Luis Román Lucambio Pérez and Dr. Benar Fux Svaiter. Their valuable suggestions helped me to improve the thesis in many ways.

In addition, I am very grateful to CNPq and TWAS, since these institutions granted jointly the scholarship for my doctorate studies. The wonderful research environment of IMPA and the excellence of its staff contributed much to bring this work to a good end.

Also I want thank all my friends at IMPA, whose presence and motivating spirits made this academic experience even more gratifying.

Finally, my enormous gratitude goes to my wife for always being there when I needed her the most, for her love, encouragement and especially for her patience through all these years. To my family for all their love, guidance, and invaluable support through my entire life.

Abstract

We analyze one-step direct methods for nonsmooth 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, which imply uniqueness of solution, which is not the case for our approach.

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.

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

Keywords: Armijo-type search, Convex minimization problem, Korpelevich's method, Nonsmooth optimization, Maximal monotone operators, Monotone variational inequalities, Point-to-set operator, Projected gradient method, Projection method, Quasi-Fejér convergence, Relaxed method, Strong convergence, Variational inequality problem, Weak convergence.

Resumo

Analisamos métodos diretos, de um passo, para problemas de desigualdades variacionais não suaves, estabelecendo convergência quando o operador é paramonotono. Resultados anteriores sobre o método exigem hipóteses muito mais fortes, como monotonia forte ou monotonia uniforme, que implicam unicidade de solução, o que não é o caso na nossa abordagem.

Apresentamos um método totalmente explícito para resolver desigualdades variacionais monótonas em espaços de Hilbert, onde projeções ortogonais sobre o conjunto viável são substituidas por projeções sobre hiperplanos adequados. Provamos convergência fraca de toda a seqüência gerada para uma solução do problema, sob as únicas hipóteses de continuidade e monotonia do operador e existência de soluções.

Introduzimos também um método direto com dois passos, como o método de Korpelevich. A vantagem do nosso método sobre o dela é que o nosso converge fortemente em espaços de Hilbert, sendo que só foi provada convergência fraca para o algoritmo de Korpelevich. Nosso método também tem a seguinte propriedade desejável: a sequência converge para a solução do problema mais próxima ao iterado inicial.

Palavras Chaves: Busca de Tipo Armijo, Problema de Minimização Convexo, Método de Korpelevich, Otimização Não Suave, Operador Monótono Maximal, Desigualdades Variacionais Monótonas, Método do Gradiente Projetado, Convergência Quase-Fejér, Convergência Forte, Convergência Fraca.

Contents

Basic Notation and Terminology

$\mathcal{H}$: a real Hilbert space,

VIP(T, C): the variational inequality problem whose objective operator is T and whose feasible set is C,

$S(T, C)$: the solution set of VIP(T, C),

$\langle \cdot, \cdot \rangle$: the inner product of the space,

$\|\cdot\|$: the norm of the space,

$\mathbb{R}^n$: the n-dimensional Euclidean space,

$\mathbb{R}_+$: the set of nonnegative real numbers,

$\mathbb{R}_{++}$: the set of positive real numbers,

$dom(g)$: the domain of a function g,

$dom(T)$: the domain of an operator T,

∂g: the subdifferential of a convex function g,

$N_C(x)$: the normal cone of a set C at a point x,

I_C: the indicator function of a set C,

$\mathcal{P}(C)$: the power set of a set C,

∂C: the boundary of a set C,

$B(x, \delta)$: the open ball with radius δ centered at x,

$co(C)$: the convex hull of a set C.

Chapter 1

Introduction

In this chapter we collect some mathematical facts, including definitions and theorems, used in sequel. Most of the material has been taken from [13], [36], [20], [64] and [60].

1.1 Elements of topology and convex analysis

Here, we introduce some standard functional analysis results. We start with some basic definitions. Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and $\mathcal{P}(\mathcal{H})$ the set of all subsets of $\mathcal{H}$. Define $\|x\| := \langle x, x \rangle^{\frac{1}{2}}$, for all $x \in \mathcal{H}$.

Definition 1.1. *A sequence $\{x^k\} \subset \mathcal{H}$ is said to be:*

i) strongly convergent to $x \in \mathcal{H}$ if and only if $\lim_{k\to\infty} \|x^k - x\| = 0$,

ii) weakly convergent to $x \in \mathcal{H}$ if and only if $\lim_{k\to\infty} \langle x^k - x, y \rangle = 0$, for all $y \in \mathcal{H}$.

Definition 1.2. *Let $f : \mathcal{H} \to \mathbb{R} \cup \{-\infty, +\infty\}$ be a real function.*

i) f is proper if $dom(f) \neq \emptyset$ and $f(x) > -\infty$ for all $x \in dom(f)$.

ii) f is convex if $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$ for all $x, y \in \mathcal{H}$ and all $\lambda \in [0, 1]$.

iii) f is concave if $-f$ is convex.

vi) For each convex function f, the set

$$\partial f(x) := \{v \in \mathcal{H} : f(y) \geq f(x) + \langle v, y - x\rangle, \ \forall y \in \mathcal{H}\}$$

is called the subdifferential of f at x.

v) f is (weakly) lower semicontinuous if for each $x \in dom(f)$, it holds that $f(x) \leq \liminf_{y \to x} f(y)$ where

$$\liminf_{y \to x} f(y) := \sup_{U \in \mathcal{U}_x} \inf_{y \in U} f(y)$$

where $\mathcal{U}_x \subset \mathcal{P}(\mathcal{H})$ denotes the family of neighborhoods of x in the (weak) topology.

Hilbert spaces enjoy the following properties, which we will use frequently in this thesis.

Proposition 1.1. *Let $\mathcal{H}$ be a Hilbert space and $f : \mathcal{H} \to \mathbb{R} \cup \{-\infty, +\infty\}$ a real function. Then*

i) If C is convex subset of $\mathcal{H}$, then C is weakly closed if and only if C is strongly closed.

ii) If C is a closed, convex and bounded subset of $\mathcal{H}$, then C is weakly compact.

iii) Every bounded sequence in $\mathcal{H}$ has a weakly convergent subsequence.

iv) If f is a weakly lower semicontinuous function, then f has at least one minimizer over each nonempty, closed, convex and bounded subset of $\mathcal{H}$.

Proof. See pages 38-50 of [13]. □

Corollary 1.1. *If f is a proper, convex and lower semicontinuous function, then f is weakly lower semicontinuous.*

Proof. It follows from Proposition 1.1. □

Proposition 1.2. *Let $f : \mathcal{H} \to \mathbb{R} \cup \{-\infty, +\infty\}$ be a convex function whose domain has nonempty interior. Then the following statements hold.*

i) For any $x \in dom(f)$, the subdifferential $\partial f(x)$ is convex and weakly closed.

ii) If f is continuous on $int(dom(f))$, then for each $x \in int(dom(f))$, the set $\partial f(x)$ is nonempty and weakly compact.

iii) If f is lower semicontinuous, then its subdifferential ∂f is locally Lipschitz on $int(dom(f))$.

Proof. See pages 6 and 8 of [20]. □

Proposition 1.3. *Assume that $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$ is a convex function whose domain has nonempty interior. Then f is locally Lipschitz on $int(dom(f))$, and consequently $\partial f(x)$ is a nonempty subset of $\mathbb{R}^n$ for each $x \in int(dom(f))$.*

Proof. See Proposition 1.2 and page 174 of [36]. □

Next we present the definition of the convex hull.

Definition 1.3. *Given $A \subset \mathcal{H}$, the convex hull of A, denoted by $co(A)$, is the set of all convex combinations of elements of A, i.e.,*

$$co(A) = \left\{ \sum_{j=1}^{k} \alpha_j x^j \; : \; x^j \in A, \; \alpha_j \in \mathbb{R}_+, \; \sum_{j=1}^{k} \alpha_j = 1, \, k = 1, 2, \ldots \right\}.$$

1.1.1 Supremum of convex functions

In this subsection we consider the particular case of $\mathcal{H} = \mathbb{R}^n$. We come now to an extremely important calculus rule. It has no equivalent in classical differential calculus, and is of constant use in optimization. We study the following situation: J is an arbitrary index set, $\{f_j\}_{j \in J}$ is a collection of finite convex functions from $\mathbb{R}^n$ to $\mathbb{R}$, and we assume that

$$f(x) := \sup\{f_j(x) \; : \; j \in J\},$$

for all $x \in \mathbb{R}^n$. It is known that f is convex, see Proposition 2.1.2 of [36]. We are interested in computing its subdifferential. At a given x, let

$$J(x) := \{j \in J \; : \; f_j(x) = f(x)\}$$

be the active index set.

Theorem 1.1. *If J is a compact set in some metric space such that the functions $j \mapsto f_j(x)$ are upper semicontinuous for each $x \in \mathbb{R}^n$, then*

$$\partial f(x) = co\{\cup \partial f_j(x) \,:\, j \in J(x)\}.$$

Proof. See Theorem 4.4.2 of [36]. □

As a special case of Theorem 1.1, when each f_j is differentiable, we have the following result.

Corollary 1.2. *Assume that J is a compact set in some metric space such that the functions $j \mapsto f_j(x)$ are upper semicontinuous for each $x \in \mathbb{R}^n$, and that each f_j is differentiable. Then*

$$\partial f(x) = co\{\nabla f_j(x) \,:\, j \in J(x)\}.$$

Proof. See Corollary 4.4.4 of [36]. □

1.2 Monotone operators

Here we state some definitions and properties of point-to-set operators.

Definition 1.4. *Let $\mathcal{H}$ be a Hilbert space.*

i) A point-to-set operator is a map $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$.

ii) The set $dom(T) := \{x \in \mathcal{H} \,:\, T(x) \neq \emptyset\}$ is called the domain of T.

iii) The set $G(T) := \{(x, v) \in \mathcal{H} \times \mathcal{H} \,:\, v \in T(x)\}$ is called the graph of T.

iv) T is locally bounded at x if there exists a neighborhood U of x such that the set

$$T(U) := \cup_{x \in U} T(x)$$

is a bounded subset of $\mathcal{H}$.

v) The graph of T is demiclosed if for all sequences $\{x^k\} \subset \mathcal{H}$, weakly convergent (strongly convergent) to $x \in \mathcal{H}$ and for all $\{v^k\} \subset \mathcal{H}$ strongly convergent (weakly convergent) to $v \in \mathcal{H}$, with $\{(x^k, v^k)\} \subset G(T)$, it holds that $v \in T(x)$, i.e. $(x, v) \in G(T)$.

Monotone operators were first introduced in [54] and [70], and can be seen as a two-way generalization: a nonlinear generalization of linear endomorphisms with positive semidefinite matrices, and a multi-dimensional generalization of nondecreasing functions of a real variable. Monotone operators are the key ingredient of monotone variational inequalities, which extend to the realm of point-to-set mappings the constrained convex minimization problem.

In order to study and develop new methods for solving variational inequalities, it is therefore essential to understand the properties of this kind of point-to-set mappings.

Definition 1.5. *Let $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ be a point-to-set operator.*

i) T is monotone if and only if $\langle u - v, x - y\rangle \geq 0$ for all x, $y \in \mathcal{H}$ and all $u \in T(x)$, $v \in T(y)$.

ii) T is maximal monotone if T is monotone and additionally $T = T'$ for all monotone operator $T' : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ such that $G(T) \subset G(T')$.

Due to the next result, we can deal mostly with maximal monotone operators, rather than just monotone ones.

Proposition 1.4. *If $\mathcal{H}$ is a Hilbert space and $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ a monotone point-to-set operator, then there exists $\hat{T} : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ maximal monotone (not necessarily unique), such that $G(T) \subset G(\hat{T})$.*

Proof. See Proposition 4.1.3 in [19]. □

We introduce next some important examples of maximal monotone operators.

Proposition 1.5. *Let $f : \mathcal{H} \to \mathbb{R} \cup \{-\infty, +\infty\}$ be a proper convex function. Then the subdifferential ∂f of f is maximal monotone.*

Proof. The fact that ∂f is monotone is straightforward from the definition. The maximality has been proved by R. T. Rockafellar in [62]. $\square$

Example 1.1. *The normality operator of a convex set C, called N_C, is defined as follows:*

$$N_C(x) := \begin{cases} w \in \mathcal{H} \,:\, \langle w, x - z \rangle \geq 0, \ \forall z \in C & \text{if} \quad x \in C \\ \emptyset & \text{if} \quad x \notin C. \end{cases}$$

It can be checked that $N_C(x) = \partial I_C(x)$, where I_C is the indicator function of C, which is defined as

$$I_C(x) := \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{otherwise.} \end{cases}$$

Thus, we have from Proposition 1.5 that N_C is a maximal monotone operator. Note that for some $x^* \in C$ and $u^* \in \mathcal{H}$, we have

$$0 \in u^* + N_C(x^*) \quad \text{if and only if} \quad \langle u^*, x - x^* \rangle \geq 0 \quad \text{for all} \quad x \in C. \tag{1.1}$$

Example 1.2. *A linear mapping $T : \mathcal{H} \to \mathcal{H}$ is maximal monotone if and only if T is positive semidefinite, i.e. $\langle T(x), x \rangle \geq 0$ for all $x \in \mathcal{H}$.*

We recall that a mapping $T : \mathcal{H} \to \mathcal{H}$ is Gâteaux differentiable at $x \in \mathcal{H}$ if for all $y \in \mathcal{H}$ the limit

$$\delta T(x)y := \lim_{t \to 0} \frac{T(x + ty) - T(x)}{t}$$

exists and is a linear and continuous function of y. We say that T is Gâteaux differentiable when it is Gâteaux differentiable at every $x \in \mathcal{H}$.

Note that when $T : \mathcal{H} \to \mathcal{H}$ is linear, it is Gâteaux differentiable and $\delta T = T$. Thus, the previous example is a particular case of the following result.

Proposition 1.6. *A Gâteaux differentiable mapping $T : \mathcal{H} \to \mathcal{H}$ is monotone if and only if its Gâteaux derivative at x, $\delta T(x)$, is positive semidefinite for all $x \in \mathcal{H}$. In this situation, T is maximal monotone.*

Proof. See Proposition 4.1.6 in [19]. □

Now we present an important maximal monotone operator, called the saddle point operator.

Definition 1.6. *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be real Hilbert spaces and S_1 and S_2 be closed and convex subsets of $\mathcal{H}_1$ and $\mathcal{H}_2$, respectively. Given $L : S_1 \times S_2 \to \mathbb{R}$, convex in its first argument and concave in the second one, the saddle point operator $T_L : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathcal{P}(\mathcal{H}_1) \times \mathcal{P}(\mathcal{H}_2)$ is defined as*

$$T_L(x, y) := (\partial_x L(x, y), \partial_y(-L(x, y)), \tag{1.2}$$

where $\partial_x L$ and $\partial_y(-L)$ denote the subdifferentials of $L(\cdot, y)$ and $-L(x, \cdot)$, respectively.

Proposition 1.7. *Saddle point operators T_L are maximal monotone.*

Proof. See Theorem 4.7.5 of [19]. □

Saddle point operators are an important family of maximal monotone operators, which has the following associated problem. Given closed and convex subsets S_1 and S_2 of $\mathcal{H}_1$ and $\mathcal{H}_2$ respectively, the saddle point problem is defined as:
find $(x^*, y^*) \in S_1 \times S_2$ such that

$$L(x^*, y) \leq L(x^*, y^*) \leq L(x, y^*) \quad \text{for all} \quad (x, y) \in S_1 \times S_2. \tag{1.3}$$

This problem is called the constrained saddle point problem $CSP(L, S_1, S_2)$. A solution (x^*, y^*) of $CSP(L, S_1, S_2)$ is called a saddle point. The connection between the saddle point problem $CSP(L, S_1, S_2)$ and the saddle point operator T_L is given the following proposition.

Proposition 1.8. *The constrained saddle point problem $(CSP(L, S_1, S_2))$ is equivalent to the following problem: find $(x^*, y^*) \in S_1 \times S_2$ such that*

$$0 \in T_L(x^*, y^*) + (N_{S_1}(x^*), N_{S_2}(y^*)) \, .$$

Proof. Suppose that

$$0 \in T_L(x^*, y^*) + (N_{S_1}(x^*), N_{S_2}(y^*)) = (\partial_x L(x^*, y^*) + N_{S_1}(x^*), \partial_y(-L(x^*, y^*)) + N_{S_2}(y^*)).$$

By (1.1), the above inclusion is equivalent to the existence of $u^* \in \partial_x L(x^*, y^*)$ such that $\langle u^*, x - x^* \rangle \geq 0$ for all $x \in S_1$, and the existence of $v^* \in \partial_y(-L(x^*, y^*))$ such that $\langle v^*, y - y^* \rangle \geq 0$ for all $y \in S_2$. Using the definition of the subdifferential, we obtain

$$L(x, y^*) - L(x^*, y^*) \geq \langle u^*, x - x^* \rangle \geq 0 \tag{1.4}$$

for all $x \in S_1$ and

$$-L(x^*, y) + L(x^*, y^*) \geq \langle v^*, y - y^* \rangle \geq 0 \tag{1.5}$$

for all $y \in S_2$. Combining (1.4) and (1.5), we obtain both inequalities in (1.3).

Conversely, using the second inequality of (1.3), we get that

$$L(x^*, y^*) \leq \min_{x \in S_1} L(x, y^*).$$

Therefore, x^* is a minimizer of $L(\cdot, y^*)$ in S_1. Since $L(\cdot, y^*)$ is a convex function and S_1 is a closed and convex set, we have

$$0 \in \partial_x L(x^*, y^*) + N_{S_1}(x^*). \tag{1.6}$$

Now, using the first inequality of (1.3), we get that

$$-L(x^*, y^*) \leq \min_{y \in S_2} -L(x^*, y).$$

Therefore, y^* is a minimizer of $-L(x^*, \cdot)$ in S_2. Since $-L(x^*, \cdot)$ is a convex function and S_2 is a closed and convex set, we have

$$0 \in \partial_y(-L(x^*, y^*)) + N_{S_2}(y^*). \tag{1.7}$$

Combining (1.6) and (1.7), we obtain that

$$0 \in T_L(x^*, y^*) + (N_{S_1}(x^*), N_{S_2}(y^*)).$$

□

1.2.1 Properties of monotone operators

In this subsection we present some results concerning maximal monotone operators in Hilbert spaces.

Lemma 1.1. *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 demiclosed.

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.

Proof. i) See Theorem 4.6.1(ii) of [19].

ii) See Proposition 4.2.1(ii) of [19].

iii) It follows easily from (i) using a compactness argument.

iv) See Theorem 4.6.3 of [19].

□

Let T_1, $T_2 : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ be maximal monotone operators. The operator $T_1 + T_2$ is defined by

$$(T_1 + T_2)(x) := \{u_1 + u_2 \ : \ u_1 \in T_1, \, u_2 \in T_2\},$$

for all $x \in dom(T_1) \cap dom(T_2)$. Then $dom(T_1 + T_2) := dom(T_1) \cap dom(T_2)$.

It is clear that if T_1 and T_2 are monotone, then $T_1 + T_2$ is also monotone. But if T_1 and T_2 are maximal, it does not necessarily follow that $T_1 + T_2$ is maximal. Some sort of condition is needed, since for example the graph of $T_1 + T_2$ can even be empty (as happens when $dom(T_1) \cap dom(T_2) = \emptyset$). The following example establishes that the sum of two maximal monotone operators may fail to be a maximal monotone operator.

First, we define the weak derivative. Let x be a function in $L^2(\mathbb{R})$. We say that x' in $L^2(\mathbb{R})$ is the weak derivative of x if

$$\int_{-\infty}^{+\infty} x(t)\varphi'(t)dt = -\int_{-\infty}^{+\infty} x'(t)\varphi(t)dt,$$

for all $\varphi \in \mathcal{C}_c^\infty(\mathbb{R})$, that is, for all infinitely differentiable functions with compact support in $\mathbb{R}$.

Example 1.3. *Define the linear operator $T_1 : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ by*

$$T_1(x) = -x'' \quad \text{with} \quad dom(T_1) = H^2(\mathbb{R}) := \left\{x \in L^2(\mathbb{R}) \,:\, x' \in L^2(\mathbb{R})\right\}.$$

Fix $f \in L^1(\mathbb{R})$ such that $f \geq 0$, and $f_{/\Omega} \notin L^2(\Omega)$ for all open set $\Omega \subset \mathbb{R}$. It has been shown in [49] that such a function exists.

Define $T_2 : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ by

$$T_2(x)(t) = f(t)x(t) \quad \text{with} \quad dom(T_2) := \left\{x \in L^2(\mathbb{R}) \,:\, fx \in L^2(\mathbb{R})\right\}.$$

The operators T_1 and T_2 are linear, self-adjoint, and maximal monotone in $L^2(\mathbb{R})$. Hence $T_1 + T_2$ is monotone. It is easy to check that the properties of f imply that $dom(T_1 + T_2) = dom(T_1) \cap dom(T_2) = \{0\}$. Note that T_1 and T_2 are point-to-point, and so the same holds for $T_1 + T_2$, in fact $(T_1 + T_2)(x) = -x'' + fx$. On the other hand, the operator $\hat{T} : L^2(\mathbb{R}) \to \mathcal{P}(L^2(\mathbb{R}))$ defined as $\hat{T}(0) = L^2(\mathbb{R})$, $\hat{T}(x) = \emptyset$ for $x \neq 0$, is monotone and $G(T_1 + T_2) \subsetneq G(\hat{T})$. Hence, $T_1 + T_2$ is not maximal. This example was proposed and extensively studied in [4].

The problem of determining conditions under which $T_1 + T_2$ is maximal turns out to be of fundamental importance in the theory of monotone operators. Results in this directions have been proved in [15] and [16]. These results were improved by R. T. Rockafellar in [61]. We state them now:

Theorem 1.2. *Let T_1, $T_2 : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ be maximal monotone operators. Suppose that any of following conditions is satisfied:*

i) $dom(T_1) \cap int(dom(T_2)) \neq \emptyset$, *ii) there exists $x \in cl(dom(T_1)) \cap cl(dom(T_2))$ such that T_2 is locally bounded at x.*

Then $T_1 + T_2$ is a maximal monotone operator.

Proof. See Theorem 1 of [61]. □

1.2.2 Some properties stronger than monotonicity

First, we define some special classes of monotone operators.

Definition 1.7. *$T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ is said to be*

i) strongly monotone if $\langle u - v, x - y\rangle \geq \omega \|x - y\|^2$ for some $\omega > 0$ and for all x, $y \in \mathcal{H}$ and all $u \in T(x)$, $v \in T(y)$,

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

iii) strictly monotone if and only if T is monotone and $\langle u - v, x - y\rangle = 0$ with x, $y \in \mathcal{H}$, $u \in T(x)$, $v \in T(y)$ implies $x = y$.

It follows from Definition 1.5 and Definition 1.7 that the following implications hold: strong monotonicity implies uniform monotonicity, which implies strict monotonicity which in turn implies monotonicity. The reverse assertions are not true in general, as will be shown with the following examples.

Example 1.4. *Take $A \in \mathbb{R}^{n\times n}$ positive semidefinite but not positive definite. Then, the linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ defined as $T(x) = Ax$ is monotone but not strictly monotone.*

Example 1.5. *Take $T : \mathbb{R} \to \mathbb{R}$ defined as $T(x) = \exp(x)$. T is strictly monotone but not uniformly monotone.*

Example 1.6. *Consider the operator $T : \mathbb{R}^n \to \mathbb{R}^n$ defined as $T(x)_j := sg(x_j)|x_j|^p$, $(1 \leq j \leq n)$, with $p > 1$. T is uniformly monotone for all $p > 1$, with*

$$\psi(t) = n^{\frac{(1-p)}{2}} t^{p+1}, \tag{1.8}$$

but is not strongly monotone.

We show next that T is uniformly monotone with $\psi(t)$ as in (1.8). We claim that

$$\frac{(s-t)\left(sg(s)|s|^p - sg(t)|t|^p\right)}{|s-t|^{p+1}} \geq 1, \tag{1.9}$$

for all s, $t \in \mathbb{R}$.

We must consider all possible signs of s, t, but it easy to check that all lead to the same calculation. Assume that $s, t \geq 0$. Taking $r = s - t$ in (1.9), this inequality is equivalent to $\frac{r}{|r|^{p+1}}((t+r)^p - t^p) \geq 1$ for all r, t such that $t \geq \max\{0, -r\}$. Looking at the problem

$$\min \varphi(t) = \tfrac{r}{|r|^{p+1}}((t+r)^p - t^p)$$

$$s.t. \quad t \geq \max\{0, -r\},$$

it follows easily that its solution is $t^* = \max\{0, -r\}$ with $\varphi(t^*) = 1$, establishing (1.9). It follows from (1.9) that

$$\begin{aligned}
\langle T(x) - T(y), x - y\rangle &= \sum_{j=1}^{n} (x_j - y_j)\left(sg(x_j)|x_j|^p - sg(y_j)|y_j|^p\right) \\
&\geq \sum_{j=1}^{n} |x_j - y_j|^{p+1} = n \sum_{j=1}^{n} \frac{1}{n}(|x_j - y_j|^2)^{\frac{p+1}{2}} \\
&\geq n\left(\frac{1}{n}\right)^{\frac{p+1}{2}} \|x - y\|^{\frac{p+1}{2}} = \psi(\|x - y\|),
\end{aligned}$$

using the convexity of $t \mapsto t^{\frac{p+1}{2}}$ in the second inequality (note that $\frac{p+1}{2} > 1$).

Now, we show that T is not strongly monotone. Take $y = 0$ and $x = (t, 0, \ldots, 0)$ with $t > 0$. Then,

$$\frac{\langle T(x) - T(y), x - y\rangle}{\|x - y\|^2} = \frac{t^{p+1}}{t^2} = t^{p-1}.$$

Taking limits with $t \to 0^+$, we get that the inequality defining strong monotonicity fails for all $\omega > 0$, using the fact that $p - 1 > 0$.

Next, we define an important class of monotone operators, called paramonotone.

Definition 1.8. *$T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ is said to be paramonotone if and only if T is monotone and $\langle u - v, x - y\rangle = 0$ with x, $y \in \mathcal{H}$, $u \in T(x)$, $v \in T(y)$ implies $u \in T(y)$ and $v \in T(x)$.*

The notion of paramonotonicity, which is in-between monotonicity and strict monotonicity, was introduced in [17], and many of its properties were established in [21] and [39]. 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 [39].

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 $\text{Rank}(J_T(x) + J_T(x)^t) = \text{Rank}(J_T(x))$ for all x, see Proposition 4.2 in [39].

It follows that affine operators $T_A : \mathbb{R}^n \to \mathbb{R}^n$ defined as $T_A(x) = Ax + b$ are paramonotone when $A \in \mathbb{R}^{n\times n}$ is positive semidefinite (not necessarily symmetric), and $\text{Rank}(A + A^t) = \text{Rank}(A)$. This situation includes cases of nonsymmetric and singular matrices, which are not strictly monotone. This also can happen for nonlinear operators, e.g. the operator defined as $T_f(x) = \partial f(x) + T_A(x)$, where f is a convex function and T_A is as above.

There exist interesting monotone operators which are not paramonotone. For instance, an operator $F : \mathbb{R}^n \to \mathbb{R}^n$ of the form $F(x) = T(x) + A(x) + b$, where $T : \mathbb{R}^n \to \mathbb{R}^n$ is monotone and $A \in \mathbb{R}^{n\times n}$ is a non-null and skew-symmetric matrix.

We also mention that an important class of monotone operators that fail to be paramonotone are the saddle point operators. We analyze a representative saddle point problem associated to the convex optimization problem.

Consider the following problem:

$$\min f(x) \quad \text{s.t.} \quad g_i(x) \leq 0 \quad (0 \leq i \leq m),$$

with convex f, $g_i : \mathbb{R}^n \to \mathbb{R}$. If f, g_i are of class $\mathcal{C}^1$, then under standard regularity conditions (e.g. [60]), this problem is equivalent to $CSP(L, \mathbb{R}^n, \mathbb{R}^m_+)$ where L is the Lagrangian function, i.e. $L(x, y) = f(x) + \sum_{i=1}^m y_i g_i(x)$. Note that L is convex in its first argument due to the convexity of f, g_i $(1 \leq i \leq m)$ and the fact that $y_i \in \mathbb{R}_+$ $(1 \leq i \leq m)$. L is concave in the second argument because it is affine as a function of y. Therefore, the saddle point T_L introduced in Definition 1.6 is given in this case by

$$T_L(x, y) = \left(\nabla f(x) + J_g(x)^t y, -g(x)\right), \tag{1.10}$$

where $g(x) = (g_1(x), \ldots, g_m(x))$ and $J_g(x)$ is the Jacobian matrix of g at x.

Proposition 1.9. *If $J_g(x) \neq 0$ for some $x \in \mathbb{R}^n$ (otherwise the convex optimization problem is unconstrained), then the saddle point operator T_L defined in* (1.10) *is not paramonotone.*

Proof. Take $\bar{x} \in \mathbb{R}^n$ such that $J_g(\bar{x}) \neq 0$. Thus, there exists $v \in \mathbb{R}^m$ such that $J_g(\bar{x})^t v \neq 0$. Write $v = y - y'$ with y, $y' \in \mathbb{R}^m_+$. It follows from (1.10) that

$$\langle T_L(\bar{x}, y) - T_L(\bar{x}, y'), (\bar{x}, y) - (\bar{x}, y') \rangle = \langle J_g(\bar{x})^t (y - y'), \bar{x} - \bar{x} \rangle - \langle g(\bar{x}) - g(\bar{x}), y - y' \rangle = 0.$$

On the other hand, $T_L(\bar{x}, y) - T_L(\bar{x}, y') = (J_g(\bar{x})^t (y - y'), 0) = (J_g(\bar{x})^t v, 0) \neq (0, 0)$, i.e. T is not paramonotone in $\mathbb{R}^n \times \mathbb{R}^m_+$. $\square$

1.3 Variational inequalities

The variational inequality problem provides a broad unifying setting for the study of optimization, equilibrium, and related problems, and serves as a useful computational framework for the solution of a host of problems in very diverse applications. Variational inequalities have been a classical subject in mathematical physics, particularly in the calculus of variations associated with the minimization of infinite-dimensional functionals. This problem was first introduced by P. Hartman and G. Stampacchia in [34].

The systematic study of the subject began in the 1960s. Variational inequalities were used as an analytic tool for studying free boundary problems defined by nonlinear partial differential operators arising from unilateral problems in elasticity and plasticity theory and in mechanics. Variational inequalities have a wide range of important applications in physics, engineering and economics. Several of them are described in [46, 25]. Some of the earliest papers on variational inequalities are [34, 50, 51, 69].

The development of the finite-dimensional variational inequalities also began in the mid-1960s but followed a different path. Unlike its infinite-dimensional counterpart, which was conceived in the area of partial differential systems, the finite-dimensional variational inequality was born in the domain of Mathematical Programming. The developments include a rich mathematical theory, a host of effective solution algorithms, and a multitude of interesting connections to numerous disciplines. The variational inequality problem is considered within optimization theory as a natural extension of minimization problems, see [33].

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

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

$$\langle u^*, x - x^* \rangle \geq 0 \qquad \forall\, x \in C. \tag{1.11}$$

The set of solutions to this problem is denoted $S(T, C)$.

A first geometric interpretation of VIP(T, C), defined by inequality (1.11), is that x^* in C is a solution of VIP(T, C) if and only if there exists $u^* \in T(x^*)$ which forms a non-obtuse angle with every vector of the form $x - x^*$ with $x \in C$. We may formalize this observation using the normality operator of C, introduced in Example 1.1. The inequality (1.11) clearly says that $x^* \in S(T, C)$ if and only if there exists $u^* \in T(x^*)$ such that $0 \in u^* + N_C(x^*)$. This observation will be used in Example 1.9.

1.3.1 Examples

There exist many problems which can be written as variational inequality problems.

Example 1.7 (convex optimization)**.** *When T is the subdifferential of a convex function f, then VIP(T, C) is equivalent to minimizing f on C.*

Example 1.8 (complementarity problem)**.** *When $T : \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n)$ and $C = \mathbb{R}^n_+$, the variational inequality problem is equivalent to the nonlinear complementarity problem, denoted CP$(T, \mathbb{R}^n_+)$, which consists of finding $x^* \in \mathbb{R}^n_+$ and $u^* \in T(x^*)$ such that $u^* \in \mathbb{R}^n_+$ and $\langle u^*, x^* \rangle = 0$.*

We prove this elementary result. First we suppose that $x^* \in S(T, \mathbb{R}^n_+)$. By taking $x = 0 \in \mathbb{R}^n_+$ in (1.11), we obtain that there exists $u^* \in T(x^*)$ such that $\langle u^*, x^* \rangle \leq 0$. Furthermore, since $x^* \in \mathbb{R}^n_+$, it follows that $2x^* \in \mathbb{R}^n_+$. Thus, by taking $x = 2x^*$ in (1.11), we obtain $\langle u^*, x^* \rangle \geq 0$.

Combining the above two inequalities, we get $\langle u^*, x^* \rangle = 0$. As a consequence, this yields $\langle u^*, x \rangle \geq 0$ for all $x \in \mathbb{R}^n_+$, and hence $u^* \in \mathbb{R}^n_+$. Therefore x^* solves $\mathrm{CP}(T, \mathbb{R}^n_+)$.

Conversely, if x^* solves $\mathrm{CP}(T, \mathbb{R}^n_+)$, then $\langle u^*, x - x^* \rangle = \langle u^*, x \rangle \geq 0$ for all $x \in \mathbb{R}^n_+$, because $u^* \in \mathbb{R}^n_+$, i.e. $x^* \in S(T, C)$.

Example 1.9 (finding zeroes of operators)**.** *The variational inequality problem is equivalent to the problem of finding a zero of* $\hat{T} := T + N_C$*, i.e. finding* $x^* \in \mathcal{H}$ *such that* $0 \in \hat{T}(x^*) = T(x^*) + N_C(x^*)$.

In effect, $x^* \in S(T, C)$ if and only if $\langle T(x^*), x - x^* \rangle \geq 0$ for all $x \in C$. Using (1.1), we get the equivalence with the fact that $-T(x^*) \in N_C(x^*)$, i.e. $0 \in T(x^*) + N_C(x^*) = \hat{T}(x^*)$.

Example 1.10 (constrained saddle point problems)**.** *When* $T = T_L$*, as in Definition 1.6, the variational inequality problem* $(VIP(T_L, S_1 \times S_2))$ *is equivalent to the constrained saddle point problem in* $S_1 \times S_2$*; see Proposition 1.8 and Example 1.9.*

In several other situations the variational inequalities formulation is also useful. We mention among them the Nash equilibrium of a n-person noncooperative game, the generalized Nash equilibrium, the traffic assignment problem, the spatial price equilibrium problem and the general equilibrium problem, see [50, 9, 30, 37, 33, 23].

1.4 Projected gradient method

In this section we deal the following smooth optimization problem:

$$\min \quad f(x) \quad \text{s.t.} \quad x \in C, \tag{1.12}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable and $C \subset \mathbb{R}^n$ is closed and convex.

Convexity of C makes it possible to use the orthogonal projection onto C for obtaining feasible directions which are also descent ones; namely a step is taken from x^k in the direction of $-\nabla f(x^k)$, the resulting vector is projected onto C, and the direction from x^k to this projection has the above mentioned properties. We remind that a point $z \in C$ is stationary for problem (1.12) if and only if $\langle \nabla f(z), x - z \rangle \geq 0$ for all $x \in C$, i.e., $z \in S(\nabla f, C)$.

We introduce next the projected gradient method. A formal description of the algorithm is the following:

Initialization: Take $x^0 \in C$.

Iterative step: If x^k is stationary, then stop. Otherwise, let

$$z^k = x^k - \beta_k \nabla f(x^k), \tag{1.13}$$

$$x^{k+1} = \alpha_k P_C(z^k) + (1 - \alpha_k)x^k, \tag{1.14}$$

where β_k, α_k are positive for all k. The coefficients β_k and α_k are called stepsizes and $P_C : \mathcal{H} \to C$ is the orthogonal projection onto C, i.e. $P_C(x) = \operatorname{argmin}_{y \in C} \|x - y\|$. Several choices are possible for the stepsizes. Before discussing them, we mention that in the unconstrained case, i.e. $C = \mathbb{R}^n$, then the method given by (1.13)-(1.14) with $\alpha_k = 1$ for all k reduces to the iteration $x^{k+1} = x^k - \beta_k \nabla f(x^k)$, called the *steepest descent* method.

Following [10] and [40], we will focus in four strategies for the stepsizes:

a) Constant stepsize: $\beta_k = \beta$ where $\beta > 0$ is a fixed number and $\alpha_k = 1$ for all k.

b) Armijo search along the boundary of C: $\alpha_k = 1$ for all k and β_k determined by

$$\beta_k = \bar{\beta} 2^{-j(k)} \tag{1.15}$$

with

$$j(k) = \min\left\{ j \geq 0 : f(z^{k,j}) - f(x^k) \leq -\delta \langle \nabla f(x^k), x^k - P_C(z^{k,j}) \rangle \right\} \tag{1.16}$$

$$z^{k,j} = x^k - \bar{\beta} 2^{-j} \nabla f(x^k), \tag{1.17}$$

for some $\bar{\beta} > 0$, and $\delta \in (0,1)$.

c) Armijo search along the feasible direction: $\{\beta_k\} \in [\hat{\beta}, \tilde{\beta}]$ for some $0 < \hat{\beta} \leq \tilde{\beta}$ and α_k determined with an Armijo rule, namely

$$\alpha_k = 2^{-j(k)} \tag{1.18}$$

with

$$\begin{aligned} j(k) = \min \{ j \geq 0 \quad : \quad & f(2^{-j} P_C(z^k) + (1 - 2^{-j}) x^k) - f(x^k) \\ & \leq -\, \delta 2^{-j} \langle \nabla f(x^k), x^k - P_C(z^k) \rangle \} \end{aligned} \tag{1.19}$$

for some $\delta \in (0, 1)$.

d) Exogenous stepsize before projecting: β_k given by

$$\beta_k = \frac{\delta_k}{\|\nabla f(x^k)\|}$$

with

$$\sum_{k=0}^{\infty} \delta_k = \infty \qquad\qquad \sum_{k=0}^{\infty} \delta_k^2 < \infty, \tag{1.20}$$

and $\alpha_k = 1$ for all k.

Several comments are in order.

Note that Strategy (b) requires a projection onto C for each step of the inner loop resulting from the Armijo search, i.e. possibly many projections for each k, while Strategy (c) demands only one projection for each outer step, i.e. for each k. Thus, Strategy (b) is competitive only when P_C is very easy to compute (e.g. when C is a halfspace, or a box, or a ball, or a subspace).

We mention that Strategy (d), as its counterpart in the unconstrained case, fails to be a descent method. Finally, it it easy to show that for Strategy (d) it holds that $\|x^{k+1} - x^k\| \leq \delta_k$ for all k, with δ_k exogenous and satisfying (1.20). In view of (1.20), this means that all stepsizes are *small*, while Strategies (c) and (b) allow for occasionally long steps. More important, Strategy (d) does not take into account the information available at iteration k for determining the stepsizes, which in general entails worse computational performance. Its redeeming feature is that its convergence properties also hold in the nonsmooth case, in which the Armijo searches given by (b) and (c) may be unsuccessful. Throughout this thesis we will work with this type of stepsize, which does not make any searches, because it allows us to obtain direct methods for solving VIP, where the orthogonal projections onto C are replaced by projections onto suitable hyperplanes, which produces very significant savings in term of computation time.

Without assuming convexity of f, the convergence results for these methods closely mirror the ones for the steepest descent method in the unconstrained case: cluster points may fail to exist, even when (1.12) has solutions, but if they exist, they are stationary and feasible, i.e. all cluster points of $\{x^k\}$ belong to $S(\nabla f, C)$. These results can be found in Section 2.3.2 of [10]. In the case of Strategy (a) it is necessary to assume Lipschitz continuity of ∇f and to choose $\bar{\beta} \in (0, \frac{2}{L})$, where L is the Lipschitz constant, in order to ensure that the cluster points of $\{x^k\}$ are stationary, see [10].

The convergence results for Strategy (b) can be found in [31]. In order to ensure existence of cluster points, it is necessary to assume that the starting iterate x^0 belongs to a bounded level set of f.

On the other hand, when f is convex, it is possible to prove for Strategies (b) and (c) convergence of the whole sequence to a minimizer of f under the sole assumption of existence of minimizers, i.e., without any additional assumption on boundedness of level sets. These results can be found in [40].

The projected gradient method under Strategy (d) keeps its good convergence properties in an arbitrary Hilbert space also when f is convex but nonsmooth, after replacing $\nabla f(x^k)$ by a subgradient u^k of f at x^k. See [3], [2] for convergence properties in this setting, which are related to results in this thesis. It is proved in these references that the whole sequence $\{x^k\}$ converges weakly to a solution of problem (1.12) under the sole assumption of existence of solutions.

1.5 Direct methods for VIP

An excellent survey of methods for finite-dimensional variational inequality problems ($\mathcal{H} = \mathbb{R}^n$) can be found in [23]. Methods for the solution of VIP(T, C) were first proposed in an infinite-dimensional setting by M. Sibony in [65], building on the previous work [14].

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

The basic idea consists of extending the projected gradient method for constrained optimization. Remember that the stationary points of the constrained optimization problem, defined as $\min f(x)$ s.t. $x \in C$, are solutions of VIP(T, C), taking $T = \nabla f$.

For the case in which T is point-to-point, an immediate extension of the method (1.13)-(1.14) to VIP(T, C), taking $\alpha_k = 1$ for all k, is the iterative procedure given by

$$x^0 \in C, \tag{1.21}$$

$$x^{k+1} = P_C(x^k - \beta_k T(x^k)). \tag{1.22}$$

It has been proved in [24] 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 (1.21)-(1.22) converges to the unique solution of VIP(T, C), provided that $\beta_k \in \left(\epsilon, \frac{2\omega}{L^2}\right)$ for all k and for some $\epsilon > 0$, where ω is the constant of the strong monotonicity which appears in Definition 1.7(i).

Ya. I. 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

$$x^0 \in C, \tag{1.23}$$

$$x^{k+1} = P_C(x^k - \beta_k u^k)\,, \tag{1.24}$$

where $u^k \in T(x^k)$, and the sequence β_k satisfies some conditions related to the function ψ in Definition 1.7(ii), is strongly convergent to a solution of 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 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 β_k, entailing a slow convergence of the method.

We remark that for the method given by (1.23)-(1.24) there is no chance to relax the assumption on T to plain monotonicity. For example, consider $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined as $T(x) = Ax$, with

$$A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

T is monotone and the unique solution of VIP(T, C) is $x^* = 0$. However, it is easy to check that $\|x^k - \beta_k T(x^k)\| > \|x^k\|$ for all $x^k \neq 0$ and all $\beta_k > 0$, and therefore the sequence generated by (1.23)-(1.24) moves away from the solution, independently of the choice of the stepsize β_k.

Thus, the scheme (1.23)-(1.24) fails to converges for arbitrary monotone operators. In order to overcome this weakness of the method, we analyze a modified approach, called the extragradient algorithm. The general iteration is given by:

Initialization: Take $x^0 \in C$.

Iterative step: Given x^k define

$$z^k = x^k - \beta_k T(x^k). \tag{1.25}$$

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

$$y^k = \alpha_k P_C(z^k) + (1 - \alpha_k)x^k, \tag{1.26}$$

$$x^{k+1} = P_C(x^k - \gamma_k T(y^k)), \tag{1.27}$$

where β_k, α_k and γ_k are positive stepsizes, for which, again several choices are possible.

a) Constant stepsizes: $\beta_k = \gamma_k = \beta$ where $\beta > 0$ is a fixed number, and $\alpha_k = 1$ for all k.

b) Armijo-type search along the boundary of C: $\alpha_k = 1$ for all k and β_k determined with an Armijo-type stepsize rule, namely

$$\beta_k = \bar{\beta} 2^{-j(k)} \tag{1.28}$$

with

$$j(k) := \min \left\{ j \geq 0 : \left\| T(x^k) - T(P_C(z^{k,j})) \right\| \leq \frac{\delta}{\bar{\beta} 2^{-j}} \left\| x^k - P_C(z^{k,j}) \right\|^2 \right\},$$

$$z^{k,j} = x^k - \bar{\beta} 2^{-j} T(x^k),$$

for some $\bar{\beta}$, and $\delta \in (0,1)$. In this approach, we take $\gamma_k = \beta_k$ for all k or

$$\gamma_k = \frac{\langle T(y^k), x^k - y^k \rangle}{\|T(y^k)\|^2}. \tag{1.29}$$

c) Armijo-type search along the feasible direction: $\{\beta_k\} \subset [\hat{\beta}, \tilde{\beta}]$ for some $0 < \hat{\beta} \leq \tilde{\beta}$, α_k determined with an Armijo-type stepsize rule, namely

$$\alpha_k = 2^{-j(k)} \tag{1.30}$$

with

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

for some $\delta \in (0,1)$, and

$$\gamma_k = \frac{\langle T(y^k), x^k - y^k \rangle}{\|T(y^k)\|^2}. \tag{1.31}$$

Several comments are in order.

Observe that Algorithm (1.23)-(1.24) can be seen as a simplification of the extragradient method, given in (1.25)-(1.27), taking $\alpha_k = 1$ and $\gamma_k = 0$ for all k.

Note that Strategy (b), as in the case of the projected gradient method, requires a projection onto C for each step of the inner loop resulting from the Armijo-type search, i.e. possibly many projections for each k, while Strategy (c) demands only one projection for each outer step, i.e. for each k. Thus, Strategy (b) is competitive only when P_C is very easy to compute (e.g. when C is a halfspace, or a box, or a ball, or a subspace).

In order to establish convergence, it must be assumed that T is Lipschitz continuous and that an estimate of the Lipschitz constant (called L) is available. It has been proved in [47] that the extragradient method with Strategy (a) is globally convergent if T is monotone and Lipschitz continuous on C and $\beta \in \left(0, \frac{1}{L}\right)$.

In Strategies (b) and (c), T is evaluated at least twice and the projection is computed at least twice per each iteration. The resulting algorithm is applicable to the whole class of monotone variational inequalities. It has the advantage that it does not require exogenous parameters.

The extragradient method with Strategy (a) was introduced by G. Korpelevich in [47], in order to overcome the already mentioned drawbacks of the method defined by (1.23)-(1.24).

Strategy (b) was presented in [45] (see also [38] and [52] for another related approach). Strategy (c) was presented in [43]. This strategy for determining the stepsizes guarantees convergence under the only assumptions of monotonicity and continuity of T and existence of solutions of $\text{VIP}(T, C)$, without assuming Lipschitz continuity of T. Also, this strategy demands only two projections onto C per iteration, unlike other variants, like (b), with projections onto C inside the inner loop for the search of the stepsize, as we mentioned earlier.

The extragradient method has generated much interest and is currently the subject of intense research activities. The extragradient method with Strategy (c) can be extended to infinite dimensional Banach spaces, achieving weak convergence under mild assumptions, see [42]. Other direct algorithms for $\text{VIP}(T, C)$, less directly related to the extragradient method, can be found in [35], [66] and [68].

1.6 Approximate projection methods for VIP

Now we will consider some computational strategies for solving the constrained optimization problem:

$$\min\ f(x) \quad \text{s.t.} \quad x \in C, \tag{1.32}$$

where C is a closed subset of $\mathbb{R}^n$.

From the computational point of view, the constrained optimization problem presents more difficulties for its solution than the unconstrained optimization problem. There exist many different strategies for solving constrained problems.

As a first option we can transform the constrained problem into a sequence of unconstrained minimization problems. The basic idea is to eliminate some or all the constraints and add to the function a penalty term that prescribes a high cost to infeasible points. Examples of methods that fall into this category are penalty methods and Augmented Lagrangian methods, which have an extensive treatment in [10] and [12].

A second option consists of replacing the original problem by a sequence of problems with simpler constraints. As an example of this strategy we mention Sequential Quadratic Programming methods (SQP), see [12], [11] and [29].

A third option consists of treating alternatingly the optimality and the feasibility issues. In these algorithms, the resolution process alternates steps that reduce the value of the objective function with steps that reduce some measure of the infeasibility, for instance $d(\cdot, C)$. Example of this strategy are the inexact restoration algorithms, see [53], [26] and [27].

When C is convex, an important instance of this third option is the projected gradient method, already discussed in Section 1.4. We can describe it in the following way. Each step consists of two phases: first it moves in the direction opposite to the gradient of f, producing a point z^k closer to the set of zeros of ∇f; see (1.13). In this phase it is possible to lose feasibility. In the second phase, given by (1.14) with $\alpha_k = 1$, the method projects the auxiliary point z^k onto the feasible set restoring the feasibility. Thus, the projected gradient method can be seen as an inexact restoration algorithm for optimization problems, where exact restoration is achieved through the projection onto C. Observe that except in special cases (e.g. when C is a halfspace, or a ball, or a subspace, or a box) the exact calculation of the orthogonal projection is a computationally nontrivial task.

The use of projections in the restoration step is possible because we assume that C is convex. We mention that inexact restoration algorithms for optimization problems, e.g. [53], are designed also for the case in which the feasible set in not convex. In the convex case, it makes sense to replace at iteration k the set C by a simpler set $C_k \supset C$, such that the orthogonal projection onto C_k is easily computable, and take

$$x^{k+1} = P_{C_k}(z^k) \tag{1.33}$$

instead of (1.14). In this case the iterates $\{x^k\}$ are not feasible in general, and hence the restoration phase is indeed inexact. This type of methods for optimization problems have been studied, e.g. in [53], [26] and [27].

In the case of the variational inequality problem the natural extension of the projected gradient method are Algorithms (1.23)-(1.24) and (1.25)-(1.27). In both of them it is necessary to compute orthogonal projections at each step.

In this setting, it also makes sense to consider the above mentioned inexact restoration strategy, replacing C at iteration k by a simpler set C_k. Convexity of C allows us to use separating hyperplanes as such simpler sets. Again, in this methods the iterates are not necessarily feasible, and hence they may be seen as inexact restoration methods for variational inequalities.

A method of this kind for solving VIP(T, C) was proposed by M. Fukushima in [28]. He considered the case in which C is of the form

$$C = \{z \in \mathbb{R}^n \,:\, g(z) \leq 0\}, \tag{1.34}$$

where $g : \mathbb{R}^n \to \mathbb{R}$ 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 (1.34) 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 (1.34) with $g(x) = \sup\{g_j(x) \,:\, j \in J\}$.

The method uses the following relaxed (or inexact) iteration:

$$x^{k+1} = P_{C_k}\left(x^k - \beta_k \frac{T(x^k)}{\|T(x^k)\|}\right), \tag{1.35}$$

where β_k is an exogenous stepsize and C_k is defined as

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

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

M. Fukushima 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 $z \in C$, $\beta > 0$, and a bounded set $D \subseteq \mathbb{R}^n$ such that

$$\langle T(x), x - z\rangle \geq \beta\|T(x)\| \quad \text{for all} \quad x \notin D. \tag{1.36}$$

Methods of this type are called explicit, 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.

In Chapter 3 we propose an extension of this method to the case of point-to-set operators, with much better convergence properties.

1.7 Proximal point methods

In this section we consider another class of methods for solving variational inequalities, where at each iteration the current iterate is obtained by solving a nontrivial subproblem. We call such algorithms implicit algorithms. One of the most important methods of this type is the proximal point algorithm.

The proximal point algorithm, whose origins can be traced back to [48] and [56], attained its basic formulation in the work of R. T. Rockafellar [63], where it is presented as an algorithm for finding zeroes of a maximal monotone point-to-set operator. As mentioned in Example 1.9, this problem is equivalent to VIP. Given a Hilbert space $\mathcal{H}$

and a maximal monotone point-to-set operator $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$, the algorithm generates a sequence $\{x^k\} \subset \mathcal{H}$, starting from some $x^0 \in \mathcal{H}$, where x^{k+1} is the unique zero of the operator T^k defined as

$$T^k(x) := T(x) + \alpha_k(x - x^k),$$

with $\{\alpha_k\}$ being a bounded sequence of positive real numbers, called regularization coefficients.

The essential fact that the algorithm is well defined, i.e. that x^{k+1} exists and is unique, is a consequence of the fundamental result due to G. Minty, proved in [55], which states that $T + \alpha I$ is onto if and only if T is maximal monotone. This procedure can be seen as a dynamic regularization of the (possibly ill-conditioned) operator T.

It has been proved in [63] that for a maximal monotone operator T, the sequence $\{x^k\}$ is weakly convergent to a zero of T when T has zeroes, and is unbounded otherwise. In [63], R. T. Rockafellar also left an open question: the strong convergence of the sequence generated by the method. This question was resolved in the negative by O. Güler in [32], who exhibited a proper closed convex function f in an infinite-dimensional Hilbert space ℓ^2, for which the proximal point algorithm (in our framework, applied to finding a zero of $T = \partial f$) converges weakly but not strongly to a minimizer of f. Naturally, the question arises whether the proximal point method can be modified, preferably in a simple way, so that strong convergence is guaranteed.

In this sense, M. Solodov and B.F. Svaiter proposed a new proximal-type algorithm in [67] which does converge strongly, if the problem has a solution. They considered the following iterative procedure:

Take

$$x^0 \in \mathcal{H}. \tag{1.37}$$

Given $\sigma \in [0, 1)$, choose $\mu_k > 0$ and find (y^k, v^k), an inexact solution of

$$0 \in T(x) + \mu_k(x - x^k) \tag{1.38}$$

with tolerance σ. Define

$$H_k := \left\{z \in \mathcal{H} \,:\, \langle z - y^k, T(y^k)\rangle \leq 0\right\},$$
$$W_k := \left\{z \in \mathcal{H} \,:\, \langle z - x^k, x^0 - x^k\rangle \leq 0\right\},$$
$$x^{k+1} := P_{H_k \cap W_k}(x^0). \tag{1.39}$$

It has been proved in [67] that strong convergence is achieved by combining the proximal point iteration (1.38) with the projection step onto the intersection of the two halfspaces H_k, W_k which contain the solution set.

We will not work with this kind of method, because we will focus our attention on direct methods (note that finding an inexact solution of the subproblem (1.38) requires the use of some auxiliary algorithm), but we mention this method because we will use a similar technique in Chapter 5, for upgrading the weak convergence of the extragradient method to strong convergence.

1.8 Contents of the thesis

1.8.1 Direct methods for VIP

In order to solve variational inequality problems, after introducing some preliminary material in Chapter 2, we will consider the Algorithm (1.23)-(1.24) in Chapter 3, which we repeat here:

$$x^{k+1} = P_C\left(x^k - \frac{\beta_k}{\eta_k}u^k\right), \tag{1.40}$$

where $u^k \in T(x^k)$ and $\eta_k = \max\{1, \|u^k\|\}$.

Assuming that T is a paramonotone operator, we will obtain that the sequence generated by (1.40) is globally convergent to some point on $S(T,C)$, if $S(T,C)$ is nonempty and the stepsizes $\{\beta_k\}$ satisfy: $\sum_{k=0}^{\infty}\beta_k = \infty$, and $\sum_{k=0}^{\infty}\beta_k^2 < \infty$.

This selection rule has been considered several times for similar methods (see [58], [3] and [2]).

Also, we will analyze an explicit algorithm given by iteration (1.35), under the assumption that T is a point-to-set paramonotone operator, which is similar to used by M. Fukushima. This algorithm is of the form:

$$x^{k+1} = P_{C_k}\left(x^k - \frac{\beta_k}{\eta_k}u^k\right), \tag{1.41}$$

where $u^k \in T(x^k)$ and $\eta_k = \max\{1, \|u^k\|\}$.

We also assume the following condition, instead of (P), introduced in (1.36):

(Q) There exist $z \in C$ and a bounded set $D \subseteq \mathbb{R}^n$ such that

$$\langle u, x - z\rangle \geq 0 \quad \text{for all} \quad x \notin D \quad \text{and for all} \quad u \in T(x). \tag{1.42}$$

The fact that condition (Q) is weaker than (P) follows from the definition of (P) and (Q), see (1.36) and (1.42). Each one of 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 1.7(ii),

$$\langle u, x - z\rangle \geq \langle v, x - z\rangle + \psi(\|x - z\|) \geq \|x - z\|\left(\frac{\psi(\|x - z\|)}{\|x - z\|} - \|v\|\right)$$

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.

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 $f(x) = \frac{1}{2}(\text{dist}(x, C))^2$, 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), because for all bounded set D, there exists $\bar{x} \notin D$ such that $T(\bar{x}) = 0$. (It suffices to take $\bar{x} \in L \setminus D$).

We analyze the method given by (1.41), relaxing the above mentioned hypotheses in [28] 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 (1.41) is bounded, the difference between consecutive iterates converges to zero, and all its cluster points belong to $S(T, C)$.

All these results appear in our paper [7].

1.8.2 An explicit method for VIP

In Chapter 4 we will analyze a new algorithm for the case in which T is a point-to-point operator, relaxing the hypotheses in [7] in two directions: we assume plain monotonicity of T instead of paramonotonicity, and we don't impose any coerciveness condition. Additionally, we obtain convergence results stronger them those in [7]; namely we get weak convergence of the whole sequence to some solution of VIP(T, C), assuming only existence of solutions, and all our 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 (1.25)-(1.27) and its variants (e.g. Strategies (a)-(c) in Section 1.5), 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 (1.25)-(1.27). In (1.25)-(1.27) 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 (1.25)-(1.27).

In connection with the algorithms which will be introduced in Chapter 3, this algorithm, which will be introduced in Chapter 4, 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 [7]. Extensions of Korpelevich method to the point-to-set case can be found in [41] and [5].

The contents of Chapter 4 appear in our paper [8].

1.8.3 An extragradient-type method with strong convergence

In Chapter 5, we will introduce a new Korpelevich-type algorithm with strong convergence in Hilbert spaces. It is related to the method by M. Solodov and B. Svaiter in [67], where a similar modification is performed upon the proximal point method for solving VIP(T, C), with the same goal, namely upgrading weak convergence to strong one. Strong convergence is forced by combining Korpelevich-type iterations with simple projection steps onto the intersection of C and two halfspaces, containing $S(T, C)$.

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

We mention that the method in [67], as all proximal point algorithm in general, requires in each step the solution of a rather hard subproblem, while our method inherits from Korpelevich's a direct nature, without subproblems to be solved, up to the projection onto the intersection of C with two half-spaces. The presence of the half-spaces does not entail any significant additional cost over the computation of the projection onto C itself. The computational cost of this projection is negligible as compared to the cost of a proximal step, for instance, and thus both Korpelevich's method and ours can be considered as direct methods for VIP(T, C).

We impose two additional conditions on T, besides maximal monotonicity: T must be point-to-point and uniformly continuous on bounded sets. We comment now on these assumptions. Uniform continuity on bounded sets holds automatically in finite dimension, due to the continuity of point-to-point maximal monotone operators (e.g. Theorem 4.6.3 in [19]). We also mention that it is required in the analysis of [42] for proving weak convergence of Korpelevich's method in infinite dimensional spaces. In connection with the possibility of considering point-to-set, rather than point-to-point operators, we mention that a variant of Korpelevich method for point-to-set maximal monotone operators was proposed in [41], but with the following serious limitation: in principle, we should replace $T(x^k)$ by some $u^k \in T(x^k)$ everywhere in the algorithm, but, due to the lack of inner continuity of T, an arbitrary $u^k \in T(x^k)$ does not work; u^k must satisfy some additional conditions, which are not adequate for computational implementation. In particular, the method cannot be applied to cases in which T is given by an "oracle", which provides just one $u \in T(x)$ for each x. This is a rather frequent situation for point-to-set operators. Thus, we have opted to present our strongly convergent method only for the point-to-point setting.

It is important to mention that many real-world problems in economics and engineering are modelled in infinite-dimensional spaces. These include optimal control and structural design problems, among others.

The contents of Chapter 5 appear in our paper [6].

Chapter 2

Preliminary material

In this chapter we present some definitions and results that are needed for the convergence analysis of the methods which we will introduce in Chapters 3–5.

2.1 The projection operator

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

Lemma 2.1. *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) $\|P_K(x) - P_K(y)\|^2 \leq \|x - y\|^2 - \|(P_K(x) - x) - (P_K(y) - y)\|^2$.

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

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

Proof. See Lemma 1 in [67]. □

Proposition 2.1. *Let $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ be a point-to-set monotone operator and K a closed and convex subset of $\mathcal{H}$. Consider the variational inequality problem $(VIP(T, K))$. If $x = P_K(x - \beta u)$ for some $\beta > 0$ and $u \in T(x)$ then $x \in S(T, K)$.*

Proof. Using Lemma 2.1(ii), we obtain

$$\langle x - (x - \beta u), z - x \rangle \geq 0 \quad \forall z \in K. \tag{2.1}$$

Using the fact that $\beta > 0$ and (2.1), we get $\langle u, z - x \rangle \geq 0$ for all $z \in K$. Since $u \in T(x)$, we conclude that $x \in S(T, K)$. □

2.2 Monotone and paramonotone operators

We also need the following results on monotone variational inequalities.

Lemma 2.2. *Let $T : \mathcal{H} \to \mathcal{P}(\mathcal{H})$ be a maximal monotone operator and K a closed and convex set. Then $S(T, K)$, if nonempty, is closed and convex.*

Proof. See Lemma 2.4(ii) of [6]. □

Proposition 2.2. *Let T be a paramonotone operator in K. Take $x \in S(T, K)$ and $x^* \in K$. If there exists $u^* \in T(x^*)$ such that $\langle u^*, x^* - x \rangle = 0$, then x^* is also solution of VIP(T, K).*

Proof. See Proposition 13 in [21]. □

2.3 Quasi-Fejér convergence

We next deal with the so called quasi-Fejér convergence and its properties.

Definition 2.1. *Let S be a nonempty subset of $\mathcal{H}$. A sequence $\{z^k\}$ in $\mathcal{H}$ is said to be quasi-Fejér convergent to S if and only if for all $z \in S$ there exist $k_0 \geq 0$ and a sequence $\{\delta_k\} \subset \mathbb{R}_+$ such that $\sum_{k=0}^{\infty} \delta_k < \infty$ and $\|z^{k+1} - z\|^2 \leq \|z^k - z\|^2 + \delta_k$ for all $k \geq k_0$.*

This definition originates in [22] and has been further elaborated in [44].

Proposition 2.3. *If $\{z^k\}$ is quasi-Fejér convergent to S then:*

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

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

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

Proof. See Proposition 1 in [3]. □

A slightly stronger result holds in the finite dimensional case: it is enough to have one accumulation point in S in order to ensure convergence of $\{z^k\}$. The proof, much easier than in the Hilbert space case, can be found in [18]. The result of Proposition 2.3(ii) in the finite dimensional case appears in Lemma 3.2.1 of [57].

2.4 Auxiliary results

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

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

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

Lemma 2.3. *There exists* $\tilde{\alpha} \in [0, 1)$ *such that* $\operatorname{dist}(\tilde{z}(x, v), C) \leq \tilde{\alpha} \operatorname{dist}(x, C)$ *for all* $x \in X \setminus C$ *and for all* $v \in \partial g(x)$, *where* $\operatorname{dist}(x, C) = \min_{y \in C} \|x - y\|$.

Proof. See Lemma 4 in [26]. □

Lemma 2.4. *Take* $\{\xi_k\}, \{\nu_k\} \subset \mathbb{R}_+$ *and* $\mu \in [0, 1)$. *If the inequalities*

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

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

Proof. See Lemma 2 in [28]. □

The next lemma will be useful for proving that all weak cluster points of the sequence generated by the method in Chapter 4 belong to

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

Lemma 2.5. *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 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

$$0 \leq \langle T((1-\alpha)x + \alpha y), y - x\rangle, \tag{2.2}$$

for all $\alpha \in (0, 1)$. By Lemma 1.1(d), if T is point-to-point and maximal monotone then T is continuous. Making $\alpha \to 0$ and using the continuity of T, we obtain from (2.2) 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 for the distance from a point to the feasible set C.

Lemma 2.6. *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

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

□

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

Proposition 2.4. *Let $\{u^k\} \subset \mathcal{H}$ be a sequence strongly convergent to $\tilde{u}$. Take non-negative 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. Since $\sum_{j=0}^{k} \zeta_{k,j} = 1$, we get that

$$\begin{aligned} \left\| \sum_{j=0}^{k} \zeta_{k,j} u^j - x^* \right\| &= \left\| \sum_{j=0}^{k} \zeta_{k,j} (u^j - x^*) \right\| \\ &\leq \sum_{j=0}^{k} \zeta_{k,j} \|u^j - x^*\|. \end{aligned} \tag{2.3}$$

Since $\lim_{k\to\infty} u^k = x^*$, given $\epsilon > 0$, take k_0 such that $\|u^k - x^*\| \leq \frac{\epsilon}{2}$ for all $k \geq k_0$. Since $\lim_{k\to\infty} \zeta_{k,j} = 0$ for all j, $\{\zeta_{k,j}\}_{k=0}^{\infty}$ is bounded for all j and $\{\|u^k - x^*\|\}$ is also bounded. Take $\zeta \geq \|u^k - x^*\|$ for all k. Since $\lim_{k\to\infty} \zeta_{k,j} = 0$ for all j, find $k_1 > k_0$ such that $\zeta_{k,j} \leq \frac{\epsilon}{2\zeta k_0}$ for all $k \geq k_1$. Thus, for all $k > k_1$

$$\begin{aligned} \sum_{j=0}^{k} \zeta_{k,j} \|u^j - x^*\| &= \sum_{j=0}^{k_0-1} \zeta_{k,j} \|u^j - x^*\| + \sum_{j=k_0}^{k} \zeta_{k,j} \|u^j - x^*\| \\ &\leq \sum_{j=0}^{k_0-1} \frac{\epsilon}{2\zeta k_0} \zeta + \frac{\epsilon}{2} \sum_{j=0}^{k} \zeta_{k,j} = \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon, \end{aligned}$$

using the fact that $\sum_{j=0}^{k} \zeta_{k,j} = 1$. Therefore, using (2.3), we get that

$$\lim_{k\to\infty} w^k = \lim_{k\to\infty} \sum_{j=0}^{k} \zeta_{k,j} u^j = x^*.$$

□

Chapter 3

Direct methods for VIP

In this chapter we introduce two algorithms for solving VIP(T, C) in the finite dimensional setting, i.e. $\mathcal{H} = \mathbb{R}^n$. Algorithm 1 is a feasible one, and does not assume any particular structure of C. It is an extension of the classical projected gradient method for constrained optimization. Under maximal monotonicity and paramonotonicity of T, it is shown that the generated sequence is globally convergent to a solution of VIP(T, C), if there exists any.

Algorithm 2 is an infeasible one, and it assumes that C is defined by convex (possibly nonsmooth) inequalities. This algorithm replaces projections onto the feasible set by easily computable projections onto suitable hyperplanes. Assuming that T is maximal monotone and paramonotone, and an extra assumption (condition (Q)), it is shown that the sequence generated is bounded, the difference between consecutive iterates converge to zero, and all its cluster points belong to $S(T, C)$, improving over previously obtained results by Ya. I. Alber and M. Fukushima (see [1] and [28]). Both methods use exogenous stepsizes.

From this point, unless otherwise stated, all results are new, to our knowledge.

3.1 Statement of Algorithm 1

Our algorithm requires an exogenous sequence $\{\beta_k\} \subset \mathbb{R}_{++}$ of stepsizes satisfying

$$\sum_{k=0}^{\infty} \beta_k = \infty, \tag{3.1}$$

and

$$\sum_{k=0}^{\infty} \beta_k^2 < \infty. \tag{3.2}$$

The algorithm is defined as:

Algorithm 1

Initialization step: Take

$$x^0 \in C.$$

Iterative step: Given x^k take $u^k \in T(x^k)$, $\eta_k := \max\{1, \|u^k\|\}$ and define

$$x^{k+1} = P_C\left(x^k - \frac{\beta_k}{\eta_k} u^k\right). \tag{3.3}$$

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

Our convergence analysis requires exogenous stepsizes. The issue of finding a similar method which uses line searches for determining the stepsizes, while preserving the convergence properties of this method, is left as an open problem. We emphasize that our method can be applied to point-to-point operators.

3.2 Convergence analysis of Algorithm 1

We first establish the validity of the stopping criterion.

Proposition 3.1. *If $x^{k+1} = x^k$ for some k, then $x^k \in S(T, C)$.*

Proof. It follows easily from Proposition 2.1. □

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

Proposition 3.2. *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

$$\langle \bar{u}, x - \bar{x} \rangle \geq 0 \quad \forall\, x \in C. \tag{3.4}$$

Then,

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

using (3.3), the fact that $\bar{x} \in S(T,C) \subset C$ and Lemma 2.1(i) in the first inequality with $K = C$, $x = x^k - \frac{\beta_k}{\eta_k} u^k$ and $y = \bar{x}$, the definition of η_k in the second one, the monotonicity of T in the third one and (3.4) in the fourth one.

Using Definition 2.1 and the fact that β_k satisfies (3.2), we conclude that the sequence $\{x^k\}$ is quasi-Fejér convergent to $S(T,C)$.

ii) Follows from (i) and Proposition 2.3(iii). □

Corollary 3.1. *The sequences* $\{x^k\}$, $\{u^k\}$ *generated by Algorithm 1 are bounded.*

Proof. For $\{x^k\}$ use Proposition 3.2(i) and Proposition 2.3(i). For $\{u^k\}$, use boundedness of $\{x^k\}$, maximal monotonicity of T and Lemma 1.1(iii). □

In our analysis, paramonotonicity of T is used for the first time in following lemma, which will be useful for proving that the sequence generated by our algorithm converges to some point belonging to $S(T,C)$.

Lemma 3.1. *Let* T *be a maximal monotone and paramonotone operator in* $\mathbb{R}^n$ *and* K *be any nonempty closed and convex set in* $\mathbb{R}^n$. *Let* $\{(z^k, v^k)\} \subset G(T)$ *be a bounded* *sequence such that all cluster points of* $\{z^k\}$ *belong to* K*. For each* $x \in S(T,K)$ *define* $\gamma_k(x) := \langle v^k, z^k - x\rangle$*. If for some* $x \in S(T,K)$ *there exists a subsequence* $\{\gamma_{j_k}(x)\}$ *of* $\{\gamma_k(x)\}$ *such that* $\lim_{k\to\infty}\gamma_{j_k}(x) \le 0$*, then there exists a cluster point of* $\{z^{j_k}\}$ *belonging to* $S(T,K)$*.*

Proof. Suppose that there exist $x \in S(T,K)$ and a subsequence $\{\gamma_{j_k}(x)\}$ of $\{\gamma_k(x)\}$ such that $\lim_{k\to\infty}\gamma_{j_k}(x) \le 0$. Let (z^*, v^*) be a cluster point of the bounded subsequence $\{(z^{j_k}, v^{j_k})\}$. Since T is maximal monotone, $v^* \in T(z^*)$ by Lemma 1.1(ii). Without loss of generality we assume that $\lim_{k\to\infty}(z^{j_k}, v^{j_k}) = (z^*, v^*)$. Therefore, $\lim_{k\to\infty}\gamma_{j_k}(x) = \lim_{k\to\infty}\langle v^{j_k}, z^{j_k} - x\rangle = \langle v^*, z^* - x\rangle \le 0$. Since $x \in S(T,K)$, there exists $u \in T(x)$ such that $\langle u, z^* - x\rangle \ge 0$, and using the monotonicity of T we obtain

$$0 \ge \lim_{k\to\infty}\gamma_{j_k}(x) = \langle v^*, z^* - x\rangle \ge \langle u, z^* - x\rangle \ge 0. \tag{3.5}$$

It follows from (3.5) that $\langle v^*, z^* - x\rangle = 0$, and we conclude from Proposition 2.2 that $z^* \in S(T,K)$. □

The following theorem is our main convergence result on Algorithm 1.

Theorem 3.1. *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 then the result follows from Proposition 3.1. Therefore, we assume that the sequence $\{x^k\}$ is infinite. By Corollary 3.1, $\{x^k\}$ has cluster points. We claim that there exists a cluster point of $\{x^k\}$ belonging to $S(T,C)$. Otherwise, in view of Lemma 3.1, 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 and $\bar{k}$. Since $\{u^k\}$ is bounded by Corollary 3.1, there exists $\theta > 1$ such that $\|u^k\| \le \theta$ for all k. Therefore

$$\eta_k = \max\{1, \|u^k\|\} \le \max\{1, \theta\} = \theta \quad \forall k, \tag{3.6}$$

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

$$\beta_k \le \frac{\rho}{\theta} \quad \text{and} \quad \langle u^k, x^k - \bar{x}\rangle \ge \rho \qquad \forall k \ge \bar{k}. \tag{3.7}$$

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

$$\begin{aligned}
\|x^{k+1} - \bar{x}\|^2 &= \left\| P_C\left(x^k - \frac{\beta_k}{\eta_k} u^k\right) - P_C(\bar{x}) \right\|^2 \leq \left\| \left(x^k - \frac{\beta_k}{\eta_k} u^k\right) - \bar{x} \right\|^2 \\
&\leq \|x^k - \bar{x}\|^2 + \beta_k^2 - 2\frac{\beta_k}{\eta_k} \langle u^k, x^k - \bar{x}\rangle \\
&\leq \|x^k - \bar{x}\|^2 - 2\beta_k \frac{\rho}{\theta} + \beta_k^2 = \|x^k - \bar{x}\|^2 - \beta_k\left(2\frac{\rho}{\theta} - \beta_k\right) \\
&\leq \|x^k - \bar{x}\|^2 - \beta_k \frac{\rho}{\theta}, \qquad (3.8)
\end{aligned}$$

using Lemma 2.1(i) in the first inequality, with $K = C$, $x = x^k - \dfrac{\beta_k}{\eta_k} u^k$ and $y = \bar{x}$, the definition of η_k in the second one, (3.6) in the third one and (3.7) in the fourth one. It follows from (3.8) that

$$\frac{\rho}{\theta}\beta_k \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2. \qquad (3.9)$$

Summing (3.9) with k between $\bar{k}$ and m,

$$\begin{aligned}
\frac{\rho}{\theta}\sum_{k=\bar{k}}^{m} \beta_k &\leq \sum_{k=\bar{k}}^{m} \left(\|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2\right) = \|x^{\bar{k}} - \bar{x}\|^2 - \|x^{m+1} - \bar{x}\|^2 \\
&\leq \|x^{\bar{k}} - \bar{x}\|^2. \qquad (3.10)
\end{aligned}$$

Taking limits in (3.10) 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 3.2(ii), $\{x^k\}$ converges to a point in $S(T, C)$. □

3.3 Statement of Algorithm 2

In this section we introduce an algorithm which eliminates the projection onto C, replacing it by the orthogonal projection onto a suitable hyperplane. We assume that $\mathcal{H} = \mathbb{R}^n$ and C is of the form given in (1.34), which we repeat here:

$$C = \{z \in \mathbb{R}^n \,:\, g(z) \leq 0\}, \qquad (3.11)$$

where $g : \mathbb{R}^n \to \mathbb{R}$ is a convex function.
The algorithm is defined as follows.

Algorithm 2

Initialization step: Take

$$x^0 \in C.$$

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

$$C_k := \left\{z \in \mathbb{R}^n \,:\, g(x^k) + \langle v^k, z - x^k\rangle \le 0\right\}. \tag{3.12}$$

Compute,

$$x^{k+1} = P_{C_k}\left(x^k - \frac{\beta_k}{\eta_k}u^k\right), \tag{3.13}$$

with β_k satisfying (3.1)-(3.2).

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. We observe that the step $x^k - \frac{\beta_k}{\eta_k}u^k$ produces a point closer than x^k to the set of zeros of T. The projection P_{C_k} produces a point which is closer to the feasible set C, without attaining this set. Thus, this method can be seen as an inexact restoration algorithm for VIP(T, C) in the sense discussed in Section 1.6. Note that Algorithm 2 can be easily implemented, because P_{C_k} has an explicit formula, given in the next proposition.

Proposition 3.3. *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 [59]. □

It follows from Proposition 3.3 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 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, Corollary 1.2 allows us to take

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

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.

3.4 Convergence analysis of Algorithm 2

For convergence of our method, we assume that T is maximal monotone, paramonotone and satisfies condition (Q) stated in (1.42), which we repeat here:

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

Observe that $\partial g(x) \neq \emptyset$ by Proposition 1.3 for all $x \in \mathbb{R}^n$, because we assume that g is convex and $\operatorname{dom}(g) = \mathbb{R}^n$.
Before establishing convergence of the algorithm, we need to ascertain the validity of the stopping criterion.

Proposition 3.4. *Take C, C_k and x^k defined by (3.11), (3.12) and (3.13) 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 (3.12) and the definition of subdifferential.

ii) Suppose that $x^{k+1} = x^k$. Then, since $x^{k+1} \in C_k$, we have $g(x^k) = g(x^k) + \langle v^k, x^{k+1} - x^k\rangle \le 0$, i.e. $x^k \in C$. Moreover, since x^{k+1} is given by (3.13), using Lemma 2.1(ii) with $x = x^k - \frac{\beta_k}{\eta_k}u^k$ and $K = C_k$, we obtain

$$\left\langle x^{k+1} - \left(x^k - \frac{\beta_k}{\eta_k}u^k\right), z - x^{k+1}\right\rangle \ge 0 \quad \forall z \in C_k. \tag{3.14}$$

Taking $x^{k+1} = x^k$ in (3.14) 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 $x^k \in S(T, C)$. $\square$

In the remainder of this section, we will suppose that the algorithm generates an infinite sequence $\{x^k\}$. The following technical lemma will be used for establishing boundedness of $\{x^k\}$.

Lemma 3.2. *Take $\hat{z} \in C$ and D as in condition (Q), let $\{x^k\}$ be a 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 3.4(i) that $\hat{z} \in C_k$ for all k, i.e. $\hat{z} = P_{C_k}(\hat{z})$. Then, in view of (3.13) and Lemma 2.1(i) with $K = C_k$, $x = x^k - \frac{\beta_k}{\eta_k} u^k$ and $y = \hat{z}$, we obtain

$$\begin{aligned} \|x^{k+1} - \hat{z}\|^2 &= \left\| P_{C_k}\left(x^k - \frac{\beta_k}{\eta_k} u^k\right) - P_{C_k}(\hat{z}) \right\|^2 \leq \left\| \left(x^k - \frac{\beta_k}{\eta_k} u^k\right) - \hat{z} \right\|^2 \\ &= \|x^k - \hat{z}\|^2 + \frac{\beta_k^2}{\eta_k^2} \|u^k\|^2 - 2\frac{\beta_k}{\eta_k} \langle u^k, x^k - \hat{z} \rangle \\ &\leq \|x^k - \hat{z}\|^2 + \beta_k^2 - 2\frac{\beta_k}{\eta_k} \langle u^k, x^k - \hat{z} \rangle. \end{aligned} \tag{3.15}$$

Thus,

i) if $x^k \in D$, applying Cauchy-Schwartz inequality in (3.15), the definition of η_k and the fact that $D \subseteq B(\hat{z}, \lambda)$, we obtain that

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

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 (3.15) that

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

using the fact that $\frac{\beta_k}{\eta_k} > 0$ for all k.

$\square$

Next, we establish some important convergence properties of Algorithm 2.

Proposition 3.5. *Let $\{x^k\}$, $\{u^k\}$ be 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 as in condition (Q), $\lambda > 0$ such that $\|x^0 - \hat{z}\| \le \lambda$ and $D \subseteq B(\hat{z}, \lambda)$, and $\bar{\beta} > 0$ such that $\beta_k \le \bar{\beta}$ for all k ($\bar{\beta}$ exists by (3.2)). Let $\sigma = \sum_{j=0}^{\infty} \beta_j^2$. Define $\bar{\lambda} := (\lambda^2 + 2\bar{\beta}\lambda + \sigma)^{1/2}$. We claim that

$$\{x^k\} \subseteq B(\hat{z}, \bar{\lambda}). \tag{3.16}$$

If $x^k \in B(\hat{z}, \lambda)$, we have $x^k \in B(\hat{z}, \bar{\lambda})$ because $\bar{\lambda} > \lambda$. Otherwise, let $\ell(k) := \max\left\{\ell < k \,:\, x^\ell \in B(\hat{z}, \lambda)\right\}$. Observe that $\ell(k)$ is well defined because $\|x^0 - \hat{z}\| \le \lambda$, so that $x^0 \in B(\hat{z}, \lambda)$. By Lemma 3.2(i),

$$\|x^{\ell(k)+1} - \hat{z}\|^2 \le \lambda^2 + \beta_{\ell(k)}^2 + 2\beta_{\ell(k)}\lambda \le \lambda^2 + 2\bar{\beta}\lambda + \beta_{l(k)}^2. \tag{3.17}$$

Iterating the inequality in Lemma 3.2(ii), since $x^j \notin D$ for j between $\ell(k)+1$ and $k-1$, we obtain

$$\|x^k - \hat{z}\|^2 \le \|x^{\ell(k)+1} - \hat{z}\|^2 + \sum_{j=\ell(k)+1}^{k-1} \beta_j^2. \tag{3.18}$$

Combining (3.17) and (3.18)

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

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 1.1(iii).

ii) For all k we have that

$$\|x^{k+1} - P_{C_k}(x^k)\| = \left\| P_{C_k}\left(x^k - \frac{\beta_k}{\eta_k} u^k\right) - P_{C_k}(x^k) \right\| \leq \frac{\beta_k}{\eta_k} \|u^k\| \leq \beta_k, \tag{3.19}$$

using (3.13) and Lemma 2.1(i) in the first inequality, and the fact that $\eta_k \geq \|u^k\|$ for all k in the second one.

We apply Lemma 2.3 with $X = B(\hat{z}, \bar{\lambda})$ and conclude that there exists $\tilde{\mu} \in [0, 1)$ such that

$$\operatorname{dist}(\tilde{z}(x, v), C) \leq \tilde{\mu} \operatorname{dist}(x, C) \tag{3.20}$$

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

By (3.16), $\{x^k\} \subseteq B(\hat{z}, \bar{\lambda})$, 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 (3.20) that

$$\operatorname{dist}(P_{C_k}(x^k), C) = \operatorname{dist}(\tilde{z}(x^k, v^k), C) \leq \tilde{\mu} \operatorname{dist}(x^k, C), \tag{3.21}$$

for all k such that $x^k \notin C$. If $x^k \in C$, (3.21) holds trivially because $C \subseteq C_k$ by Proposition 3.4(i). Observe that

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

using (3.19) and (3.21) in the second inequality. Therefore, using Lemma 2.4 with $\nu_k = \beta_k$, $\mu = \tilde{\mu}$ and $\xi_k = \operatorname{dist}(x^k, C)$, we obtain $\lim_{k \to \infty} \operatorname{dist}(x^k, C) = 0$, establishing (ii).

iii) Using (3.19), we get

$$\|x^{k+1} - x^k\| \leq \|x^{k+1} - P_{C_k}(x^k)\| + \|P_{C_k}(x^k) - x^k\| \leq \beta_k + \operatorname{dist}(x^k, C). \tag{3.22}$$

Since $\lim_{k\to\infty} \beta_k = 0$ by (3.2), it follows from (ii) and (3.22) that $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$.

iv) Follows from (ii). □

The following lemma will be useful for proving that all cluster points of the sequence generated by Algorithm 2 belong to $S(T, C)$.

Lemma 3.3. *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*

$$\operatorname{dist}(z^{j_k+1}, K) > \operatorname{dist}(z^{j_k}, K) \tag{3.23}$$

and

$$\operatorname{dist}(z^{j_k}, K) > \zeta. \tag{3.24}$$

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

$$U := \{x \in \mathbb{R}^n \,:\, \operatorname{dist}(x, K) \leq 2\,\zeta\}\,. \tag{3.25}$$

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 (3.25). Thus, $\operatorname{dist}(z^{j_k+1}, K) > 2\,\zeta \geq \operatorname{dist}(z^{j_k}, K)$ by definition of U, so that (3.23) holds.

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

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

for all $k \geq \tilde{k}$, and hence $\{z^{j_k}\}_{k\geq\tilde{k}}$ satisfies (3.23) and (3.24). □

Paramonotonicity of T is used for the first time in the convergence analysis of Algorithm 2, in the following theorem.

Theorem 3.2. *If T is paramonotone and $S(T, C) \neq \emptyset$, then all cluster points of any sequence $\{x^k\}$ generated by Algorithm 2 solve VIP(T, C).*

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

$$\gamma_k(x) := \langle u^k, x^k - x\rangle. \tag{3.26}$$

Since $C \subset C_k$ by Proposition 3.4(i), we get from Lemma 2.1(i) and the definition of η_k that

$$
\begin{aligned}
\|x^{k+1}-x\|^2 &= \left\| P_{C_k}\left(x^k - \frac{\beta_k}{\eta_k}u^k\right) - P_{C_k}(x)\right\|^2 \leq \left\|\left(x^k - \frac{\beta_k}{\eta_k}u^k\right) - x\right\|^2 \\
&= \|x^k - x\|^2 + \frac{\beta_k^2}{\eta_k^2}\|u^k\|^2 - 2\frac{\beta_k}{\eta_k}\,\langle u^k, x^k - x\rangle \\
&\leq \|x^k - x\|^2 - \beta_k\left(2\frac{\gamma_k(x)}{\eta_k} - \beta_k\right). \qquad (3.27)
\end{aligned}
$$

First we prove that $\{x^k\}$ has cluster points in $S(T,C)$. Since $\{(x^k,u^k)\}$ is bounded by Proposition 3.5(i), in view of Lemma 3.3 it suffices to prove that $\{\gamma_k(x)\}$ has a nonpositive cluster point for some $x \in S(T,C)$. Assume that this is not true, and take any $\bar{x} \in S(T,C)$. Our assumption implies that $\{\gamma_k(\bar{x})\}$ 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

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

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 (3.27), we obtain

$$\|x^{k+1} - \bar{x}\|^2 \leq \|x^k - \bar{x}\|^2 - \beta_k(2\bar{\rho} - \beta_k) \qquad (3.28)$$

for all $k \geq \bar{k}$. Since $\lim_{k\to\infty} \beta_k = 0$ by (3.2), there exists $k' \geq \bar{k}$ such that $\beta_k \leq \bar{\rho}$ for all $k \geq k'$. So, we get from (3.28), for all $k \geq k'$,

$$\bar{\rho}\beta_k \leq \|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2. \qquad (3.29)$$

Summing (3.29) with k between k' and m, we obtain:

$$
\begin{aligned}
\bar{\rho}\sum_{k=k'}^{m} \beta_k &\leq \sum_{k=k'}^{m}\left(\|x^k - \bar{x}\|^2 - \|x^{k+1} - \bar{x}\|^2\right) \leq \|x^{k'} - \bar{x}\|^2 - \|x^{m+1} - \bar{x}\|^2 \\
&\leq \|x^{k'} - \bar{x}\|^2. \qquad (3.30)
\end{aligned}
$$

Taking limits in (3.30) with $m \to \infty$, we contradict assumption (3.1). 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 2.2 and $\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0$ by Proposition 3.5(iii), we invoke Lemma 3.3 to obtain a subsequence $\{x^{j_k}\}$ of $\{x^k\}$ and a real number $\zeta > 0$ such that

$$\operatorname{dist}(x^{j_k}, S(T, C)) > \zeta, \tag{3.31}$$

and

$$\operatorname{dist}(x^{j_k+1}, S(T, C)) > \operatorname{dist}(x^{j_k}, S(T, C)). \tag{3.32}$$

Take $\gamma_k(x)$ as defined by (3.26). Note that $\{\gamma_k(x)\}$ is bounded by Proposition 3.5(i). Define $\gamma : S(T, C) \to \mathbb{R}$ as

$$\gamma(x) := \liminf_{k\to\infty} \gamma_{j_k}(x). \tag{3.33}$$

We claim that $\gamma(x) > 0$ for all $x \in S(T, C)$. Otherwise, by Lemma 3.1 $\{x^{j_k}\}$ has a cluster point in $S(T, C)$, in contradiction with (3.31). We claim now that γ is continuous in $S(T, C)$. Take x, $x' \in S(T, C)$. Note that $\gamma_{j_k}(x) = \langle u^{j_k}, x^{j_k} - x\rangle = \langle u^{j_k}, x^{j_k} - x'\rangle + \langle u^{j_k}, x' - x\rangle \le \gamma_{j_k}(x') + \theta\|x - x'\|$. Thus, $\gamma(x) \le \gamma(x') + \theta\|x - x'\|$, where θ is a upper bound of $\{\|u^k\|\}$. Reversing the role of x, x', we obtain $|\gamma(x) - \gamma(x')| \le \theta\|x - x'\|$, 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 $\gamma(x) \ge \gamma(x^*) > 0$ for all $x \in V \cap S(T, C)$, using the claim above.

Take $\hat{k}$ such that

$$\gamma_{j_k}(x) \ge \frac{\gamma(x^*)}{2}, \tag{3.34}$$

and

$$\beta_{j_k} < \frac{\gamma(x^*)}{\theta}, \tag{3.35}$$

for all $k \ge \hat{k}$. Note that $\hat{k}$ exists because, for all large enough k, (3.34) holds by virtue of (3.33), and (3.35) because $\lim_{k\to\infty} \beta_k = 0$. In view of (3.27), we get, for all

$x \in V \cap S(T, C)$ and all $k \geq \hat{k}$,

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

using (3.34) in the second inequality and (3.35) in the third one. It follows that

$$\text{dist}(x^{j_k+1}, V \cap S(T, C)) \leq \text{dist}(x^{j_k}, V \cap S(T, C))$$

for all $k \geq \hat{k}$, in contradiction with (3.32). 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 3.2. *If T is paramonotone and $S(T, C) \neq \emptyset$, then any 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 3.5(i), Proposition 3.5(iii) and Theorem 3.2. □

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 [28] for obtaining a similar result.

Chapter 4

An explicit algorithm for VIP

In this chapter we introduce an algorithm for solving VIP(T, C), which replaces projections onto the feasible set by easily computable projections onto suitable hyperplanes. We assume that C satisfies Slater's condition, and that T is point-to-point and maximal monotone. Unlike the previous chapter, in this one we will work in an arbitrary Hilbert space. The algorithm presented here, called Algorithm 3, has higher computational demands than the previous algorithms. These algorithms (Algorithm 1 and 2) are applicable to the general case where T is a point-to-set operator. Unlike these two, Algorithm 3 requires that the operator be point-to-point. On the other hand, we will obtain convergence assuming only that T is monotone, while the methods in Chapter 3 require that T be paramonotone. It is important to have methods that do not require paramonotonicity, because there exist important problems where the operator is not paramonotone. A very relevant example is the constrained saddle point problem (CSP) presented in Subsection 1.2.2. As shown there, this problem is equivalent to the constrained convex optimization problem under suitable regularity conditions (see Proposition 1.7, Proposition 1.8 and Proposition 1.9 in Chapter 1).

4.1 Statement of Algorithm 3

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 1.1(iii), because T and ∂g are maximal monotone operators. We also assume that C is of the form given in (3.11), which we repeat here:

$$C = \{z \in \mathcal{H} \,:\, g(z) \leq 0\}, \tag{4.1}$$

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

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

$$\sum_{k=0}^{\infty} \beta_k = \infty, \tag{4.2}$$

$$\sum_{k=0}^{\infty} \beta_k^2 < \infty. \tag{4.3}$$

The algorithm is defined as follows.

Algorithm 3

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 $y^{k,0}$, $y^{k,1}$, $\ldots$. Take $y^{k,0} = z^k$, choose $v^{k,0} \in \partial g(y^{k,0})$. For $j = 0, 1, \ldots$, let

$$C_{k,j} := \left\{z \in \mathcal{H} \,:\, g(y^{k,j}) + \langle v^{k,j}, z - y^{k,j}\rangle \leq 0\right\}, \tag{4.4}$$

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

$$y^{k,j+1} := P_{C_{k,j}}(y^{k,j}). \tag{4.5}$$

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

$$j(k) := \min\left\{ j \geq 0 : \frac{g(y^{k,j})\, \|y^{k,j} - w\|}{g(y^{k,j}) - g(w)} \leq \theta\beta_k \right\}. \tag{4.6}$$

Let

$$\tilde{y}^k := y^{k,j(k)}. \tag{4.7}$$

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

$$C_k := C_{k,j(k)} = \left\{z \in \mathcal{H} \,:\, g(\tilde{y}^k) + \langle \tilde{v}^k, z - \tilde{y}^k\rangle \leq 0\right\} . \tag{4.8}$$

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

$$z^{k+1} := P_{C_k}\left(\tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k)\right). \tag{4.9}$$

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

$$\sigma_k := \sum_{j=0}^{k} \frac{\beta_j}{\eta_j} = \sigma_{k-1} + \frac{\beta_k}{\eta_k}, \tag{4.10}$$

$$x^{k+1} := \left(1 - \frac{\beta_k}{\eta_k \sigma_k}\right) x^k + \frac{\beta_k}{\eta_k \sigma_k} \tilde{y}^k. \tag{4.11}$$

Unlike other projection methods, Algorithm 3 generates a sequence $\{x^k\}$ which is not necessarily contained in the set C, which also happens with Algorithm 2. This method can be seen as an inexact restoration algorithm. The restoration requires possibly several consecutive projections onto hyperplanes, as opposed to Algorithm 2, which demands only one projection. As a matter of fact, it is necessary to project onto separating hyperplanes until the current point is close to C. The computational cost per iteration of Algorithm 3 is higher that the similar cost of Algorithm 2, due to the inner loop in (4.5)-(4.7) and the additional step given in (4.11), but, on the other hand, we establish convergence of Algorithm 3 assuming only monotonicity of T, while our analysis of Algorithm 2 require paramonotonicity of T. As will be shown in the next section, the generated sequence converges to some solution of VIP(T, C).

Algorithm 3 can be easily implemented, because $P_{C_{k,j}}$ and P_{C_k} have explicit formulae, which we present next. It follows from Proposition 3.3, (4.4), (4.5), (4.8) and (4.9) 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 3 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, Corollary 1.2 allows us to take

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

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

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.

4.2 Convergence analysis of Algorithm 3

For convergence of our method, we assume that T is point-to-point and maximal monotone, and hence continuous by Lemma 1.1(iv). Observe that $\partial g(x) \neq \emptyset$ for all $x \in \mathcal{H}$, because we assume that g is convex and $\text{dom}(g) = \mathcal{H}$.

First we establish the validity of the stopping criterion and the fact that Algorithm 3 is well defined.

Proposition 4.1. *Take C, $C_{k,j}$, C_k, $\tilde{y}^k$, z^k and x^k defined by* (4.1), (4.4), (4.8), (4.7), (4.9) *and* (4.11) *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 (4.4), (4.8) 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 (4.9), using Lemma 2.1(ii) with $x = \tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k)$ and $K = C_k$, we obtain

$$\left\langle z^{k+1} - \left(\tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k) \right), z - z^{k+1} \right\rangle \geq 0 \quad \forall z \in C_k. \tag{4.12}$$

Taking $z^{k+1} = \tilde{y}^k$ in (4.12) 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

$$\liminf_{j\to\infty} \frac{g(y^{k,j})\, \|y^{k,j} - w\|}{g(y^{k,j}) - g(w)} \geq \theta\beta_k > 0. \tag{4.13}$$

Taking into account the inner loop in j given in (4.5) i.e. $y^{k,j+1} = P_{C_{k,j}}(y^{k,j})$ for each k, we obtain, for each $x \in C$,

$$\begin{aligned} \|y^{k,j+1} - x\|^2 &= \|P_{C_{k,j}}(y^{k,j}) - P_{C_{k,j}}(x)\|^2 \leq \|y^{k,j} - x\|^2 - \|y^{k,j+1} - y^{k,j}\|^2 \\ &\leq \|y^{k,j} - x\|^2, \end{aligned} \tag{4.14}$$

using Lemma 2.1(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 2.3(i). It follows that $\tau := \dfrac{1}{-g(w)} \sup\limits_{0 \leq j \leq \infty} \|y^{k,j} - w\|$ is finite and also,

$$g(y^{k,j}) > 0 \qquad \text{for all} \quad j. \tag{4.15}$$

Using (4.14), we get

$$\lim_{j\to\infty} \|y^{k,j+1} - y^{k,j}\| = \lim_{j\to\infty} \|P_{C_{k,j}}(y^{k,j}) - y^{k,j}\| = 0. \tag{4.16}$$

Since $y^{k,j+1}$ belongs to $C_{k,j}$, we have from (4.4) that

$$g(y^{k,j}) \leq \langle v^{k,j}, y^{k,j} - y^{k,j+1} \rangle \leq \|v^{k,j}\| \|y^{k,j} - y^{k,j+1}\|, \tag{4.17}$$

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 (4.16) and (4.17),

$$\liminf_{j\to\infty} g(y^{k,j}) \leq 0. \tag{4.18}$$

It follows from (4.15) and (4.18) 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 (4.13). 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 3.

Proposition 4.2. *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,

$$\frac{1}{\eta_k} \leq 1 \tag{4.19}$$

and

$$\frac{\|T(\tilde{y}^k)\|}{\eta_k} \leq 1. \tag{4.20}$$

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

$$\begin{aligned}
\|\tilde{y}^k - \bar{x}\| &= \|y^{k,j(k)} - \bar{x}\| = \|P_{C_{k,j(k)-1}}(y^{k,j(k)-1}) - P_{C_{k,j(k)-1}}(\bar{x})\| \\
&\leq \|y^{k,j(k)-1} - \bar{x}\| = \|P_{C_{k,j(k)-2}}(y^{k,j(k)-2}) - P_{C_{k,j(k)-2}}(\bar{x})\| \\
&\leq \|y^{k,j(k)-2} - \bar{x}\| \leq \cdots \leq \|y^{k,0} - \bar{x}\| = \|z^k - \bar{x}\|, \qquad (4.21)
\end{aligned}$$

using Lemma 2.1(i) and (4.5).
Let $\tilde{\theta} = 1 + \theta\, T(\bar{x}) \geq 1 + \theta \dfrac{T(\bar{x})}{\eta_k}$, by (4.19). Then

$$\begin{aligned}
\|\tilde{y}^{k+1} - \bar{x}\|^2 &\leq \|z^{k+1} - \bar{x}\|^2 = \left\| P_{C_k}\left(\tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k)\right) - P_{C_k}(\bar{x}) \right\|^2 \\
&\leq \left\| \tilde{y}^k - \frac{\beta_k}{\eta_k} T(\tilde{y}^k) - \bar{x} \right\|^2 \\
&= \|\tilde{y}^k - \bar{x}\|^2 + \frac{\|T(\tilde{y}^k)\|^2}{\eta_k^2} \beta_k^2 - 2\frac{\beta_k}{\eta_k} \langle T(\tilde{y}^k), \tilde{y}^k - \bar{x}\rangle \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 - 2\, \frac{\beta_k}{\eta_k} \langle T(\bar{x}), \tilde{y}^k - \bar{x}\rangle \\
&= \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 - 2\, \frac{\beta_k}{\eta_k} \big(\langle T(\bar{x}), \tilde{y}^k - P_C(\tilde{y}^k)\rangle + \langle T(\bar{x}), P_C(\tilde{y}^k) - \bar{x}\rangle\big) \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 + 2\, \frac{\beta_k}{\eta_k} \langle T(\bar{x}), P_C(\tilde{y}^k) - \tilde{y}^k\rangle \\
&\leq \|\tilde{y}^k - \bar{x}\|^2 + \beta_k^2 + \frac{\beta_k}{\eta_k} \|T(\bar{x})\| \|P_C(\tilde{y}^k) - \tilde{y}^k\| \leq \|\tilde{y}^k - \bar{x}\|^2 + \tilde{\theta}\beta_k^2 \\
&\leq \|z^k - \bar{x}\|^2 + \tilde{\theta}\beta_k^2, \qquad (4.22)
\end{aligned}$$

using (4.21) in the first inequality, Lemma 2.1(i) in the second one, the monotonicity of T and (4.20) in the third one, the definition of $S(T, C)$ in the fourth one, Cauchy-Schwartz inequality in the fifth one, Lemma 2.6 and the definition of $j(k)$ in the sixth one, and (4.21) in the last one.

Using Definition 2.1, (4.22) and (4.3), we conclude that the sequences $\{\tilde{y}^k\}$ and $\{z^k\}$ are quasi-Fejér convergent to $S(T, C)$. □

Proposition 4.3. *Let $\{z^k\}$, $\{\tilde{y}^k\}$ and $\{x^k\}$ be the sequences generated by Algorithm 3. 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 4.2 and Proposition 2.3(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 (4.11).

ii) Apply (4.11) recursively.

iii) It follows from Lemma 2.6 and (4.6)-(4.7) that

$$\operatorname{dist}(\tilde{y}^k, C) \leq \frac{g(\tilde{y}^k)\|\tilde{y}^k - w\|}{g(\tilde{y}^k) - g(w)} \leq \theta\beta_k. \tag{4.23}$$

Define

$$\tilde{x}^{k+1} := \frac{1}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}P_C(\tilde{y}^j). \tag{4.24}$$

Since $\frac{1}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j} = 1$ by (4.10), we get from the convexity of C that $\tilde{x}^{k+1} \in C$. Let

$$\tilde{\beta} := \sum_{j=0}^{\infty}\beta_j^2. \tag{4.25}$$

Note that $\tilde{\beta}$ is finite by (4.3). Then

$$\begin{aligned}
\operatorname{dist}(x^{k+1}, C) &\leq \|x^{k+1} - \tilde{x}^{k+1}\| = \left\|\frac{1}{\sigma_k}\left(\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}(\tilde{y}^j - P_C(\tilde{y}^j))\right)\right\| \\
&\leq \frac{1}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}\|\tilde{y}^j - P_C(\tilde{y}^j)\| = \frac{1}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j}\operatorname{dist}(\tilde{y}^j, C) \\
&\leq \frac{\theta}{\sigma_k}\sum_{j=0}^{k}\frac{\beta_j^2}{\eta_j} \leq \frac{\theta}{\sigma_k}\sum_{j=0}^{k}\beta_j^2 \leq \theta\frac{\tilde{\beta}}{\sigma_k},
\end{aligned} \tag{4.26}$$

using the fact that $\tilde{x}^{k+1}$ belongs to C in the first inequality, (ii) and (4.24) in the first equality, convexity of $\|\cdot\|$ in the second inequality, (4.23) in the third one, (4.19) in the fourth one and (4.25) 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,

$$\lim_{k\to\infty}\sigma_k = \lim_{k\to\infty}\sum_{j=0}^{k}\frac{\beta_j}{\eta_j} \geq \lim_{k\to\infty}\frac{1}{\gamma}\sum_{j=0}^{k}\beta_j = \infty, \tag{4.27}$$

using that $\eta_j = \max\{1, \|T(\tilde{y}^j)\|\} \leq \max\{1, \gamma\} \leq \gamma$ for all j in the first inequality and (4.2) in the last equality. Thus, taking limits in (4.26), we get, using (4.27), that $\lim_{k\to\infty} \operatorname{dist}(x^k, C) = 0$, establishing (iii).

iv) Follows from (iii). □

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

Theorem 4.1. *If $S(T, C) \neq \emptyset$ then all weak cluster points of the sequence $\{x^k\}$ generated by Algorithm 3 solve VIP(T, C).*

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

$$\begin{aligned} \|z^{j+1} - x\|^2 &= \left\| P_{C_j}\left(\tilde{y}^j - \frac{\beta_j}{\eta_j} T(\tilde{y}^j)\right) - P_{C_j}(x) \right\|^2 \leq \left\| \left(\tilde{y}^j - \frac{\beta_j}{\eta_j} T(\tilde{y}^j)\right) - x \right\|^2 \\ &= \|\tilde{y}^j - x\|^2 + \frac{\|T(\tilde{y}^j)\|^2}{\eta_j^2}\beta_j^2 - 2\frac{\beta_j}{\eta_j}\langle T(\tilde{y}^j), \tilde{y}^j - x\rangle \\ &\leq \|z^j - x\|^2 + \beta_j^2 + 2\frac{\beta_j}{\eta_j}\langle T(x), x - \tilde{y}^j\rangle, \end{aligned} \tag{4.28}$$

using Lemma 2.1(i) in the first inequality, and the monotonicity of T and (4.20) in the last inequality. Summing (4.28) from 0 to $k-1$ and dividing by σ_{k-1}, we obtain from Proposition 4.3(ii)

$$\frac{(\|z^k - x\|^2 - \|z^0 - x\|^2)}{\sigma_{k-1}} \leq \frac{\sum_{j=0}^{k-1}\beta_j^2}{\sigma_{k-1}} + \langle T(x), x - x^k\rangle. \tag{4.29}$$

Let $\hat{x}$ be a weak cluster point of $\{x^k\}$. Existence of $\hat{x}$ is guaranteed by Proposition 4.3(i). Note that $\hat{x} \in C$ by Proposition 4.3(iv).

By (4.2), (4.3), (4.27) and boundedness of $\{z^k\}$, taking limits in (4.29) with $k \to \infty$, we obtain that $\langle T(x), x - \hat{x}\rangle \geq 0$ for all $x \in C$. Using Lemma 2.5, we get that $\hat{x} \in S(T, C)$. Therefore, all weak cluster points of $\{x^k\}$ solve VIP(T, C). □

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. These properties were first introduced in [8]. The next lemma will be useful for proving that the sequence generated by our algorithm converges weakly to some point belongs to $S(T, C)$.

Lemma 4.1. *If a sequence $\{x^k\}$ is quasi-Fejér convergent to a closed and convex set S, then*

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

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

Proof. i) The sequence $\{\text{dist}(x^k, S)\}$ is bounded, because $0 \le \text{dist}(x^k, S) \le \|x^k - x\|$ for all $x \in S$, and $\{\|x^k - x\|\}$ converges for all $x \in S$, by Proposition 2.3(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

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

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. that the sequence $\{\text{dist}(x^k, S)\}$ converges.

ii) We will prove that $\{u^k\} := \{P_S(x^k)\}$ is a Cauchy sequence, hence strongly convergent. Using Lemma 2.1(i) with $K = S$, $x = x^q$ and $y = u^p$, we get

$$\|u^q - u^p\|^2 = \|P_S(x^q) - P_S(u^p)\|^2 \leq \|x^q - u^p\|^2 - \|x^q - u^q\|^2. \tag{4.30}$$

Since $\{x^k\}$ is quasi-Fejér convergent to S and $p < q$, we get from (4.30) that

$$\begin{aligned} \|u^q - u^p\|^2 &\leq \|x^p - u^p\|^2 - \|x^q - u^q\|^2 + \sum_{j=p}^{q} \delta_j \\ &\leq \operatorname{dist}(x^p, S)^2 - \operatorname{dist}(x^q, S)^2 + \sum_{j=p}^{\infty} \delta_j. \end{aligned} \tag{4.31}$$

By (i), $\{\operatorname{dist}(x^k, S)\}$ converges, and using the fact $\sum_{j=0}^{\infty} \delta_j < \infty$, we obtain from (4.31) that $\{u^k\}$ is a Cauchy sequence.

□

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

Theorem 4.2. *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 2.2. By Proposition 4.2, $\{\tilde{y}^k\}$ is quasi-Fejér convergent to $S(T, C)$. Therefore, it follows from Lemma 4.1(ii) that $\{P_{S(T,C)}(\tilde{y}^k)\}$ is strongly convergent. Let

$$x^* = \lim_{k\to\infty} P_{S(T,C)}(\tilde{y}^k) = \lim_{k\to\infty} u^k. \tag{4.32}$$

By Proposition 4.3(i) and Theorem 4.1, $\{x^k\}$ 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 2.1(ii) we have that $\langle \hat{x} - u^j, \tilde{y}^j - u^j \rangle \leq 0$ for all j. Define $\xi = \sup_{0\leq j\leq\infty} \|\tilde{y}^j - u^j\|$. By Proposition 4.3(i), $\xi < \infty$. Then,

$$\langle \hat{x} - x^*, \tilde{y}^j - u^j \rangle \leq \langle u^j - x^*, \tilde{y}^j - u^j \rangle \leq \xi \|u^j - x^*\| \qquad \forall j. \tag{4.33}$$

Multiplying (4.33) by $\dfrac{\beta_j}{\eta_j \sigma_{k-1}}$ and summing from 0 to $k - 1$, we get from Proposition 4.3(ii)

$$\left\langle \hat{x} - x^*, x^k - \frac{1}{\sigma_{k-1}} \sum_{j=0}^{k-1} \frac{\beta_j}{\eta_j} u^j \right\rangle \leq \frac{\xi}{\sigma_{k-1}} \sum_{j=0}^{k-1} \frac{\beta_j}{\eta_j} \|u^j - x^*\|. \tag{4.34}$$

Define $\zeta_{k,j} := \frac{1}{\sigma_k} \frac{\beta_j}{\eta_j}$ ($k \geq 0$, $0 \leq j \leq k$). In view of (4.27), $\lim_{k\to\infty} \zeta_{k,j} = 0$ for all j. By (4.10), $\sum_{j=0}^{k} \zeta_{k,j} = 1$ for all k. Then, using (4.32) and Proposition 2.4 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

$$\lim_{k\to\infty} \frac{1}{\sigma_{k-1}} \sum_{j=0}^{k-1} \frac{\beta_j}{\eta_j} u^j = x^*, \tag{4.35}$$

and hence

$$\lim_{k\to\infty} \frac{1}{\sigma_{k-1}} \sum_{j=0}^{k-1} \frac{\beta_j}{\eta_j} \|u^j - x^*\| = 0, \tag{4.36}$$

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

From (4.35) and (4.36), since $\lim_{k\to\infty} x^{i_k} = \hat{x}$, taking limits in (4.34) 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 4.1 no subsequence of $\{x^k\}$ can be bounded, and hence $\lim_{k\to\infty} \|x^k\| = \infty$. □

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

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

Assuming (H), we can replace the left hand side of the inequality in (4.6) by $\tilde{g}(y^{k,j})$, and all our convergence results are preserved; in fact only the proof of Proposition 4.1(iii) has to modified.

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

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

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

Chapter 5

An extragradient-type method with strong convergence

In this chapter we introduce a new iterative method for solving VIP(T, C), which generates a strongly convergent sequence to some point belonging to $S(T, C)$. This is different from the cases of Korpelevich's method and the algorithm in the previous chapters (Algorithms 3), for which only weak convergence has been established.

It is clear that weak and strong convergence are only distinguishable in the infinite-dimensional setting. Therefore, from now on we will work with VIP(T, C) in infinite-dimensional Hilbert spaces. We assume in this section that T is maximal monotone, point-to-point, and uniformly continuous on bounded sets.

We mention that we know no examples in which of Korpelevich algorithm converge weakly but not strongly.

5.1 Statement of Algorithm 4

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

Algorithm 4

Initialization step. Take

$$x^0 \in C. \tag{5.1}$$

Iterative step. Given x^k define

$$z^k := x^k - \beta_k T(x^k). \tag{5.2}$$

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

$$\begin{aligned} j(k) := \min \Big\{ \, j \geq 0 \quad : \quad & \langle T(2^{-j} P_C(z^k) + (1 - 2^{-j})x^k), x^k - P_C(z^k)\rangle \\ & \geq \frac{\delta}{\beta_k} \|x^k - P_C(z^k)\|^2 \Big\}, \end{aligned} \tag{5.3}$$

$$\alpha_k := 2^{-j(k)}, \tag{5.4}$$

$$y^k := \alpha_k P_C(z_k) + (1 - \alpha_k)x^k. \tag{5.5}$$

Define

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

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

$$x^{k+1} := P_{H_k \cap W_k \cap C}(x^0). \tag{5.6}$$

5.2 Convergence analysis of Algorithm 4

First, we establish that Algorithm 4 is well defined.

Proposition 5.1. *Suppose that Algorithm 4 generates an infinite sequence. Then,*

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

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

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

Proof. i) Follows from (5.6).

ii) Assume by contradiction that the minimum in (5.3) is not achieved. In this case, for all $\alpha > 0$, it holds that

$$\langle T(y^k(\alpha)), x^k - P_C(z^k)\rangle < \frac{\delta}{\beta_k}\|x^k - P_C(z^k)\|^2, \tag{5.7}$$

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

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

using Lemma 2.1(iii) in the first inequality, (5.2) in the equality and (5.7) in the second inequality, after taking limits with $\alpha \to \infty$, in view of the continuity of T. Since $\|x^k - P_C(z^k)\| > 0$ by the stopping criterion and $\delta \in (0,1)$, we arrive at a contradiction.

iii) Follows from (5.1) and (5.5), taking into account that $\alpha_k \in [0,1]$ for all $k \ge 0$ by (5.3) and (5.4). □

Next, we establish some properties of Algorithm 4.

Proposition 5.2. *For all k,*

$$\|x^{k+1} - x^0\|^2 \ge \|x^k - x^0\|^2 + \|x^{k+1} - x^k\|^2, \tag{5.8}$$

and

$$\|x^{k+1} - x^k\| \ge \alpha_k\frac{\delta}{\tilde{\beta}}\frac{\|x^k - P_C(z^k)\|^2}{\|T(y^k)\|}. \tag{5.9}$$

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

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

which implies (5.8).

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

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

using (5.2)-(5.5) in the second inequality and the fact that $\beta_k \leq \tilde{\beta}$ for all k in the third one. $\square$

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

Theorem 5.1. *Suppose that Algorithm 4 generates an infinite sequence* $\{x^k\}$*. Then either* $\{x^k\}$ *is bounded and each of its weak cluster points belongs to* $S(T, C) \neq \emptyset$*, or* $S(T, C) = \emptyset$ *and* $\lim_{k\to\infty} \|x^k\| = \infty$.

Proof. If $\{x^k\}$ is bounded, we obtain from (5.8) that the sequence $\{\|x^k - x^0\|\}$ is nondecreasing and bounded, hence convergent. By (5.8) again, $0 \leq \|x^{k+1} - x^k\|^2 \leq \|x^{k+1} - x^0\|^2 - \|x^k - x^0\|^2$, and we conclude that

$$\lim_{k\to\infty} \|x^{k+1} - x^k\| = 0. \tag{5.10}$$

It follows from (5.9) and (5.10) that

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

The sequence $\{P_C(z^k)\}$ is bounded, using boundedness of $\{x^k\}$ and (5.2), and (5.3)-(5.5) imply that $\{y^k\}$ is bounded. It follows from the uniform continuity of T that $\{T(y^k)\}$ is also bounded. Thus,

$$\lim_{k\to\infty} \alpha_k \left\|x^k - P_C(x^k)\right\| = 0. \tag{5.11}$$

Since $\{x^k\}$ is bounded, there exists a subsequence $\{x^{i_k}\}$ of $\{x^k\}$ that converges weakly to some x^*.

We consider now two cases.

Case 1. Suppose that $\{\alpha_k\}$ does not converge to 0, i.e. there exists a subsequence $\{\alpha_{i_k}\}$ of $\{\alpha_k\}$ and some $\alpha > 0$ such that $\alpha_{i_k} \geq \alpha$ for all k. In this case, we define $w^k := P_C(z^k)$ and it follows from (5.11) that

$$\lim_{k\to\infty} \|x^{i_k} - w^{i_k}\| = 0. \tag{5.12}$$

Since T is uniformly continuous, we have

$$\lim_{k\to\infty} \|T(x^{i_k}) - T(w^{i_k})\| = 0. \tag{5.13}$$

Let x^* be a weak cluster point of $\{x^{i_k}\}$. By (5.12), it is also a weak cluster point of $\{w^{i_k}\}$. Without loss of generality, we assume that $\{x^{i_k}\}$ and $\{w^{i_k}\}$ converge weakly to x^*. Let $N_C(x)$ be the normal cone to C at $x \in C$, i.e., $N_C(x) = \{z \in \mathcal{H} : \langle x - y, z\rangle \geq 0 \quad \forall y \in C\}$. Define

$$\hat{T}(x) := T(x) + N_C(x). \tag{5.14}$$

It is known that $\hat{T}$, as given in (5.14), is maximal monotone and that $0 \in \hat{T}(x)$ if and only if $x \in S(T, C)$; see [61] and Example 1.1 in Section 1.2.

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

$$\langle x - y, u - T(x)\rangle \geq 0 \quad \forall y \in C. \tag{5.15}$$

On the other hand, since $w^k = P_C(x^k - \beta_k T(x^k))$ and $x \in C$, it follows from Lemma 2.1(ii), with $K = C$ and $x = x^k - \beta_k T(x^k)$, that

$$\langle x - w^k, x^k - \beta_k T(x^k) - w^k\rangle \leq 0 \quad \forall\, x \in C \text{ and } k \geq 0. \tag{5.16}$$

Since β_k is positive for all k, we get from (5.16)

$$\left\langle x - w^k, \frac{x^k - w^k}{\beta_k} - T(x^k)\right\rangle \leq 0 \quad \forall\, x \in C \text{ and } k \geq 0. \tag{5.17}$$

Thus,

$$\begin{aligned}
\langle x - w^k, u\rangle &\geq \langle x - w^k, T(x)\rangle \geq \langle x - w^k, T(x)\rangle + \left\langle x - w^k, \frac{x^k - w^k}{\beta_k} - T(x^k)\right\rangle \\
&= \langle x - w^k, T(x) - T(w^k)\rangle + \langle x - w^k, T(w^k) - T(x^k)\rangle \\
&\quad + \left\langle x - w^k, \frac{x^k - w^k}{\beta_k}\right\rangle \\
&\geq \langle x - w^k, T(w^k) - T(x^k)\rangle + \left\langle x - w^k, \frac{x^k - w^k}{\beta_k}\right\rangle \\
&\geq -\|x - w^k\| \left(\|T(w^k) - T(x^k)\| + \frac{1}{\beta_k}\|w^k - x^k\|\right) \\
&\geq -\|x - w^k\| \left(\|T(w^k) - T(x^k)\| + \frac{1}{\hat{\beta}}\|w^k - x^k\|\right),
\end{aligned} \tag{5.18}$$

using (5.15) with $y = w^k$ because $w^k \in C$ in the first inequality, (5.17) in the second inequality, the monotonicity of T in the third one, Cauchy-Schwartz inequality in the fourth one and the fact that $\beta_k \geq \hat{\beta} > 0$ for all k in the last one.

Now, using (5.12) and (5.13), we obtain that the subsequences $\{w^{i_k} - x^{i_k}\}$ and $\{T(w^{i_k}) - T(x^{i_k})\}$ strongly converge to zero. Then, we can take limits with $k \to \infty$ in (5.18) over the subsequence with superindices $\{i_k\}$ and, using that $\{w^{i_k}\}$ converges weakly to x^*, we obtain that

$$\langle x - x^*, u\rangle \geq 0 \qquad \forall\, (x, u) \in G(\hat{T}). \tag{5.19}$$

Since $\hat{T}$ is maximal monotone, it follows from (5.19) that $(x^*, 0) \in G(\hat{T})$ i.e. $0 \in \hat{T}(x^*) = T(x^*) + N_C(x^*)$ and hence $x^* \in S(T, C)$.

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

$$\hat{y}^k = 2\alpha_k P_C(z^k) + (1 - 2\alpha_k)x^k, \tag{5.20}$$

it follows from the definition of $j(k)$ in (5.3) that

$$\langle T(\hat{y}^k), x^k - P_C(z^k)\rangle < \frac{\delta}{\beta_k}\, \|x^k - P_C(z^k)\|^2. \tag{5.21}$$

Note that $\hat{y}^k - x^k = 2\alpha_k(P_C(x^k) - x^k)$ by (5.20). Since, as discussed above, $\{P_C(z^k)\}$ and $\{x^k\}$ are bounded, it follows from the assumption of this case that $\lim_{k\to\infty} \|\hat{y}^k - x^k\| = 0$. Thus, we get from (5.21),

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

using (5.2) in the second equality, and Cauchy-Schwartz inequality and Lemma 2.1(iii) in the second inequality. Now, an elementary rearrangement yields

$$\begin{aligned}
\|T(\hat{y}^k) - T(x^k)\|\|x^k - P_C(x^k)\| &\geq \frac{(1-\delta)}{\beta_k}\|x^k - P_C(x^k)\|^2 \\
&\geq \frac{(1-\delta)}{\tilde{\beta}}\|x^k - P_C(x^k)\|^2, \qquad (5.22)
\end{aligned}$$

using the fact that $\beta_k \leq \tilde{\beta}$ for all k in the second inequality. Since $\|x^k - P_C(x^k)\| > 0$ for all k because $x^k \notin S(T, C)$ for all k in view of Proposition 2.1, it follows from (5.22) that

$$\|T(\hat{y}^k) - T(x^k)\| \geq \frac{(1-\delta)}{\tilde{\beta}} \|x^k - P_C(x^k)\| \geq 0. \tag{5.23}$$

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

$$\lim_{k\to\infty} \|T(\hat{y}^k) - T(x^k)\| = 0. \tag{5.24}$$

Taking limits with $k \to \infty$ in (5.23) and using (5.24), we get

$$0 \geq \lim_{k\to\infty} \|x^k - P_C(x^k)\| \geq 0.$$

Therefore, $\lim_{k\to\infty} \|x^k - P_C(x^k)\| = \lim_{k\to\infty} \|x^k - w^k\| = 0$. From here on, we can proceed as in the previous case, from (5.12) on, considering the whole sequences $\{x^k\}$, $\{w^k\}$ instead of $\{x^{i_k}\}$, $\{w^{i_k}\}$, in order to complete the proof of the first assertion.

Suppose now that $S(T, C) = \emptyset$. Using the previous assertion in this proposition, we obtain that $\{x^k\}$ is unbounded. Since the sequence $\{\|x^k - x^0\|\}$ is nondecreasing by (5.8), it follows that $\lim_{k\to\infty} \|x^k - x^0\| = \infty$ and so $\lim_{k\to\infty} \|x^k\| = \infty$. □

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

$$Y := \{x \in \mathcal{H} \, : \, \langle z - x, x^0 - x\rangle \leq 0 \quad \forall z \in S(T, C)\}. \tag{5.25}$$

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

Proposition 5.3. *If $x^k \in Y$ then*

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

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

Proof. i) Note that

$$\begin{aligned} \langle T(y^k), x^* - y^k\rangle &= \langle T(y^k) - T(x^*), x^* - y^k\rangle + \langle T(x^*), x^* - y^k\rangle \\ &\leq \langle T(x^*), x^* - y^k\rangle \leq 0, \end{aligned} \tag{5.26}$$

for any $x^* \in S(T, C)$, using the monotonicity of T in the first inequality, and the definition of $S(T, C)$ together with Proposition 5.1(iii) in the second inequality. It follows from (5.26) that $S(T, C) \subseteq H_k$.

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

ii) Since $S(T, C) \subseteq H_k \cap W_k \cap C$ and $S(T, C)$ is nonempty, it follows that $H_k \cap W_k \cap C$ is nonempty. Thus the next iterate x^{k+1} is well defined, in view of (5.6). By Lemma 2.1(ii), we have that

$$\langle z - x^{k+1}, x^0 - x^{k+1} \rangle \leq 0 \quad \forall z \in H_k \cap W_k \cap C. \tag{5.27}$$

Since $S(T, C) \subseteq H_k \cap W_k \cap C$ for all k, (5.27) holds for all $z \in S(T, C)$, and so $x^{k+1} \in Y$ by (5.25).

□

Corollary 5.1. *Algorithm 4 is well defined and generates infinite sequences $\{x^k\}$, $\{y^k\}$ and $\{u^k\}$ such that $\{x^k\} \subset Y$ and $S(T, C) \subseteq H_k \cap W_k \cap C$ for all k.*

Proof. It is enough to observe that $x^0 \in Y$ and apply inductively Proposition 5.3.

□

Corollary 5.2. *The sequence $\{x^k\}$ generated by Algorithm 4 is bounded and each of its weak cluster points belong to $S(T, C)$.*

Proof. If the solution set is nonempty, in view of (5.6) we have that $\|x^{k+1} - x^0\| \leq \|z - x^0\|$ for all $z \in H_k \cap W_k \cap C$. Since $S(T, C) \subseteq H_k \cap W_k \cap C$ by Corollary 5.1, it follows that $\|x^{k+1} - x^0\| \leq \|x^* - x^0\|$ for all $x^* \in S(T, C)$. Thus, $\{x^k\}$ is bounded, and by Theorem 5.1, all its weak cluster points belong to $S(T, C)$. □

Finally, we can now state and prove our main result on Algorithm 4.

Theorem 5.2. *Assume that $S(T, C) \neq \emptyset$ and let $\{x^k\}$ be a sequence generated by Algorithm 4. Define $x^* = P_{S(T,C)}(x^0)$. Then $\{x^k\}$ converges strongly to x^*.*

Proof. Note that x^*, the orthogonal projection of x^0 onto $S(T, C)$, exists because the solution set $S(T, C)$ is nonempty by assumption, and closed and convex by Lemma 2.2. By the definition of x^{k+1}, we have that

$$\|x^{k+1} - x^0\| \leq \|z - x^0\| \qquad \forall\, z \in H_k \cap W_k \cap C. \tag{5.28}$$

Since $x^* \in S(T,C) \subseteq H_k \cap W_k \cap C$ for all k, it follows from (5.28) that

$$\|x^k - x^0\| \leq \|x^* - x^0\|, \tag{5.29}$$

for all k. By Corollary 5.2, $\{x^k\}$ is bounded and each of its weak cluster points belongs to $S(T,C)$. Let $\{x^{i_k}\}$ be any weakly convergent subsequence of $\{x^k\}$, and let $\hat{x} \in S(T,C)$ be its weak limit. Observe that

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

where the inequality follows from (5.29). By the weak convergence of $\{x^{i_k}\}$ to $\hat{x}$, we obtain

$$\limsup_{k\to\infty} \|x^{i_k} - x^*\|^2 \leq 2(\|x^* - x^0\|^2 - \langle \hat{x} - x^0, x^* - x^0\rangle). \tag{5.30}$$

Applying Lemma 2.1(ii) with $K = S(T,C)$, $x = x^0$ and $z = \hat{x} \in S(T,C)$, and taking into account that x^* is the projection of x^0 onto $S(T,C)$, we have that

$$\langle x^0 - x^*, \hat{x} - x^*\rangle \leq 0. \tag{5.31}$$

Now, using (5.31) we have

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

It follows that

$$\langle \hat{x} - x^0, x^* - x^0\rangle \geq \|x^* - x^0\|^2. \tag{5.32}$$

Combining (5.32) with (5.30), we conclude that $\{x^{i_k}\}$ converges strongly to x^*. Thus, we have shown that every weakly convergent subsequence of $\{x^k\}$ converges strongly to x^*. Hence, the whole sequence $\{x^k\}$ converges strongly to $x^* \in S(T,C)$. $\square$

Bibliography

[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] Attouch, H., Baillon, J.B., Théra, M. Variational sum of monotone operators. *Journal of Convex Analysis* **1** (1994) 1-29.

[5] 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.

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

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

[9] Bensoussan, A. Points de Nash dans le cas de fonctionnelles quadratiques et jeux différentiels linéaires à N personnes. *SIAM Journal on Control and Optimization* **12** (1974) 460-499.

[10] Bertsekas, D. *Nonlinear Programming*. Athena Scientific, Belmont (1995).

[11] Boggs, B.T., Tolle, J.W. Sequential quadratic programming. *Acta Numerica* **4** (1996) 151.

[12] Bonnans, J.F., Gilbert, J.Ch., Lemarechal, C. *Numerical Optimization: Theoretical and Practical Aspects*. Springer, Berlin (2006).

[13] Brezis, H. *Analyse fonctionnelle. Théorie et application*. Masson, Paris (1983).

[14] Brezis, H., Sibony, M. Méthodes d'approximation et d'iteration pour les opérateurs monotones. *Archive for Rational Mechanics and Analysis* **28** (1968) 59-82.

[15] Browder, F.E. Nonlinear operators and nonlinear equations of evolution in Banach spaces. *Proceedings of Symposia in Pure Mathematics, American Mathematical Society* **18** (1976).

[16] Browder, F.E., Hess, P. Nonlinear mappings of monotone type in Banach spaces. *Journal of Functional Analysis* **11** (1972) 251-294.

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

[18] 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.

[19] Burachik, R.S., Iusem, A.N. *Set-Valued Mappings and Enlargements of Monotone Operators*. Springer, Berlin (2008).

[20] Butnariu, D., Iusem, A.N. *Totally convex functions for fixed points computation and infinite dimensional optimization.* Kluwer, Dordrecht (2000).

[21] 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.

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

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

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

[25] Ferris, M.C., Pang, J.S. Engineering and economic applications of complementarity problems. *SIAM Review* **39** (1997) 669-713.

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

[27] Fukushima, M. On the convergence of a class of outer approximation algorithms for convex programs. *Journal of Computational and Applied Mathematics* **10** (1984) 147-156.

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

[29] Fukushima, M., Luo, Z.Q., Tseng, P. A sequential quadratically constrained quadratic programming method for differentiable convex minimization. *SIAM Journal on Optimization* **13** (2003) 1098-1119.

[30] Gabay, D., Moulin, H. On the uniqueness and stability of Nash equilibria in noncooperative games. In Bensoussan, A., Kleindorfer, P., Tapiero, C.S., editors, *Applied Stochastic in Econometrics and Management Science.* North-Holland (1980) 271-292.

[31] Gafni, E.N., Bertsekas, D. Convergence of a gradient projection method. Technical Report LIDS-P-1201, Laboratory for Information and Decision Systems, M.I.T. (1982).

[32] Güler, O. New proximal point algorithms for convex minimization. *SIAM Journal on Optimization* **2** (1992) 649-664.

[33] Harker, P.T., Pang, J.S. Finite dimensional variational inequalities and nonlinear complementarity problems: a survey of theory, algorithms and applications. *Mathematical Programming* **48** (1990) 161-220.

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

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

[36] Hiriart-Urruty, J.-B., Lemaréchal, C. *Convex Analysis and Minimization Algorithms*. Springer, Berlin (1993).

[37] Ichiishi, T. *Game Theory for Economic Analysis.* Academic Press, New York (1983).

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

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

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

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

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

[43] 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.

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

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

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

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

[48] Krasnoselskii, M.A. Two observations about the method of succesive approximations. *Uspekhi Matematicheskikh Nauk* **10** (1955) 123-127.

[49] Lapidus, M. Formules de Trotter et calcul opérationnel de Feynman. *Thèse d'État, Université Paris VI* (1986).

[50] Lions, P.L., Stampacchia, G. Variational inequalities. *Communications on Pure and Applied Mathematics* **20** (1967) 493-519.

[51] Mancino, O.G., Stampacchia, G. Convex Programming and variational inequalities. *Journal of Optimization Theory and Applications* **9** (1972) 3-23.

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

[53] Martínez, J.M., Pilotta E.A. Inexact restoration algorithm for constrained optimization. *Journal of Optimization Theory and Applications* **104** (2000) 135-163.

[54] Minty, G. Monotone networks. *Proceedings of the Royal Society* **257** (1960) 194-212.

[55] Minty, G. A theorem on monotone sets in Hilbert spaces. *Journal of Mathematical Analysis and Applications* **11** (1967) 434-439.

[56] Moreau, J. Proximité et dualité dans un espace hilbertien. *Bulletin de la Societé Mathématique de France* **93** (1965) 273-299.

[57] Nesterov, Y.E. *Effective Methods in Nonlinear Programming*. Moscow (1989).

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

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

[60] Rockafellar, R.T. *Convex Analysis.* Princeton, New York (1970).

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

[62] Rockafellar, R.T. On the maximal monotonicity of subdifferential mapping. *Pacific Journal of Mathematics* **33** (1970) 209-216.

[63] Rockafellar, R.T. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization* **14** (1976) 877-898.

[64] Rockafellar, R.T., Wets, R.J-B. *Variational Analysis.* Springer, Berlin (1998).

[65] Sibony, M. Méthodes itératives pour les équations et inéquations aux dérivées partielles non linéaires de type monotone. *Calcolo* **7** (1970) 65-183.

[66] 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.

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

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

[69] Stampacchia, G. Formes bilinéaires coercives sur les ensembles convexes. *Comptes Rendus de l'Académie des Sciences de Paris* **258** (1964) 4413-4416.

[70] Zarantonello, E.H. Solving functional equations by contractive averaging. Report 160, Mathematics Research Center, University of Wisconsin, Madison (1960).