Pure and Applied Functional Analysis
Volume 3, Number 3, 2018, 403–415

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

A NEW BREGMAN PROJECTION METHOD FOR SOLVING VARIATIONAL INEQUALITIES IN HILBERT SPACES

AVIV GIBALI

Abstract. In this paper we are concern with solving variational inequalities for monotone and Lipschitz mappings in real Hilbert spaces. Motivated by the works of Popov [23], Malitsky and Semenov [22] and Semenov [24], we propose an extension of the subgradient extragradient method (Censor et al [6–8]) with Bregman projections which calls for only one evaluation of the variational inequalities associated mapping $\mathcal{F}$ per each iteration. Two numerical experiments are given which demonstrate the algorithm performances. Our result generalize and extend several existing results in the literature.

1. Introduction

In this paper we focus on the classical Variational Inequality (VI) of Fichera [12, 13] and Stampacchia [25] (see also Kinderlehrer and Stampacchia [18]) which consists of finding a point $x^* \in C$ such that

$$\langle \mathcal{F}(x^*), x - x^* \rangle \geq 0 \quad \text{for all } x \in C, \tag{1.1}$$

where C is non-empty, closed convex subset of the Hilbert space $\mathcal{H}$ and $\mathcal{F}\colon \mathcal{H} \to \mathcal{H}$ is a given mapping. we denote the solution set of (1.1) as $\mathrm{Sol}(\mathcal{F}, C)$.

This problem plays an important role as a modelling tool in various fields such as Optimization Theory, Nonlinear Analysis, differential equations and more. For an extensive and excellent books on theory, algorithms and applications to VIs see Facchinei and Pang book [11], Kinderlehrer and Stampacchia [18].

One fundamental example which can be reformulated as a variational inequality is the following constrained minimization.

Example 1.1. Let $C \subset \mathcal{H}$ be a nonempty, closed and convex subset of real Hilbert space $\mathcal{H}$ and let $f : \mathcal{H} \to \mathbb{R}$ be a continuously differentiable function which is convex on C. Then x^* is a minimizer of f over C iff x^* solves the VI

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0 \text{ for all } x \in C, \tag{1.2}$$

where ∇f is the gradient of f.

2010 *Mathematics Subject Classification.* 47J20, 58E35, 65K15.
Key words and phrases. Variational inequalities, subgradient extragradient method, Bregman distance, Hilbert spaces.
This work is supported by the EU FP7 IRSES program STREVCOMS, grant no. PIRSES-GA-2013-612669.

One of the simplest iterative scheme for solving constrained minimization problems is the well-known *Projected Gradient Method* ([15, 20]), given the current iterate x^k, the next iterate x^{k+1} is calculated as follows.

$$x^{k+1} = P_C(x^k - \gamma \nabla f(x^k)), \tag{1.3}$$

where P_C denoted the orthogonal projections onto C (explained further) and γ is some positive number. This of course led to introduce an iterative method for solving VIs. The convergence of such algorithm has been studied by a number of authors, for example, Dafermos [10] shows that, if ∇f is strongly monotone on C then the sequence $\{x^k\}_{k=0}^{\infty}$, generated by (1.3), is a globally converges to the unique solution of (1.1). It appears that if the strong monotonicity assumption is dropped, then the situation becomes more complicated, and quite different from the case of convex optimization. In order to deal with this situation, Korpelevich [19] (also Antipin [1]) proposed the *Extragradient Method* which converges for monotone mappings. In this method, per each iteration, in order to get the next iterate x^{k+1}, two orthogonal projections onto C are calculated, according to the following iterative step. Given the current iterate x^k, calculate the next iterate x^{k+1} via

$$\begin{cases} y^k = P_C(x^k - \gamma \mathcal{F}(x^k)) \\ x^{k+1} = P_C(x^k - \gamma \mathcal{F}(y^k)) \end{cases} \tag{1.4}$$

where P_C denoted the orthogonal projections onto C (explained further), $\gamma \in (0, 1/L)$, and L is the Lipschitz constant of $\mathcal{F}$ (or γ is replaced by a sequence of $\{\gamma_k\}_{k=1}^{\infty}$ which is updated by some adaptive procedure, see for example [17]).

Although convergence of the extragradient method is guaranteed under the assumptions of Lipschitz continuity and monotonicity (even pseudo-monotonicity), there is still the need to calculate two evaluations of $\mathcal{F}$ and two projections onto the VI feasibility set C. Regarding the projection, if the set C is a general closed and convex subset, then there is the need to compute two projections per each iteration, which translated to a minimum norm problem

$$\min\{\|x - (x^k - \gamma \mathcal{F}(x^k))\| \mid \text{for all } \ x \in C\}, \tag{1.5}$$

and this might effect the computationally of the method. So, one step in the direction of simplifying the extragradient method is Censor et. al. [6–8] *Subgradient Extragradient Method*. In this method, the second orthogonal projection onto the feasible set is replaced by an easy computed projection onto some constructible set. Given the current iterate x^k, calculate the next iterate x^{k+1} via

$$\begin{cases} y^k = P_C(x^k - \gamma \mathcal{F}(x^k)) \\ x^{k+1} = P_{T_k}(x^k - \gamma \mathcal{F}(y^k)) \end{cases} \tag{1.6}$$

where P_{T_k} is the orthogonal projection onto the set T_k defined as

$$T_k := \begin{cases} \left\{ w \in \mathcal{H} \mid \left\langle (x^k - \gamma \mathcal{F}(x^k)) - y^k, w - y^k \right\rangle \leq 0 \right\}, & \text{if } x^k - \gamma \mathcal{F}(x^k) \neq y^k, \\ \mathcal{H}, & \text{if } x^k - \gamma \mathcal{F}(x^k) = y^k. \end{cases} \tag{1.7}$$

Observe that both the extragradient and the subgradient extragradient methods, require two evaluations of $\mathcal{F}$ per each iteration. Popov [23] proposed a modification of the extragradient method that uses only one evaluation of $\mathcal{F}$ per each iteration. Following Popov's work, Malitsky and Semenov [22] proposed a modification of the subgradient extragradient method which requires only one evaluation of $\mathcal{F}$ per each iteration. Recently, Semenov [24] used Popov's idea and extended the extragradient method using Bregman projections, which generalize the orthogonal metric projection. Following these developments, we propose an extension of the subgradient extragradient method in the spirit of Popov with Bregamn projections in real Hilbert spaces. In the next subsection we provide more details and descriptions of the above methods.

The paper is organized as follows. In Section 2 we present definitions and notions that will be need for the rest of the paper. In Section 3 our two new extensions are presented and analysed. In section 4 a numerical example is given which demonstrate our algorithm performances. Final remarks are given in Section 5.

1.1. **Relation to previous work.** Let $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a bi-function and $C \subseteq \mathbb{R}^n$ and $Q \subseteq \mathbb{R}^m$. The saddle-point problem consists of finding a point $(x^*, y^*) \in C \times Q$ such that

$$f(x^*, y) \le f(x^*, y^*) \le f(x, y^*) \tag{1.8}$$

for all $x \in C$ and $y \in Q$.

One of the simplest gradient methods for solving (1.8) is presented by Arrow, Hurwicz and Uzawa (AHU) in 1958 [2]. Under the assumption that f is differentiable, convex-concave, and its gradient is Lipschitz gradient and the set of saddle point is non-empty, the iterative method of AHU converges in Euclidean spaces. As the assumptions for saddle points problem is quite rigid, Korpelevich in 1976 [19] proposed the extragradient method (1.4) which converges under weaker assumption than the AHU method and it is actually a modification of the gradient method by using extrapolation and hence two evaluations per each iteration.

As mentioned before, the extragradient method for solving VIs, requires two evaluations of the associated mapping $\mathcal{F}$ as well as two orthogonal projections per each iteration. So, Popov in 1980 [23] proposed a modification of the extragradient method in Euclidean spaces that uses only one evaluation of $\mathcal{F}$ per each iteration. Given the current iterates $x^k, y^k \in C$, calculate the next iterate x^{k+1}, y^{k+1} via

$$\begin{cases} x^{k+1} = P_C(x^k - \gamma \mathcal{F}(y^k)) \\ y^{k+1} = P_C(x^{k+1} - \gamma \mathcal{F}(y^k)) \end{cases} \tag{1.9}$$

where $\gamma \in (0, 1/3L)$, and L is the Lipschitz constant of $\mathcal{F}$.

Gradient methods and in particular extragradient methods have been studied, modified and extended intensively in the last decades, and among all the many developments which are introduced, there is the subgradient extragradient method 1.6 of Censor et al. [6–8]. The subgradient extragradient method requires one orthogonal projection onto the feasible set C and one easily computable projection onto a constructible set. The drawback of the method is the need to evaluate $\mathcal{F}$ at two different points per each iteration. So, in the spirit of Popov, Malitsky and Semenov [22] proposed the following method. Given the current iterates x^k, y^k, y^{k-1}, calculate the next iterate x^{k+1}, y^{k+1} via

$$\begin{cases} x^{k+1} = P_{T_k}(x^k - \gamma\mathcal{F}(y^k)) \\ y^{k+1} = P_C(x^{k+1} - \gamma\mathcal{F}(y^k)) \end{cases} \tag{1.10}$$

where P_{T_k} is the orthogonal projection onto the set T_k (slightly different from (1.11)) defined as

(1.11)

$$T_k := \begin{cases} \left\{w \in \mathcal{H} \mid \left\langle (x^k - \gamma\mathcal{F}(y^{k-1})) - y^k, w - y^k \right\rangle \leq 0\right\}, & \text{if } x^k - \gamma\mathcal{F}(y^{k-1}) \neq y^k, \\ \mathcal{H}, & \text{if } x^k - \gamma\mathcal{F}(y^{k-1}) = y^k. \end{cases}$$

Under the assumption of monotonicity and L-Lipschitz continuity of $\mathcal{F}$, with $\gamma \in (0, 1/3L)$ weak convergence in real Hilbert spaces is proved in [22, Theorem 1]. Very recently, Semenov [24] introduced a new modification of extragradient method (a mirror descent variant) for solving VIs in Euclidean spaces with pseudo-monotone mapping $\mathcal{F}$. Semenov proposed method is actually the extragradient method (1.4) when the Euclidean distances are replaced with the generalized Bregman distances.

Following the above developments, we wish to present a subgradient extragradient method with Bregman projections in real Hilbert spaces, which generalizes the above methods.

2. Preliminaries

Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and the induced norm $\|\cdot\|$, and let C be a nonempty, closed and convex subset of $\mathcal{H}$. We write $x^k \rightharpoonup x$ to indicate that the sequence $\left\{x^k\right\}_{k=0}^{\infty}$ converges weakly to x and $x^k \to x$ to indicate that the sequence $\left\{x^k\right\}_{k=0}^{\infty}$ converges strongly to x.

We now recall some definitions and properties of mappings and operators.

Definition 2.1. Let $\mathcal{F} : \mathcal{H} \to \mathcal{H}$ be some mapping.

- The mapping $\mathcal{F}$ is called `Lipschitz-continuous` on $\mathcal{H}$ with constant $L > 0$, iff there exists $L > 0$ such that

$$\|\mathcal{F}(x) - \mathcal{F}(y)\| \leq L\|x - y\| \text{ for all } x, y \in \mathcal{H}. \tag{2.1}$$

- The mapping $\mathcal{F}$ is called `monotone` on $\mathcal{H}$ iff

$$\langle \mathcal{F}(x) - \mathcal{F}(y), x - y \rangle \geq 0 \quad \text{for all } x, y \in \mathcal{H}. \tag{2.2}$$

- The mapping $\mathcal{F}$ is called `hemi-continuous` iff for any $x, y, z \in \mathcal{H}$, the function $t \mapsto \langle z, \mathcal{F}(tx + (1-t)y) \rangle$ of $[0, 1]$ into $\mathbb{R}$ is continuous.

Definition 2.2. Let $f : \mathcal{H} \to \mathbb{R}$ be a convex differentiable function.

- The `domain` of the function f, denoted by $\text{dom}f$ an defined as

$$\text{dom}f := \{x \in \mathcal{H} \mid f(x) < +\infty\} \tag{2.3}$$

When $\text{dom}f \neq \emptyset$, we say that f is `proper`.

- The `subdifferential set` of f at a point x, denote by $\partial f(x)$ is defined as

$$\partial f(x) := \{\xi \in \mathcal{H} \mid f(y) - f(x) \geq \langle \xi, y - x\rangle \text{ for all } y \in \mathcal{H}\} \tag{2.4}$$

an element $\xi \in \partial f(x)$ is called `subgradient`. In case that the function f is continuously differentiable then $\partial f(x) = \{\nabla f(x)\}$, this is the `gradient` of f.

- The `Fenchel conjugate` function of f is the convex function $f^* : \mathcal{H} \to \mathbb{R}$ defined by

$$f^*(\xi) := sup\{\langle \xi, x\rangle - f(x) \mid x \in \mathcal{H}\}. \tag{2.5}$$

- The function f is called `Legendre` iff it satisfies the following two conditions.
  (1) int dom $f \neq \emptyset$ and the subdifferential ∂f is single-valued on its domain.
  (2) int dom$f^* \neq \emptyset$ and ∂f^* is single-valued on its domain.
- The function f is called `strongly convex` with constant $\sigma > 0$, iff

$$f(x) - f(x) \geq \langle \nabla f(x), y - x\rangle + \frac{\sigma}{2}\|y - x\|^2. \tag{2.6}$$

- The function f is called `weakly-weakly continuous` iff

$$x^k \rightharpoonup x \quad \Longrightarrow \quad f(x^k) \rightharpoonup f(x). \tag{2.7}$$

Let C be a closed convex subset of $\mathcal{H}$. For every element $x \in \mathcal{H}$, there exists a unique nearest point in C, denoted by $P_C(x)$ such that

$$\|x - P_C(x)\| = \min\{\|x - y\| \mid y \in C\}. \tag{2.8}$$

The operator P_C is called the *metric projection* of x onto C and some of its properties are summarized in the next lemma, see e.g., [14].

Lemma 2.3. Let $C \subseteq \mathcal{H}$ be a closed convex set, P_C fulfils the following:
(1) $\langle x - P_C(x), y - P_C(x)\rangle \leq 0$ for all $x \in \mathcal{H}$ and $y \in C$;
(2) $\|P_C(x) - y\|^2 \leq \|x - y\|^2 - \|x - P_C(x)\|^2$ for all $x \in \mathcal{H}$, $y \in C$;

Definition 2.4. Given some function $f : \mathcal{H} \to \mathbb{R}$, the bi-function D_f : dom$f \times$ intdom$f \to [0, +\infty)$, which is defined by

$$D_f(x, y) := f(x) - f(y) - \langle \nabla f(y), x - y\rangle, \tag{2.9}$$

is called the `Bregman distance` (see for example [3, 9]).

For different choices of the function f, the Bregman distance generates some known distances, for example, for $f(x) = \|x\|^2$, we obtain the squared Euclidean distance, that is $D_f(x, y) = \|x - y\|^2$. Another useful generalization is when $f(x) = -\sum_{i=1}^{n} x_i \log(x_i)$ is the *Shannon's entropy* for $x \in \mathbb{R}^n_{++} := \{w \in \mathbb{R}^n \mid w_i > 0\}$, then we obtain the *Kullback–Leibler cross entropy* from statistics, that is

$$D_f(x, y) = \sum_{i=1}^{n} \left(x_i \log\left(\frac{x_i}{y_i}\right) - 1\right) + \sum_{i=1}^{n} y_i. \tag{2.10}$$

The Bregman distance fulfils the following important property, which is called the *three point identity*.

Corollary 2.5. *For any* $x \in$ dom$f$ *and* $y, z \in$ intdom$f$,

$$D_f(x, y) + D_f(y, z) - D_f(x, z) = \langle \nabla f(z) - \nabla f(y), x - y\rangle. \tag{2.11}$$

The *Bregman projection* (see e.g., [3]) with respect to f of $x \in \text{int dom} f$ onto a nonempty, closed and convex set $C \subset \text{intdom} f$ is defined as the unique vector $\Pi_C(x) \in C$, which satisfies

$$\Pi_C(x) := \inf\{D_f(y,x) \mid y \in C\}. \tag{2.12}$$

The Bregman projection has a variational characterization (see for example [4, Corollary 4.4]), similarly to the metric projection in Hilbert spaces.

Corollary 2.6.

$$\bar{x} = \Pi_C(x) \quad \Leftrightarrow \quad \langle \nabla f(x) - \nabla f(\bar{x}), y - \bar{x} \rangle \leq 0 \textit{ for all } y \in C \tag{2.13}$$

Note that by the definition of the Bregman distance and (2.6) we get that

$$D_f(x,y) \geq \frac{1}{2}\|x - y\|^2. \tag{2.14}$$

Next lemma is an analogue for Bregman distance of the celebrated Opial's lemma.

Lemma 2.7. Let $\{x^k\}_{k=0}^{\infty}$ be a sequence in $\mathcal{H}$ such that $x^k \rightharpoonup x$. Assume that $f\colon \mathcal{H} \to \mathbb{R}$ is a strongly convex, differential function with weakly-weakly continuous ∇f, (see (2.7)). Then for all $y \neq x$

$$\liminf_{k\to\infty} D_f(x, x^k) < \liminf_{k\to\infty} D_f(y, x^k). \tag{2.15}$$

Proof. Using Corollary 2.5, we have

$$D_f(y, x^k) = D_f(y, x) + D_f(x, x^k) + \left\langle \nabla f(x) - \nabla f(x^k), y - x \right\rangle. \tag{2.16}$$

Since for all $y \neq x$ $D_f(y,x) > 0$ and $\nabla f(x^k) \rightharpoonup \nabla f(x)$ as $k \to \infty$, we obtain the desired. □

Lemma 2.8. Let M be a closed convex set in $\mathcal{H}$, $\{x^k\}_{k=0}^{\infty}$ be a sequence in $\mathcal{H}$. Suppose that the following two conditions hold.

(1) All weak cluster points of $\{x^k\}_{k=0}^{\infty}$ lie in M;
(2) For all $z \in M$ there exist $\lim_{k\to\infty} D_f(z, x^k)$.

Then $\{x^k\}_{k=0}^{\infty}$ weakly converges to some element of M.

Proof. On the contrary assume that the sequence $\{x^k\}_{k=0}^{\infty}$ has at least two weak cluster points $\bar{x} \in \text{Sol}(\mathcal{F}, C)$ and $\tilde{x} \in \text{Sol}(\mathcal{F}, C)$ such that $\bar{x} \neq \tilde{x}$. Let $\{x^{k_n}\}_{n=0}^{\infty}$ be a sequence such that $x^{k_n} \rightharpoonup \bar{x}$ as $n \to \infty$. Then by Lemma 2.7 we have

$$\begin{aligned} \lim_{k\to\infty} D_f(\bar{x}, x^k) &= \lim_{n\to\infty} D_f(\bar{x}, x^{k_n}) = \liminf_{n\to\infty} D_f(\bar{x}, x^{k_n}) \\ &< \liminf_{n\to\infty} D_f(\tilde{x}, x^{k_n}) = \lim_{n\to\infty} D_f(\tilde{x}, x^{k_n}) \\ &= \lim_{k\to\infty} D_f(\tilde{x}, x^k). \end{aligned} \tag{2.17}$$

We can now proceed analogously to the proof that

$$\lim_{k\to\infty} D_f(\tilde{x}, x^k) < \lim_{k\to\infty} D_f(\bar{x}, x^k), \tag{2.18}$$

which is impossible, and hence we conclude that the sequence $\{x^k\}_{k=0}^{\infty}$ converges to some $x^* \in M$, and the desired result is obtained. □

A useful result showing the relation between a primal and a dual variational inequality for continuous, monotone operators is given next. One direction can be found in [26, Lemma 7.1.7] and the other can easily obtained from the monotonicity.

Corollary 2.9. *Let $C \subseteq \mathcal{H}$ be a nonempty and convex subset and $\mathcal{F}$ be a hemi-continuous mapping of C into $\mathcal{H}$. Let ζ be an element of C such that*

$$\langle \mathcal{F}(x), x - \zeta \rangle \geq 0, \quad \text{for all } x \in C. \tag{2.19}$$

Then,

$$\langle \mathcal{F}(\zeta), x - \zeta \rangle \geq 0, \quad \text{for all } x \in C. \tag{2.20}$$

An elementary useful result for our analysis is given next.

Corollary 2.10. *Let $\{a_k\}_{k=0}^{\infty}$, $\{b_k\}_{k=0}^{\infty}$ be two nonnegative real sequences such that*

$$a_{k+1} \leq a_k - b_k. \tag{2.21}$$

Then $\{a_k\}_{k=0}^{\infty}$ is bounded and $\sum_{k=0}^{\infty} b_k < \infty$.

3. The algorithm

In this section we present our iterative extension of the subgradient extragradient method using Popov [23], Malitsky and Semenov [22] and Semenov [24] techniques with Bregman projections. The convergence analysis uses similar arguments as in Semenov [24].

Algorithm 3.1. Choose $x^0, y^0 \in \mathcal{H}$ and $\lambda > 0$. Given the current iterates x^k and y^k and also y^{k-1}, if $\nabla f(x^k) - \lambda \mathcal{F}(y^{k-1}) \neq \nabla f(y^k)$, construct the set

$$T_k := \{w \in \mathcal{H} \mid \langle \nabla f(x^k) - \lambda \mathcal{F}(y^{k-1}) \nabla f(y^k), w - y^k \rangle \leq 0\} \tag{3.1}$$

and if $\nabla f(x^k) - \lambda \mathcal{F}(y^{k-1}) = \nabla f(y^k)$, take $T_k = \mathcal{H}$.

Now, compute the next iterates via

$$\begin{cases} x^{k+1} = \Pi_{T_k}((\nabla f)^{-1}(\nabla f(x^k) - \lambda \mathcal{F}(y^k))) \\ y^{k+1} = \Pi_C((\nabla f)^{-1}(\nabla f(x^{k+1}) - \lambda \mathcal{F}(y^k))). \end{cases} \tag{3.2}$$

3.1. **Convergence.** For the convergence of Algorithm 3.1, we assume that the following conditions hold.

Condition 3.2. The solution set of (1.1), denoted by $\mathrm{Sol}(\mathcal{F}, C)$, is nonempty.

Condition 3.3. The mapping $\mathcal{F}$ is monotone and Lipschitz-continuous with constant $L > 0$.

Condition 3.4. The function $f \colon \mathcal{H} \to \mathbb{R}$ is differential and strongly convex (that is (2.6)), and its gradient ∇f is weakly-weakly continuous (that is (2.7)).

Lemma 3.5. Assume that Conditions 3.2–3.4 hold. Let $\{x^k\}_{k=0}^{\infty}$ and $\{y^k\}_{k=0}^{\infty}$ be two sequences generated by Algorithm 3.1, $\lambda \in (0, \frac{\sqrt{2}-1}{L})$, and let $z \in \mathrm{Sol}(\mathcal{F}, C)$. Then

$$D_f(z, x^{k+1}) \leq D_f(z, x^k) - \alpha D_f(x^{k+1}, y^k) - \beta D_f(y^k, x^k) + \gamma D_f(x^k, y^{k-1}), \tag{3.3}$$

where $\alpha = 1 - \lambda L(1+\sqrt{2})$, $\beta = 1 - \sqrt{2}\lambda L$, and $\gamma = \lambda L$.

Proof. By Corollary 2.6 we have

$$\left\langle \nabla f(x^k) - \lambda \mathcal{F}(y^k) - \nabla f(x^{k+1}), z - x^{k+1} \right\rangle \leq 0, \tag{3.4}$$

or equivalently

$$\left\langle \nabla f(x^k) - \nabla f(x^{k+1}), z - x^{k+1} \right\rangle - \lambda \left\langle \mathcal{F}(y^k), z - x^{k+1} \right\rangle \leq 0. \tag{3.5}$$

Using Corollary 2.5, (3.5) can written as

$$D_f(z, x^{k+1}) \leq D_f(z, x^k) - D_f(x^{k+1}, x^k) + \lambda \left\langle \mathcal{F}(y^k), z - x^{k+1} \right\rangle. \tag{3.6}$$

Following Corollary 2.9, we can add $\langle \mathcal{F}(y^k), y^k - z \rangle \geq 0$ to (3.6) and obtain the following.

$$\begin{aligned} D_f(z, x^{k+1}) &\leq D_f(z, x^k) - D_f(x^{k+1}, x^k) + \lambda \left\langle \mathcal{F}(y^k), y^k - x^{k+1} \right\rangle \\ &\leq D_f(z, x^k) - D_f(x^{k+1}, x^k) + \lambda \left\langle \mathcal{F}(y^k) - \mathcal{F}(y^{k-1}), y^k - x^{k+1} \right\rangle \\ &+ \lambda \left\langle \mathcal{F}(y^{k-1}), y^k - x^{k+1} \right\rangle. \end{aligned} \tag{3.7}$$

Since $y^k = \Pi_C((\nabla f)^{-1}(\nabla f(x^k) - \lambda \mathcal{F}(y^{k-1})))$, by Corollary 2.6 we get

$$\begin{aligned} \lambda \left\langle \mathcal{F}(y^{k-1}), y^k - x^{k+1} \right\rangle &\leq \left\langle \nabla f(x^k) - \nabla f(y^k), y^k - x^{k+1} \right\rangle \\ &= D_f(x^{k+1}, x^k) - D_f(x^{k+1}, y^k) - D_f(y^k, x^{k+1}) \end{aligned} \tag{3.8}$$

Now wish to estimate $\langle \mathcal{F}(y^k) - \mathcal{F}(y^{k-1}), y^k - x^{k+1} \rangle$. Using the Cauchy-Schwarz inequality and the L-Lipschitz continuity of $\mathcal{F}$, we get that.

$$\begin{aligned} \lambda \left\langle \mathcal{F}(y^k) - \mathcal{F}(y^{k-1}), y^k - x^{k+1} \right\rangle &\leq \lambda L \left\| y^k - y^{k-1} \right\| \left\| x^{k+1} - y^k \right\| \\ &\leq \lambda L \Big(\frac{1}{2\sqrt{2}} \left\| y^k - y^{k-1} \right\|^2 + \frac{1}{\sqrt{2}} \left\| x^{k+1} - y^k \right\|^2 \Big) \\ &\leq \lambda L \frac{1}{2\sqrt{2}} \Big((2+\sqrt{2}) \left\| y^k - x^k \right\|^2 + \sqrt{2} \left\| x^k - y^{k-1} \right\|^2 \Big) \\ &\quad + \frac{\lambda L}{\sqrt{2}} \left\| x^{k+1} - y^k \right\|^2 \\ &= \lambda L \frac{1+\sqrt{2}}{2} \left\| y^k - x^k \right\|^2 + \frac{\lambda L}{2} \left\| x^k - y^{k-1} \right\|^2 \\ &\quad + \frac{\lambda L}{\sqrt{2}} \left\| x^{k+1} - y^k \right\|^2 \\ &\leq \lambda L(1+\sqrt{2}) D_f(y^k, x^k) + \lambda L D_f(x^k, y^{k-1}) \\ &\quad + \sqrt{2} \lambda L D_f(x^{k+1}, y^k). \end{aligned} \tag{3.9}$$

In (3.9) we used two basic inequalities: $ab \leq \frac{\varepsilon^2}{2} a^2 + \frac{1}{2\varepsilon^2} b^2$ and $(a+b)^2 \leq \sqrt{2} a^2 + (2+\sqrt{2} b^2)$ (see also [24]). Moreover, in the last inequality we used (2.14).

Now, applying (3.8) and (3.9) to (3.7) and taking into account that $\lambda L \le 1 - \sqrt{2}\lambda L$, we get that

$$\begin{aligned} D_f(z, x^{k+1}) &\le D_f(z, x^k) - \lambda L(1+\sqrt{2})D_f(y^k, x^k) \\ &\quad - (1-\sqrt{2}\lambda L)D_f(x^{k+1}, y^k) + \lambda L D_f(x^k, y^{k-1}) \\ &= D_f(z, x^k) - \alpha D_f(y^k, x^k) - \beta D_f(x^{k+1}, y^k) + \gamma D_f(x^k, y^{k-1}). \end{aligned} \tag{3.10}$$

And the proof is complete. □

Remark 3.6. It is worth mentioning that in Popov's method [23], the step-size is chosen such that $\lambda < 1/3L$. Here we use estimations for λ which appeared first in Malitsky [21] and is an improvement to the interval $(0, \frac{\sqrt{2}-1}{L})$.

We are now ready to prove the weak convergence theorem of Algorithm 3.1.

Theorem 3.7. *Assume that Conditions 3.2–3.4 hold, and let* $\lambda \in (0, \frac{\sqrt{2}-1}{L})$. *Then any two sequences* $\{x^k\}_{k=0}^{\infty}$ *and* $\{y^k\}_{k=0}^{\infty}$ *generated by Algorithm 3.1 converge weakly to a solution of the variational inequality (1.1).*

Proof. We start by showing that the sequence $\{x^k\}_{k=0}^{\infty}$ is bounded. Fix any $z \in \mathrm{Sol}(\mathcal{F}, C)$ and for $k \ge 2$ let

$$a_k = D_f(z, x^k) + \gamma D_f(x^k, y^{k-1}) \tag{3.11}$$

$$b_k = \alpha D_f(y^k, x^k) + (\beta - \gamma) D_f(x^{k+1}, y^k) \tag{3.12}$$

where α, β, γ are defined as in Lemma 3.5. Hence, inequality (3.3) can be rewritten as $a_{k+1} \le a_k - b_k$. By Corollary 2.10, we conclude that $\{a_k\}_{k=0}^{\infty}$ is bounded and $\lim_{k\to\infty} D_f(y^k, x^k) = 0$.

Due to (2.14), we get that the sequence $\{x^k\}_{k=0}^{\infty}$ is bounded and $\|x^k - y^k\| \to 0$, $\|x^{k+1} - y^k\| \to 0$ as $k \to \infty$. Consequently, we also have $\|x^{k+1} - x^k\| \to 0$.

Since $\{x^k\}_{k=0}^{\infty}$ is bounded, there exist a subsequence $\{x^{k_i}\}_{i=0}^{\infty}$ of $\{x^k\}_{k=0}^{\infty}$ such that $\{x^{k_i}\}_{i=0}^{\infty}$ converges weakly to some $x^* \in \mathcal{H}$. It is clear that $\{y^{k_i}\}_{i=0}^{\infty}$ also converges to $x^* \in \mathcal{H}$. It is now left to show that $x^* \in \mathrm{Sol}(\mathcal{F}, C)$.

From Corollary 2.6 it follows that

$$\left\langle \nabla f(x^{k_i+1}) - \nabla f(x^{k_i}) + \lambda \mathcal{F}(y^{k_i}), y - x^{k_i+1} \right\rangle \ge 0 \quad \text{for all } y \in C. \tag{3.13}$$

From this we conclude that for all $y \in C$

$$\begin{aligned} 0 &\le \left\langle \nabla f(x^{k_i+1}) - \nabla f(x^{k_i}), y - x^{k_i+1} \right\rangle + \lambda \left\langle \mathcal{F}(y^{k_i}), y - y^{k_i} \right\rangle \\ &\qquad + \lambda \left\langle \mathcal{F}(y^{k_i}), y^{k_i} - x^{k_i+1} \right\rangle \\ &\le (\nabla f(x^{k_i+1}) - \nabla f(x^{k_i}), y - x^{k_i+1}) + \lambda \left\langle \mathcal{F}(y), y - y^{k_i} \right\rangle \\ &\qquad + \lambda \left\langle \mathcal{F}(y^{k_i}), y^{k_i} - x^{k_i+1} \right\rangle. \end{aligned} \tag{3.14}$$

Taking the limit as $i \to \infty$ in (3.14), using the weakly-weakly continuity of ∇f and

$$\lim_{i\to\infty} \left\| x^{k_i+1} - x^{k_i} \right\| = \lim_{i\to\infty} \left\| y^{k_i+1} - y^{k_i} \right\| = 0, \tag{3.15}$$

we obtain

$$(\mathcal{F}(y), y - x^*) \geq 0 \quad \text{for all } y \in C. \tag{3.16}$$

Following Lemma 2.9, we get that $x^* \in$Sol$(\mathcal{F}, C)$.

Finally, we prove that the sequence $\{x^k\}_{k=0}^{\infty}$ converges weakly to $x^* \in$Sol$(\mathcal{F}, C)$. Since the sequence $\{a_k\}_{k=0}^{\infty}$ is monotone and bounded, we conclude that it converges. Since the sequence $D_f(x^k, y^{k-1}) \to 0$, we get that the sequence $\{D_f(z, x^k)\}_{k=0}^{\infty}$ also converges. By Corollary 2.8, we deduce that $\{x^k\}_{k=0}^{\infty}$ converges weakly to some point $x^* \in$Sol$(\mathcal{F}, C)$, and the proof is complete. □

4. Numerical experiments

In this section we present two numerical experiments which demonstrate our algorithm (Algorithm 3.1) performances.

Consider the Hilbert space $\mathcal{H} = L^2([0,1])$ with norm $\|x\| := \left(\int_0^1 |x(t)|^2 dt \right)^{\frac{1}{2}}$ and inner product $\langle x, y \rangle := \int_0^1 x(t)y(t)dt,\ x, y \in \mathcal{H}$. Let C be the unit ball in $\mathcal{H}$, that is $C := \{x \in \mathcal{H} \mid \|x\| \leq 1\}$. We define the 2-Lipschitz continuous and monotone mapping $\mathcal{F} : C \to \mathcal{H}$ as $(\mathcal{F}x)(t) = \max(0, x(t))$, see [16]. It can be easily verified that the VI with the above $\mathcal{F}$ and C has a unique solution which is $0 \in L^2([0,1])$.

For the algorithm implementation, we used the Euclidean distances, and hence recall the orthogonal projections onto C and the half-space $H := \{x \in \mathcal{H} \mid \langle a, x \rangle \leq b\}$ with $0 \neq a \in \mathcal{H}$ and $b \in \mathbb{R}$, see e.g., [5].

$$P_C(x) = \begin{cases} \frac{x}{\|x\|}, & \text{if } \|x\| > 1, \\ x, & \text{if } \|x\| \leq 1, \end{cases} \tag{4.1}$$

and

$$P_H(x) = \begin{cases} x + \frac{b - \langle a, x \rangle}{\|a\|^2} a, & \text{if } \langle a, x \rangle > b; \\ x, & \text{if } \langle a, x \rangle \leq b, \end{cases} \tag{4.2}$$

The parameters used in our experiments are: $\lambda = 0.1$ and the stopping criterion $||x^k - y^k|| < 10^{-5}$. We present numerical illustrations for Algorithm 3.1 for two different starting points $x^1(t)$. The results are presented in Table 1 and in Figures 1-2.

Case I: $x^1(t) = \frac{1}{600}\left[sin(-3t) + cos(-10t)\right]$.

Case II: $x^1(t) = \frac{1}{525}\left[t^2 - e^{-t}\right]$.

	No. of Iterations	CPU Time
Case I	51	1.1793×10^{-3}
Case II	51	1.2381×10^{-3}

TABLE 1. Algorithm 3.1 with different starting points $x^1(t)$

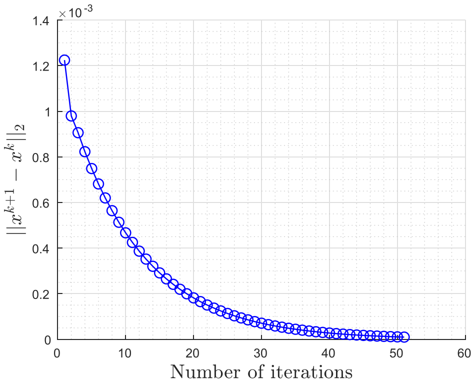

FIGURE 1. Case I

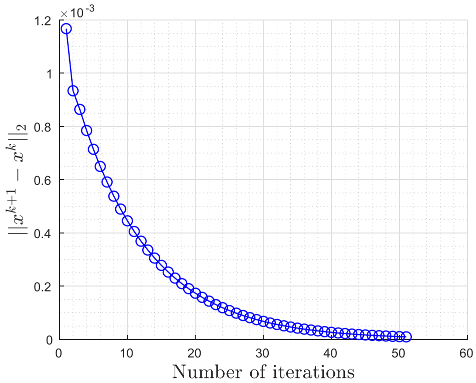

FIGURE 2. Case II

5. Conclusions

In this work we present an extension of the subgradient extragradient method for solving variational inequalities with monotone and Lipschitz continuous mappings in real Hilbert spaces using Bregman projections. The motivation of this generalization is the works of Popov [23], Malitsky and Semenov [22] and Semenov [24] and its main advantage these and other existing results is the need to evaluate the VI associated mapping $\mathcal{F}$ only once per each iteration. The usage of the Bregman distance allows flexibility in choosing the projection type (orthogonal projection, subgradient projection and more) to be computed in the new method.

Our result open new directions for future investigations, for example extensions to Banach spaces, line-search approaches as well as replacing the first projection onto C by an easily computable projection.

6. Acknowledgements

This work is based on discussions between Dr. Aviv Gibali and Dr. Yura Malitsky started already in 2014. Hence wish to express our deep gratitude to **Dr. Yura Malitsky** for his time and efforts which helped improving this manuscript and make it suitable for publication.

References

[1] A. S. Antipin, *On a method for convex programs using a symmetrical modification of the Lagrange function*, Ekon. Mat. Metody **12** (1976), 1164–1173.

[2] K. J. Arrow, L. Hurwicz and H. Uzawa, *Studies in Linear and Non-Linear Programming*, Stanford University Press, Stanford 1958.

[3] L. M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Comput. Math. and Math. Phys. **7** (1967), 200–217.
[4] D. Butnariu and E. Resmerita, *Bregman distances, totally convex functions and a method for solving operator equations in Banach spaces*, Abstr. Appl. Anal. **2006** (2006), 1–39.
[5] A. Cegielski, *Iterative Methods for Fixed Point Problems in Hilbert Spaces*, Lecture Notes in Mathematics, vol. 2057, Springer, Berlin 2012.
[6] Y. Censor, A. Gibali and S. Reich, *The subgradient extragradient method for solving variational inequalities in Hilbert space*, J. Optim. Theory Appl. **148** (2011), 318–335.
[7] Y. Censor, A. Gibali and S. Reich, *Strong convergence of subgradient extragradient methods for the variational inequality problem in Hilbert space*, Optim. method Soft. **6** (2011), 827–845.
[8] Y. Censor, A. Gibali and S. Reich, *Extensions of Korpelevich's extragradient method for solving the variational inequality problem in Euclidean space*, Optimization **61** (2012), 1119–1132.
[9] Y. Censor and A. Lent, *An iterative row-action method for interval convex programming*, J. Optim. Theory Appl. **34** (1981), 321–353.
[10] S. Dafermos, *Traffic equilibrium and variational inequalities*, Transp. Sci. **14** (1980), 42–54.
[11] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume I and Volume II,* Springer-Verlag, New York, NY, USA, 2003.
[12] G. Fichera,*Sul problema elastostatico di Signorini con ambigue condizioni al contorno*, Atti Accad. Naz. Lincei, VIII. Ser., Rend., Cl. Sci. Fis. Mat. Nat. **34** (1963), 138–142.
[13] G. Fichera, *Problemi elastostatici con vincoli unilaterali: il problema di Signorini con ambigue condizioni al contorno*, Atti Accad. Naz. Lincei, Mem., Cl. Sci. Fis. Mat. Nat., Sez. I, VIII. Ser. **7** (1964), 91–140.
[14] K. Goebel and S. Reich, *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings*, Marcel Dekker, New York and Basel, 1984.
[15] A. A. Goldstein, *Convex programming in Hilbert space*, Bull. Am. Math. Soc. **70** (1964), 709–710.
[16] D. V. Hieu, P. K. Anh and D. Muu, *Modified hybrid projection methods for finding common solutions to variational inequality problems*, Comput. Optim. Appl. **66** (2017), 75–96.
[17] E. N. Khobotov, *Modification of the extragradient method for solving variational inequalities and certain optimization problems.* USSR Comput. Math. Math. Phys. **27** (1987) 120–127.
[18] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications*, Academic Press, New York-London, 1980.
[19] G. M. Korpelevich, *The extragradient method for finding saddle points and other problems.* Ekon. Mate. Metody **12** (1976) 747–756.
[20] E. S. Levitin and B. T. Polyak, *Constrained minimization problems*, USSR Comp. Math. Math. Phy. **6** (1966), 1–50.
[21] Yu. V. Malitsky, *Projected reflected gradient methods for monotone variational inequalities*, SIAM J. Optim. **25** (2015), 502–520.
[22] Yu.V. Malitsky and V. V. Semenov, *An extragradient algorithm for monotone variational inequalities*, Cybernet. Systems Anal. **50** (2014), 271–277.
[23] L. C. Popov, *A modification of the Arrow-Hurwicz method for finding saddle points*, Math. Notes **28** (1980), 845–848.
[24] V. V. Semenov, *A version of the mirror descent method to solve variational inequalities*, Cybern. Syst. Anal. **53** (2017), 234–243.
[25] G. Stampacchia, *Formes bilineaires coercitives sur les ensembles convexes*, C. R. Acad. Sci., Paris **258** (1964), 4413–4416.
[26] W. Takahashi, *Nonlinear Functional Analysis*, Yokohama Publishers, Yokohama, 2000.

Manuscript received February 8 2018
revised February 15 2018

Aviv Gibali
Department of Mathematics, ORT Braude College, Karmiel 2161002, Israel

The Center for Mathematics and Scientific Computation, University of Haifa, Mt. Carmel, Haifa 3498838, Israel
E-mail address: avivg@braude.ac.il