Two Bregman Projection Methods for Solving Variational Inequality Problems in Hilbert Spaces with Applications to Signal Processing

Lateef Olakunle Jolaoso 1,*, Maggie Aphane 1 and Safeer Hussain Khan 2

1 Department of Mathematics and Applied Mathematics, Sefako Makgatho Health Sciences University, P.O. Box 94, Medunsa 0204, South Africa; maggie.aphane@smu.ac.za
2 Department of Mathematics, Statistics and Physics, Qatar University, Doha 2713, Qatar; safeer@qu.edu.qa
* Correspondence: lateef.jolaoso@smu.ac.za

Received: 14 November 2020; Accepted: 25 November 2020; Published: 5 December 2020

Abstract: Studying Bregman distance iterative methods for solving optimization problems has become an important and very interesting topic because of the numerous applications of the Bregman distance techniques. In this paper, two different extragradient methods were proposed for studying pseudomonotone variational inequality problems using Bregman distance in real Hilbert spaces. The first algorithm uses a fixed stepsize which depends on a prior estimate of the Lipschitz constant of the cost operator. The second algorithm uses a self-adaptive stepsize which does not require prior estimate of the Lipschitz constant of the cost operator. Some convergence results were proved for approximating the solutions of pseudomonotone variational inequality problem under standard assumptions. Moreover, some numerical experiments were also given to illustrate the performance of the proposed algorithms using different convex functions such as the Shannon entropy and the Burg entropy. In addition, an application of the result to a signal processing problem is also presented.

Keywords: bregman divergence; popov extragradient; subgradient; variational inequalities; weak convergence; strong convergence

1. Introduction

Let $H$ be a real Hilbert space with norm $\| \cdot \|$ and inner product $\langle \cdot, \cdot \rangle$. We study the Variational Inequality Problem (shortly, VIP) of the form:

\[
\text{Find } v \in C \text{ such that } \langle A v, u - v \rangle \geq 0, \quad \forall u \in C, \tag{1}
\]

where $A : H \to H$ is a given operator and $C$ is a nonempty closed convex subset of $H$. We shall denote the set of solutions of the VIP (1) by $\Omega_{VI}$. The VIP has drawn much attention from researchers because of its importance as a core model for studying many mathematical problems which include convex programming, equilibrium problem, inclusion problem, split feasibility problem, complementarity problem, minimization problem, etc, see e.g., [1,2]. Recently, VIP has also been used for studying optimization problems arising in machine learning, signal processing and linear inverse problem; see, for instance [3–5].

Due to its importance, many researchers have developed several iterative methods for solving the VIP (1) in Hilbert and Banach spaces, see for example [6–22] and references therein. One of the simplest methods known for solving the VIP is the Extragradient Method (EM) introduced by Korpelevich [23]...
for a case where \( A \) is monotone and Lipschitz operator in a finite-dimensional space as follows: Given \( a_0 \in C \subseteq \mathbb{R}^n \), compute

\[
\begin{align*}
    b_n &= P_C(a_n - \gamma Aa_n), \\
    a_{n+1} &= P_C(a_n - \gamma Ab_n), \quad n \geq 1,
\end{align*}
\]

(2)

where \( \gamma \in (0, 1/L) \), \( A : H \rightarrow H \) is monotone and \( L \)-Lipschitz continuous which has been extended to infinite-dimensional spaces by many authors, e.g., \([8,9,13,19,24,25]\). However, the EM requires computing the projection onto \( C \) twice per each iteration which is not cost effective in practice. As a result of this drawback, there has been an increasing effort in finding appropriate method for solving the VIP. Censor et al. \([13]\) introduced a Subgradient Extragradient Method (SEM) by replacing the second projection in SEM with projection onto a half-space as follows:

Algorithm SEM: Given \( a_0 \in C, \gamma \in (0, \frac{1}{L}) \), compute

\[
\begin{align*}
    a_0 &\in C, \\
    b_n &= P_C(a_n - \gamma Aa_n), \\
    a_{n+1} &= P_{T_n}(a_n - \gamma Ab_n), \\
    \text{where } &\quad T_n := \{ x \in H : \langle a_n - \gamma Aa_n - b_n, x - b_n \rangle \geq 0 \} \quad n \geq 0.
\end{align*}
\]

(3)

Note that the set \( T_n \) is a half-space whose projection can be explicitly calculated (see for instance \([26]\)). Under some mild assumptions, Censor et al. \([13]\) obtained a weak convergence result for solving VIP using (3), and by modifying the SEM with Halpern iterative scheme (see \([27]\)), they proved the strong convergences of the SEM under certain mild conditions (see for instance \([14,15]\)).

It is worthy to mention that both the EM and SEM require two evaluations of \( A \) per each iteration. Popov \([28]\) later proposed a nice modification of the EM that required the evaluation of \( A \) at only one point in each iteration. Malitsky and Semenov \([29]\) modified the Popov’s extragradient method by replacing the second projection onto \( C \) with a projection onto a half-space. Further extension of Popov’s extragradient method was introduced by Nomirovskii et al. \([30]\) by using Bregman distance function \([31]\) in a finite-dimensional setting. Recently, Gibali \([32]\) extended the methods of Malitsky and Semenov \([29]\) and Nemirovskii et al. \([30]\) to infinite-dimensional Hilbert space using a Bregman distance function. A weak convergence result is proved for the sequence generated by these methods provided the stepsize satisfies the condition \( \lambda \in \left(0, \frac{\sqrt{2} - 1}{L}\right) \), where \( A \) is monotone and \( L \)-Lipschitz continuous.

An obvious disadvantage in the above-mentioned results is that the stepsize of the algorithms requires a prior estimate of the Lipschitz constant \( L \). It is known that the Lipschitz constant are too difficult to estimate in general. Even when it is possible, the estimate is often too small which affects the convergence of the algorithm. An effort in solving this problem for the case of monotone VIP was recently introduced by Hieu and Cholamjiak \([33]\). They introduced an extragradient method with Bregman distance which does not require a prior estimate of the Lipschitz constant. However, their algorithm can only be applicable to monotone VIP but not pseudomonotone. When the cost operator is not monotone, the result of \([33]\) can not be applied to such VIP, for instance, see Example 3.12 and 4.2 of \([19]\).

Motivated by these results, in the present paper, we first introduce two Bregman extragradient methods of Popov’s type for solving pseudomonotone variational inequalities in real Hilbert spaces. In the first algorithm, the stepsize requires a prior estimate of the Lipschitz constant \( L \). This extends the results of \([30,32]\) from monotone VIP to pseudomonotone VIP in real Hilbert spaces. In the second algorithm, the stepsize is selected self-adaptively and does not require a prior estimate of the Lipschitz constant. This improves Algorithm 1 of the paper and also the results of \([30,32]\). It also improves the method of \([33,34]\) from the extragradient type method with Bregman distance to the subgradient extragradient method with Bregman distance. More so, it extends the result of \([33,34]\) from monotone VIP to pseudomonotone VIP. We further provide some numerical examples using
different types of convex functions and a numerical application of our result to the Least Absolute Shrinkage and Selection Operator (LASSO) problem in compressing sensing. We denote the strong and weak convergence of a sequence \( \{a_n\} \subseteq H \) to a point \( x \in H \) by \( a_n \to x \) and \( a_n \rightharpoonup x \) respectively.

The paper is organized as follows: We first recall some basic definitions and results in Section 2. We present our algorithm and discuss its convergence in Section 3. We give some numerical experiments in Section 4. We conclude with a final remark in Section 5.

2. Preliminaries

In this section, we recall some basic definitions and concepts that are needed for establishing our results.

**Definition 1.** [20] Let \( A : C \rightarrow H \) be an operator. Then \( A \) is called

(i) monotone if \( \langle Au - Az, u - z \rangle \geq 0 \) for all \( u, z \in C \);

(ii) \( \alpha \)-strongly pseudomonotone if

\[
\langle Au, z - u \rangle \geq 0 \Rightarrow \langle Az, z - u \rangle \geq \alpha \|u - z\|^2 \quad \forall u, z \in C, \alpha > 0;
\]

(iii) pseudomonotone if for all \( u, z \in C \),

\[
\langle Au, z - u \rangle \geq 0 \Rightarrow \langle Az, z - u \rangle \geq 0;
\]

(iv) Lipschitz continuous if there exists a constant \( L > 0 \) such that

\[
\|Au - Az\| \leq L\|u - z\| \quad \forall u, z \in C;
\]

(v) weakly sequentially continuous if for each sequence \( \{a_n\} \), we have:

\[
a_n \rightharpoonup p \Rightarrow Ap_n \rightharpoonup Ap.
\]

**Definition 2.** A function \( f : H \rightarrow \mathbb{R} \cup \{+\infty\} \) is said to be proper if its effective domain, i.e., \( \text{dom } f = \{v \in H : f(v) < +\infty\} \) is nonempty. The Fenchel conjugate function of \( f \) is the convex function \( f^* : H \rightarrow \mathbb{R} \) defined by \( f^*(b) = \sup\{\langle b, x \rangle - f(x) : x \in H\} \). \( f \) is said to be subdifferentiable at a point \( v \in E \) if the set \( \partial f(v) = \{e \in H : f(y) - f(v) \geq \langle e, y - v \rangle, \forall y \in H\} \) is nonempty. An element \( e \in \partial f(x) \) is called subgradient of \( f \). If \( f \) is continuously differentiable, then \( \partial f(x) = \{\nabla f(x)\} \), where \( \nabla f(x) \) denotes the gradient of \( f \) at \( x \); \( f \) is called Legendre if

(i) \( \text{int(dom } f) \neq \emptyset \) and \( \partial f \) is single-valued on its domain

(ii) \( \text{int(dom } f^*) \neq \emptyset \) and \( \partial f^* \) is single-valued on its domain.

The function \( f \) is said to be strongly convex with strong convexity constant \( \sigma > 0 \), if

\[
f(u) \geq f(v) - \langle \nabla f(v), u - v \rangle + \frac{\sigma}{2}\|u - v\|^2, \quad \forall u \in \text{dom } f \text{ and } v \in \text{int}(\text{dom } f).
\]

A function \( f : H \rightarrow \mathbb{R} \) is Gâteaux differentiable at \( u \in \text{int}(\text{dom } f) \) if the limit

\[
f^\prime(u, v) = \lim_{\lambda \to 0^+} \frac{f(u + \lambda v) - f(u)}{\lambda} \quad \forall v \in H,
\]

exists for any \( v \in H \). The gradient of \( f \) at \( u \) is a linear function \( \nabla f(u) \) defined by \( \langle \nabla f(u), y \rangle = f^\prime(u, v) \forall v \in H \). When the limit in (5) holds for all \( u \in \text{int}(\text{dom } f) \), we say that \( f \) is Gâteaux differentiable. More so, when (5) is attained uniformly for any \( v \in H \) with \( \|v\| = 1 \), we say that \( f \) is Fréchet differentiable. It is well-known that \( f \) is Fréchet differentiable if and only if the gradient \( \nabla f \) is norm-to-norm continuous at \( u \) (see [35]).
Given a Gâteaux differentiable function $f : H \to \mathbb{R} \cup \{+\infty\}$. The Bregman distance corresponding to the function $f$ is defined by (see [31])

$$\Delta_f(v, u) = f(v) - f(u) - \langle \nabla f(u), v - u \rangle \quad \forall v \in \text{dom } f \text{ and } u \in \text{int}(\text{dom } f).$$  \hfill (6)

**Remark 1.** The Bregman distance was introduced by Bregman [31] as a nice and effective method for designing and analyzing feasibility and optimization algorithms. It is easy to see that for any $v \in \text{dom } f$, $u, w \in \text{int}(\text{dom } f)$, we have

$$\Delta_f(v, w) = \Delta_f(v, u) + \Delta_f(u, w) + \langle \nabla f(u) - \nabla f(w), v - u \rangle. \hfill (7)$$

Moreover, when $f$ is strongly convex with strong convexity modulus $\sigma > 0$, then the Bregman distance satisfies the following identity:

$$\Delta_f(v, u) \geq \frac{\sigma}{2} \|v - u\|^2. \hfill (8)$$

It is important we emphasize that various types of function $f$ gives different Bregman distance. We give the following important examples of some practical important types of function $f$ and their corresponding Bregman distance (see, e.g., [33,36]):

**Example 1.** (i) If $f(v) = \frac{1}{2} \|v\|^2$, then

$$\Delta_f(v, u) = \frac{1}{2} \|v - u\|^2,$$

which is the squared Euclidean distance (SE).

(ii) If $f(v) = \sum_{i=1}^n a_i \log(v_i)$ called Shannon entropy, then its corresponding Bregman distance is given as

$$\Delta_f(v, u) = \sum_{i=1}^n \left( v_i \log \left( \frac{v_i}{u_i} \right) + u_i - v_i \right).$$

This distance is called Kullback–Leibler distance (KL) and as become a very important tool in several areas of applied mathematics such as machine learning.

(iii) If $f(v) = -\sum_{i=1}^n \log(v_i)$ called Burg entropy, then its corresponding Bregman distance is given as

$$\Delta_f(v, u) = \sum_{i=1}^n \left( \log \left( \frac{v_i}{u_i} \right) + \frac{v_i}{u_i} - 1 \right).$$

This is called Itakura–Saito distance (IS), which is very important in information theory.

(iv) If $f(v) = \frac{1}{2} v^T D v$ where $v^T$ is stands for the transpose of $v \in \mathbb{R}^n$, and $D = \text{diag}(1, 2, \ldots, n) \in \mathbb{R}^n$ is a positive semidefinite matrix, then

$$\Delta_f(v, u) = \frac{1}{2} (v - u)^T D (v - u),$$

which is called the Mahalanobis distance (MD) used for cluster analysis.

**Definition 3.** [37] The Bregman projection with respect to $f$ of $v \in \text{int}(\text{dom } f)$ onto a nonempty closed convex set $C \subset \text{int}(\text{dom } f)$ is the unique vector $\Pi_C(v) \in C$ satisfying

$$\Pi_C(v) := \inf \{ \Delta_f(v, u) : u \in C \}.$$
Similar with the metric projection, the Bregman projection is characterized by the following identities (see [37]):

\[ z = \Pi_C(v) \iff \langle \nabla f(v) - \nabla f(z), u - z \rangle \leq 0, \quad \forall u \in C, \quad (9) \]

and

\[ \Delta_f(u, \Pi_C(v)) + \Delta_f(\Pi_C(v), v) \leq \Delta_f(u, v) \quad \forall u \in C, v \in \text{int(dom } f). \quad (10) \]

**Lemma 1.** [38] Let \( E \) be a Banach space and let \( f : E \to (-\infty, \infty] \) be a proper strictly convex function so that it is Gâteaux differentiable and \( \{a_n\} \) is a sequence in \( E \) such that \( a_n \rightharpoonup u \) for some \( u \in E \). Then

\[ \limsup_{n \to \infty} \Delta_f(u, a_n) < \limsup_{n \to \infty} \Delta_f(v, a_n), \]

for all \( v \) in the interior of \( \text{dom } f \) with \( u \neq v \).

The following result is well-known; see, e.g., [30,32].

**Lemma 2.** Let \( \{a_n\} \) and \( \{b_n\} \) be two non-negative real sequences such that

\[ a_{n+1} \leq a_n - b_n. \]

Then \( \{a_n\} \) is bounded and \( \sum_{n=0}^{\infty} b_n < \infty \).

**Definition 4.** [39,40] The Minty Variational Inequalities (MVI) are defined as finding a point \( \bar{x} \in C \) such that

\[ \langle Ay, y - \bar{x} \rangle \geq 0, \quad \forall y \in C. \quad (11) \]

We denote by \( M(C, A) \), the set of solution of (11). Some existence results for the MVIP have been presented in [39]. The assumption that \( M(C, A) \neq \emptyset \) has already been used for solving \( VI(C, A) \) in finite dimensional spaces (see e.g. [41]).

**Lemma 3.** [40] Consider the VI (1). If the mapping \( h : [0, 1] \to E^* \) defined as \( h(t) = A(tx + (1-t)y) \) is continuous for all \( x, y \in C \) (i.e., \( h \) is hemicontinuous), then \( M(C, A) \subset \text{VI}(C, A) \). Moreover, if \( A \) is pseudomonotone, then \( \text{VI}(C, A) \) is closed, convex and \( \text{VI}(C, A) = M(C, A) \).

3. Main Results

In this section, we present our algorithms and their convergence analysis. In the sequel, we assume that the following assumptions hold.

**Assumption 1.** (A1) The feasible set \( C \subseteq H \) is a nonempty, closed and convex set.
(A2) The mapping \( A : C \to H \) is pseudomonotone, L-Lipschitz continuous and weakly sequentially continuous on \( H \).
(A3) The solution set \( \text{Sol} := \Omega_{VI} \) is nonempty.
(A4) The function \( f : H \to \mathbb{R} \cup \{+\infty\} \) uniformly Fréchet differentiable, Legendre, strongly coercive and strongly convex with a strong convexity modulus \( \sigma > 0 \).

3.1. Bregman Projection Method with Fixed Stepsize

We present our first algorithm as follows:

**Algorithm BPMFS**
Initialization: Pick \(a_0, b_0 \in H\) arbitrarily and let \(\lambda > 0\). Set \(n = 0\).

**Step 1:** Compute

\[
\begin{align*}
a_1 &= \Pi_C(\nabla f^*(\nabla f(a_0) - \lambda Ab_0)), \\
b_1 &= \Pi_C(\nabla f^*(\nabla f(a_1) - \lambda Ab_0)).
\end{align*}
\]

**Step 2:** Given the current iterates \(a_n, b_n\) and \(b_{n-1}\), calculate \(a_{n+1}\) and \(b_{n+1}\) as follows:

\[
\begin{align*}
a_{n+1} &= \Pi_{T_n}(\nabla f^*(\nabla f(a_n) - \lambda Ab_n)), \\
b_{n+1} &= \Pi_C(\nabla f^*(\nabla f(a_{n+1}) - \lambda Ab_n)),
\end{align*}
\]

where

\[
T_n := \{z \in H : \langle \nabla f(a_n) - \lambda Ab_{n-1} - \nabla f(b_n), z - b_n \rangle \leq 0\}.
\]

If \(a_{n+1} = a_n\) and \(b_{n+1} = b_{n-1}\) or \(Ab_n = 0\) : STOP. Otherwise, set \(n \leftarrow n + 1\) and repeat Step 2.

**Remark 2.** The projection for \(a_{n+1}\) in Step 2 can be calculated explicitly using the projection formula in [26]. Algorithm BPMFS is constructed based on the concepts of [28,29] using Bregman projections. As in [29], the main task of Algorithm BPMFS is to compute one projection for \(b_{n+1}\) onto \(C\) and one evaluation of \(A\) at the current approximation \(b_n\). More so, it is easy to see from (12) that \(C \subset T_n\). Indeed from (12), we have that

\[
b_n = \Pi_C(\nabla f^*(\nabla f(a_n) - \lambda Ab_{n-1})),
\]

hence, using (9), we obtain

\[
\langle \nabla f(a_n) - \lambda Ab_{n-1} - \nabla f(b_n), \omega - b_n \rangle \leq 0 \quad \forall \omega \in C.
\]

It follows from (14) that \(C \subset T_n\).

**Remark 3.** Moreover, if for some \(n \geq 0\), \(a_{n+1} = a_n\) and \(b_n = b_{n-1}\) in Algorithm BPMFS. Then \(b_n\) is a solution of the VI (1).

**Proof.** From \(a_{n+1} = a_n\), this implies that

\[
\langle \nabla f(a_{n+1}) - \nabla f(a_n) + \lambda Ab_n, z - a_{n+1} \rangle \geq 0 \quad \forall z \in T_n.
\]

Hence

\[
\lambda \langle Ab_n, z - a_n \rangle \geq 0 \quad \forall z \in T_n.
\]

Equivalently

\[
\langle Ab_n, z - b_n \rangle \geq \langle Ab_n, a_n - b_n \rangle \quad \forall z \in T_n.
\]

Since \(C \subset T_n\), we get

\[
\langle Ab_n, z - b_n \rangle \geq \langle Ab_n, a_n - b_n \rangle \quad \forall z \in C.
\]

On the other hand, it follows from the definition of \(T_n\) that

\[
\langle \nabla f(a_n) - \lambda Ab_{n-1} - \nabla f(b_n), z - b_n \rangle \leq 0 \quad \forall z \in T_n.
\]

Substituting \(b_{n-1} = b_n\) and \(z = a_n\) into the above inequality, we get

\[
\langle \nabla f(a_n) - \lambda Ab_n - \nabla f(b_n), a_n - b_n \rangle \leq 0.
\]
Then we have

$$\lambda \langle Ab_n, a_n - b_n \rangle \geq 0.$$  \hspace{1cm} (16)

Combining (15) and (16), we get

$$\langle Ab_n, z - b_n \rangle \geq 0 \quad \forall z \in C.$$

Therefore $b_n$ is a solution of the VI (1). \qed

**Lemma 4.** Let $\{a_n\}$ and $\{b_n\}$ be the sequences generated from Algorithm BPMFS. Then the following inequality holds:

$$\Delta_f(z, a_{n+1}) \leq \Delta_f(z, a_n) - \left(1 - \frac{\lambda L \sqrt{2}}{\sigma}\right) \Delta_f(a_{n+1}, b_n) - \left(1 - \frac{\lambda L}{\sigma} (1 + \sqrt{2})\right) \Delta_f(b_n, a_n) + \frac{\lambda L}{\sigma} \Delta_f(a_n, b_{n-1}),$$ \hspace{1cm} (17)

for any $z \in \text{Sol}$.

**Proof.** Since $z \in \text{Sol} \subset C \subset T_n$ and $a_{n+1} = \Pi_{T_n}(\nabla f^*(\nabla f(a_n) - \lambda Ab_n))$, we have from (9)

$$\langle \nabla f(a_{n+1}) - (\nabla f(a_n) - \lambda Ab_n), z - a_{n+1} \rangle \geq 0.$$

This implies that

$$\langle \nabla f(a_{n+1}) - \nabla f(a_n) + \lambda Ab_n, z - a_{n+1} \rangle \geq 0,$$

thus we get

$$\lambda \langle Ab_n, z - a_{n+1} \rangle \geq \langle \nabla f(a_{n+1}) - \nabla f(a_n), a_{n+1} - z \rangle.$$ \hspace{1cm} (18)

By the three point identity (7), we have

$$\langle \nabla f(a_{n+1}) - \nabla f(a_n), a_{n+1} - z \rangle = \Delta_f(z, a_n) - \Delta_f(z, a_{n+1}) - \Delta_f(a_{n+1}, a_n).$$

This implies that

$$\lambda \langle Ab_n, z - a_{n+1} \rangle \geq \Delta_f(z, a_n) - \Delta_f(z, a_{n+1}) - \Delta_f(a_{n+1}, a_n).$$

Thus

$$\Delta_f(z, a_{n+1}) \leq \Delta_f(z, a_n) - \Delta_f(a_{n+1}, a_n) + \lambda \langle Ab_n, z - a_{n+1} \rangle.$$ \hspace{1cm} (19)

Since $A$ is pseudomonotone and $b_n \in C$, it follows that $\langle Ab_n, z - b_n \rangle \leq 0$. Then this implies that

$$\langle Ab_n, z - a_{n+1} \rangle = \langle Ab_n, b_n - a_{n+1} \rangle + \langle Ab_n, z - b_n \rangle \leq \langle Ab_n, b_n - a_{n+1} \rangle.$$ \hspace{1cm} (20)

Combining (19) and (20), we get

$$\Delta_f(z, a_{n+1}) \leq \Delta_f(z, a_{n+1}) - \Delta_f(a_{n+1}, a_n) + \lambda \langle Ab_n, b_n - a_{n+1} \rangle + \lambda \langle Ab_n, z - b_n \rangle \leq \Delta_f(z, a_n) - \Delta_f(a_{n+1}, a_n) + \lambda \langle Ab_n, b_n - a_{n+1} \rangle = \Delta_f(z, a_n) - \Delta_f(a_{n+1}, a_n) + \lambda \langle Ab_n, b_n - a_{n+1} \rangle + \lambda \langle Ab_n - Ab_{n-1}, b_n - a_{n+1} \rangle.$$ \hspace{1cm} (21)
However, we have
\[
\lambda \langle Ab_{n-1}, b_n - a_{n+1} \rangle = \langle \nabla f(a_n) - \lambda Ab_{n-1} - \nabla f(b_n), a_{n+1} - b_n \rangle \\
+ \langle \nabla f(b_n) - \nabla f(a_n), a_{n+1} - b_n \rangle.
\]

Since \( a_{n+1} \in T_n \) and by the definition of \( T_n \), we get
\[
\lambda \langle Ab_{n-1}, b_n - a_{n+1} \rangle \leq \langle \nabla f(b_n) - \nabla f(a_n), a_{n+1} - b_n \rangle.
\]

Using the three points identity (7) in (22), we obtain
\[
\lambda \langle Ab_{n-1}, b_n - a_{n+1} \rangle \leq \Delta_f(a_{n+1}, a_n) - \Delta_f(a_{n+1}, b_n) - \Delta_f(b_n, a_n).
\]

Combining (21) and (23), we get
\[
\Delta_f(z, a_{n+1}) \leq \Delta_f(z, a_n) - \Delta_f(a_{n+1}, b_n) - \Delta_f(b_n, a_n) + \lambda \langle Ab_n - Ab_{n-1}, b_n - a_{n+1} \rangle.
\]

Note that
\[
\lambda \langle Ab_{n-1} - Ab_n, a_{n+1} - b_n \rangle \leq \lambda \|Ab_{n-1} - Ab_n\| \|a_{n+1} - b_n\|
\]
\[
\leq \lambda L \|b_{n-1} - b_n\| \|a_{n+1} - b_n\|
\]
\[
\leq \lambda L \left\{ \frac{1}{2\sqrt{2}} \|b_{n-1} - b_n\|^2 + \frac{1}{\sqrt{2}} \|a_{n+1} - b_n\|^2 \right\}
\]
\[
\leq \frac{\lambda L}{2\sqrt{2}} \left( 2 + \sqrt{2} \right) \|a_n - b_n\|^2 + \frac{\lambda L}{\sqrt{2}} \|a_{n-1} - b_n\|^2
\]
\[
\leq \frac{\lambda L}{\sigma} (1 + \sqrt{2}) \Delta_f(b_n, a_n) + \frac{\lambda L}{\sigma} \Delta_f(a_n, b_{n-1}) + \frac{\lambda L}{\sigma} \Delta_f(a_{n+1}, b_n),
\]
where in (25), we have used the following basic inequalities with (8):
\[
ab \leq \frac{c^2}{2} a^2 + \frac{1}{2c^2} b^2 \text{ and } (a + b)^2 \leq \sqrt{2} a^2 + (2 + \sqrt{2}) b^2.
\]

It follows from (24) and (25) that
\[
\Delta_f(z, a_{n+1}) \leq \Delta_f(z, a_n) - \left( 1 - \frac{\lambda L \sqrt{2}}{\sigma} \right) \Delta_f(a_{n+1}, b_n) - \left( 1 - \frac{\lambda L}{\sigma} \frac{1}{1 + \sqrt{2}} \right) \Delta_f(b_n, a_n)
\]
\[
+ \frac{\lambda L}{\sigma} \Delta_f(a_n, b_{n-1}).
\]

Hence, we obtained the desired result. \( \square \)

**Theorem 1.** Suppose that Assumption 1 holds, and let \( \lambda \in \left( 0, \frac{(\sqrt{2} - 1)\sigma}{L} \right) \). Then, the sequences \( \{a_n\} \) and \( \{b_n\} \) generated by Algorithm BPMFS converges weakly to a solution of the VI (I).

**Proof.** Let \( z \in Sol \). First, we show that the sequence \( \{a_n\} \) is bounded. From Lemma 4, we obtain
\[
\Delta_f(z, a_{n+1}) \leq \Delta_f(z, a_n) + \frac{\lambda L}{\sigma} \Delta_f(a_n, b_{n-1}).
\]
Now let $\Gamma_n = \Delta f(z, a_n) + \frac{\lambda}{p} \Delta f(a_n, b_{n-1})$ and $\Lambda_n = \left(1 - \frac{\lambda(1+\sqrt{2})}{\rho}\right) \left(\Delta f(b_n, a_n) + \Delta f(a_{n+1}, b_n)\right)$. Note that (17) can be re-written as

$$a_{n+1} \leq \Gamma_n - \Lambda_n.$$ 

Hence, it follows from Lemma 2 that $\{a_n\}$ is bounded and $\lim_{n \to \infty} \Delta f(b_n, a_n) = 0$. Consequently, $\{a_n\}$ is bounded and $\|a_n - b_n\| \to 0$, $\|a_{n+1} - b_n\| \to 0$ as $n \to \infty$. This implies that $\|a_{n+1} - a_n\| \to 0$ as $n \to \infty$. Furthermore

$$\|b_{n+1} - b_n\| \leq \|a_{n+1} - a_n\| + \|a_{n+1} - a_n\| + \|a_n - b_n\| \to 0, \quad \text{as } n \to \infty. \quad (26)$$

Consequent

$$\lim_{n \to \infty} \|a_{n+1} - b_{n+1}\| \leq \lim_{n \to \infty} (\|a_{n+1} - b_n\| + \|b_{n+1} - b_n\|) = 0.$$ 

Since $f$ is uniformly Fréchet differentiable, the $\nabla f$ is norm-to-norm uniformly continuous on bounded subsets of $E$. Thus, we have

$$\lim_{n \to \infty} \|\nabla f(a_{n+1}) - \nabla f(b_{n+1})\| = 0. \quad (27)$$

Since $\{a_n\}$ is bounded, then we can choose a subsequence $\{a_{n_k}\}$ of $\{a_n\}$ such that $a_{n_k} \to p$. Since $\|a_n - b_n\| \to 0$, then $b_{n_k} \to p$. Since $\{b_n\} \subset C$, thus $p \in C$. We now show that $p \in \text{Sol}$. Since $b_{n+1} = \Pi_C(\nabla f^*(\nabla f(a_{n+1}) - \lambda \nabla f(b_{n+1})))$, we have

$$\langle \nabla f(a_{n+1}) - \lambda \nabla f(b_{n+1}), v - b_{n+1} \rangle \leq 0, \quad \forall v \in C.$$ 

Equivalently

$$\langle \nabla f(a_{n+1}) - \nabla f(b_{n+1}), v - b_{n+1} \rangle \leq \langle \lambda \nabla f(b_{n+1}), v - b_{n+1} \rangle \quad \forall v \in C.$$ 

This implies that

$$\left\langle \frac{\nabla f(a_{n+1}) - \nabla f(b_{n+1})}{\lambda}, v - b_{n+1} \right\rangle + \langle \nabla f(a_{n+1}) - \nabla f(b_{n+1}), v - b_{n+1} \rangle \leq \langle \nabla f(b_{n+1}), v - b_{n+1} \rangle \quad \forall v \in C. \quad (28)$$

Hence, we have

$$\left\langle \frac{\nabla f(b_{n+1}) - \nabla f(b_{n+1})}{\lambda}, v - b_{n+1} \right\rangle + \langle \nabla f(b_{n+1}), v - b_{n+1} \rangle \leq \langle \nabla f(b_{n+1}), v - b_{n+1} \rangle \quad \forall v \in C.$$ 

Taking limits of the above inequality as $k \to \infty$, we get

$$\liminf_{k \to \infty} \langle \nabla f(b_{n_k}), v - b_{n_k} \rangle \geq 0 \quad \forall v \in C. \quad (29)$$

We choose a sequence $\{\epsilon_k\}$ of positive numbers such that $\{\epsilon_k\}$ is decreasing and convergence to 0. For each $k \geq 1$, we denote by $N_k$ the smallest positive integer such that

$$\langle \nabla f(b_{n_k}), v - b_{n_k} \rangle + \epsilon_k \geq 0 \quad \forall k \geq N_k.$$ 

Since $\{\epsilon_k\}$ is decreasing, it is easy to see that the sequence $\{N_k\}$ is non-decreasing. Thus, for each $k \geq 1$, since $\{b_n\} \subset C$, we have $\nabla f_{N_k} \neq 0$ and by setting

$$\theta_{N_k} = \frac{\nabla f_{N_k}}{\|\nabla f_{N_k}\|^2},$$
we have \( \langle Ab_{N_k}, \theta_{N_k} \rangle = 1 \) for each \( k \geq 1 \). Now, it follows from (29) that for each \( k \geq 1 \),
\[
\langle Ab_{N_k}, x + \epsilon_k \theta_{N_k} - b_{N_k} \rangle \geq 0.
\]
Since \( A \) is pseudomonotone, we get
\[
\langle A(v + \epsilon_k \theta_{N_k}), v + \epsilon_k \theta_{N_k} - b_{N_k} \rangle \geq 0.
\]
This implies that
\[
\langle Ax, v - b_{N_k} \rangle \geq \langle Av - A(x + \epsilon_k \theta_{N_k}), v + \epsilon_k \theta_{N_k} - b_{N_k} \rangle - \epsilon_k \langle Ax, \theta_{N_k} \rangle.
\]
Now we show that \( \lim_{k \to \infty} \epsilon_k \theta_{N_k} = 0 \). Indeed, since \( a_{n_k} \to p \) and \( \lim_{k \to \infty} \| a_{n_k} - b_{N_k} \| = 0 \), we obtain that \( b_{N_k} \to p \) as \( k \to \infty \). Since \( \{b_{n_k}\} \subset C \) and \( A \) is weakly sequentially continuous on \( H \), it follows that \( \{Ab_{N_k}\} \) converges weakly to \( Ap \). We can suppose that \( Ap \neq 0 \) (otherwise, \( p \) is a solution). Since the norm is sequentially weakly lower semicontinuous, we have
\[
0 < \| Ap \| \leq \liminf_{k \to \infty} \| Ab_{n_k} \|.
\]
Since \( \{b_{N_k}\} \subset \{b_{n_k}\} \) and \( \epsilon_k \to 0 \) as \( k \to \infty \), we get
\[
0 \leq \limsup_{k \to \infty} \| \epsilon_k \theta_{N_k} \| = \limsup_{k \to \infty} \left( \frac{\epsilon_k}{\| Ab_{n_k} \|} \right) \leq \limsup_{k \to \infty} \epsilon_k \frac{\| Ab_{n_k} \|}{\liminf_{k \to \infty} \| Ab_{n_k} \|} = 0,
\]
which implies that \( \lim_{k \to \infty} \epsilon_k \theta_{N_k} = 0 \). Thus, we get
\[
\liminf_{k \to \infty} \langle Ax, v - b_{N_k} \rangle \geq 0.
\]
Hence for all \( v \in C \), we get
\[
\langle Ax, v - p \rangle = \lim_{k \to \infty} \langle Ax, x - b_{N_k} \rangle = \liminf_{k \to \infty} \langle Ax, v - b_{N_k} \rangle \geq 0.
\]
Therefore by Lemma 3, \( p \in \text{Sol} = \Omega_{VI} \).

Finally, we show that \( p \) is unique. Assume the contrary, i.e., there exists a subsequence \( \{a_{n_j}\} \) of \( \{a_n\} \) such that \( a_{n_j} \to q \) with \( q \neq p \). Following similar argument has above, we get \( q \in \Omega_{VI} \). It follows from the Bregman opial-like property of \( H \) (more precisely, Lemma 1) that
\[
\lim_{n \to \infty} \Delta_f(p, a_n) = \lim_{k \to \infty} \Delta_f(p, a_{n_k}) < \lim_{k \to \infty} \Delta_f(q, a_{n_k}) = \lim_{j \to \infty} \Delta_f(q, a_{n_j}) < \lim_{j \to \infty} \Delta_f(p, a_{n_j}) = \lim_{n \to \infty} \Delta_f(p, a_n),
\]
which is a contradiction. Thus, we have \( p = q \) and the desired result follows. This completes the proof. \( \square \)

**Remark 4.** We note that Algorithm BPMFS converges strongly to a solution of the VI (1) if the operator \( A : E \to E^* \) is \( \gamma \)-strongly pseudomonotone and \( L \)-Lipschitz continuous.

**Proof.** Indeed, let \( p \in \text{Sol} \). Following similar argument in the proof of Lemma 4 till (19), we have
\[
\Delta_f(p, a_{n+1}) \leq \Delta_f(p, a_n) - \Delta_f(a_{n+1}, a_n) + \lambda \langle Ab_n, p - a_{n+1} \rangle.
\] (30)
Since $p$ is a solution of the VI (1), we have $\langle Ap, v - p \rangle \geq 0$ for all $v \in C$. Using the strong pseudomonotonicity of $A$, we obtain $\langle Ax, v - p \rangle \geq \gamma \|x - p\|^2$ for all $x \in C$. Taking $v = b_n \in C$, we get

$$\langle Ab_n, p - b_n \rangle \leq -\gamma \|b_n - p\|^2.$$ 

Hence, we have

$$\langle Ab_n, p - a_{n+1} \rangle = \langle Ab_n, p - b_n \rangle + \langle Ab_n, b_n - a_{n+1} \rangle \leq -\gamma \|b_n - p\|^2 + \langle Ab_n, b_n - a_{n+1} \rangle.$$  \hspace{1cm} (31)

Therefore (30) and (31), we get

$$\Delta_f(p, a_{n+1}) \leq \Delta_f(p, a_n) - \Delta_f(a_{n+1}, a_n) + \lambda \langle Ab_n, b_n - a_{n+1} \rangle - \gamma \lambda \|b_n - p\|^2.$$ 

Using (7), we have

$$\Delta_f(p, a_{n+1}) \leq \Delta_f(p, a_n) - \Delta_f(a_{n+1}, b_n) - \Delta_f(b_n, a_n) - \langle \nabla f(b_n) - \nabla f(a_n), a_{n+1} - b_n \rangle$$

$$+ \lambda \langle Ab_n, b_n - a_{n+1} \rangle - \gamma \lambda \|b_n - p\|^2$$

$$= \Delta_f(p, a_n) - \Delta_f(a_{n+1}, b_n) - \Delta_f(b_n, a_n) + \langle \nabla f(a_n) - \nabla f(b_n), a_{n+1} - b_n \rangle$$

$$- \gamma \lambda \|b_n - p\|^2.$$ \hspace{1cm} (32)

Note that

$$\langle \nabla f(a_n) - \lambda Ab_n - \nabla f(b_n), a_{n+1} - b_n \rangle = \langle \nabla f(a_n) - \lambda Ab_n - \nabla f(b_n), a_{n+1} - b_n \rangle$$

$$+ \lambda \langle Ab_n - Ab_n, a_{n+1} - b_n \rangle.$$ 

Since $a_{n+1} \in T_n$, then we have

$$\langle \nabla f(a_n) - \lambda Ab_n - \nabla f(b_n), a_{n+1} - b_n \rangle \leq 0.$$ 

Hence, we have

$$\langle \nabla f(a_n) - \lambda Ab_n - \nabla f(b_n), a_{n+1} - b_n \rangle \leq \lambda \langle Ab_n - Ab_n, a_{n+1} - b_n \rangle$$

$$\leq \lambda \|b_n - b_n\| \|a_{n+1} - b_n\|$$

$$\leq \lambda \left\{ \frac{1}{2\sqrt{2}} \|b_n - b_n\|^2 + \frac{1}{\sqrt{2}} \|a_{n+1} - b_n\|^2 \right\}$$

$$\leq \frac{\lambda L}{2\sqrt{2}} \left( (2 + \sqrt{2}) \|a_n - b_n\| + \sqrt{2} \|a_n - b_{n-1}\|^2 \right)$$

$$+ \frac{\lambda L}{\sqrt{2}} \|a_{n+1} - b_n\|^2$$

$$\leq \frac{\lambda L}{\sigma} (1 + \sqrt{2}) \Delta_f(b_n, a_n) + \frac{\lambda L}{\sigma} \Delta_f(a_n, b_{n-1})$$

$$+ \frac{\lambda L \sqrt{2}}{\sigma} \Delta_f(a_{n+1}, b_n).$$ \hspace{1cm} (33)
Combining (32) and (33), we get

\[
\Delta_f(p, a_{n+1}) \leq \Delta_f(p, a_n) - \left(1 - \frac{\lambda L \sqrt{2}}{\sigma}\right) \Delta_f(a_{n+1}, b_n) - \left(1 - \frac{\lambda L}{\sigma}(1 + \sqrt{2})\right) \Delta_f(b_n, a_n) + \frac{\lambda L}{\sigma} \Delta_f(a_n, b_{n-1}) - \gamma \lambda \|b_n - p\|^2. 
\] (34)

Following similar argument as in the Proof of Theorem 1, we have that the sequence \(\{a_n\}\) is bounded and \(\|b_n - a_n\| \to 0, \|a_{n+1} - b_n\| \to 0\) as \(n \to \infty\). \(\|a_{n+1} - a_n\| \to 0\) as \(n \to \infty\). Since \(f\) is continuous on bounded sets and \(\nabla f\) is weakly–weakly continuous, we get \(\|f(a_{n+1}) - f(a_n)\| \to 0\) and \(\|\nabla f(a_{n+1}) - \nabla f(a_n)\| \to 0\) as \(n \to \infty\). Therefore

\[
\Delta_f(p, a_n) - \Delta_f(p, a_{n+1}) = f(p) - f(a_n) - (\nabla f(a_n), p - a_n) - (f(a_{n+1}), p - a_{n+1})
\]

\[
= f(a_{n+1}) - f(a_n) + (\nabla f(a_{n+1}), p - a_{n+1}) - (\nabla f(a_n), p - a_n)
\]

\[
= f(a_{n+1}) - f(a_n) + (\nabla f(a_{n+1}) - \nabla f(a_n), p - a_n) + (\nabla f(a_{n+1}), a_n - a_{n+1}).
\]

Thus, we have that

\[
\lim_{n \to \infty} (\Delta_f(p, a_n) - \Delta_f(p, a_{n+1})) = 0.
\]

Note that

\[
\Delta_f(a_n, b_{n-1}) = f(a_n) - f(b_{n-1}) - (\nabla f(b_{n-1}), a_n - b_{n-1}) \to 0.
\]

It follows from (34) that

\[
\gamma \lambda \|b_n - p\|^2 \leq \Delta_f(p, a_n) - \Delta_f(p, a_{n+1}) + \frac{\lambda L}{\sigma} \Delta_f(a_n, b_{n-1}) \to 0.
\]

Therefore, we get

\[
\lim_{n \to \infty} \|b_n - p\| = 0.
\]

Since \(\|a_n - b_n\| \to 0\), then \(\|a_n - p\| \to 0\) as \(n \to \infty\). This completes the proof. \(\square\)

3.2. Bregman Projection Method with Self-Adaptive Stepsize

Next, we propose a Bregman subgradient extragradient method whose convergence does not require a prior estimate of the Lipschitz constant of the cost operator \(A\). The importance of the second algorithm is to cater for the situations when estimating the Lipschitz constant of \(A\) is very difficult. So, the algorithm uses a self-adaptive process for selecting its stepsize and does not require a prior estimate of the Lipschitz constant of \(A\).

Algorithm BPMSAS

Initialization: Pick \(a_0, b_0 \in H, \lambda_0 > 0\) arbitrarily and \(\alpha \in (0, 1)\).

**Step 1:** Compute

\[
a_1 = \Pi_C(\nabla f^*(\nabla f(a_0) - \lambda_0 Ab_0)),
\]

\[
b_1 = \Pi_C(\nabla f^*(\nabla f(a_1) - \lambda_0 Ab_0)).
\]

**Step 2:** Given the current iterates \(a_n, b_n\) and \(b_{n-1}\), and \(\lambda_n\), calculate \(a_{n+1}\) and \(b_{n+1}\) as follows:

\[
a_{n+1} = \Pi_{T_{\alpha}}(\nabla f^*(\nabla f(a_n) - \lambda_n Ab_n)),
\]

\[
b_{n+1} = \Pi_C(\nabla f^*(\nabla f(a_{n+1}) - \lambda_{n+1} Ab_n)). 
\] (35)
Theorem 2. Suppose that Assumption 1 holds, and let
\[
\lambda_{n+1} = \begin{cases} 
\min \left\{ \lambda_n, \frac{\|b_n - b_{n-1}\|}{\|Ab_n - Ab_{n-1}\|} \right\}, & \text{if } Ab_n = Ab_{n-1}, \\
\lambda_n, & \text{otherwise,}
\end{cases}
\]
(36)
and
\[
T_n := \{ z \in H : \langle \nabla f(a_n) - \lambda_n Ab_n - \nabla f(b_n), z - b_n \rangle \leq 0 \}.
\]
If \( a_{n+1} = a_n \) and \( b_{n+1} = b_{n-1} \) or \( Ab_n = 0 \) : STOP. Otherwise, set \( n \leftarrow n + 1 \) and repeat Step 2.

Remark 5. Note that in Algorithm BPMSAS, the stepsize \( \lambda_n \) is chosen by a self-adaptive process. This means that the stepsize is improved along with the iterations. Moreso, the sequence \( \{\lambda_n\} \) is monotonically non-increasing and bounded below by \( \min\{\lambda_0, \frac{\sigma}{2}\} \). Then, we can say that the limit \( \lim_{n \to \infty} \lambda_n \) exist, which we denote by \( \lambda \), i.e., \( \lim_{n \to \infty} \lambda_n = \lambda > 0 \).

We now give the convergence of Algorithm BPMSAS.

Theorem 2. Suppose that Assumption 1 holds, and let \( \alpha \in \left( 0, (\sqrt{2} - 1)\sigma \right) \). Then, the sequences \( \{a_n\} \) and \( \{b_n\} \) generated by Algorithm BPMSAS converges weakly to a solution of the VI (1).

Proof. Let \( z \in VI(C, A) \). Following similar argument as in the proof of (24), we obtain
\[
\Delta_f(z, a_{n+1}) \leq \Delta_f(z, a_n) - \Delta_f(a_{n+1}, b_n) - \Delta_f(b_n, a_n) - \Delta_f(b_n, a_{n+1}) + \lambda_n \langle Ab_n - Ab_{n-1}, b_n - a_n \rangle.
\]
Now, using Cauchy-Schwartz inequality and the definition of \( \lambda_{n+1} \) in (36), we get
\[
\begin{align*}
\Delta_f(z, a_{n+1}) & \leq \Delta_f(z, a_n) - \Delta_f(a_{n+1}, b_n) - \Delta_f(b_n, a_n) \\
& \quad + \frac{\lambda_n}{\lambda_{n+1}} \|Ab_n - Ab_{n-1}\| \|b_n - a_{n+1}\| \\
& \leq \Delta_f(z, a_n) - \Delta_f(a_{n+1}, b_n) - \Delta_f(b_n, a_n) \\
& \quad + \frac{\lambda_n}{\lambda_{n+1}} \left( \frac{1}{\sqrt{2}} \|b_{n-1} - b_n\|^2 + \sqrt{2} \|b_n - a_{n+1}\|^2 \right) \\
& \leq \Delta_f(z, a_n) - \Delta_f(a_{n+1}, b_n) - \Delta_f(b_n, a_n) \\
& \quad + \frac{\lambda_n a}{2\lambda_{n+1}} \|b_{n-1} - b_n\|^2 + \frac{(\sqrt{2} + 1)\lambda_n a}{2\lambda_{n+1}} \|a_n - b_n\|^2 + \frac{\lambda_n a \sqrt{2}}{2\lambda_{n+1}} \|b_n - a_{n+1}\|^2,
\end{align*}
\]
(37)
where in (37) and (38), we have used the identities: \( 2ab \leq \frac{1}{\sqrt{2}} a^2 + \sqrt{2} b^2 \) and \( (a + b)^2 \leq \sqrt{2} a^2 + (2 + \sqrt{2}) b^2 \) respectively. Since \( f \) is strongly convex and \( \sigma > 0 \), it follows from (8) that
\[
\begin{align*}
\Delta_f(z, a_{n+1}) & \leq \Delta_f(z, a_n) - \Delta_f(a_{n+1}, b_n) - \Delta_f(b_n, a_n) \\
& \quad + \frac{\lambda_n a}{\sigma \lambda_{n+1}} \Delta_f(a_n, b_{n-1}) + \frac{(\sqrt{2} + 1)\lambda_n a}{\sigma \lambda_{n+1}} \Delta_f(b_n, a_n) + \frac{\lambda_n a \sqrt{2}}{\sigma \lambda_{n+1}} \Delta_f(a_{n+1}, b_n) \\
& = \Delta_f(z, a_n) - \left( 1 - \frac{\lambda_n a \sqrt{2}}{\sigma \lambda_{n+1}} \right) \Delta_f(a_{n+1}, b_n) - \left( 1 - \frac{(\sqrt{2} + 1)\lambda_n a}{\sigma \lambda_{n+1}} \right) \Delta_f(b_n, a_n) \\
& \quad + \frac{\lambda_n a}{\sigma \lambda_{n+1}} \Delta_f(a_n, b_{n-1}).
\end{align*}
\]
(39)
Now put
\[ \Gamma_n = \Delta f(z, a_n) + \frac{\alpha \lambda_n}{\sigma \lambda_{n+1}} \Delta f(a_n, b_{n-1}), \]
and
\[ \Lambda_n = \left( 1 - \frac{\sqrt{2} + 1}{\sigma \lambda_{n+1}} \right) \left( \Delta f(b_n, a_n) + \Delta f(a_{n+1}, b_n) \right). \]
Then (39) can be rewritten as
\[ a_{n+1} \leq \Gamma_n - \Lambda_n - \Omega_n, \]
where \( \Omega_n = \frac{\lambda_n}{\lambda_{n+1}} \Delta f(a_{n+1}, b_n) \). It follows from Lemma 2 that \( \{a_n\} \) is bounded. Note that since \( \lim_{n \to \infty} \lambda_n = \lambda > 0 \) and \( \alpha \in (0, (\sqrt{2} - 1)\sigma) \), there exists a number \( n_0 \geq 0 \) such that
\[ \lim_{n \to \infty} \left( 1 - \frac{\sqrt{2} + 1}{\sigma \lambda_{n+1}} \right) = 1 - \frac{\sqrt{2} + 1}{\sigma} > 0. \]
This implies that there exists \( c > 0 \) such that
\[ \left( 1 - \frac{\sqrt{2} + 1}{\sigma \lambda_{n+1}} \right) \geq c > 0, \quad \forall n \geq n_0. \]
Hence from Lemma 2, we obtain
\[ \sum_{n=0}^{\infty} \epsilon (\Delta f(b_n, a_n) + \Delta f(a_{n+1}, b_n)) < \infty, \]
which implies that
\[ \lim_{n \to \infty} \Delta f(b_n, a_n) = \lim_{n \to \infty} \Delta f(a_{n+1}, b_n) = 0. \]
Consequently
\[ \lim_{n \to \infty} \|a_n - b_n\| = \lim_{n \to \infty} \|a_{n+1} - b_n\| = 0. \]
Hence
\[ \lim_{n \to \infty} \|a_{n+1} - a_n\| = \lim_{n \to \infty} \|b_{n+1} - b_n\| = \lim_{n \to \infty} \|a_{n+1} - b_{n+1}\| = 0. \]
Since \( f \) is uniformly Fréchet differentiable, then \( \nabla f \) is norm-to-norm continuous and therefore
\[ \lim_{n \to \infty} \|\nabla f(a_{n+1}) - \nabla f(b_{n+1})\| = 0. \]
Since \( \{a_n\} \) is bounded, then there exists a subsequence \( \{a_{n_k}\} \) of \( \{a_n\} \) such that \( a_{n_k} \to \bar{x}. \)
Following similar argument from (27) of Theorem 1, we get that \( \bar{x} \in VI(C, A) \). This completes the proof. \( \square \)

4. Numerical Examples
In this section, we test the performance of Algorithm BPMFS and Algorithm BPMSAS for solving some VI available in the literature. All numerical computation are carried out using MATLAB program on a PC (with Intel(R) Core (TM) i7-5600U CPU @ 3.40GHz, RAM 8.00GB) in MATLAB 9.7 2019(a).
Example 1 are respectively: BPMFS and the SEM in terms of number of iteration and CPU time taken for computation. This is due to the BPMFS (named, Alg. 1), Algorithm BPMSAS (named, Alg. 2) and the SEM (Algorithm (3)) using the following parameters: for Algorithm BPMFS and SEM, we choose $\lambda = 0.3L$ while for Algorithm BPMSAS, we choose $\lambda_0 = 0.3$. The computation is stopped when $D_n = \|a_n - 0\| < 10^{-4}$ is satisfied. We plot the graphs of $D_n$ against number of iteration in each case for each iteration. The numerical results are shown in Table 1 and Figure 1. The numerical results show that Algorithm BPMSAS has computational advantage than Algorithm BPMFS and the SEM in terms of number of iteration and CPU time taken for computation. This is due to the fact that the stepsize is improve in each iteration compare to the Algorithm BPMFS and SEM which used fixed stepsize. More so, Algorithm BPMFS performs better than the SEM.

Table 1. Computation result for Example 2.

| $m$   | Iter | Alg. 1 | Alg. 2 | SEM |
|-------|------|--------|--------|-----|
| 50    | 29   | 0.0050 | 0.0022 | 0.0083 |
| 100   | 32   | 0.0165 | 0.0101 | 0.0203 |
| 150   | 35   | 0.0161 | 0.0075 | 0.0211 |
| 200   | 37   | 0.0321 | 0.0146 | 0.0385 |

Example 2. First, we consider the variational inequality problem for a finite dimensional space $\mathbb{R}^m$. The cost operator $A : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is defined by $Ax = Gx + q$, such that

$$G = SS^T + Q + R,$$

where $S$ is a $m \times m$ matrix, $Q$ is a $m \times m$ skew-symmetric matrix, $R$ is a $m \times m$ diagonal matrix, whose diagonal entries are positive (so, $G$ is positive definite) and $q$ is a vector in $\mathbb{R}^m$. The feasible set $C$ is the closed convex polyhedron which is defined by $C = \{x \in \mathbb{R}^m : Bx \leq b\}$, where $B$ is a $l \times m$ ($l = 10$) matrix and $b$ is a non-negative vector in $\mathbb{R}^m$. It is easy to see that $A$ is monotone (thus, pseudomonotone) and $L$-Lipschitz continuous with $L = \|G\|$. For experimental purpose, all the entries of $S, Q, R, B$ and $b$ are generated randomly such that their properties are satisfied and the starting points $a_0,b_0$ are generated randomly in $\mathbb{R}^m$ ($m = 50, 100, 150, 200$). The Bregman projection onto the feasible set $C$ is easily computed using the \texttt{fmincon} solver in MATLAB Optimization toolbox. We set $q = 0 \in \mathbb{R}^m$, in this case, we know that the unique solution of the VI is $\{0\}$. We choose $f(x) = \frac{1}{2} \|x\|^2$ which is strongly convex with modulus $1$. We test Algorithm BPMFS (named, Alg. 1), Algorithm BPMSAS (named, Alg. 2) and the SEM (Algorithm (3)) against the number of iteration in each case for each iteration. The numerical results are shown in Table 1 and Figure 1. The numerical results show that Algorithm BPMSAS has computational advantage than Algorithm BPMFS and the SEM in terms of number of iteration and CPU time taken for computation. This is due to the fact that the stepsize is improve in each iteration compare to the Algorithm BPMFS and SEM which used fixed stepsize. More so, Algorithm BPMFS performs better than the SEM.

Example 3. Here, we compare the performance of Algorithm BPMFS and Algorithm BPMSAS for different convex functions. Let $C$ be defined by

$$C = \{ x = (a_1, \ldots, a_m)^T : \|x\| \leq 1 \text{ and } a_i > 0, \ i = 1, \ldots, m \}.$$  

Note that the corresponding values of $\nabla f(x)$ and $\nabla f^*(x) = \nabla f^{-1}(x)$ for the convex function given in Example 1 are respectively:

(i) $\nabla f(x) = x$ and $\nabla f^*(x) = x$ for SE,

(ii) $\nabla f(x) = \left(1 + \log(a_1), \ldots, 1 + \log(a_m)\right)^T$ and $\nabla f^*(x) = \left(\exp(a_1 - 1), \ldots, \exp(a_m - 1)\right)^T$ for KL,

(iii) $\nabla f(x) = -\left(\frac{1}{a_1}, \ldots, \frac{1}{a_m}\right)^T$ and $\nabla f^*(x) = -\left(\frac{1}{\exp(a_1 - 1)}, \ldots, \frac{1}{\exp(a_m - 1)}\right)^T$ for IS,

(iv) $\nabla f(x) = Qx$ and $\nabla f^*(x) = Q^{-1}(x)$ for MD.

We test the algorithms using similar parameters as in Example 1 with $\|a_n - 0\| < 10^{-4}$ as stopping criterion. The numerical results are shown in Table 2. The computation result shows that the IS has the best performance while the MD has the poorest performance for both Algorithm BPMFS and Algorithm BPMSAS.
Figure 1. Example 2, Top Left: $m = 50$; Top Right: $m = 100$, Bottom Left: $m = 150$; Bottom Right: $M = 200$.

Table 2. Example 3 Numerical result showing number of iteration and CPU time (sec) for various types of function $f$ using Algorithm BPMFS and Algorithm BPMSAS.

|                | $m = 50$ |  | $m = 100$ |  | $M = 150$ |  | $m = 200$ |  |
|----------------|----------|---|-----------|---|-----------|---|-----------|---|
|                | Iter.    | CPU | Iter.      | CPU | Iter.      | CPU | Iter.      | CPU |
| SE Alg. 1      | 23       | 0.0049 | 26       | 0.0099 | 18       | 0.0153 | 18       | 0.0163 |
| Alg. 2         | 15       | 0.0021 | 16       | 0.0039 | 16       | 0.0109 | 16       | 0.0138 |
| KL Alg. 1      | 15       | 0.0044 | 17       | 0.0081 | 18       | 0.0150 | 18       | 0.00160 |
| Alg. 2         | 15       | 0.0040 | 15       | 0.0058 | 15       | 0.00109 | 15       | 0.0143 |
| IS Alg. 1      | 14       | 0.0048 | 14       | 0.0104 | 14       | 0.0133 | 12       | 0.0183 |
| Alg. 2         | 7        | 0.0028 | 5        | 0.0025 | 4        | 0.0041 | 3        | 0.0057 |
| MD Alg. 1      | 24       | 0.0104 | 27       | 0.0419 | 28       | 0.0960 | 30       | 0.1910 |
| Alg. 2         | 12       | 0.0094 | 12       | 0.0254 | 12       | 0.0423 | 15       | 0.0698 |

Example 4. Next, we apply our algorithms to solve the Least Absolute Selection and Shrinkage Operator (LASSO) problem [42]. Let $z$ be a vector in $\mathbb{R}^M$, $B$ be a matrix of order $M \times N$ ($M << N$) and $\lambda > 0$. The LASSO problem is given by

$$
\min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \| Bx - z \|_2^2 + \lambda \| x \|_1 \right\}, \quad (40)
$$

where $\| x \|_2$ is the Euclidean norm of $x$ and $\| x \|_1 = \sum_{i=1}^{N} |a_i|$ is the $l_1$-norm of $x$. This problem has been considered as a great tool in several branches of science and engineering. Clearly, (40) is a convex unconstrained minimization problem which appears in compress sensing and image reconstruction, where the original signal (or image) is sparse in some orthogonal basis by the process

$$
z = Bx + \xi,
$$
where $x$ is the original signal (or image), $B$ is the blurring operator, $\xi$ is a noise and $z$ is the degraded or blurred data which needs to be recovered. Several iterative methods for solving (40) have been introduced with the earliest being the projection method by Figureido et al. [43]. Equivalently, the LASSO problem (40) can be expressed as a VI by setting $A = B^T(Bx - z)$ (see, [44]). We apply our algorithms to solve the signal problem with $N = 4096$ which is generated by the uniform distribution on the interval $[-2, 2]$ with 160 non-zero elements. The matrix $B$ is generated by the normal distribution with mean zero and variance one while the observation $z$ with $M = 512$ is generated by Gaussian noise with SNR = 40. The initial points $a_0, y_0$ are picked randomly and in Algorithm 1, we choose $\lambda = 0.35$, in Algorithm 2, $\lambda_0 = 0.02$ and in SEM, $\lambda = 0.35$. We compare the performance of each algorithm with respect to the number of iterations, CPU time and MSE values, see Figures 2 and 3.

Figure 2. Sparse signal recovery. From the top to the bottom is the original signal, the measurement, and the reconstructed signal using Algorithms 1 and 2 and SEM.

Figure 3. Graphs of objective function value against number of iteration (Top) and objective function against CPU time (sec) (Bottom) for Algorithms 1 and 2 and SEM.
5. Conclusions

In this paper, we introduce two Bregman subgradient extragradient method for solving variational inequalities with pseudomonotone and Lipschitz continuous operator in a real Hilbert space. We established some weak convergence results for the sequence generated by the algorithms. We also gives some numerical experiments to illustrate the performance of the proposed algorithms. The numerical results show that the proposed methods perform better than the usual subgradient extragradient method and their efficiency depends on the type of convex function used in the algorithms.

Author Contributions: Conceptualization, L.O.J.; methodology, L.O.J.; validation, L.O.J., M.A., S.H.K.; formal analysis, L.O.J., S.H.K.; writing—original draft preparation, L.O.J.; writing—review and editing, M.A., S.H.K.; visualization, L.O.J.; supervision, L.O.J., M.A., S.H.K.; project administration, M.A.; funding acquisition, M.A.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Sefako Makgatho Health Sciences University Postdoctoral research fund and and the APC was funded by Department of Mathematics and Applied Mathematics, Sefako Makgatho Health Sciences University, Pretoria, South Africa.

Acknowledgments: The authors acknowledge with thanks, the Department of Mathematics and Applied Mathematics at the Sefako Makgatho Health Sciences University for making their facilities available for the research.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Glowinski, R.; Lions, J.L.; Trémoliexeres, R. Numerical Analysis of Variational Inequalities; North-Holland: Amsterdam, The Netherlands, 1981.
2. Kinderlehrer, D.; Stampacchia, G. An Introduction to Variational Inequalities and Their Applications; Academic Press: New York, NY, USA, 1980.
3. Combettes, P.L.; Pesquet, J.-C. Deep neural structures solving variational inequalities. Set-Valued Var. Anal. 2020, 28, 491–518. [CrossRef]
4. Juditsky, A.; Nemirovski, A. Solving variational inequalities with monotone operators in domains given by linear minimization oracles. arXiv 2016, arXiv:1312.1073v2.
5. Luo, M.J.; Zhang, Y. Robust solution to box-constrained stochastic linear variational inequality problem. J. Ineq. Appl. 2017, 2017, 253. [CrossRef]
6. Abass, H.A.; Aremu, K.O.; Jolaoso, L.O.; Mewomo, O.T. An inertial forward-backward splitting algorithm for approximating solutions of certain optimization problems. J. Nonlinear Funct. Anal. 2020, 2020, 6.
7. Alakoya, T.O.; Jolaoso, L.O.; Mewomo, O.T. Modified Inertial Subgradient Extragradient Method with Self-Adaptive Stepsize for Solving Monotone Variational Inequality and Fixed Point Problems. Available online: https://www.tandfonline.com/doi/abs/10.1080/02331934.2020.1723586. (accessed on 1 November 2020)
8. Cruz, J.Y.B.; Iusem, A.N. A strongly convergent direct method for monotone variational inequalities in Hilbert spaces. Numer. Funct. Anal. Optim. 2009, 30, 23–36. [CrossRef]
9. Cai, G.; Gibali, A.; Iyiola, O.S.; Shehu, Y. A new double-projection method for solving variational inequalities in Banach space. J. Optim Theory Appl. 2018, 178, 219–239. [CrossRef]
10. Cegielski, A. Iterative Methods for Fixed Point Problems in Hilbert Spaces; Lecture Notes in Mathematics; Springer: Heidelberg, Germany, 2012; Volume 2057.
11. Ceng, L.C.; Teboulle, M.; Yao, J.C. Weak convergence of an iterative method for pseudomonotone variational inequalities and fixed-point problems. J. Optim. Theory Appl. 2010, 146, 19–31. [CrossRef]
12. Censor, Y.; Gibali, A.; Reich, S. Algorithms for the split variational inequality problem. Numer. Algorithms 2012, 59, 301–323. [CrossRef]
13. Censor, Y.; Gibali, A.; Reich, S. Extensions of Korpelevich’s extragradient method for variational inequality problems in Euclidean space. Optimization 2012, 61, 119–1132. [CrossRef]
14. Censor, Y.; Gibali, A.; Reich, S. The subgradient extragradient method for solving variational inequalities in Hilbert spaces. *J. Optim. Theory Appl.* 2011, 148, 318–335. [CrossRef]
15. Censor, Y.; Gibali, A.; Reich, S. Strong convergence of subgradient extragradient methods for the variational inequality problem in Hilbert space. *Optim. Methods Softw.* 2011, 26, 827–845. [CrossRef]
16. Chidume, C.E.; Nnakwe, M.O. Convergence theorems of subgradient extragradient algorithm for solving variational inequalities and a convex feasibility problem. *Fixed Theory Appl.* 2018, 2018, 16. [CrossRef]
17. Iusem, A.N. An iterative algorithm for the variational inequality problem. *Comput. Appl. Math.* 1994, 13, 103–114.
18. Jolaoso, L.O.; Alakoya, T.O.; Taiwo, A.; Mewomo, O.T. An inertial extragradient method via viscosity approximation approach for solving equilibrium problem in Hilbert spaces. *Optimization* 2020. [CrossRef]
19. Jolaoso, L.O.; Aphane, M. Weak and strong convergence Bregman extragradient schemes for solving pseudo-monotone and non-Lipschitz variational inequalities. *J. Inequalities Appl.* 2020, 2020, 195. [CrossRef]
20. Jolaoso, L.O.; Taiwo, A.; Alakoya, T.O.; Mewomo, O.T. A unified algorithm for solving variational inequality and fixed point problems with application to split equality problems. *Comput. Appl. Math.* 2020, 39, 38. [CrossRef]
21. Kanzow, C.; Shehu, Y. Strong convergence of a double projection-type method for monotone variational inequalities in Hilbert spaces. *J. Fixed Point Theory Appl.* 2018, 20. [CrossRef]
22. Khanh, P.D.; Vuong, P.T. Modified projection method for strongly pseudomonotone variational inequalities. *J. Glob. Optim.* 2014, 58, 341–350. [CrossRef]
23. Korpelevich, G.M. The extragradient method for finding saddle points and other problems. *Ekon. Mat. Metody.* 1976, 12, 747–756. (In Russian)
24. Apostol, R.Y.; Semenov, A.A.G.V. Iterative algorithms for monotone bilevel variational inequalities. *J. Comput. Appl. Math.* 2012, 107, 3–14.
25. Kassay, G.; Reich, S.; Sabach, S. Iterative methods for solving system of variational inequalities in reflexive Banach spaces. *SIAM J. Optim.* 2011, 21, 1319–1344. [CrossRef]
26. Bauschke, H.H.; Combettes, P.L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*; (CMS Books in Mathematics); Springer: New York, NY, USA, 2011.
27. Halpern, B. Fixed points of nonexpanding maps. *Proc. Am. Math. Soc.* 1967, 73, 957–961. [CrossRef]
28. Popov, L.C. A modification of the Arrow-Hurwicz method for finding saddle point. *Math. Notes* 1980, 28, 845–848. [CrossRef]
29. Malitsky, Y.V.; Semenov, V.V. An extragradient algorithm for monotone variational inequalities. *Cybern. Syst. Anal.* 2014, 50, 271–277. [CrossRef]
30. Nomirovskii, D.A.; Rublyov, B.V.; Semenov, V.V. Convergence of two-step method with Bregman divergence for solving variational inequalities. *Cybern. Syst. Anal.* 2019, 55, 359–368. [CrossRef]
31. Bregman, L.M. The relaxation method for finding common points of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* 1967, 7, 200–217. [CrossRef]
32. Gibali, A. A new Bregman projection method for solving variational inequalities in Hilbert spaces. *Pure Appl. Funct. Anal.* 2018, 3, 403–415.
33. Hieu, D.V.; Cholamjiak, P. Modified Extragradient Method with Bregman Distance for Variational Inequalities. *Appl. Anal.* 2020. Available online: https://www.tandfonline.com/doi/abs/10.1080/00036811.2020.1757078 (accessed on 1 November 2020).
34. Denisov, S.V.; Semenov, V.V.; Stetsykyn, P.I. Bregman extragradient method with monotone rule of step adjustment. *Cybern. Syst. Anal.* 2019, 55, 377–383. [CrossRef]
35. Bauschke, H.H.; Borwein, J.M.; Combettes, P.L. Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces. *Commun. Contemp. Math.* 2001, 3, 615–647. [CrossRef]
36. Beck, A. First-Order Methods in Optimization, Society for Industrial and Applied Mathematics. *Philadelphia* 2017. Available online: https://www.worldcat.org/title/first-order-methods-in-optimization/oclc/1002692951 (accessed on 1 November 2020).
37. Reich, S.; Sabach, S. A strong convergence theorem for proximal type- algorithm in reflexive Banach spaces. *J. Nonlinear Convex Anal.* 2009, 10, 471–485.
38. Huang, Y.Y.; Jeng, J.C.; Kuo, T.Y.; Hong, C.C. Fixed point and weak convergence theorems for point-dependent \( \lambda \)-hybrid mappings in Banach spaces. *Fixed Point Theory Appl.* 2011, 2011, 105. [CrossRef]
39. Lin, L.J.; Yang, M.F.; Ansari, Q.H.; Kassay, G. Existence results for Stampacchia and Minty type implicit variational inequalities with multivalued maps. *Nonlinear Anal. Theory Methods Appl.* 2005, 61, 1–19. [CrossRef]

40. Mashreghi, J.; Nasri, M. Forcing strong convergence of Korpelevich’s method in Banach spaces with its applications in game theory. *Nonlinear Anal.* 2010, 72, 2086–2099. [CrossRef]

41. Solodov, M.V.; Svaiter, B.F. A new projection method for variational inequality problems. *SIAM J. Control Optim.* 1999, 37, 765–776. [CrossRef]

42. Xiao, Y.; Zhu, H. A conjugate gradient method to solve convex constrained monotone equations with applications in compressive sensing. *J. Math. Anal. Appl.* 2013, 405, 310–319. [CrossRef]

43. Figueiredo, M.A.T.; Nowak, R.D.; Wright, S.J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Signal Process.* 2007, 1, 586–597. [CrossRef]

44. Shehu, Y.; Iyiola, O.S.; Ogbuisi, F.U. Iterative method with inertial terms for nonexpansive mappings: Applications to compressed sensing. *Numer. Algorithms* 2020, 83, 1321–1347. [CrossRef]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).