

Halpern-type relaxed inertial algorithms with Bregman divergence for solving variational inequalities

Lateef Olakunle Jolaoso¹, Pongsakorn Sunthrayuth^{2,*},
Prasit Cholamjiak³ and Yeol Je Cho⁴

¹Department of Mathematics and Applied Mathematics,
Sefako Makgatho Health Sciences University, Pretoria, South Africa

²Department of Mathematics and Computer Science, Faculty of Science and Technology,
Rajamangala University of Technology Thanyaburi (RMUTT)
Thanyaburi, Pathumthani, 12110, Thailand

³School of Science, University of Phayao, Phayao 56000, Thailand

⁴Department of Mathematics Education and the RINS,
Gyeongsang National University, Jinju, Korea

November 16, 2021

Abstract

It is well-known that the use of Bregman divergence is an elegant and effective technique for solving many problems in applied sciences. In this paper, we introduce and analyze two new inertial-like algorithms with Bregman divergence for solving pseudomonotone variational inequalities in a real Hilbert space. The first algorithm is inspired by Halpern-type iteration and subgradient extragradient method and the second algorithm is inspired by Halpern-type iteration and Tseng's extragradient method. Under suitable conditions, the strong convergence theorems of the algorithms are established without assuming the Lipschitz continuity and the sequential weak continuity of any mapping. Finally, several numerical experiments with various types of Bregman divergence are also performed to illustrate the theoretical analysis. The results presented in this paper improve and generalize the related works in the literature.

Keywords: Bregman divergence; Hilbert space; Strong convergence; Variational inequality problem; Pseudomonotone mapping

2010 Mathematics Subject Classification: 47H09, 47H10, 47J25, 47J05.

*Corresponding author

E-mail: pongsakorn_su@rmutt.ac.th (P. Sunthrayuth).

1 Introduction

In 1966, Hartman and Stampacchia [21] first introduced the variational inequality problem (VIP) for used in the study of partial differential equations with unilateral boundary conditions and free boundary value problems of elliptic type from mechanics. The VIP has been intensively and wildly studied and it has been found that it also can be applied to real world problems such as equilibrium problems, optimal control problems, machine learning, signal processing and linear inverse problems (see, for example, [14, 15, 26, 27, 31, 34, 46]).

Throughout this paper, we assume that H is a real Hilbert space, whose inner product and norm are denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. Let C be a nonempty, closed and convex subset of H and $A : C \rightarrow H$ be a given mapping. The *variational inequality problem* is to find a point $z \in C$ such that

$$\langle Az, x - z \rangle \geq 0 \tag{1.1}$$

for all $x \in C$. We denote by $VI(C, A)$ the solution set of VIP (1.1). A concrete example of VIP is the problem of solving a system of equations. It is to see that if $C = H = \mathbb{R}^m$, then $z \in VI(C, A)$ if and only if $Az = 0$. Another example of VIP is the constrained optimization problem. In fact, we set $A := \nabla f$, where ∇f is the gradient of a continuously differentiable convex function f and if $z \in VI(C, A)$, then z solves the following minimization problem:

$$\min_{x \in C} f(x), \tag{1.2}$$

where C is a closed and convex subset of \mathbb{R}^m . It is also known that VIP can equivalently be rewritten as the following fixed point problem involving the metric projection P_C of H onto C :

$$z = P_C(z - \lambda Az), \tag{1.3}$$

where $\lambda > 0$.

There are various methods for solving VIP. One well-known method to solve the problem is the *extra-gradient method* (EGM), which was originally introduced by Antipin [4] for solving the saddle point problem and later extended by Korpelevich [32] to VIP in the finite dimensional Euclidean space. The EGM is of the following form:

$$\begin{cases} x_1 \in \mathbb{R}^m, \\ y_n = P_C(x_n - \lambda Ax_n), \\ x_{n+1} = P_C(x_n - \lambda Ay_n), \end{cases} \tag{1.4}$$

where A is a monotone and L -Lipschitz continuous mapping and $\lambda \in \left(0, \frac{1}{L}\right)$. The algorithm converges to a point of $VI(C, A)$ provided that $VI(C, A)$ is nonempty. In recent years, the EGM was widely extended to infinite dimensional Hilbert spaces by many authors (see, for example, [12, 25, 40, 50]). It is remarked that this method requires calculating two projections onto C and two evaluations of A in each iteration. However, this may be difficult when the feasible set C has complicated structures.

In order to overcome some disadvantages of the EGM, Censor et al. [12] replaced the second projection onto C of the EGM by a projection onto a half space, which significantly reduces the difficulty of calculations.

This method is called the *subgradient extragradient method* (SEGM) and it is of the following form:

$$\begin{cases} x_1 \in H, \\ y_n = P_C(x_n - \lambda Ax_n), \\ x_{n+1} = P_{T_n}(x_n - \lambda Ay_n), \\ T_n := \{x \in H : \langle x_n - \lambda Ax_n - y_n, x - y_n \rangle \leq 0\}, \end{cases} \quad (1.5)$$

where $\lambda \in \left(0, \frac{1}{L}\right)$. The weak convergence of SEGM was established provided that $VI(C, A)$ is nonempty.

Tseng [50] proposed a single projection method, which is some modifications of the EGM. This method is known as *Tseng's extragradient method* (TEGM) and it is of the following form:

$$\begin{cases} x_1 \in H, \\ y_n = P_C(x_n - \lambda Ax_n), \\ x_{n+1} = y_n - \lambda(Ay_n - Ax_n), \end{cases} \quad (1.6)$$

where A is monotone and L -Lipschitz continuous and $\lambda \in \left(0, \frac{1}{L}\right)$. He proved that this method converges weakly to a point of $VI(C, A)$. Note that this method only requires calculating one projection onto the feasible set C in each iteration, which is simple than the original EGM.

Another one of method to overcome the disadvantages of EGM is the *Popov's subgradient extragradient method* (PSEGM), which was introduced by Malitsky and Semenov [36]. They improved the EGM by combining the advantages of SEGM and the Popov's extragradient method introduced by Popov [40]. The method is of the following form:

$$\begin{cases} y_0, x_1, y_1 \in H, \\ y_{n+1} = P_C(x_{n+1} - \lambda Ay_n), \\ x_{n+1} = P_{T_n}(x_n - \lambda Ay_n), \\ T_n := \{x \in H : \langle x_n - \lambda Ay_{n-1} - y_n, x - y_n \rangle \leq 0\}. \end{cases} \quad (1.7)$$

It was proved that the PSEM converges weakly to point of $VI(C, A)$ provided $\lambda \in \left(0, \frac{1}{3L}\right)$. The advantages of the PSEGM are computing one projection onto the feasible set C and one evaluation of the mapping A in each iteration.

On the other hand, the inertial technique was introduced to speed up the convergence rate of algorithms by Polyak [39] in 1964. This technique originates from an implicit discretization method of the second-order dynamical systems (heavy ball with friction) in solving the smooth convex minimization problem. For approximating the null point of a maximal monotone operator A , Alvarez and Attouch [1] introduced the following *inertial proximal point algorithm* (IPPA):

$$\begin{cases} x_0, x_1 \in H, \\ y_n = x_n + \theta_n(x_n - x_{n-1}), \\ x_{n+1} = J_{\lambda_n}^A(y_n), \end{cases} \quad (1.8)$$

where $J_{\lambda_n}^A$ is the resolvent operator of A for $\lambda_n > 0$ and $\theta_n(x_n - x_{n-1})$ is called the *inertial extrapolation* with $\theta_n \in [0, 1)$. In recent years, the inertial technique has been applied to improves the performance of the algorithms for solving the VIP and related optimization problems.

Chbani and Riahi [13] proposed a new type of inertial term is known as *relaxed inertial*, whose structure is a convex combination of two iterates x_{n-1} and x_n , that is,

$$y_n = (1 - \theta_n)x_n + \theta_n x_{n-1} = x_n + \theta_n(x_{n-1} - x_n). \quad (1.9)$$

They also proposed two modifications of IPPA with relaxed inertial (1.9) for solving the equilibrium problem. Under suitable conditions, they obtained both weak and strong convergence of the algorithms to a solution of the problem.

In general, many algorithms based on the Halpern iteration algorithm [20], viscosity approximation algorithm [37], hybrid projection algorithm [38], and the shrinking projection algorithm [30] have been usually constructed to provide the strong convergence.

In 2019, Thong et al. [48] applied the inertial technique in (1.8) with the SEGM (1.5) for solving the monotone VIP in a real Hilbert space. They proposed two algorithms, which the first algorithm is based on the hybrid projection method and it is of the following form:

$$\left\{ \begin{array}{l} x_0, x_1 \in C, \\ u_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = P_C(u_n - \lambda A u_n), \\ z_n = \alpha_n u_n + (1 - \alpha_n)(y_n - \lambda(A y_n - A u_n)), \\ C_n = \{w \in H : \|z_n - w\| \leq \|u_n - w\|\}, \\ Q_n = \{w \in H : \langle w - x_n, x_1 - x_n \rangle \leq 0\}, \\ x_{n+1} = P_{C_n \cap Q_n}(x_1). \end{array} \right. \quad (1.10)$$

They also proposed another algorithm, which is based on the shrinking projection method. The algorithm is of the following form:

$$\left\{ \begin{array}{l} C_1 = C, \\ x_0, x_1 \in C, \\ u_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = P_C(u_n - \lambda A u_n), \\ z_n = \alpha_n u_n + (1 - \alpha_n)(y_n - \lambda(A y_n - A u_n)), \\ C_{n+1} = \{w \in C_n : \|z_n - w\| \leq \|u_n - w\|\}, \\ x_{n+1} = P_{C_{n+1}}(x_1), \end{array} \right. \quad (1.11)$$

where $\{\theta_n\}$ is a bounded real sequence and $\{\alpha_n\}$ is a sequence in $[0, 1)$ with $0 \leq \alpha_n \leq \alpha < 1$. They proved that the sequences $\{x_n\}$ generated by (1.10) and (1.11) converge strongly to a point in $VI(C, A)$ provided $\lambda \in \left(0, \frac{1}{L}\right)$. However, the hybrid (shrinking) projection method requires constructing the sets C_n and Q_n (C_{n+1}), and computing a projection of x_1 onto the set $C_n \cap Q_n$ (C_{n+1}), which make calculating at each iteration even more complicated.

It is worth noticing that the stepsizes of all methods as above are required a prior knowledge of the Lipschitz constant of the cost operators, which is very difficult to estimate. Even when it can be estimate, it is often too small which decelerate the rate of convergence of the methods.

In order to overcome this disadvantage, Thong and Vuong [47] proposed a modification of TEGM with a linesearch procedure for solving VIP involving a pseudomonotone, which is a generalization of the monotone

mappings. To be more precise, they proposed the following algorithm:

Algorithm A: TEGM for pseudomonotone VIP

(Step 0) Given $\gamma > 0$, $l \in (0, 1)$ and $\mu \in (0, 1)$. Let $x_1 \in H$ be arbitrary.

(Step 1) Compute

$$y_n = P_C(x_n - \lambda_n A x_n),$$

where $\lambda_n := \gamma l^{m_n}$ and m_n is the smallest non-negative integer m satisfying

$$\gamma l^m \|A x_n - A y\| \leq \mu \|x_n - y_n\|.$$

(Step 2) Compute

$$x_{n+1} = y_n - \lambda_n (A y_n - A x_n).$$

Update $n := n + 1$ go to Step 1.

They proved that if $A : H \rightarrow H$ is pseudomonotone and satisfies the following additional assumptions:

(A1) A is L -Lipschitz continuous;

(A2) A is sequentially weakly continuous,

then the sequence $\{x_n\}$ generated by Algorithm A converges weakly to a point of $VI(C, A)$.

Very recently, Khanh et al. [29] also proposed the following SEGM for solving pseudomonotone VIP in a Hilbert space:

Algorithm B: SEGM for pseudomonotone VIP

(Step 0) Given $\gamma > 0$, $l \in (0, 1)$ and $\mu \in (0, 1)$. Let $x_1 \in H$ be arbitrary.

(Step 1) Compute

$$y_n = P_C(x_n - \lambda_n A x_n),$$

where $\lambda_n := \gamma l^{m_n}$ and m_n is the smallest non-negative integer m satisfying

$$\gamma l^m \|A x_n - A y\| \leq \mu \|x_n - y_n\|.$$

(Step 2) Construct the half-space

$$T_n := \{x \in H : \langle x_n - \lambda_n A x_n - y_n, x - y_n \rangle \leq 0\}$$

and compute

$$x_{n+1} = P_{T_n}(x_n - \lambda_n A y_n).$$

Update $n := n + 1$ go to Step 1.

The weak convergence of the sequence $\{x_n\}$ generated by Algorithm B was also established under Assumptions (A1) and (A2). Note that Assumptions (A1) and (A2) are standard assumption, which often assumed in many recent works (see, for example, [8, 14, 29, 33, 46, 47, 49]). However, these assumptions may be stringent in practice.

Nowadays, the usages of the Bregman divergence arise frequently in many applied sciences such as in machine learning [2], clustering [7] and optimization [9]. Recently, Gibali [19] proposed a nice extension of PSEGM with the Bregman divergence, which is a generalization of the Squared Euclidean distance, Kullback–Leibler divergence, Itakura–Saito divergence, Squared Mahalanobis divergence and so on. This allows for applications to more solving VIP that the metric projection depends on the structure of generally feasible sets, which obtained from the various choices of functions.

Motivated and inspired by the above works, in this paper, we proposed two new relaxed inertial algorithms with Bregman divergence for solving pseudomonotone VIP, which provide strong convergence in Hilbert spaces. The first one, we combine SEGM and Halpern-type iteration, and in the second one, we combine TEGM and Halpern-type iteration. Some numerical experiments with various types of Bregman divergence are provided to show the effectiveness of the algorithms and some numerical experiments to the image deblurring problem are also presented.

The main contributions of this paper are highlighted as follows:

- Our algorithms are based on the Bregman divergence, which allow for more flexibility in solving the problem.
- It is know that any inertial algorithm based on Bregman divergence requires to use the hybrid projection method or the shrinking projection method, which ensures to obtain the strong convergence. In this situation, our algorithms add relaxed inertial term which are allowed to get the strong convergence theorem without the using two mentioned methods.
- The inertial parameter of our algorithms contains a computation procedure of the gradient of f at two iterates x_{n-1} and x_n . This approach is quite new and different from many recent works related to inertial algorithms for solving VIP (see, for example, [3, 14, 46, 49]).
- Our strong convergence theorems are established without assuming standard Assumptions (A1) and (A2), which are more relaxed than many recent works related to pseudomonotone VIP (see, for example, [8, 14, 29, 33, 46, 47]).

The rest of the paper is organized as follows: In section 2, we provide some useful preliminaries, which will be used to prove our main results. In section 3, we present our main results and finally, some numerical examples to illustrate the effectiveness of the obtained results are provided in Section 4.

2 Preliminaries

Let H be a real Hilbert space with the inner product $\langle \cdot, \cdot \rangle$ and the induced norm $\| \cdot \|$. The following notations are adopted throughout the paper:

- \mathbb{R} denotes the set of all real numbers;
- \mathbb{N} denotes the set of all positive integers;
- $x_n \rightharpoonup x$ denotes the weak convergence of the sequence $\{x_n\}$ to x ;
- $x_n \rightarrow x$ denotes the strong convergence of the sequence $\{x_n\}$ to x .

Let $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$ be the extend real-valued function. We denote the domain of f by $\text{dom}f$, that is,

$$\text{dom}f := \{x \in H : f(x) < +\infty\}.$$

A function f is said to be *proper* if $\text{dom}f \neq \emptyset$ and it is said to be *lower semicontinuous* if the set $\{x \in H : f(x) \leq r\}$ is closed for all $r \in \mathbb{R}$. A function f is said to be *convex* if for any $x, y \in \text{dom}f$ and $t \in [0, 1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \quad (2.1)$$

and it is also said to be *strictly convex* if the strict inequality holds in (2.1) for all $x, y \in \text{dom}f$ with $x \neq y$ and $t \in (0, 1)$.

Throughout this paper, we assume that $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper, semicontinuous and convex function. The *subdifferential* of f at x defined by

$$\partial f(x) = \{u \in H : f(y) - f(x) \geq \langle u, y - x \rangle \quad \forall y \in H\}.$$

The *conjugate function* of f is the function f^* on H defined by

$$f^*(x^*) = \sup_{x \in H} \{\langle x^*, x \rangle - f(x)\}$$

for all $(x, x^*) \in H \times H$. It is known that $x^* \in \partial f(x)$ is equivalent to $f(x) + f^*(x^*) = \langle x^*, x \rangle$ (see [45, Theorem 7.4.5]). We also know that if f is a proper, lower semicontinuous and convex function, then $f^* : H \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper, lower semicontinuous and convex function (see [45, Theorem 7.4.2]).

A function f is said to be *Gâteaux differentiable* at $x \in \text{int}(\text{dom}f)$ if there is $\nabla f \in H$ such that

$$\lim_{t \rightarrow 0} \frac{f(x + ty) - f(x)}{t} = \langle \nabla f(x), y \rangle \quad (2.2)$$

for each $y \in H$. When the limit (2.2) is attained uniformly for $\|y\| = 1$, we say that f is *Fréchet differentiable* at x . A function f is said to be *Gâteaux differentiable (Fréchet differentiable)* if it is Gâteaux differentiable everywhere (Fréchet differentiable everywhere) and f is said to be *uniformly Fréchet differentiable* (or, equivalently f is uniformly smooth) on a subset C of H if the limit (2.2) is attained uniformly for $x \in C$ and $\|y\| = 1$. We also know that if f is uniformly Fréchet differentiable and bounded on bounded subsets of H , then ∇f is uniformly continuous on bounded subsets of H (see [42, Proposition 2]).

Definition 2.1. A function $f : H \rightarrow \mathbb{R}$ is said to be:

- (1) *uniformly convex* with modulus ϕ if

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - t(1-t)\phi(\|x - y\|),$$

for all $x, y \in \text{dom}f$ and $t \in (0, 1)$, where ϕ is an increasing function vanishing only at 0;

- (2) *strongly convex* with a constant $\sigma > 0$ if

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) - \frac{\sigma}{2}t(1-t)\|x - y\|^2 \quad (2.3)$$

for all $x, y \in \text{dom}f$ and $t \in (0, 1)$.

We know that f is uniformly convex if and only if f^* is Fréchet differentiable and ∇f^* is uniformly continuous (see [52, Theorem 3.5.10]). Obviously, f is strongly convex with a constant σ if and only if it is uniformly convex with modulus $\phi(s) = \frac{\sigma}{2}s^2$ and it is also equivalent to the following inequality (see [6, Theorem 5.24]):

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

for all $x \in \text{dom}f$ and $y \in \text{int}(\text{dom}f)$.

A function f is said to be *Legendre* if f is essentially smooth and essentially strictly convex in the sense of [43, Section 26]. If f is additionally assumed to be Gâteaux differentiable, then the bifunction $D_f : \text{dom}f \times \text{int}(\text{dom}f) \rightarrow [0, +\infty)$ defined by

$$D_f(x, y) := f(x) - f(y) - \langle \nabla f(y), x - y \rangle$$

is called the *Bregman divergence (distance)* with respect to f [9].

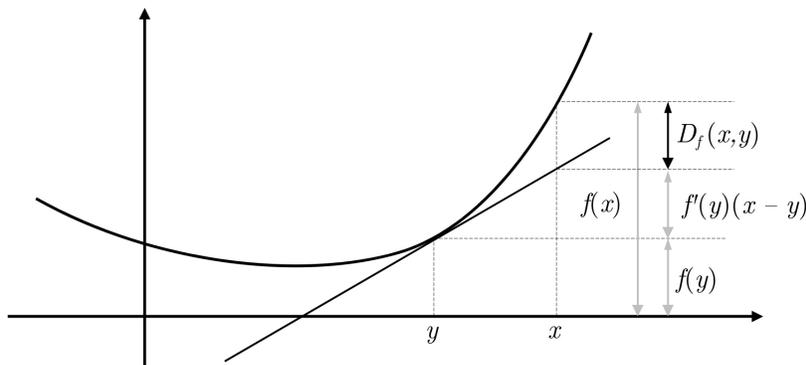


Figure 1: Bregman divergence with respect to f

In mathematics and other disciplines, the Bregman divergence is one kind of measurement of the difference between two points (or distribution in statistics) on a differentiable convex function of Legendre type. Note that the Bregman divergence is not a usual metric because it is asymmetric and does not satisfy the triangle inequality. The Bregman divergence with respect to various types of f can be seen as follows [5, 23]:

Example 2.2. Let $x = (x_1, x_2, \dots, x_m)^T$ and $y = (y_1, y_2, \dots, y_m)^T$ be two points in \mathbb{R}^m .

(i) *The Kullback–Leibler divergence*

$$D_f^{KL}(x, y) = \sum_{i=1}^m \left(x_i \ln \left(\frac{x_i}{y_i} \right) + y_i - x_i \right)$$

generated by the function $f^{KL}(x) = \sum_{i=1}^m x_i \ln x_i$ with its domain $\text{dom}f^{KL} = \{x \in \mathbb{R}^m : x_i > 0, i = 1, 2, \dots, m\}$ and gradient $\nabla f^{KL}(x) = (1 + \ln(x_1), 1 + \ln(x_2), \dots, 1 + \ln(x_m))^T$. In statistics, the Kullback–Leibler divergence is used to measure the difference between two probability distributions.

(ii) *The Itakura–Saito divergence*

$$D_f^{IS}(x, y) = \sum_{i=1}^m \left(\frac{x_i}{y_i} - \ln \left(\frac{x_i}{y_i} \right) - 1 \right)$$

generated by the function $f^{IS}(x) = -\sum_{i=1}^m \ln x_i$ with its domain $\text{dom}f^{IS} = \{x \in \mathbb{R}^m : x_i > 0, i = 1, 2, \dots, m\}$ and gradient $\nabla f^{IS}(x) = -\left(\frac{1}{x_1}, \frac{1}{x_2}, \dots, \frac{1}{x_m}\right)^T$. In signal processing, the Itakura–Saito divergence is used to measure the difference between original spectrum and approximation of that spectrum.

(iii) *The Bregman divergence*

$$D_f^{FD}(x, y) = \sum_{i=1}^m \left(x_i \ln \left(\frac{x_i}{y_i} \right) + (1 - x_i) \ln \left(\frac{1 - x_i}{1 - y_i} \right) \right).$$

generated by the Fermi–Dirac entropy function $f^{FD}(x) = \sum_{i=1}^m \left(x_i \ln x_i + (1 - x_i) \ln(1 - x_i) \right)$ with its domain $\text{dom}f^{FD} = \{x \in \mathbb{R}^m : 0 < x_i < 1, i = 1, 2, \dots, m\}$ and gradient

$$\nabla f^{FD}(x) = \left(\ln \left(\frac{x_1}{1 - x_1} \right), \ln \left(\frac{x_2}{1 - x_2} \right), \dots, \ln \left(\frac{x_m}{1 - x_m} \right) \right)^T.$$

(iv) *The Squared Mahalanobis divergence*

$$D_f^{SM}(x, y) = \frac{1}{2}(x - y)^T Q(x - y)$$

generated by the function $f^{SM}(x) = \frac{1}{2}x^T Qx$ with its domain $\text{dom}f^{SM} = \mathbb{R}^m$ and gradient $\nabla f^{SM}(x) = Qx$, where $Q = \text{diag}(1, 2, \dots, m)$. The Squared Mahalanobis divergence is used to measure the difference between standard deviation and mean in a normal distribution.

(v) *The Squared Euclidean divergence*

$$D_f^{SE}(x, y) = \frac{1}{2}\|x - y\|^2$$

generated by the function $f^{SE}(x) = \frac{1}{2}\|x\|^2$ with its domain $\text{dom}f^{SE} = \mathbb{R}^m$ and gradient $\nabla f^{SE}(x) = x$.

We know that if f is strongly convex, then for any $x \in \text{dom}f$ and $y \in \text{int}(\text{dom}f)$,

$$D_f(x, y) \geq \frac{\sigma}{2}\|x - y\|^2. \quad (2.4)$$

We also know that if f is twice continuously differentiable, then it is strongly convex if and only if $\nabla^2 f(x) \succeq \sigma I$ for all $x \in \text{dom}f$, where $\nabla^2 f(x)$ is the Hessian matrix at x , I is the identity matrix and the notation \succeq means that $\nabla^2 f(x) - \sigma I$ is positive semi-definite (see [6]).

The following important properties follow from the definition of Bregman divergence:

(i) (two-point identity) for any $x, y \in \text{int}(\text{dom}f)$,

$$D_f(x, y) + D_f(y, x) = \langle \nabla f(x) - \nabla f(y), x - y \rangle; \quad (2.5)$$

(ii) (three-point identity) for any $x \in \text{dom}f$ and $y, z \in \text{int}(\text{dom}f)$,

$$D_f(x, y) = D_f(x, z) - D_f(y, z) + \langle \nabla f(z) - \nabla f(y), x - y \rangle. \quad (2.6)$$

The *Bregman projection* with respect to f of $y \in \text{int}(\text{dom}f)$ is the unique point in C , denoted by Π_C^f , defined by

$$\Pi_C^f(y) := \arg \min\{D_f(x, y) : x \in C\}.$$

In particular, if $f(x) = \frac{1}{2}\|x\|^2$, then Π_C^f reduces to the metric projection P_C .

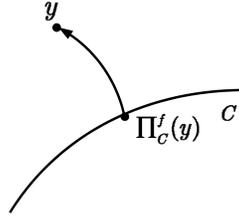


Figure 2: Bregman projection with respect to f

It is known that Π_C^f is continuous (see [5, Theorem 4.3]). Moreover, Π_C^f has the following properties (see [10]): for each $y \in H$,

$$\langle \nabla f(\Pi_C^f(y)) - \nabla f(y), x - \Pi_C^f(y) \rangle \geq 0 \quad \forall x \in C \quad (2.7)$$

and

$$D_f(x, \Pi_C^f(y)) + D_f(\Pi_C^f(y), y) \leq D_f(x, y) \quad \forall x \in C. \quad (2.8)$$

The property (2.8) is also called the *generalized Pythagorean theorem*.

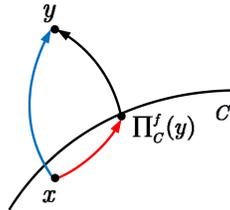


Figure 3: Generalized Pythagorean theorem

Let $f : H \rightarrow \mathbb{R}$ be a Legendre function. Let $V_f : H \times H \rightarrow [0, +\infty)$ associated with f , be defined by

$$V_f(x, x^*) = f(x) - \langle x^*, x \rangle + f^*(x^*)$$

for all $(x, x^*) \in H \times H$. We know the following properties:

- (i) V_f is nonnegative and convex in the second variable.

(ii) $V_f(x, x^*) = D_f(x, \nabla f^*(x^*))$ for all $(x, x^*) \in H \times H$.

(iii) $V_f(x, x^*) + \langle y^*, \nabla f^*(x^*) - x \rangle \leq V_f(x^* + y^*, x)$ for all $(x, x^*) \in H \times H$ and $y^* \in H$.

Since V_f is convex in the second variable. Thus for all $z \in H$,

$$D_f\left(z, \nabla f^*\left(\sum_{i=1}^N t_i \nabla f(x_i)\right)\right) \leq \sum_{i=1}^N t_i D_f(z, x_i), \quad (2.9)$$

where $\{x_i\}_{i=1}^N \subset H$ and $\{t_i\}_{i=1}^N$ with $\sum_{i=1}^N t_i = 1$.

Definition 2.3. A mapping $A : C \rightarrow H$ is said to be:

(1) *monotone* if $\langle Ax - Ay, x - y \rangle \geq 0$ for all $x, y \in C$;

(2) *pseudomonotone* if $\langle Ax, y - x \rangle \geq 0$, we have $\langle Ay, y - x \rangle \geq 0$ for all $x, y \in C$;

(3) *L-Lipschitz continuous* if there exists a constant $L > 0$ such that $\|Ax - Ay\| \leq L\|x - y\|$ for all $x, y \in C$;

(4) *sequentially weakly continuous* on C if for each sequence $\{x_n\} \subset C$ such that $x_n \rightharpoonup x$, we have $Ax_n \rightharpoonup Ax$.

Remark 2.4. It is observe that every monotone mapping is a pseudomonotone mapping but converse is not true. The example of a pseudomonotone mapping but not necessarily monotone can be found in [28].

Lemma 2.5. ([16]) *Let C be a nonempty, closed and convex subset of H and A be a pseudomonotone and continuous of C into H . Then z is a solution of VIP if and only if*

$$\langle Ax, x - z \rangle \geq 0 \quad \forall x \in C.$$

Lemma 2.6. *For any $a, b \in \mathbb{R}$ and $\epsilon > 0$. Then the following inequality holds:*

$$2ab \leq \frac{a^2}{\epsilon} + \epsilon b^2.$$

Proof. Since $0 \leq \left(\frac{1}{\sqrt{\epsilon}}a - \sqrt{\epsilon}b\right)^2 = \frac{a^2}{\epsilon} - 2ab + \epsilon b^2$, we have $2ab \leq \frac{a^2}{\epsilon} + \epsilon b^2$. □

Following the proof line as in Lemma 2 of [17], we obtain the following result:

Lemma 2.7. *For $x \in H$ and $\alpha \geq \beta > 0$, the following inequalities hold:*

$$\left\| \frac{x - \Pi_C^f \nabla f^*(\nabla f(x) - \alpha Ax)}{\alpha} \right\| \leq \left\| \frac{x - \Pi_C^f \nabla f^*(\nabla f(x) - \beta Ax)}{\beta} \right\|$$

and

$$\|x - \Pi_C^f \nabla f^*(\nabla f(x) - \beta Ax)\| \leq \|x - \Pi_C^f \nabla f^*(\nabla f(x) - \alpha Ax)\|.$$

Lemma 2.8. ([24]) *Let H_1 and H_2 be two real Hilbert spaces. Suppose $A : H_1 \rightarrow H_2$ is uniformly continuous on bounded subsets of H_1 and M is a bounded subset of H_1 . Then $A(M)$ is bounded.*

The following lemmas are useful in our proofs.

Lemma 2.9. ([35]) *Let $\{a_n\}$ be a sequence of real numbers such that there exists a subsequence $\{n_i\}$ of $\{n\}$ such that $a_{n_i} < a_{n_i+1}$ for all $i \in \mathbb{N}$. Then there exists an increasing sequence $\{m_k\} \subset \mathbb{N}$ such that $\lim_{k \rightarrow \infty} m_k = \infty$ and the following properties are satisfied by all (sufficiently large) numbers $k \in \mathbb{N}$:*

$$a_{m_k} \leq a_{m_k+1} \quad \text{and} \quad a_k \leq a_{m_k+1}.$$

In fact, $m_k := \max\{j \leq k : a_j \leq a_{j+1}\}$.

Lemma 2.10. ([13]) *Let $\{a_n\}$, $\{\gamma_n\}$, $\{\delta_n\}$ and $\{t_n\}$ be sequences of nonnegative real numbers such that $\{\gamma_n\} \subset [0, \frac{1}{2}]$, $\limsup_{n \rightarrow \infty} s_n \leq 0$, $\sum_{n=n_0}^{\infty} \delta_n < \infty$, $\sum_{n=n_0}^{\infty} t_n = \infty$ and for each $n \geq n_0$ (where n_0 is a positive integer)*

$$a_{n+1} \leq (1 - t_n - \gamma_n)a_n + \gamma_n a_{n-1} + t_n s_n + \delta_n.$$

Then $\lim_{n \rightarrow \infty} a_n = 0$.

3 Algorithms and their convergence

In this section, we propose two relaxed inertial algorithms for solving pseudomonotone variational inequalities. In order to establish the convergence of the algorithms, the following assumptions are need:

Assumption 1 The feasible set C is a closed and convex subset of a real Hilbert space H .

Assumption 2 The function $f : H \rightarrow \mathbb{R}$ is σ -strongly convex, Legendre which is bounded and uniformly Fréchet differentiable on bounded subsets of H .

Assumption 3 The mapping $A : H \rightarrow H$ is pseudomonotone and uniformly continuous which satisfies the following condition:

$$\text{if for each } \{q_n\} \subset H \text{ such that } q_n \rightharpoonup q \text{ and } \liminf_{n \rightarrow \infty} \|Aq_n\| = 0 \text{ implies } Aq = 0. \quad (3.1)$$

Assumption 4 The solution set of VIP is nonempty, that is, $VI(C, A) \neq \emptyset$.

Assumption 5 The positive sequence $\{\xi_n\}$ satisfies $\xi_n = o(\alpha_n)$, that is, $\lim_{n \rightarrow \infty} \frac{\xi_n}{\alpha_n} = 0$, where $\{\alpha_n\} \subset (0, 1)$ such that $\lim_{n \rightarrow \infty} \alpha_n = 0$ and $\sum_{n=1}^{\infty} \alpha_n = \infty$.

Remark 3.1. We remark on Assumption 3 in the following aspects:

(1) When H is a finite-dimensional Hilbert space, it suffices to assume that the mapping A is continuous pseudomonotone and it is not necessary to assume (3.1).

(2) The uniform continuity is weaker than the Lipschitz continuity. Clearly, if A is Lipschitz continuous, then A is uniformly continuous but the converse is not true. Let $A : [0, \infty) \rightarrow [0, \infty)$ be a mapping define by $Ax = \sqrt{x}$. For each $\epsilon > 0$, let $\delta = \epsilon^2$ and $|x - y| < \delta$, where $x, y \geq 0$. To estimate $|Ax - Ay|$, we consider possible two cases of x, y . In the case $x, y \in [0, \delta)$. Using the fact that A is strictly increasing, we have $|Ax - Ay| < A(\delta) - A(0) < \sqrt{\delta} = \epsilon$. Otherwise, in the case $x \notin [0, \delta)$ or $y \notin [0, \delta)$, we have $\max\{x, y\} \geq \delta$. It follows that

$$|Ax - Ay| = |\sqrt{x} - \sqrt{y}| = \left| \frac{x - y}{\sqrt{x} + \sqrt{y}} \right| \leq \frac{|x - y|}{\sqrt{\max\{x, y\}}} < \frac{\delta}{\sqrt{\delta}} = \sqrt{\delta} = \epsilon.$$

Thus A is uniformly continuous. For each $n \in \mathbb{N}$, we have $\left| A\left(\frac{1}{n}\right) - A(0) \right| = \sqrt{\frac{1}{n}} = \sqrt{n} \left| \frac{1}{n} - 0 \right|$. Thus A is not Lipschitz continuous.

(3) Note that (3.1) is weaker than the sequential weak continuity of the mapping A . Indeed, let $A : \ell_2 \rightarrow \ell_2$ be a mapping define by $Ax = x\|x\|$ for all $x \in \ell_2$. Let $\{q_n\} \subset \ell_2$ such that $q_n \rightharpoonup q$ and $\liminf_{n \rightarrow \infty} \|Aq_n\| = 0$. By the weak lower semicontinuity of norm, we have $\|q\| \leq \liminf_{n \rightarrow \infty} \|q\|$. It follows that

$$\|Aq\| = \|q\|^2 \leq \liminf_{n \rightarrow \infty} \|q\|^2 = \liminf_{n \rightarrow \infty} \|Aq_n\| = 0.$$

This implies that $\|Aq\| = 0$. To show that A is not sequentially weakly continuous, choose $q_n = e_n + e_1$, where $\{e_n\}$ is a standard basis of ℓ_2 , that is, $e_n = (0, 0, \dots, 1, \dots)$ with 1 at the n -th position. It is clear that $q_n \rightharpoonup e_1$ and $Aq_n = A(e_n + e_1) = (e_n + e_1)\|e_n + e_1\| \rightharpoonup \sqrt{2}e_1$ but $Ae_1 = e_1\|e_1\| = e_1$. Hence A is not sequentially weakly continuous.

(4) If A is monotone, then (3.1) can be removed.

Now, we propose the first algorithm, which combines the Halpern-type iteration and the subgradient extragradient method. The algorithm is shown as below.

Algorithm 1: Relaxed inertial subgradient extragradient algorithm for VIP

(Step 0) Given $\theta \in (0, 1/2]$, $\gamma > 0$, $l \in (0, 1)$ and $\mu \in (0, \sigma)$, where σ is a constant given by (2.4). Let $x_0, x_1, u \in H$ be arbitrary.

(Step 1) Given the current iterates x_{n-1} and x_n ($n \geq 1$). Choose θ_n such that $0 \leq \theta_n \leq \bar{\theta}_n$, where

$$\bar{\theta}_n = \begin{cases} \min \left\{ \frac{\xi_n}{\|\nabla f(x_{n-1}) - \nabla f(x_n)\|}, \theta \right\} & \text{if } x_{n-1} \neq x_n, \\ \theta & \text{otherwise.} \end{cases} \quad (3.2)$$

Set $u_n = \nabla f^*(\nabla f(x_n) + \theta_n(\nabla f(x_{n-1}) - \nabla f(x_n)))$ and compute

$$y_n = \Pi_C^f \nabla f^*(\nabla f(u_n) - \lambda_n A u_n),$$

where $\lambda_n = \gamma l^{m_n}$, with m_n is the smallest nonnegative integer m satisfying

$$\gamma l^m \|A u_n - A y_n\| \leq \mu \|u_n - y_n\|. \quad (3.3)$$

If $u_n = y_n$ or $A y_n = 0$, then stop and y_n is a solution of VIP. Otherwise, go to Step 2.

(Step 2) Construct the half-space

$$T_n := \{x \in H : \langle \nabla f(u_n) - \lambda_n A u_n - \nabla f(y_n), x - y_n \rangle \leq 0\}$$

and compute

$$z_n = \Pi_{T_n}^f \nabla f^*(\nabla f(u_n) - \lambda_n A y_n).$$

(Step 3) Compute

$$x_{n+1} = \nabla f^*(\alpha_n \nabla f(u) + (1 - \alpha_n) \nabla f(z_n)).$$

Update $n := n + 1$ go to Step 1.

Remark 3.2. If $f(x) = \frac{1}{2}\|x\|^2$ and $\theta_n = 0$, then Algorithm 1 reduces to the following one:

$$\begin{cases} y_n = P_C(x_n - \lambda_n A x_n), \\ z_n = P_{T_n}(x_n - \lambda_n A y_n), \\ x_{n+1} = \alpha_n u + (1 - \alpha_n) z_n, \end{cases} \quad (3.4)$$

where $\lambda_n = \gamma l^{m_n}$, with m_n is the smallest nonnegative integer m satisfying

$$\gamma l^m \|A x_n - A y_n\| \leq \mu \|x_n - y_n\| \quad (3.5)$$

and

$$T_n := \{x \in H : \langle x_n - \lambda_n A x_n - y_n, x - y_n \rangle \leq 0\}.$$

The Algorithm (3.4) is a modification of SEGМ without the relaxed inertial term for pseudomonotone VIP with a non-Lipschitz mapping.

(2) From (3.2), it is easy to see that $\theta_n \|\nabla f(x_{n-1}) - \nabla f(x_n)\| \leq \xi_n$ for all $n \in \mathbb{N}$. Since $\lim_{n \rightarrow \infty} \frac{\xi_n}{\alpha_n} = 0$,

it follows that

$$\lim_{n \rightarrow \infty} \frac{\theta_n}{\alpha_n} \|\nabla f(x_{n-1}) - \nabla f(x_n)\| \leq \lim_{n \rightarrow \infty} \frac{\xi_n}{\alpha_n} = 0.$$

Lemma 3.3. *The Armijo-line search rule (3.3) is well-defined.*

Proof. If $u_n \in VI(C, A)$, then $u_n = \Pi_C^f \nabla f^*(\nabla f(u_n) - \gamma A u_n)$ and $m_n = 0$. In this case, we consider $u_n \notin VI(C, A)$ and assume that the contrary for all m . Thus we have

$$\gamma l^m \|A u_n - A(\Pi_C^f \nabla f^*(\nabla f(u_n) - \gamma l^m A u_n))\| > \mu \|u_n - \Pi_C^f \nabla f^*(\nabla f(u_n) - \gamma l^m A u_n)\|.$$

This implies that

$$\|A u_n - A(\Pi_C^f \nabla f^*(\nabla f(u_n) - \gamma l^m A u_n))\| > \mu \frac{\|u_n - \Pi_C^f \nabla f^*(\nabla f(u_n) - \gamma l^m A u_n)\|}{\gamma l^m}. \quad (3.6)$$

We now consider two possible cases of u_n , that is, $u_n \in C$ and $u_n \notin C$. If $u_n \in C$, then $u_n = \Pi_C^f(u_n)$. By the continuity of Π_C^f , we have

$$\lim_{m \rightarrow \infty} \|u_n - \Pi_C^f \nabla f^*(\nabla f(u_n) - \gamma l^m A u_n)\| = 0$$

and by the uniform continuity of A , we have

$$\lim_{m \rightarrow \infty} \|A u_n - A(\Pi_C^f \nabla f^*(\nabla f(u_n) - \gamma l^m A u_n))\| = 0. \quad (3.7)$$

Combining (3.6) and (3.7), we get

$$\lim_{m \rightarrow \infty} \frac{\|u_n - \Pi_C^f \nabla f^*(\nabla f(u_n) - \gamma l^m A u_n)\|}{\gamma l^m} = 0.$$

Also, by the uniform continuity of ∇f , we have

$$\lim_{m \rightarrow \infty} \frac{\|\nabla f(u_n) - \nabla f(\Pi_C^f \nabla f^*(\nabla f(u_n) - \gamma l^m A u_n))\|}{\gamma l^m} = 0. \quad (3.8)$$

Let $v_n = \Pi_C^f \nabla f^*(\nabla f(u_n) - \gamma l^m A u_n)$. From (2.7), we have

$$\langle \nabla f(v_n) - \nabla f(u_n) + \gamma l^m A u_n, x - v_n \rangle \geq 0 \quad \forall x \in C.$$

This implies that

$$\left\langle \frac{\nabla f(v_n) - \nabla f(u_n)}{\gamma l^m}, x - v_n \right\rangle + \langle A u_n, x - v_n \rangle \geq 0 \quad \forall x \in C. \quad (3.9)$$

Letting $m \rightarrow \infty$ in (3.9) and by (3.8), we have

$$\langle A u_n, x - u_n \rangle \geq 0 \quad \forall x \in C.$$

That is $u_n \in VI(C, A)$. This is a contradiction. On the other hand, if $u_n \notin C$, then

$$\lim_{m \rightarrow \infty} \|u_n - \Pi_C^f \nabla f^*(\nabla f(u_n) - \gamma l^m A u_n)\| = \lim_{m \rightarrow \infty} \|u_n - \Pi_C^f(u_n)\| > 0 \quad (3.10)$$

and

$$\lim_{m \rightarrow \infty} \gamma l^m \|A u_n - A(\Pi_C^f \nabla f^*(\nabla f(u_n) - \gamma l^m A u_n))\| = 0. \quad (3.11)$$

Combining (3.6), (3.10) and (3.11), we get a contradiction. \square

Remark 3.4. (1) We note that the pseudomonotonicity of the mapping is not used in the proof of Lemma 3.3.

(2) It is obvious that $0 < \lambda_n \leq \gamma$ for all $n \in \mathbb{N}$.

Lemma 3.5. *Suppose that Assumptions 1-4 are satisfied. Let $\{u_n\}$ generated by Algorithm 1. If there exists a subsequence $\{u_{n_k}\}$ of $\{u_n\}$ such that $\{u_{n_k}\}$ converges weakly to $v \in H$ and $\lim_{k \rightarrow \infty} \|u_{n_k} - y_{n_k}\| = 0$, then $v \in VI(C, A)$.*

Proof. Let $\{u_{n_k}\}$ be a subsequence of $\{u_n\}$ such that $u_{n_k} \rightharpoonup v \in H$. Since $\lim_{k \rightarrow \infty} \|u_{n_k} - y_{n_k}\| = 0$ and $\{y_{n_k}\} \subset C$, we have $y_{n_k} \rightarrow v \in C$. By the definition of y_{n_k} and (2.7), we have

$$\langle \nabla f(y_{n_k}) - \nabla f(u_{n_k}) + \lambda_{n_k} Au_{n_k}, x - y_{n_k} \rangle \geq 0 \quad \forall x \in C,$$

which implies that

$$\lambda_{n_k} \langle Au_{n_k}, x - y_{n_k} \rangle \geq \langle \nabla f(u_{n_k}) - \nabla f(y_{n_k}), x - y_{n_k} \rangle \quad \forall x \in C.$$

Hence

$$\langle Au_{n_k}, x - u_{n_k} \rangle \geq \left\langle \frac{\nabla f(u_{n_k}) - \nabla f(y_{n_k})}{\lambda_{n_k}}, x - y_{n_k} \right\rangle + \langle Au_{n_k}, y_{n_k} - u_{n_k} \rangle \quad \forall x \in C. \quad (3.12)$$

We now consider two possible cases. In the first case, we assume that $\liminf_{k \rightarrow \infty} \lambda_{n_k} > 0$. By the weakly convergent of $\{u_{n_k}\}$, we have $\{u_{n_k}\}$ is bounded and since A is uniformly continuous, it follows from Lemma 2.8 that $\{Au_{n_k}\}$ is bounded. Moreover, ∇f is uniformly continuous, we have $\lim_{k \rightarrow \infty} \|\nabla f(u_{n_k}) - \nabla f(y_{n_k})\| = 0$. Taking limit inferior as $k \rightarrow \infty$ in (3.12), we have

$$\liminf_{k \rightarrow \infty} \langle Au_{n_k}, x - u_{n_k} \rangle \geq 0 \quad \forall x \in C.$$

In the second case, we assume that $\liminf_{k \rightarrow \infty} \lambda_{n_k} = 0$. Let $w_{n_k} = \Pi_C^f \nabla f^*(\nabla f(u_{n_k}) - \lambda_{n_k} l^{-1} Au_{n_k})$. Clearly, that $\lambda_{n_k} l^{-1} > \lambda_{n_k}$. Then from Lemma 2.7, we have

$$\|u_{n_k} - w_{n_k}\| \leq \frac{1}{l} \|u_{n_k} - y_{n_k}\| \rightarrow 0 \text{ as } k \rightarrow \infty.$$

Moreover, we have

$$\|Au_{n_k} - Aw_{n_k}\| \rightarrow 0 \text{ as } k \rightarrow \infty. \quad (3.13)$$

By the Armijo linesearch rule (3.3), we have

$$\lambda_{n_k} l^{-1} \|Au_{n_k} - Aw_{n_k}\| > \mu \|u_{n_k} - w_{n_k}\|.$$

That is,

$$\frac{1}{\mu} \|Au_{n_k} - Aw_{n_k}\| > \frac{\|u_{n_k} - w_{n_k}\|}{\lambda_{n_k} l^{-1}}. \quad (3.14)$$

Combining (3.13) and (3.14), we get

$$\lim_{k \rightarrow \infty} \frac{\|u_{n_k} - w_{n_k}\|}{\lambda_{n_k} l^{-1}} = 0$$

and hence

$$\lim_{k \rightarrow \infty} \frac{\|\nabla f(u_{n_k}) - \nabla f(w_{n_k})\|}{\lambda_{n_k} l^{-1}} = 0.$$

Moreover, we have

$$\langle \nabla f(w_{n_k}) - \nabla f(u_{n_k}) + \lambda_{n_k} l^{-1} Au_{n_k}, x - w_{n_k} \rangle \geq 0 \quad \forall x \in C.$$

It follows that

$$\langle Au_{n_k}, x - u_{n_k} \rangle \geq \left\langle \frac{\nabla f(u_{n_k}) - \nabla f(w_{n_k})}{\lambda_{n_k} l^{-1}}, x - w_{n_k} \right\rangle + \langle Au_{n_k}, w_{n_k} - u_{n_k} \rangle \quad \forall x \in C. \quad (3.15)$$

Taking limit inferior as $k \rightarrow \infty$ in (3.15), we have

$$\liminf_{k \rightarrow \infty} \langle Au_{n_k}, x - u_{n_k} \rangle \geq 0 \quad \forall x \in C.$$

On the other hand, we observe that

$$\langle Ay_{n_k}, x - y_{n_k} \rangle = \langle Ay_{n_k} - Au_{n_k}, x - u_{n_k} \rangle + \langle Au_{n_k}, x - u_{n_k} \rangle + \langle Ay_{n_k}, u_{n_k} - y_{n_k} \rangle.$$

Again, since A is uniformly continuous, $\lim_{k \rightarrow \infty} \|u_{n_k} - y_{n_k}\| = 0$ and $\liminf_{k \rightarrow \infty} \langle Au_{n_k}, x - u_{n_k} \rangle \geq 0$, we have

$$\liminf_{k \rightarrow \infty} \langle Ay_{n_k}, x - y_{n_k} \rangle \geq 0.$$

Next, we show that $v \in VI(C, A)$. In order to show this, we consider two possible cases as follows.

Case 1. Suppose that $\liminf_{k \rightarrow \infty} \|Au_{n_k}\| = 0$. Since $u_{n_k} \rightarrow v$ and by (3.1), we have $Av = 0$. Hence $v \in VI(C, A)$.

Case 2. Suppose that $\liminf_{k \rightarrow \infty} \|Au_{n_k}\| > 0$. Let $\{\epsilon_k\}$ be a positive real sequence such that $\epsilon_k \rightarrow 0$ as $k \rightarrow \infty$. For each ϵ_k , we denote by N_k the smallest positive integer such that

$$\langle \widehat{Ay_{n_k}}, x - y_{n_k} \rangle + \epsilon_k \geq 0 \quad \forall k \geq N_k, \quad (3.16)$$

where $\widehat{Ay_{n_k}}$ is the unique vector of Ay_{n_k} , that is, $\widehat{Ay_{n_k}} = \frac{Ay_{n_k}}{\|Ay_{n_k}\|}$. Since, for each $k \geq 1$, $Ay_{n_k} \neq 0$ (otherwise, $y_{n_k} \in VI(C, A)$), it follows from (3.16) that

$$\langle Ay_{n_k}, x - y_{n_k} \rangle + \|Ay_{n_k}\| \epsilon_k \geq 0 \quad \forall k \geq N_k, \quad (3.17)$$

Setting $v_{n_k} = \frac{Ay_{n_k}}{\|Ay_{n_k}\|^2}$, we have $\langle Ay_{n_k}, v_{n_k} \rangle = 1$. Thus we can write (3.17) as

$$\langle Ay_{n_k}, x + \epsilon_k \|Ay_{n_k}\| v_{n_k} - y_{n_k} \rangle \geq 0 \quad \forall k \geq N_k.$$

The pseudomonotonicity of A implies that

$$\langle A(x + \epsilon_k \|Ay_{n_k}\| v_{n_k}), x + \epsilon_k \|Ay_{n_k}\| v_{n_k} - y_{n_k} \rangle \geq 0 \quad \forall k \geq N_k. \quad (3.18)$$

Since $\epsilon_k \rightarrow 0$, $\{\|Ay_{n_k}\| v_{n_k}\}$ is bounded. By the continuity of A , we have

$$\langle Ax, x - v \rangle \geq 0 \quad \forall x \in C.$$

By Lemma 2.5, we get $v \in VI(C, A)$ □

Lemma 3.6. *Suppose that Assumptions 1-4 are satisfied. Then the sequence $\{x_n\}$ generated by Algorithm 1 satisfies the following inequality:*

$$D_f(p, z_n) \leq D_f(p, u_n) - \left(1 - \frac{\mu}{\sigma}\right) D_f(y_n, u_n) - \left(1 - \frac{\mu}{\sigma}\right) D_f(z_n, y_n) \quad \forall p \in VI(C, A).$$

In particular, if $\mu \in (0, \sigma)$, then $D_f(p, z_n) \leq D_f(p, u_n)$.

Proof. Let $p \in VI(C, A)$. By the definition of the Bregman divergence, we have

$$\begin{aligned} D_f(p, z_n) &= D_f(p, \Pi_{T_n}^f \nabla f^*(\nabla f(u_n) - \lambda_n A y_n)) \\ &\leq D_f(p, \nabla f^*(\nabla f(u_n) - \lambda_n A y_n)) - D_f(z_n, \nabla f^*(\nabla f(u_n) - \lambda_n A y_n)) \\ &= V_f(p, \nabla f(u_n) - \lambda_n A y_n) - V_f(z_n, \nabla f(u_n) - \lambda_n A y_n) \\ &= f(p) - \langle \nabla f(u_n) - \lambda_n A y_n, p \rangle + f^*(\nabla f(u_n) - \lambda_n A y_n) - f(z_n) + \langle \nabla f(u_n) - \lambda_n A y_n, z_n \rangle \\ &\quad - f^*(\nabla f(u_n) - \lambda_n A y_n) \\ &= f(p) - \langle \nabla f(u_n), p \rangle + \lambda_n \langle A y_n, p \rangle - f(z_n) + \langle f(u_n), z_n \rangle - \lambda_n \langle A y_n, z_n \rangle \\ &= f(p) - \langle \nabla f(u_n), p \rangle + f(u_n) - f(z_n) + \langle \nabla f(u_n), z_n \rangle - f(u_n) + \lambda_n \langle A y_n, p \rangle - \lambda_n \langle A y_n, z_n \rangle \\ &= D_f(p, u_n) - D_f(z_n, u_n) - \lambda_n \langle A y_n, z_n - p \rangle. \end{aligned} \quad (3.19)$$

Using the fact that $\langle A p, y_n - p \rangle \geq 0$ and the pseudomonotonicity of A , we have $\langle A y_n, y_n - p \rangle \geq 0$. It follows that

$$\langle A y_n, z_n - p \rangle = \langle A y_n, y_n - p \rangle + \langle A y_n, z_n - y_n \rangle \geq \langle A y_n, z_n - y_n \rangle. \quad (3.20)$$

Combining (3.19) and (3.20), we have

$$D_f(p, z_n) \leq D_f(p, u_n) - D_f(z_n, u_n) + \lambda_n \langle A y_n, y_n - z_n \rangle. \quad (3.21)$$

Then using (2.5) and (2.6), we get

$$\begin{aligned} D_f(p, z_n) &\leq D_f(p, u_n) - D_f(z_n, y_n) + D_f(u_n, y_n) - \langle \nabla f(y_n) - \nabla f(u_n), z_n - u_n \rangle + \lambda_n \langle A y_n, y_n - z_n \rangle \\ &= D_f(p, u_n) - D_f(z_n, y_n) - D_f(y_n, u_n) + \langle \nabla f(u_n) - \nabla f(y_n), u_n - y_n \rangle \\ &\quad - \langle \nabla f(y_n) - \nabla f(u_n), z_n - u_n \rangle + \lambda_n \langle A y_n, y_n - z_n \rangle \\ &= D_f(p, u_n) - D_f(z_n, y_n) - D_f(y_n, u_n) + \langle \nabla f(u_n) - \nabla f(y_n), z_n - y_n \rangle + \lambda_n \langle A y_n, y_n - z_n \rangle \\ &= D_f(p, u_n) - D_f(z_n, y_n) - D_f(y_n, u_n) + \langle \nabla f(u_n) - \lambda_n A u_n - \nabla f(y_n), z_n - y_n \rangle \\ &\quad + \lambda_n \langle A u_n - A y_n, z_n - y_n \rangle. \end{aligned} \quad (3.22)$$

It is clear that $z_n \in T_n$ and hence

$$\langle \nabla f(u_n) - \lambda_n A u_n - \nabla f(y_n), z_n - y_n \rangle \leq 0. \quad (3.23)$$

Combining (3.22) and (3.23), we have

$$\begin{aligned} D_f(p, z_n) &\leq D_f(p, u_n) - D_f(z_n, y_n) - D_f(y_n, u_n) + \lambda_n \langle A u_n - A y_n, z_n - y_n \rangle \\ &\leq D_f(p, u_n) - D_f(z_n, y_n) - D_f(y_n, u_n) + \lambda_n \|A u_n - A y_n\| \|z_n - y_n\| \\ &\leq D_f(p, u_n) - D_f(y_n, u_n) - D_f(z_n, y_n) + \mu \|u_n - y_n\| \|z_n - y_n\| \\ &\leq D_f(p, u_n) - D_f(y_n, u_n) - D_f(z_n, y_n) + \frac{\mu}{2} \|u_n - y_n\|^2 + \frac{\mu}{2} \|z_n - y_n\|^2 \\ &\leq D_f(p, u_n) - \left(1 - \frac{\mu}{\sigma}\right) D_f(y_n, u_n) - \left(1 - \frac{\mu}{\sigma}\right) D_f(z_n, y_n). \end{aligned} \quad (3.24)$$

Since $\mu \in (0, \sigma)$, we have $1 - \frac{\mu}{\sigma} = \frac{\sigma - \mu}{\sigma} > 0$. Consequently,

$$\left(1 - \frac{\mu}{\sigma}\right)D_f(y_n, x_n) + \left(1 - \frac{\mu}{\sigma}\right)D_f(z_n, y_n) \geq 0.$$

Then from (3.24), we obtain

$$D_f(p, z_n) \leq D_f(p, u_n). \quad (3.25)$$

□

Now, we prove strong convergence theorem of Algorithm 1.

Theorem 3.7. *Suppose that Assumptions 1-5 are satisfied. Then the sequence $\{x_n\}$ generated by Algorithm 1 converges strongly to $z \in VI(C, A)$, where $z = \Pi_{VI(C, A)}^f(u)$.*

Proof. First, we show that $\{x_n\}$ is bounded. Let $p \in VI(C, A)$. From (2.9), we have

$$\begin{aligned} D_f(p, u_n) &= D_f(p, \nabla f^*(\nabla f(x_n) + \theta_n(\nabla f(x_{n-1}) - \nabla f(x_n)))) \\ &= D_f(p, \nabla f^*((1 - \theta_n)\nabla f(x_n) + \theta_n\nabla f(x_{n-1}))) \\ &\leq (1 - \theta_n)D_f(p, x_n) + \theta_n D_f(p, x_{n-1}). \end{aligned} \quad (3.26)$$

It follows from (3.25) and (3.26) that

$$\begin{aligned} D_f(p, x_{n+1}) &\leq \alpha_n D_f(p, u) + (1 - \alpha_n)D_f(p, z_n) \\ &\leq \alpha_n D_f(p, u) + (1 - \alpha_n)D_f(p, u_n) \\ &\leq \alpha_n D_f(p, u) + (1 - \alpha_n)(1 - \theta_n)D_f(p, x_n) + (1 - \alpha_n)\theta_n D_f(p, x_{n-1}) \\ &\leq \alpha_n D_f(p, u) + (1 - \alpha_n) \max\{D_f(p, x_n), D_f(p, x_{n-1})\} \\ &\leq \max\{D_f(p, u), D_f(p, x_n), D_f(p, x_{n-1})\}. \end{aligned}$$

By induction, we deduce that

$$D_f(p, x_n) \leq \max\{D_f(p, u), D_f(p, x_1), D_f(p, x_0)\}.$$

This implies that $\{D_f(p, x_n)\}$ is bounded. From the relation $D_f(x, y) \geq \frac{\sigma}{2}\|x - y\|^2$ for all $x, y \in H$, we can see that $\{x_n\}$ is bounded and consequently $\{u_n\}$, $\{y_n\}$ and $\{z_n\}$ are bounded. Let $z = \Pi_{VI(C, A)}^f(u)$. From Lemma 3.6 and (3.26), we have

$$\begin{aligned} D_f(z, x_{n+1}) &\leq \alpha_n D_f(z, u) + (1 - \alpha_n)D_f(z, z_n) \\ &\leq \alpha_n D_f(z, u) + (1 - \alpha_n)D_f(z, u_n) - (1 - \alpha_n)\left(1 - \frac{\mu}{\sigma}\right)D_f(y_n, u_n) - (1 - \alpha_n)\left(1 - \frac{\mu}{\sigma}\right)D_f(z_n, y_n) \\ &\leq \alpha_n D_f(z, u) + (1 - \alpha_n)(1 - \theta_n)D_f(z, x_n) + (1 - \alpha_n)\theta_n D_f(z, x_{n-1}) - (1 - \alpha_n)\left(1 - \frac{\mu}{\sigma}\right)D_f(y_n, u_n) \\ &\quad - (1 - \alpha_n)\left(1 - \frac{\mu}{\sigma}\right)D_f(z_n, y_n). \end{aligned}$$

This implies that

$$\begin{aligned} &(1 - \alpha_n)\left(1 - \frac{\mu}{\sigma}\right)D_f(y_n, u_n) + (1 - \alpha_n)\left(1 - \frac{\mu}{\sigma}\right)D_f(z_n, y_n) \\ &\leq D_f(z, x_n) - D_f(z, x_{n+1}) + (1 - \alpha_n)\theta_n(D_f(z, x_{n-1}) - D_f(z, x_n)) + \alpha_n K, \end{aligned} \quad (3.27)$$

where $K = \sup_{n \geq 1} \{|D_f(z, u) - D_f(z, x_n)|\}$. We consider the following two possible cases to prove $\lim_{n \rightarrow \infty} D_f(z, x_n) = 0$.

Case 1. There exists an $N \in \mathbb{N}$ such that $D_f(z, x_{n+1}) \leq D_f(z, x_n)$ for all $n \geq N$. This gives $\{D_f(z, x_n)\}$ is convergent and consequently

$$\lim_{n \rightarrow \infty} (D_f(z, x_n) - D_f(z, x_{n+1})) = \lim_{n \rightarrow \infty} (D_f(z, x_{n-1}) - D_f(z, x_n)) = 0.$$

Then (3.27) implies that $\lim_{n \rightarrow \infty} D_f(y_n, u_n) = \lim_{n \rightarrow \infty} D_f(z_n, y_n) = 0$. Hence

$$\lim_{n \rightarrow \infty} \|\nabla f(y_n) - \nabla f(u_n)\| = \lim_{n \rightarrow \infty} \|\nabla f(z_n) - \nabla f(y_n)\| = 0.$$

Thus we have

$$\begin{aligned} \|\nabla f(z_n) - \nabla f(u_n)\| &\leq \|\nabla f(z_n) - \nabla f(y_n)\| + \|\nabla f(y_n) - \nabla f(u_n)\| \\ &\rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned} \tag{3.28}$$

Note that

$$\begin{aligned} \|\nabla f(x_{n+1}) - \nabla f(u_n)\| &\leq \|\nabla f(x_{n+1}) - \nabla f(z_n)\| + \|\nabla f(z_n) - \nabla f(u_n)\| \\ &= \alpha_n \|\nabla f(u) - \nabla f(z_n)\| + \|\nabla f(z_n) - \nabla f(u_n)\|. \end{aligned}$$

It follows from (3.28) that

$$\lim_{n \rightarrow \infty} \|\nabla f(x_{n+1}) - \nabla f(u_n)\| = 0. \tag{3.29}$$

Since $\alpha_n \in (0, 1)$, we have $\theta_n \|\nabla f(x_{n-1}) - \nabla f(x_n)\| \leq \frac{\theta_n}{\alpha_n} \|\nabla f(x_{n-1}) - \nabla f(x_n)\| \rightarrow 0$ as $n \rightarrow \infty$. Thus we have

$$\|\nabla f(u_n) - \nabla f(x_n)\| = \theta_n \|\nabla f(x_{n-1}) - \nabla f(x_n)\| \rightarrow 0 \text{ as } n \rightarrow \infty. \tag{3.30}$$

It follows from (3.29) and (3.30) that

$$\begin{aligned} \|\nabla f(x_{n+1}) - \nabla f(x_n)\| &\leq \|\nabla f(x_{n+1}) - \nabla f(u_n)\| + \|\nabla f(u_n) - \nabla f(x_n)\| \\ &\rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

Hence

$$\lim_{n \rightarrow \infty} \|x_{n+1} - x_n\| = 0. \tag{3.31}$$

In fact, since $\{x_n\}$ is bounded, we assume that there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $x_{n_k} \rightharpoonup v$ and

$$\limsup_{n \rightarrow \infty} \langle \nabla f(u) - \nabla f(z), x_n - z \rangle = \lim_{k \rightarrow \infty} \langle \nabla f(u) - \nabla f(z), x_{n_k} - z \rangle.$$

From (3.30), we have $\|u_n - x_n\| \rightarrow 0$ and hence $u_{n_k} \rightharpoonup v$. Since $\|\nabla f(y_{n_k}) - \nabla f(u_{n_k})\| \rightarrow 0$, we have $\|y_{n_k} - u_{n_k}\| \rightarrow 0$. By Lemma 3.5, we get $v \in VI(C, A)$. Then from (2.7), we obtain

$$\limsup_{n \rightarrow \infty} \langle \nabla f(u) - \nabla f(z), x_n - z \rangle = \langle \nabla f(u) - \nabla f(z), v - z \rangle \leq 0.$$

Also, from (3.31), we obtain

$$\limsup_{n \rightarrow \infty} \langle \nabla f(u) - \nabla f(z), x_{n+1} - z \rangle \leq 0. \tag{3.32}$$

By the properties of V_f , we get

$$\begin{aligned}
D_f(z, x_{n+1}) &= V_f(z, \alpha_n \nabla f(u) + (1 - \alpha_n) \nabla f(z_n)) \\
&\leq V_f(z, \alpha_n \nabla f(u) + (1 - \alpha_n) \nabla f(z_n) - \alpha_n (\nabla f(u) - \nabla f(z))) + \alpha_n \langle \nabla f(u) - \nabla f(z), x_{n+1} - z \rangle \\
&= V_f(z, \alpha_n \nabla f(z) + (1 - \alpha_n) \nabla f(z_n)) + \alpha_n \langle \nabla f(u) - \nabla f(z), x_{n+1} - z \rangle \\
&= D_f(z, \nabla f^*(\alpha_n \nabla f(z) + (1 - \alpha_n) \nabla f(z_n))) + \alpha_n \langle \nabla f(u) - \nabla f(z), x_{n+1} - z \rangle \\
&\leq \alpha_n D_f(z, z) + (1 - \alpha_n) D_f(z, z_n) + \alpha_n \langle \nabla f(u) - \nabla f(z), x_{n+1} - z \rangle \\
&\leq (1 - \alpha_n) ((1 - \theta_n) D_f(z, x_n) + \theta_n D_f(z, x_{n-1})) + \alpha_n \langle \nabla f(u) - \nabla f(z), x_{n+1} - z \rangle \\
&= (1 - \alpha_n - (1 - \alpha_n) \theta_n) D_f(z, x_n) + (1 - \alpha_n) \theta_n D_f(z, x_{n-1}) \\
&\quad + \alpha_n \langle \nabla f(u) - \nabla f(z), x_{n+1} - z \rangle.
\end{aligned} \tag{3.33}$$

Using Lemma 2.10 and (3.32), we obtain $\lim_{n \rightarrow \infty} D_f(z, x_n) = 0$ and hence $x_n \rightarrow z$ as $n \rightarrow \infty$.

Case 2. There exists a subsequence $\{D_f(z, x_{n_i})\}$ of $\{D_f(z, x_n)\}$ such that $D_f(z, x_{n_i}) \leq D_f(z, x_{n_i+1})$ for all $i \in \mathbb{N}$. It follows from Lemma 2.9 that there exists a nondecreasing sequence $\{m_k\}$ of \mathbb{N} such that $\lim_{k \rightarrow \infty} m_k = \infty$ and the following inequalities hold for all $k \in \mathbb{N}$:

$$D_f(z, x_{m_k}) \leq D_f(z, x_{m_k+1}) \tag{3.34}$$

and

$$D_f(z, x_k) \leq D_f(z, x_{m_k+1}). \tag{3.35}$$

From (3.27), we have

$$\begin{aligned}
&(1 - \alpha_{m_k}) \left(1 - \frac{\mu}{\sigma}\right) D_f(y_{m_k}, u_{m_k}) + (1 - \alpha_{m_k}) \left(1 - \frac{\mu}{\sigma}\right) D_f(z_{m_k}, y_{m_k}) \\
&\leq D_f(z, x_{m_k}) - D_f(z, x_{m_k+1}) + (1 - \alpha_{m_k}) \theta_{m_k} (D_f(z, x_{m_k-1}) - D_f(z, x_{m_k})) + \alpha_{m_k} K \\
&\leq \alpha_{m_k} K,
\end{aligned}$$

where $K > 0$. Then we obtain

$$\lim_{k \rightarrow \infty} D_f(y_{m_k}, u_{m_k}) = \lim_{k \rightarrow \infty} D_f(z_{m_k}, y_{m_k}) = 0.$$

Hence

$$\lim_{k \rightarrow \infty} \|\nabla f(y_{m_k}) - \nabla f(u_{m_k})\| = \lim_{k \rightarrow \infty} \|\nabla f(z_{m_k}) - \nabla f(y_{m_k})\| = 0.$$

Using the same arguments as in the proof of Case 1, we can show that

$$\begin{aligned}
\lim_{k \rightarrow \infty} \|\nabla f(z_{m_k}) - \nabla f(u_{m_k})\| &= 0, \quad \lim_{k \rightarrow \infty} \|\nabla f(x_{m_k+1}) - \nabla f(u_{m_k})\| = 0, \\
\lim_{k \rightarrow \infty} \|\nabla f(u_{m_k}) - \nabla f(x_{m_k})\| &= 0, \quad \lim_{k \rightarrow \infty} \|\nabla f(x_{m_k+1}) - \nabla f(x_{m_k})\| = 0
\end{aligned}$$

and

$$\limsup_{k \rightarrow \infty} \langle \nabla f(u) - \nabla f(z), x_{m_k+1} - z \rangle \leq 0. \tag{3.36}$$

Also, from (3.33) and (3.34), we can show that

$$\begin{aligned}
D_f(z, x_{m_k+1}) &\leq (1 - \alpha_{m_k} - (1 - \alpha_{m_k}) \theta_{m_k}) D_f(z, x_{m_k}) + (1 - \alpha_{m_k}) \theta_{m_k} D_f(z, x_{m_k-1}) \\
&\quad + \alpha_{m_k} \langle \nabla f(u) - \nabla f(z), x_{m_k+1} - z \rangle \\
&\leq (1 - \alpha_{m_k}) D_f(z, x_{m_k}) + \alpha_{m_k} \langle \nabla f(u) - \nabla f(z), x_{m_k+1} - z \rangle \\
&\leq (1 - \alpha_{m_k}) D_f(z, x_{m_k+1}) + \alpha_{m_k} \langle \nabla f(u) - \nabla f(z), x_{m_k+1} - z \rangle.
\end{aligned}$$

Since $\alpha_{m_k} > 0$ and from (3.35), we have

$$D_f(z, x_k) \leq D_f(z, x_{m_k+1}) \leq \langle \nabla f(u) - \nabla f(z), x_{m_k+1} - z \rangle. \quad (3.37)$$

Combining (3.36) and (3.37), we get

$$\limsup_{k \rightarrow \infty} D_f(z, x_k) \leq 0.$$

This gives $\limsup_{k \rightarrow \infty} D_f(z, x_k) = 0$ and hence $x_k \rightarrow z$ as $k \rightarrow \infty$. From above two cases, we can conclude that the sequence $\{x_n\}$ converges strongly to $z = \Pi_{VI(C,A)}^f(u)$. We thus complete the proof. \square

Next, we propose the second relaxed inertial algorithm, which combines the Halpern-type iteration and the Tseng's extragradient method. The algorithm is of the following form:

Algorithm 2: Relaxed inertial Tseng's extragradient algorithm for VIP

(Step 0) Given $\theta \in (0, 1/2]$, $\gamma > 0$, $l \in (0, 1)$ and $\mu \in (0, \sigma)$, where σ is a constant given by (2.4). Let $x_0, x_1, u \in H$ be arbitrary.

(Step 1) Given the current iterates x_{n-1} and x_n ($n \geq 1$). Choose θ_n such that $0 \leq \theta_n \leq \bar{\theta}_n$, where

$$\bar{\theta}_n = \begin{cases} \min \left\{ \frac{\xi_n}{\|\nabla f(x_{n-1}) - \nabla f(x_n)\|}, \theta \right\} & \text{if } x_{n-1} \neq x_n, \\ \theta & \text{otherwise.} \end{cases}$$

Set $u_n = \nabla f^*(\nabla f(x_n) + \theta_n(\nabla f(x_{n-1}) - \nabla f(x_n)))$ and compute

$$y_n = \Pi_C^f \nabla f^*(\nabla f(u_n) - \lambda_n A u_n),$$

where $\lambda_n = \gamma l^{m_n}$, with m_n is the smallest nonnegative integer m satisfying

$$\gamma l^m \|A u_n - A y_n\| \leq \mu \|u_n - y_n\|.$$

If $u_n = y_n$ or $A y_n = 0$, then stop and y_n is a solution of VIP. Otherwise, go to Step 2.

(Step 2) Compute

$$z_n = \nabla f^*(\nabla f(y_n) - \lambda_n (A y_n - A u_n)).$$

(Step 3) Compute

$$x_{n+1} = \nabla f^*(\alpha_n \nabla f(u) + (1 - \alpha_n) \nabla f(z_n)).$$

Update $n := n + 1$ go to Step 1.

Remark 3.8. If $f(x) = \frac{1}{2}\|x\|^2$ and $\theta_n = 0$, then Algorithm 2 reduces to the following one:

$$\begin{cases} y_n = P_C(x_n - \lambda_n A x_n), \\ z_n = y_n - \lambda_n (A y_n - A x_n), \\ x_{n+1} = \alpha_n u + (1 - \alpha_n) z_n, \end{cases} \quad (3.38)$$

where λ_n is defined in (3.5). The Algorithm (3.38) is a modification of TEGM without the relaxed inertial term for pseudomonotone VIP with a non-Lipschitz mapping.

Lemma 3.9. *Suppose that Assumptions 1-4 are satisfied. Then the sequence $\{x_n\}$ generated by Algorithm 2 satisfies the following inequality:*

$$D_f(p, z_n) \leq D_f(p, u_n) - \left(1 - \frac{\mu}{\sigma}\right) D_f(y_n, u_n) - \left(1 - \frac{\mu}{\sigma}\right) D_f(z_n, y_n) \quad \forall p \in VI(C, A).$$

In particular, if $\mu \in (0, \sigma)$, then $D_f(p, z_n) \leq D_f(p, u_n)$.

Proof. Let $p \in VI(C, A)$. By the definition of Bregman divergence, we have

$$\begin{aligned} D_f(p, z_n) &= D_f(p, \nabla f^*(\nabla f(y_n) - \lambda_n(Ay_n - Au_n))) \\ &= f(p) - f(z_n) - \langle \nabla f(y_n) - \lambda_n(Ay_n - Au_n), p - z_n \rangle \\ &= f(p) - f(z_n) - \langle \nabla f(y_n), p - z_n \rangle + \lambda_n \langle Ay_n - Au_n, p - z_n \rangle \\ &= f(p) - f(y_n) - \langle \nabla f(y_n), p - y_n \rangle + \langle \nabla f(y_n), p - y_n \rangle + f(y_n) - f(z_n) - \langle \nabla f(y_n), p - z_n \rangle \\ &\quad + \lambda_n \langle Ay_n - Au_n, p - z_n \rangle \\ &= f(p) - f(y_n) - \langle \nabla f(y_n), p - y_n \rangle - f(z_n) + f(y_n) + \langle \nabla f(y_n), z_n - y_n \rangle + \lambda_n \langle Ay_n - Au_n, p - z_n \rangle \\ &= D_f(p, y_n) - D_f(z_n, y_n) + \lambda_n \langle Ay_n - Au_n, p - z_n \rangle. \end{aligned} \quad (3.39)$$

From (2.2), we see that

$$D_f(p, y_n) = D_f(p, u_n) - D_f(y_n, u_n) + \langle \nabla f(u_n) - \nabla f(y_n), p - y_n \rangle. \quad (3.40)$$

Substituting (3.40) into (3.39), we get

$$\begin{aligned} D_f(p, z_n) &= D_f(p, u_n) - D_f(y_n, u_n) - D_f(z_n, y_n) + \langle \nabla f(u_n) - \nabla f(y_n), p - y_n \rangle \\ &\quad + \lambda_n \langle Ay_n - Au_n, p - z_n \rangle. \end{aligned} \quad (3.41)$$

By the definition of y_n , we have

$$\langle \nabla f(u_n) - \lambda_n Au_n - \nabla f(y_n), p - y_n \rangle \leq 0,$$

which implies that

$$\langle \nabla f(u_n) - \nabla f(y_n), p - y_n \rangle \leq \lambda_n \langle Au_n, p - y_n \rangle. \quad (3.42)$$

Substituting (3.42) into (3.41), we get

$$\begin{aligned} D_f(p, z_n) &\leq D_f(p, u_n) - D_f(y_n, u_n) - D_f(z_n, y_n) + \lambda_n \langle Au_n, p - y_n \rangle + \lambda_n \langle Ay_n - Au_n, p - z_n \rangle \\ &= D_f(p, u_n) - D_f(y_n, u_n) - D_f(z_n, y_n) + \lambda_n \langle Au_n, p - y_n \rangle + \lambda_n \langle Ay_n, p - z_n \rangle - \lambda_n \langle Au_n, p - z_n \rangle \\ &= D_f(p, u_n) - D_f(y_n, u_n) - D_f(z_n, y_n) + \lambda_n \langle Au_n, z_n - y_n \rangle + \lambda_n \langle Ay_n, p - z_n \rangle \\ &= D_f(p, u_n) - D_f(y_n, u_n) - D_f(z_n, y_n) + \lambda_n \langle Au_n, z_n - y_n \rangle - \lambda_n \langle Ay_n, y_n - p \rangle + \lambda_n \langle Ay_n, y_n - z_n \rangle \\ &= D_f(p, u_n) - D_f(y_n, u_n) - D_f(z_n, y_n) + \lambda_n \langle Au_n - Ay_n, z_n - y_n \rangle - \lambda_n \langle Ay_n, y_n - p \rangle. \end{aligned}$$

Since $p \in VI(C, A)$ and $y_n \in C$, we have $\langle Ap, y_n - p \rangle \geq 0$, which implies by the pseudomonotonicity of A that $\langle Ay_n, y_n - p \rangle \geq 0$. From (2.4), we have

$$\begin{aligned} D_f(p, z_n) &\leq D_f(p, u_n) - D_f(y_n, u_n) - D_f(z_n, y_n) + \lambda_n \langle Au_n - Ay_n, z_n - y_n \rangle \\ &\leq D_f(p, u_n) - D_f(y_n, u_n) - D_f(z_n, y_n) + \lambda_n \|Au_n - Ay_n\| \|z_n - y_n\| \\ &\leq D_f(p, u_n) - D_f(y_n, u_n) - D_f(z_n, y_n) + \mu \|u_n - y_n\| \|z_n - y_n\| \\ &\leq D_f(p, u_n) - D_f(y_n, u_n) - D_f(z_n, y_n) + \frac{\mu}{2} \|u_n - y_n\|^2 + \frac{\mu}{2} \|z_n - y_n\|^2 \\ &\leq D_f(p, u_n) - \left(1 - \frac{\mu}{\sigma}\right) D_f(y_n, u_n) - \left(1 - \frac{\mu}{\sigma}\right) D_f(z_n, y_n). \end{aligned} \quad (3.43)$$

Since $\mu \in (0, \sigma)$, we have $1 - \frac{\mu}{\sigma} = \frac{\sigma - \mu}{\sigma} > 0$. This implies that

$$\left(1 - \frac{\mu}{\sigma}\right)D_f(y_n, x_n) + \left(1 - \frac{\mu}{\sigma}\right)D_f(z_n, y_n) \geq 0.$$

Then from (3.43), we obtain

$$D_f(p, z_n) \leq D_f(p, u_n).$$

□

Theorem 3.10. *Suppose that Assumptions 1-5 are satisfied. Then the sequence $\{x_n\}$ generated by Algorithm 2 converges strongly to $z \in VI(C, A)$, where $z = \Pi_{VI(C, A)}^f(u)$.*

Proof. The proof of theorem is quite similar to that of Theorem 3.7, so we omit it here. □

Next, we also utilize Algorithm 1 and Algorithm 2 to finding a common solution of VIP and fixed point problem (FPP) of a Bregman quasi-nonexpansive mapping. Let C be a nonempty subset of H and $S : C \rightarrow C$ be a mapping with a fixed point set is nonempty, that is, $F(S) := \{x \in C : x = Sx\} \neq \emptyset$. A point $z \in C$ is called an *asymptotic fixed point* of S [41], if C contains a sequence $\{x_n\}$, which converges weakly to z and $\lim_{n \rightarrow \infty} \|x_n - Sx_n\| = 0$. We denote $\widehat{F}(S)$ by the set of asymptotic fixed points of S . A mapping S is said to be *Bregman quasi-nonexpansive* [11] if $F(S) \neq \emptyset$ and $D_f(v, Sx) \leq D_f(v, x)$ for all $v \in F(S)$ and $x \in C$.

Algorithm 3: Relaxed inertial subgradient extragradient algorithm for VIP and FPP

(Step 0) Given $\theta \in (0, 1/2]$, $\gamma > 0$, $l \in (0, 1)$ and $\mu \in (0, \sigma)$, where σ is a constant given by (2.4). Let $x_0, x_1, u \in H$ be arbitrary.

(Step 1) Given the current iterates x_{n-1} and x_n ($n \geq 1$). Choose θ_n such that $0 \leq \theta_n \leq \bar{\theta}_n$, where $\bar{\theta}_n$ is defined by (3.2). Set $u_n = \nabla f^*(\nabla f(x_n) + \theta_n(\nabla f(x_{n-1}) - \nabla f(x_n)))$ and compute

$$y_n = \Pi_C^f \nabla f^*(\nabla f(u_n) - \lambda_n A u_n),$$

where λ_n is defined in (3.3).

(Step 2) Construct the half-space

$$T_n := \{x \in H : \langle \nabla f(u_n) - \lambda_n A u_n - \nabla f(y_n), x - y_n \rangle \leq 0\}$$

and compute

$$z_n = \Pi_{T_n}^f \nabla f^*(\nabla f(u_n) - \lambda_n A y_n).$$

(Step 3) Compute

$$x_{n+1} = \nabla f^*(\alpha_n \nabla f(u) + (1 - \alpha_n)(\beta_n \nabla f(z_n) + (1 - \beta_n) \nabla f(Sz_n))).$$

Update $n := n + 1$ go to Step 1.

Theorem 3.11. *Suppose that Assumptions 1-5 are satisfied. Let $S : H \rightarrow H$ be a Bregman quasi-nonexpansive mapping such that $F(S) = \widehat{F}(S)$ and $\{\beta_n\} \subset (0, 1)$ such that $\liminf_{n \rightarrow \infty} \beta_n(1 - \beta_n) > 0$. Suppose that $\Omega := VI(C, A) \cap F(S) \neq \emptyset$, then the sequence $\{x_n\}$ generated by Algorithm 3 converges strongly to $z \in \Omega$, where $z = \Pi_{\Omega}^f(u)$.*

Proof. As proved in Theorem 3.7, we have $\{x_n\}$ is bounded and consequently $\{u_n\}$, $\{y_n\}$ and $\{z_n\}$ are bounded. Let $z \in \Omega$ and $w_n = \nabla f^*(\beta_n \nabla f(z_n) + (1 - \beta_n) \nabla f(Sz_n))$ for all $n \geq 1$. Since f is uniformly Fréchet differentiable, we have f is uniformly smooth (see [52, p. 207]). This implies that f^* is uniformly convex (see [52, Theorem 3.5.5]). By the property of V_f and Lemma 3.6, we have

$$\begin{aligned}
D_f(z, w_n) &= V_f(z, \beta_n \nabla f(x_n) + (1 - \beta_n) \nabla f(Sz_n)) \\
&= f(z) - \langle z, \beta_n \nabla f(x_n) + (1 - \beta_n) \nabla f(Sz_n) \rangle + f^*(\beta_n \nabla f(x_n) + (1 - \beta_n) \nabla f(Sz_n)) \\
&\leq \beta_n f(z) + (1 - \beta_n) f(z) - \beta_n \langle z, \nabla f(z_n) \rangle - (1 - \beta_n) \langle z, \nabla f(Sz_n) \rangle + \beta_n f^*(\nabla f(z_n)) \\
&\quad + (1 - \beta_n) f^*(\nabla f(Sz_n)) - \beta_n (1 - \beta_n) \phi^*(\|\nabla f(z_n) - \nabla f(Sz_n)\|) \\
&= \beta_n (f(z) - \langle z, \nabla f(z_n) \rangle + f^*(\nabla f(z_n))) + (1 - \beta_n) (f(z) - \langle z, \nabla f(Sz_n) \rangle + f^*(\nabla f(Sz_n))) \\
&\quad - \beta_n (1 - \beta_n) \phi^*(\|\nabla f(z_n) - \nabla f(Sz_n)\|) \\
&= \beta_n D_f(z, z_n) + (1 - \beta_n) D_f(z, Sz_n) - \beta_n (1 - \beta_n) \phi^*(\|\nabla f(z_n) - \nabla f(Sz_n)\|) \\
&\leq \beta_n D_f(z, z_n) + (1 - \beta_n) D_f(z, z_n) - \beta_n (1 - \beta_n) \phi^*(\|\nabla f(z_n) - \nabla f(Sz_n)\|) \\
&= D_f(z, z_n) - \beta_n (1 - \beta_n) \phi^*(\|\nabla f(z_n) - \nabla f(Sz_n)\|) \\
&\leq D_f(z, u_n) - \left(1 - \frac{\mu}{\sigma}\right) D_f(y_n, u_n) - \left(1 - \frac{\mu}{\sigma}\right) D_f(z_n, y_n) - \beta_n (1 - \beta_n) \phi^*(\|\nabla f(z_n) - \nabla f(Sz_n)\|).
\end{aligned}$$

From (3.26), we have

$$\begin{aligned}
D_f(z, w_n) &\leq (1 - \theta_n) D_f(z, x_n) + \theta_n D_f(p, x_{n-1}) - \left(1 - \frac{\mu}{\sigma}\right) D_f(y_n, u_n) - \left(1 - \frac{\mu}{\sigma}\right) D_f(z_n, y_n) \\
&\quad - \beta_n (1 - \beta_n) \phi^*(\|\nabla f(z_n) - \nabla f(Sz_n)\|).
\end{aligned} \tag{3.44}$$

It follows that

$$\begin{aligned}
D_f(z, x_{n+1}) &\leq \alpha_n D_f(z, u) + (1 - \alpha_n) D_f(z, w_n) \\
&\leq \alpha_n D_f(z, u) + (1 - \alpha_n) (1 - \theta_n) D_f(p, x_n) + \theta_n D_f(p, x_{n-1}) - \left(1 - \frac{\mu}{\sigma}\right) D_f(y_n, u_n) - \left(1 - \frac{\mu}{\sigma}\right) D_f(z_n, y_n) \\
&\quad - \beta_n (1 - \beta_n) \phi^*(\|\nabla f(z_n) - \nabla f(Sz_n)\|).
\end{aligned}$$

This implies that

$$\begin{aligned}
&(1 - \alpha_n) \left(1 - \frac{\mu}{\sigma}\right) D_f(y_n, u_n) + (1 - \alpha_n) \left(1 - \frac{\mu}{\sigma}\right) D_f(z_n, y_n) + \beta_n (1 - \beta_n) \phi^*(\|\nabla f(z_n) - \nabla f(Sz_n)\|) \\
&\leq D_f(z, x_n) - D_f(z, x_{n+1}) + (1 - \alpha_n) \theta_n (D_f(z, x_{n-1}) - D_f(z, x_n)) + \alpha_n K,
\end{aligned}$$

where $K = \sup_{n \geq 1} \{|D_f(z, u) - D_f(z, x_n)|\}$. Obviously, by following the methods used in the proof of Theorem 3.7, we have

$$\lim_{n \rightarrow \infty} D_f(y_n, u_n) = \lim_{n \rightarrow \infty} D_f(z_n, y_n) = \lim_{n \rightarrow \infty} \phi^*(\|\nabla f(z_n) - \nabla f(Sz_n)\|) = 0.$$

Hence $\lim_{n \rightarrow \infty} \|\nabla f(y_n) - \nabla f(u_n)\| = \lim_{n \rightarrow \infty} \|\nabla f(z_n) - \nabla f(y_n)\| = 0$. By the property of ϕ^* , we have $\lim_{n \rightarrow \infty} \|\nabla f(z_n) - \nabla f(Sz_n)\| = 0$ and hence $\lim_{n \rightarrow \infty} \|z_n - Sz_n\| = 0$. Moreover, we can show that

$$\lim_{n \rightarrow \infty} \|\nabla f(z_n) - \nabla f(u_n)\| = 0 \tag{3.45}$$

and

$$\lim_{n \rightarrow \infty} \|\nabla f(u_n) - \nabla f(x_n)\| = 0. \tag{3.46}$$

It follows from (3.45) and (3.46) that

$$\begin{aligned} \|\nabla f(z_n) - \nabla f(x_n)\| &\leq \|\nabla f(z_n) - \nabla f(u_n)\| + \|\nabla f(u_n) - \nabla f(x_n)\| \\ &\rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

Hence

$$\lim_{n \rightarrow \infty} \|z_n - x_n\| = 0. \quad (3.47)$$

Since $\{x_n\}$ is bounded, there exists a subsequence of $\{x_{n_k}\}$ of $\{x_n\}$ such that $x_{n_k} \rightharpoonup v$. From (3.47), we also get $z_{n_k} \rightharpoonup v$ and since $\|z_n - Sz_n\| \rightarrow 0$, we have $v \in \widehat{F}(S) = F(S)$. In the rest of the proof, we follow the lines of the proof of Theorem 3.7 and hence it is omitted. \square

Algorithm 4: Relaxed inertial Tseng's extragradient algorithm for VIP and FPP

(Step 0) Given $\theta \in (0, 1/2]$, $\gamma > 0$, $l \in (0, 1)$ and $\mu \in (0, \sigma)$, where σ is a constant given by (2.4). Let $x_0, x_1, u \in H$ be arbitrary.

(Step 1) Given the current iterates x_{n-1} and x_n ($n \geq 1$). Choose θ_n such that $0 \leq \theta_n \leq \bar{\theta}_n$, where $\bar{\theta}_n$ is defined by (3.2). Set $u_n = \nabla f^*(\nabla f(x_n) + \theta_n(\nabla f(x_{n-1}) - \nabla f(x_n)))$ and compute

$$y_n = \Pi_C^f \nabla f^*(\nabla f(u_n) - \lambda_n A u_n),$$

where λ_n is defined in (3.3).

(Step 2) Compute

$$z_n = \nabla f^*(\nabla f(y_n) - \lambda_n (A y_n - A u_n)).$$

(Step 3) Compute

$$x_{n+1} = \nabla f^*(\alpha_n \nabla f(u) + (1 - \alpha_n)(\beta_n \nabla f(x_n) + (1 - \beta_n) \nabla f(S z_n))).$$

Update $n := n + 1$ go to Step 1.

Theorem 3.12. *Suppose that Assumptions 1-5 are satisfied. Let $S : H \rightarrow H$ be a Bregman quasi-nonexpansive mapping such that $F(S) = \widehat{F}(S)$ and $\{\beta_n\} \subset (0, 1)$ such that $\liminf_{n \rightarrow \infty} \beta_n(1 - \beta_n) > 0$. Suppose that $\Omega := VI(C, A) \cap F(S) \neq \emptyset$, then the sequence $\{x_n\}$ generated by Algorithm 4 converges strongly to $z \in \Omega$, where $z = \Pi_\Omega^f(u)$.*

4 Numerical experiments

In this section, we provide some numerical experiments with a non-Euclidean distance to illustrate the convergence behavior of the proposed algorithms. Let $H = \mathbb{R}^m$, then $\nabla f^* = (\nabla f)^{-1}$. The following lists are values of $(\nabla f)^{-1}$ for various functions in Example 2.2.

- (i) For $f^{KL}(x)$, we have $(\nabla f^{KL})^{-1}(x) = (\exp(x_1 - 1), \exp(x_2 - 1), \dots, \exp(x_m - 1))^T$.
- (ii) For $f^{IS}(x)$, we have $(\nabla f^{IS})^{-1}(x) = -\left(\frac{1}{x_1}, \frac{1}{x_2}, \dots, \frac{1}{x_m}\right)^T$.

(iii) For $f^{FD}(x)$, we have $(\nabla f^{FD})^{-1}(x) = \left(\frac{\exp(x_1)}{1+\exp(x_1)}, \frac{\exp(x_2)}{1+\exp(x_2)}, \dots, \frac{\exp(x_m)}{1+\exp(x_m)} \right)^T$.

(iv) For $f^{SM}(x)$, we have $(\nabla f^{SM})^{-1}(x) = Q^{-1}x$.

(v) For $f^{SE}(x)$, we have $(\nabla f^{SE})^{-1}(x) = x$.

Note that each f satisfies Assumption 2 (see [5, 23]). Also, we can calculate explicitly the Hessian matrix as follows:

$$\nabla^2 f^{KL}(x) = \begin{pmatrix} \frac{1}{x_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_m} \end{pmatrix}, \quad \nabla^2 f^{IS}(x) = \begin{pmatrix} \frac{1}{x_1^2} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_2^2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_m^2} \end{pmatrix},$$

$$\nabla^2 f^{FD}(x) = \begin{pmatrix} \frac{1}{x_1(1-x_1)} & 0 & \cdots & 0 \\ 0 & \frac{1}{x_2(1-x_2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{x_m(1-x_m)} \end{pmatrix}, \quad \nabla^2 f^{SM}(x) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & m \end{pmatrix}$$

and

$$\nabla^2 f^{SE}(x) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Let the feasible set C is given by

$$C = \{x = (x_1, x_2, \dots, x_m)^T \in \mathbb{R}^m : \|x\| \leq 1 \text{ and } x_i \geq a > 0 \text{ for } i = 1, 2, \dots, m\},$$

where $a < \frac{1}{\sqrt{m}}$. Then it is easy to check that $\nabla^2 f^{KL}(x) \succeq I$, $\nabla^2 f^{IS}(x) \succeq I$, $\nabla^2 f^{FD}(x) \succeq I$, $\nabla^2 f^{SM}(x) \succeq I$ and $\nabla^2 f^{SE}(x) \succeq I$ for all $x \in C$. This implies that all functions are strongly convex on C with $\sigma = 1$.

Example 4.1. Let $A : \mathbb{R}^m \rightarrow \mathbb{R}^m$ ($m = 100$) be an operator given by

$$Ax = \frac{1}{\|x\|^2 + 1} \arg \min_{y \in \mathbb{R}^m} \left\{ \frac{\|y\|^4}{4} + \frac{1}{2} \|x - y\|^2 \right\}.$$

Then A is continuous pseudomonotone but not monotone. We choose $\theta = 0.333$, $\gamma = 2$, $l = 0.5$, $\mu = 0.38$, $\alpha_n = \frac{1}{n+1}$, $\xi_n = \alpha_n^2$ and two cases for θ_n , that is, $\theta_n = \theta_n^{max} := \bar{\theta}_n$ and $\theta_n = \theta_n^{min} := 0$. Note that when $\theta_n = \theta_n^{min}$, the Algorithms 1 and 2 are modified SEGMA and modified TEGMA without inertial terms, respectively. We use $E_n = \|u_n - y_n\| < 10^{-5}$ as the stopping criterion and the starting points x_0, x_1 are generated randomly in \mathbb{R}^m . In this experiments, we compare Algorithm 1 and Algorithm 2 with Algorithm 1 and Algorithm 2 without the inertial terms. The numerical results of our methods have been reported in the Table 1 and Figures 4.

Remark 4.2. From aforementioned numerical results as above, we summarize the performance of our methods as follows:

Table 1: Numerical results for Example 4.1

Bregman divergence	Alg. 1 ($\theta_n = \theta_n^{\min}$)		Alg. 1 ($\theta_n = \theta_n^{\max}$)		Alg. 2 ($\theta_n = \theta_n^{\min}$)		Alg. 2 ($\theta_n = \theta_n^{\max}$)	
	Iter.	Time	Iter.	Time	Iter.	Time	Iter.	Time
D_f^{KL}	25	0.0092	18	0.0041	16	0.0309	12	0.0124
D_f^{IS}	70	0.0163	65	0.0160	46	0.0395	38	0.0219
D_f^{FD}	15	0.0127	11	0.0097	8	0.0219	6	0.0068
D_f^{SM}	4	0.0082	2	0.0022	4	0.0156	3	0.0052

- (1) The Algorithm 1 and Algorithm 2 with relaxed inertial terms ($\theta_n = \theta_n^{\max}$) have a good running effect than the algorithms without relaxed inertial terms ($\theta_n = \theta_n^{\min}$) in each the Bregman divergence. This assured that adding the relaxed inertial term to algorithms has some effect like the classical inertial algorithms for solving the problem.
- (2) The Algorithm 1 and Algorithm 2 with the Bregman divergence D_f^{SM} have a number of iterations and elapsed times less than the algorithms with the Bregman divergences D_f^{KL} , D_f^{IS} and D_f^{FD} . This is because the structure of D_f^{SM} is not complicated to perform.

Next, we provide numerical experiments to illustrate the performance of our algorithms in solving the image deblurring problem and also compare them with Algorithm A and Algorithm B. Now, let $f(x) = \frac{1}{2}\|x\|^2$ for all $x \in H$.

Example 4.3. *The digital image restoration problem plays an important role in many applications of science and engineering such as film restoration, image and video coding, medical and astronomical imaging, etc. [18, 44, 51]. Restoring an image from a degraded one is typically an ill-posed inverse problem, which can be modelled by the following linear equation:*

$$b = Bx + v, \quad (4.1)$$

where $x \in \mathbb{R}^N$ is the original image, $b \in \mathbb{R}^M$ is the degraded image, $B \in \mathbb{R}^{M \times N}$ is the blurring matrix and v is an additive noise. An efficient method for recovering the original image is the ℓ_1 -norm regularized least square method given by

$$\min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \|Bx - b\|_2^2 + \lambda \|x\|_1 \right\}, \quad (4.2)$$

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 . Our main task is to restore the original image x given the data of the blurred image b . Several iterative algorithms have been introduced for treating such problems with the earliest being the projection method by Figureido et al. [18]. More so, the least square problem (4.2) can be expressed as a variational inequality problem by setting $A = B^T(Bx - b)$. It is known that the operator A in this case is monotone and $\|B^T B\|$ -Lipschitz continuous (hence it is pseudomonotone and uniformly continuous). We consider the grey scale image of M pixels wide and N pixel height, each value is known to be in the range $[0, 255]$. The quality of the restored image is measured by the signal-to-noise ratio defined by

$$SNR = 20 \log_{10} \left(\frac{\|x\|_2}{\|x - x^*\|_2} \right),$$

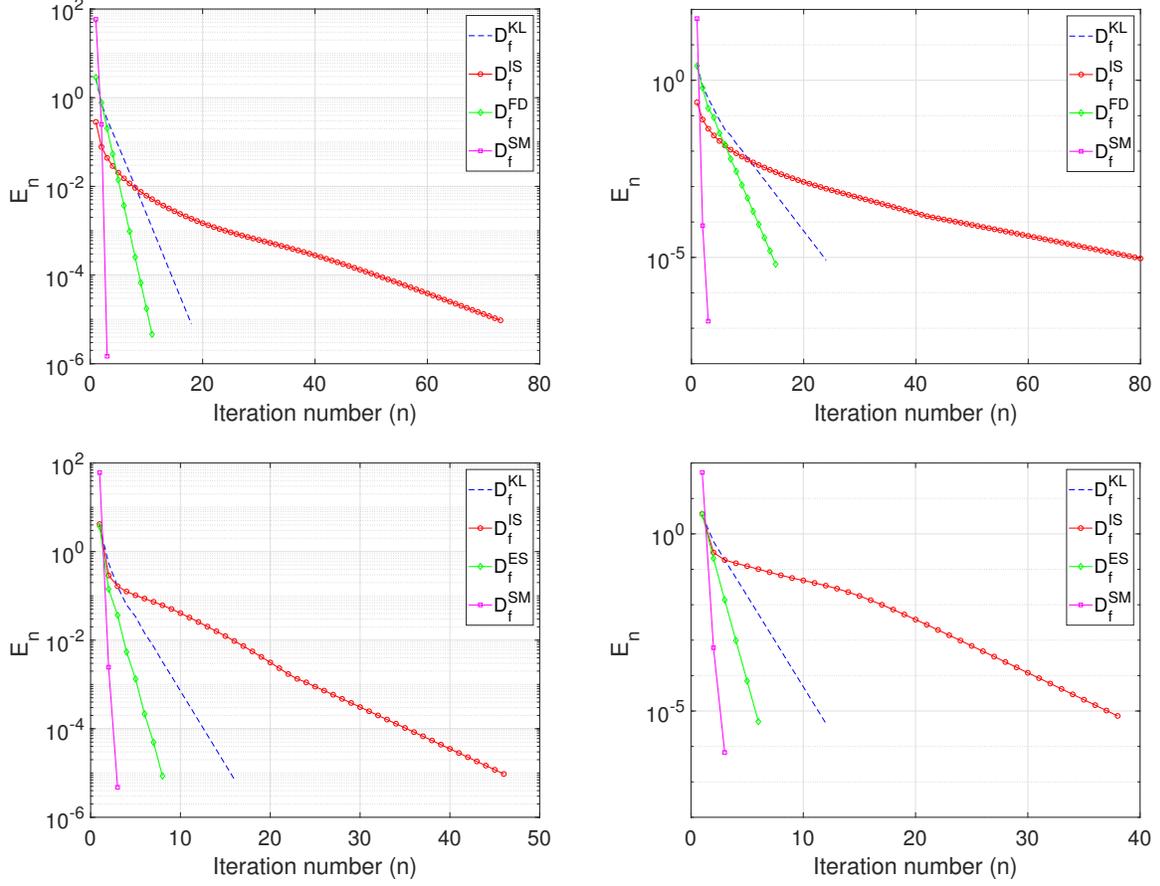


Figure 4: Numerical result for Example 4.1. Top: Algorithm 1, Left ($\theta_n = \theta_n^{\min}$), Right ($\theta_n = \theta_n^{\max}$); Bottom: Algorithm 2, Left ($\theta_n = \theta_n^{\min}$), Right ($\theta_n = \theta_n^{\max}$).

where x is the original image and x^* is the restored image. Note that the larger the value of SNR, the better the quality of the restored image. In our experiments, we use the grey test image *Pout* (291×240) and *Cameraman* (256×256), each test image is degraded by Gaussian 7×7 blur kernel with standard deviation 4. We choose $\gamma = 2, l = 0.36, \mu = 0.64, x_0 = \mathbf{0} \in \mathbb{R}^{\mathbf{D}}$ and $x_1 = \mathbf{1} \in \mathbb{R}^{\mathbf{D}}$, where $\mathbf{D} = M \times N$. Also, we choose $\alpha_n = \frac{1}{200(n+1)}, \xi_n = \alpha_n^2, \theta = 0.0266$ and $\theta_n = \theta_n^{\max} := \bar{\theta}_n$. Figure 5 and 6 show the original, blurred and restored image by using the Algorithms 1, 2, A and B. Also, Figure 7 shows the graph of SNR against number of iterations for each test image using the algorithms. More so, we report the time (in seconds) for each algorithm in Table 2. The computational results shows that Algorithms 1 and 2 are more efficient for restoring the degraded image than Algorithms A and B.

5 Conclusions

In this paper, by using the Halpern iteration method, we have proposed and analysed two relaxed inertial type algorithms with Bregman divergence for approximating solutions of pseudomonotone variational inequalities

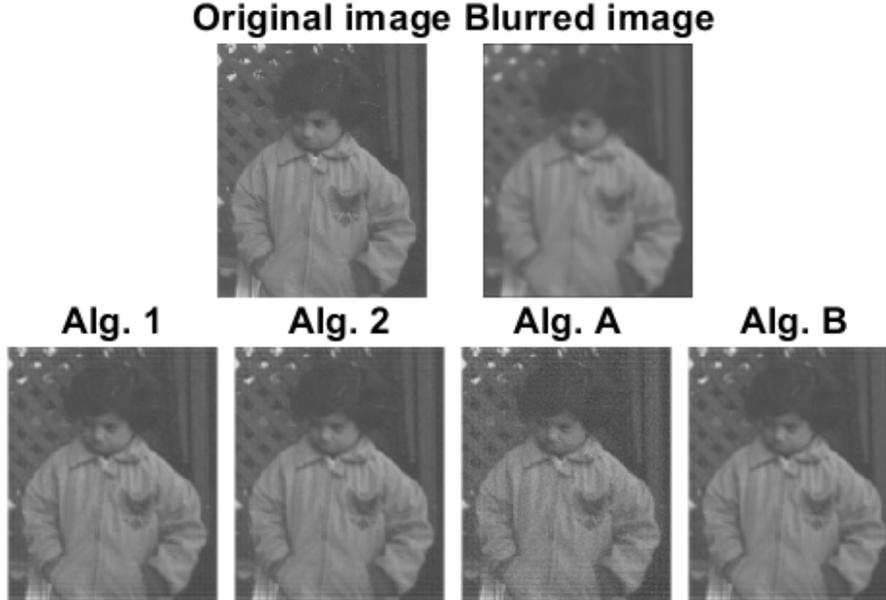


Figure 5: Example 4.3, Top shows original image of Pout (left) and degraded image of Pout (right); Bottom shows recovered image by Algorithm 1, Algorithm 2, Algorithm A and Algorithm B.

Table 2: Computational result for Example 4.3

Algorithms	Pout		Cameraman	
	Time (secs)	SNR	Time (secs)	SNR
Alg. 1	28.6139	34.2679	26.0414	31.0415
Alg. 2	26.9383	34.3372	24.6394	31.0582
Alg. A	38.6904	34.0122	26.5851	33.3580
Alg. B	45.6154	32.5071	36.8937	29.6873

in real Hilbert spaces. The strong convergence of the sequences generated by the proposed algorithms are established without assuming the Lipschitz continuity and the sequential weak continuity of the cost operator. Some numerical experiments have been performed to illustrate the performance and efficiency of the proposed methods in comparison with some existing methods.

Acknowledgments

L.O. Jolaoso was supported by the Postdoctoral research funding from the Research Office, Sefako Makgatho Health Sciences University, South Africa and P. Cholaamjiak was supported by Thailand Science Research and Innovation under the project IRN62W0007.

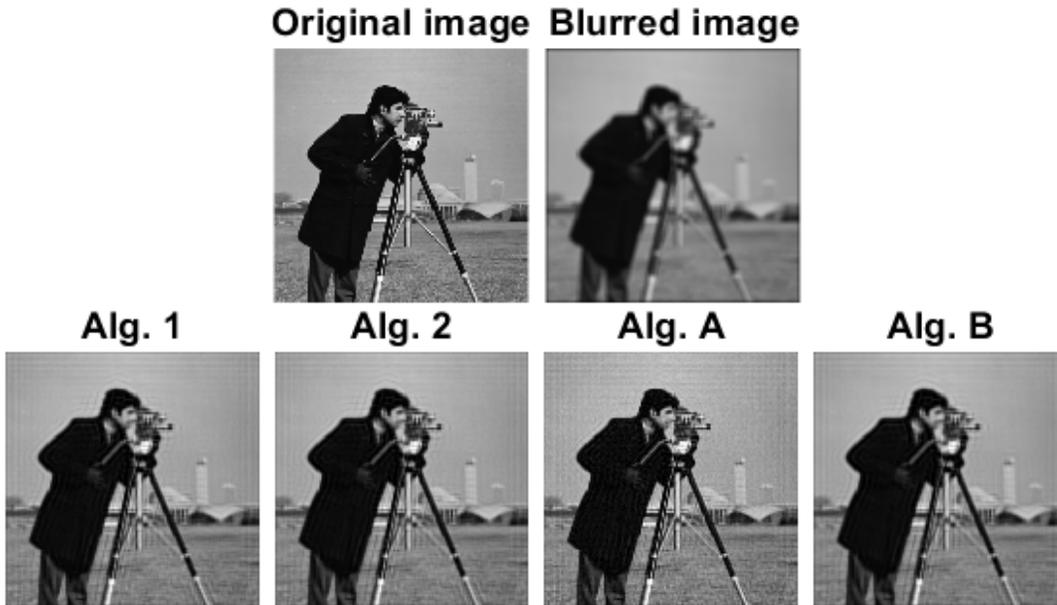


Figure 6: Example 4.3, Top shows original image of Cameraman (left) and degraded image of Cameraman (right); Bottom shows recovered image by Algorithm 1, Algorithm 2, Algorithm A and Algorithm B.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

- [1] F. Alvarez and H. Attouch, An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping, *Set-Valued Anal.*, 9, pp. 3–11 (2001).
- [2] E. Amid, M. K. Warmuth, R. Anil and T. Koren, Robust Bi-Tempered Logistic Loss Based on Bregman Divergences, *Conference on Neural Information Processing Systems*, pp. 14987–14996 (2019).
- [3] P.K. Anh, D.V. Thong and N.T. Vinh, Improved inertial extragradient methods for solving pseudomonotone variational inequalities, *Optimization* (2020), DOI: 10.1080/02331934.2019.1616191
- [4] A.S. Antipin, On a method for convex programs using a symmetrical modification of the Lagrange function, *Ekonomika i Matematicheskie Metody*, 12, pp. 1164–1173 (1976).
- [5] H.H. Bauschke, X. Wang, J. Ye and X. Yuan, Bregman distances and Chebyshev sets, *Journal of Approximation Theory*, 159, pp. 3–25 (2009).
- [6] A. Beck, *First-Order Methods in Optimization*, Society for Industrial and Applied Mathematics, Philadelphia (2017).

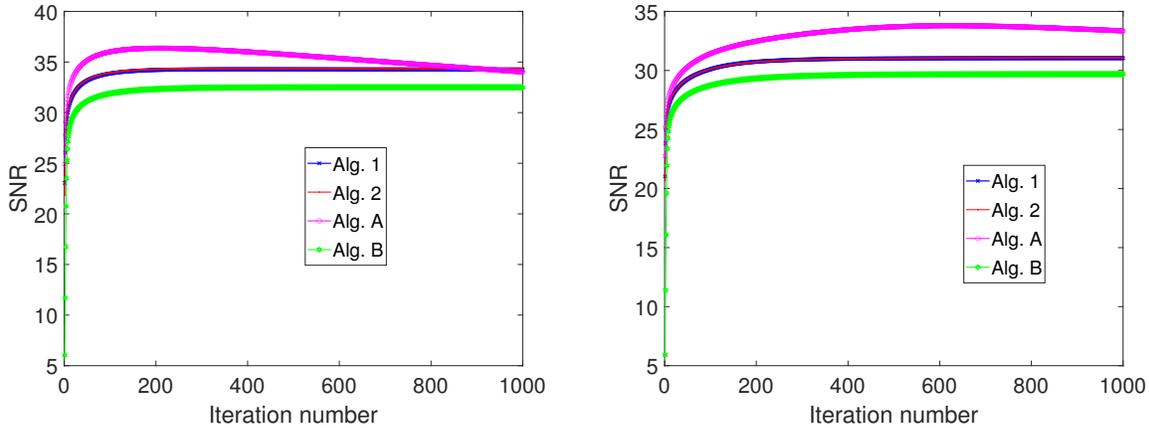


Figure 7: Example 4.3: Graphs of SNR values against number of iteration for Pout (Left) and Cameraman (Right).

- [7] A. Banerjee, S. Merugu, I.S. Dhillon J. and Ghosh, Clustering with Bregman divergences, *Journal of machine learning research*, 6(Oct), pp. 1705–1749 (2005).
- [8] R. I. Bot, E. R. Csetnek and P. T. Vuong, The forward–backward–forward method from continuous and discrete perspective for pseudo-monotone variational inequalities in Hilbert spaces, *European Journal of Operational Research*, 287, pp. 49–60 (2020).
- [9] L.M. Bregman, The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming, *USSR Computational Mathematics and Mathematical Physics*, 7(3), pp. 200–217 (1967).
- [10] 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), (Art. ID 84919), pp. 1–39 (2006).
- [11] D. Butnariu, S. Reich and A.J. Zaslavski, Asymptotic behavior of relatively nonexpansive mappings in Banach spaces, *J. Appl. Anal.* 7, pp. 151–174 (2001)
- [12] Y. Censor, A. Gibali and S. Reich, The subgradient extragradient method for solving variational inequalities in Hilbert space, *J. Optim. Theory Appl.*, 148, pp. 318–335 (2011).
- [13] Z. Chbani and H. Riahi, Weak and strong convergence of an inertial proximal method for solving Ky Fan minimax inequalities, *Optimization letters*, 7, pp. 185–206 (2013).
- [14] P. Cholamjiak, D.V. Thong and Y.J. Cho, A Novel Inertial Projection and Contraction Method for Solving Pseudomonotone Variational Inequality Problems, *Acta Applicandae Mathematicae*, 169, pp. 217–245 (2020).
- [15] P.L. Combettes and J.C. Pesquet, Deep neural structures solving variational inequalities, *Set-Valued Var. Anal.*, 28, pp. 491–518 (2020).
- [16] R.W. Cottle and J.C. Yao, Pseudo-monotone complementarity problems in Hilbert space, *J. Optim. Theory Appl.*, 75, pp. 281–295 (1992).

- [17] S. V. Denisov, V. V. Semenov and L. M. Chabak, Convergence of the Modified Extragradient Method for Variational Inequalities with Non-Lipschitz Operators, *Cybernetics and Systems Analysis*, 51, pp. 757–765 (2015).
- [18] Figueiredo MAT, Nowak RD, Wright SJ, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems, *IEEE J Sel Top Signal Process*, 1, pp. 586–597 (2007).
- [19] A. Gibali, A new Bregman projection method for solving variational inequalities in Hilbert spaces, *Pure Appl Funct Anal.*, 3, pp. 403–415 (2018).
- [20] B. Halpern, Fixed points of nonexpanding maps, *Bull. Am. Math. Soc.* 73, pp. 95–961 (1967).
- [21] P. Hartman and G. Stampacchia, On some non-linear elliptic differential-functional equations, *Acta Math.*, 115, pp. 271–310 (1966).
- [22] D.V. Hieu, Y.J. Cho, Y.B. Xiao and P. Kumam, Relaxed extragradient algorithm for solving pseudomonotone variational inequalities in Hilbert spaces, *Optimization*, 69, pp. 2279–2304 (2020).
- [23] D.V. Hieu and P. Cholamjiak, Modified extragradient method with Bregman distance for variational inequalities, *Applicable Analysis* (2020), <https://doi.org/10.1080/00036811.2020.1757078>
- [24] A.N. Iusem and R. Gárciga Otero, Inexact versions of proximal point and augmented Lagrangian algorithms in Banach spaces, *Numer. Funct. Anal. Optim.*, 22, pp. 609–640 (2001).
- [25] A.N. Iusem and B.F. Svaiter, A variant of Korpelevich’s method for variational inequalities with a new search strategy, *Optimization*, 42, pp. 309–321 (1997).
- [26] L.O. Jolaoso, An inertial projection and contraction method with a line search technique for variational inequality and fixed point problems, *Optimization* (2021), <https://doi.org/10.1080/02331934.2021.1901289>
- [27] A. Juditsky and A. Nemirovski, Solving variational inequalities with monotone operators on domains given by Linear Minimization Oracles, *Mathematical Programming*, 156, pp. 221–256 (2016).
- [28] P.D. Khanh and P.T. Vuong, Modified projection method for strongly pseudo-monotone variational inequalities, *Journal of Global Optimization*, 58, pp. 341–350 (2014).
- [29] P.Q. Khanh, D.V. Thong and N.T. Vinh, Versions of the Subgradient Extragradient Method for Pseudomonotone Variational Inequalities, *Acta Applicandae Mathematicae*, 170, pp. 319–345 (2020).
- [30] Y. Kimura, K. Nakajo and W. Takahashi, Convexity of the set of fixed points of a quasi-pseudocontractive type Lipschitz mapping and the shrinking projection method, *Sci. Math. Jpn.* 70, pp. 213–220 (2009).
- [31] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and their Applications*, Academic press Inc., New York (1980).
- [32] G.M. Korpelevich, The extragradient method for finding saddle points and other problems, *Matecon*, 12, pp. 747–756 (1976).
- [33] L. Liu and X. Qin, Strong convergence of an extragradient-like algorithm involving pseudo-monotone mappings, *Numerical Algorithms*, 83, pp. 1577–1590 (2020).

- [34] M.J. Luo and Y. Zhang, Robust solutions to box-constrained stochastic linear variational inequality problem, *Journal of Inequalities and Applications*, 2017, Article number: 253 (2017).
- [35] P.E. Maingé, Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization, *Set-Valued Anal.*, 16, pp. 899–912 (2008).
- [36] Y.V. Malitsky and V.V. Semenov, An extragradient algorithm for monotone variational inequalities, *Cybern. Syst. Anal.*, 50, pp. 271–277 (2014).
- [37] A. Moudafi, Viscosity approximation methods for fixed-points problems, *J. Math. Anal. Appl.* 241, pp. 46–55 (2000).
- [38] K. Nakajo and W. Takahashi, Strong convergence theorems for nonexpansive mappings and nonexpansive semigroups, *J. Math. Anal. Appl.* 279, pp. 372–379 (2003).
- [39] B.T. Polyak, Some methods of speeding up the convergence of iterative methods, *Zh. Vychisl. Mat. Mat. Fiz.*, 4, pp. 1–17 (1964).
- [40] L.D. Popov, A modification of the Arrow-Hurwicz method for searching for saddle points, *Mat. Zametki*, 28, pp. 777–784 (1980).
- [41] S. Reich, A weak convergence theorem for the alternating method with Bregman distances, in: *Theory and Applications of Nonlinear Operators of Accretive and Monotone Type*, Marcel Dekker, New York, pp. 313–318 (1996).
- [42] S. Reich and S. Sabach, A strong convergence theorem for a proximal-type algorithm in reflexive Banach spaces, *J. Nonlinear Convex Anal.*, 10, pp. 471–485 (2009).
- [43] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
- [44] Y. Shehu, O.S. Iyiola, F.U. Ogbuisi, Iterative method with inertial terms for nonexpansive mappings, Applications to compressed sensing, *Numer. Algorithms*, 83, pp. 1321–1347 (2020).
- [45] W. Takahashi, *Introduction to nonlinear and convex analysis*, Yokohama publishers (2009).
- [46] B. Tan, X. Qin and J.C. Yao, Two modified inertial projection algorithms for bilevel pseudomonotone variational inequalities with applications to optimal control problems, *Numerical Algorithms* (2021), <https://link.springer.com/article/10.1007/s11075-021-01093-x>
- [47] D.V. Thong and P.T. Vuong, Modified Tseng’s extragradient methods for solving pseudo-monotone variational inequalities, *Optimization*, 68, pp. 2207–2226 (2019).
- [48] D.V. Thong, N.T. Vinh and D.V. Hieu, Accelerated hybrid and shrinking projection methods for variational inequality problems, *Optimization*, vol. 68, issue 5, pp. 981–998 (2019).
- [49] D.V. Thong, X.H. Li, Q.L. Dong, Y.J. Cho and T.M. Rassias, An inertial Popov’s method for solving pseudomonotone variational inequalities, *Optimization Letters*, 15, pp. 757–777 (2021).
- [50] P. Tseng, A modified forward-backward splitting method for maximal monotone mapping, *SIAM J. Control. Optim.*, 38, pp. 431–446 (2000).

- [51] Y. Xiao, H. Zhu, A conjugate gradient method to solve convex constrained monotone equations with applications in compressive sensing, *J. Math. Anal. Appl.*, 405, pp. 310–319 (2013).
- [52] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, Publishing Co., Inc., River Edge, NJ, 2002.