Pure and Applied Functional Analysis
Volume 8, Number 3, 2023, 881–902

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

AN INERTIAL ITERATIVE ALGORITHM FOR APPROXIMATING SOLUTIONS TO VARIATIONAL INCLUSION PROBLEMS IN BANACH SPACES

OLAWALE KAZEEM OYEWOLE AND SIMEON REICH*

ABSTRACT. We consider the problem of approximating an element in the solution set of a variational inclusion. We propose an inertial iterative method for solving such problems for the sum of two monotone operators in the framework of real reflexive Banach spaces. In order to achieve strong convergence of the generated approximating sequences, we use a modified Halpern method. Some numerical experiments and applications of our main result are also presented.

1. INTRODUCTION

Let K be a nonempty, closed and convex subset of a real Banach space E with dual E^*. Let $A : E \to E^*$ be a single-valued nonlinear mapping and $B : E \to 2^{E^*}$ be a set-valued mapping. We consider the following Variational Inclusion Problem (VIP): Find a point $x^* \in E$ such that

$$0 \in Ax^* + Bx^*. \tag{1.1}$$

The point x^* is called a zero of the sum $A+B$. We denote by Γ the set of solutions of (1.1), that is, $\Gamma = (A+B)^{-1}(0) = \{x \in E : 0 \in Ax + Bx\}$. It is well known that the VIP (1.1) finds application in solving problems such as convex minimization problems, variational inequality problems and convex-concave saddle point problems. The VIP also comes in handy in real life applications such as compressed sensing, image processing and restoration, computer vision, machine learning and signal processing. For this reason, the VIP has received considerable amount of study with several iterative methods introduced for approximating its solutions in Hilbert and Banach spaces (see [24, 32, 35, 50]). One of the most popular methods for solving the VIP (1.1) is the forward-backward splitting method given by: $x_1 \in E$ and

$$x_{n+1} = (I + \lambda B)^{-1}(I - \lambda A)x_n, \ n \geq 1.$$

As the name implies, this method involves an explicit forward step with respect to the mapping A which is followed by an implicit backward step with respect to the

2020 *Mathematics Subject Classification.* 47H09, 49J25, 65K10, 90C25.
Key words and phrases. Banach space, inertial technique, image processing, monotone operator, variational inclusion problem.
* Simeon Reich was partially supported by the Israel Science Foundation (Grant 820/17), by the Fund for the Promotion of Research at the Technion (Grant 2001893) and by the Technion General Research Fund (Grant 2016723).

mapping B. The splitting method is known to yield weak convergence to a solution of the VIP. For more literature on the forward backward method, the following references are useful: [33, 34, 50]. In the setting of Banach spaces, Fang and Huang [27] introduced the concept of H-accretive operators and defined a corresponding resolvent operator in the framework of q-uniformly smooth Banach spaces. We note that in the result of Fang and Huang [27], the mappings A and B are defined from $E \to E$ and from $E \to 2^E$, respectively; thus their proposed method cannot be used to solve the VIP (1.1). Inspired by this drawback, Xia and Huang [52] introduced another operator which they termed an H-monotone operator. By using this operator, they introduced a method for approximating a solution to the VIP (1.1), where E is a uniformly smooth Banach space. In 2020 Ogbuisi and Izuchukwu [38] considered the problem of approximating a solution to (1.1) in the setting of a reflexive Banach space. They introduced a resolvent operator and presented its properties. Furthermore, they proposed a hybrid iterative algorithm for finding the zeroes of (1.1), which are also fixed points of a suitable nonlinear mapping.

In another direction, the inertial method for approximating a solution to the VIP (1.1) was introduced by Alvarez and Attouch [3]. They introduced the following method for finding a zero of (1.1), where $A = 0$ and B is a maximal monotone operator. Given $x_0, x_1 \in E$, let

$$\begin{cases} y_n = x_n + \beta_n(x_n - x_{n-1}) \\ x_{n+1} = (I + \lambda_n B)^{-1} y_n, \ n \geq 0. \end{cases} \tag{1.2}$$

They established the weak convergence of the sequences generated by this method to a solution of the VIP (1.1) in a real Banach space under some mild assumptions on $\{\beta_n\}$ and $\{\lambda_n\}$. The step concerning y_n in the above algorithm is called the extrapolation step and is based on the heavy ball method. For $B = 0$ the backward step reduces to the identity and the sequence $\{x_n\}$ given by

$$x_{n+1} = (x_n - \lambda_n A x_n + \beta_n(x_n - x_{n-1})), \ n \geq 0,$$

is usually referred to as the heavy ball method. It is an explicit finite difference discretization of the so-called heavy ball with friction dynamical system

$$\frac{d^2x}{dt^2} + \gamma \frac{dx}{dt} + \nabla f(x(t)) = 0. \tag{1.3}$$

Equation (1.3) arises from the application of Newton's law to a point subjected to a frictional force determined by a constant coefficient of friction $\gamma > 0$ and the velocity $\frac{dx}{dt}$, and the gravitational force f. This is the reason why the term *heavy ball method* is used. The term $\beta_n(x_n - x_{n-1})$ is called the inertial force. Thus, an inertial-type algorithm uses the previous iterates to obtain the next one. The term $\{\beta_n\}$ controls the momentum $x_n - x_{n-1}$. Polyak [40] suggested that this combination of the previous iterates was a very good way of accelerating iterative algorithms (see also [33]).

The inertial method for approximating a solution to the VIP (1.1) has been well studied in Hilbert space (see [4, 22, 25, 34]). Inertia-based algorithms have also been considered in the framework of Banach spaces. We observe the following points of reference which are present in most of these results:

(i) To obtain strong convergence, hybrid algorithms are employed (see [20, 21, 48]).
(ii) The direction of the momentum $x_n - x_{n-1}$ was changed in [1, 5, 30]. That is, $\beta_n(x_{n-1} - x_n)$ was used in place of $\beta_n(x_n - x_{n-1})$.

While seeking a way to avoid these two features, we came across the result of Adamu et al. [2], where they considered an inertial-type algorithm for approximating a solution to the VIP (1.1) without the highlighted features in the framework of 2-uniformly convex Banach spaces.

Motivated by the literature cited above, in particular, the papers of Polyak [40], Ogbuisi and Izuchukwu [38], and Adamu et al. [2], we propose in the present paper an inertia-based iterative algorithm for approximating a zero of the sum of two monotone operators in a class of reflexive Banach spaces which is more general than the class of uniformly convex Banach spaces. Our method combines the inertial technique and the modified Halpern iterative process. Using this method, we have established a strong convergence result for approximating a solution to the VIP.

The rest of our paper is organized as follows: first, we recall some useful definitions and preliminary results in Section 2. In Section 3 we introduce our proposed method, state our main result and present its convergence analysis. Some applications of our main result are presented in Section 4. In Section 5 we present the results of numerical experiments which demonstrate the efficiency of our method. We provide some concluding remarks in Section 6.

2. Preliminaries

In this section we give some definitions and preliminary results which are used in our convergence analysis. Let K be a nonempty, closed and convex subset of a real Banach space E with norm $\|\cdot\|$ and dual space E^*. We denote the weak and strong convergence of a sequence $\{x_n\} \subset E$ to a point $x \in E$ by $x_n \rightharpoonup x$ and $x_n \to x$, respectively.

A function $f : E \to (-\infty, +\infty]$ is said to be

(i) proper if $\mathrm{dom}(f) = \{x \in E : f(x) < \infty\} \neq \emptyset$;
(ii) Gâteaux differentiable at $x \in E$ if there exists an element in E^*, denoted by $f'(x)$ or by $\nabla f(x)$, such that

$$\lim_{t\to 0} \frac{f(x+ty) - f(x)}{t} = \langle y, f'(x)\rangle \quad \forall y \in E,$$

where $f'(x)$ or $\nabla f(x)$ is called the Gâteaux differential or gradient of f at x;
(iii) strongly convex with a strong convexity constant $\rho > 0$ if

$$f(x) \geq f(y) + \langle \nabla f(y), x - y\rangle + \frac{\rho}{2}\|x - y\|^2 \quad \forall\, x, y \in E;$$

(iv) Fréchet differentiable at x if the limit in (ii) above exists uniformly on the unit sphere of E.

Lemma 2.1 ([43]). *Let $f : E \to \mathbb{R}$ be uniformly Fréchet differentiable and bounded on bounded subsets of E. Then ∇f is uniformly continuous on bounded subsets of E from the strong topology of E to the strong topology of E^*.*

The subdifferential set of f at a point x, denoted by ∂f, is defined by

$$\partial f(x) := \{x^* \in E^* : f(x) - f(y) \leq \langle y - x, x^* \rangle \ \ \forall y \in E\}.$$

Every element $x^* \in \partial f(x)$ is called a subgradient of f at x. If f is continuously differentiable, then $\partial f(x) = \{\nabla f(x)\}$ is the gradient of f at x. The Fenchel conjugate of f is the convex functional $f^* : E^* \to \mathbb{R} \cup \{+\infty\}$ defined by $f^*(x^*) = \sup\{\langle x^*, x\rangle - f(x) : x \in E\}$. We note that $\nabla f(\nabla f^*(x^*)) = x^*$ for all $x^* \in E^*$. Let E be a reflexive Banach space. The function f is said to be Legendre if it satisfies the following two conditions (see [6]):

(L1) int dom$(f) \neq \emptyset$ and ∂f is single-valued on its domain;
(L2) int dom$(f^*) \neq \emptyset$ and ∂f^* is single-valued on its domain.

Let f be a strongly convex and Gâteaux differentiable function. The function D_f: dom $(f)\times$int dom (f)$\to [0, \infty)$, defined by

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

is called the Bregman distance induced the function f. It is worth mentioning that the bifunction D_f is not a metric in the usual sense because it does not satisfy the symmetry and the triangle inequality properties. However, it does possess the following important property called the three point identity:

$$D_f(x, y) + D_f(y, z) - D_f(x, z) = \langle \nabla f(z) - \nabla f(y), x - y\rangle, \tag{2.1}$$

where $x \in$ dom(f) and $y, z \in$ int dom(f). Also, from the strong convexity of f and the definition of the Bregman distance it follows that

$$D_f(x, y) \geq \frac{\rho}{2}\|x - y\|^2. \tag{2.2}$$

The Bregman distance function has been widely used by many authors in the literature (see [6, 14, 15, 41] and references therein).

Remark 2.2. The following important examples of Bregman distance functions can be found, for instance, in [8].

(i) If $f(x) = \frac{1}{2}\| \cdot \|^2$, then $D_f(x, y) = \frac{1}{2}\|x - y\|^2$.
(ii) Let $f(x) = -\sum_{j=1}^{m} x_j \log(x_j)$. This is Shannon's entropy for the nonnegative orthant $\mathbb{R}^m_{++} := \{x \in \mathbb{R}^m : x_j > 0\}$. It induces the Kullback-Leibler cross entropy defined by

$$D_f(x, y) := \sum_{j=1}^{m} \left(x_j \log\left(\frac{x_j}{y_j}\right) - 1\right) + \sum_{j=1}^{m} y_j. \tag{2.3}$$

We also require the functional $V_f : E \times E^* \to [0, +\infty]$ associated with the function f, defined by

$$V(x, x^*) = f(x) - \langle x, x^*\rangle + f^*(x^*) \ \forall\ x \in E,\ x^* \in E^*.$$

We have $V_f(x, x^*) = D_f(x, \nabla f^*(x^*))$ for all $x \in E$ and $x^* \in E^*$. Moreover, by the subdifferential inequality, we also have

$$V_f(x, x^*) + \langle \nabla f^*(x^*) - x, y^*\rangle \leq V_f(x, x^* + y^*)$$

for all $x \in E$ and $x^*, y^* \in E^*$ (see [31]).

Definition 2.3 ([13,16]). The bifunction v_f : int dom$(f) \times [0,+\infty)$, defined by

$$v_f(x,t) := \inf\{D_f(y,x) : y \in dom(f), \|y-x\| = t\},$$

is called the modulus of total convexity of f at x. The function f is called totally convex at $x \in$ int dom(f) if $v_f(x,t)$ is positive for all $t > 0$. The modulus of total convexity of f on K is the bifunction v_f : *int dom*$(f) \times [0,+\infty)$, defined by

$$v_f(K,t) := \inf\{v_f(x,t) : x \in K \cap int\ dom(f)\}.$$

The function f is said to be totally convex on bounded subsets if $v_f(K,t) > 0$ for any nonempty and bounded subset K and any $t > 0$. Also, f is said to be strongly coercive if $\lim_{\|x\|\to+\infty} \left|\frac{f(x)}{\|x\|}\right| = +\infty$.

Definition 2.4 ([11]). Let K be a nonempty, closed and convex subset of a reflexive real Banach space E. The Bregman projection of $x \in$ int dom(f) onto $K \subset$ int dom(f) is the unique vector $\Pi_K x \in K$ which satisfies

$$D_f(\Pi_K x, x) = \inf\{D_f(y,x) : y \in K\}.$$

Lemma 2.5 ([16]). *Let K be a nonempty, closed and convex subset of E and $x \in E$. Let $f : E \to \mathbb{R}$ be a Gâteaux differentiable and totally convex function. Then*

(i) $q = \Pi_K x$ *if and only if* $\langle \nabla f(x) - \nabla f(q), y - q\rangle \leq 0$ *for all* $y \in K$;
(ii) $D_f(y, \Pi_K x) + D_f(\Pi_K(x), x) \leq D_f(y,x)$ *for all* $y \in K$.

Proposition 2.6 ([13]). *If $x \in$ int dom(f), then the following assertions are equivalent:*

(i) *the function f is totally convex at x;*
(ii) *f is sequentially consistent at x, that is, for any sequence $\{y_n\} \subset$ dom(f), we have*

$$\lim_{n\to\infty} D_f(y_n, x) = 0 \Rightarrow \lim_{n\to\infty} \|y_n - x\| = 0.$$

We also recall (see [13]) that the function f is called sequentially consistent if for any two sequences $\{x_n\}$ and $\{y_n\}$ in E such that the first one is bounded, we have

$$\lim_{n\to\infty} D_f(x_n, y_n) = 0 \Rightarrow \lim_{n\to\infty} \|x_n - y_n\| = 0. \tag{2.4}$$

Proposition 2.7 ([13]). *If dom(f) contains at least two points, then the function f is totally convex on bounded sets if and only if the function f is sequentially consistent.*

Proposition 2.8 ([45]). *Let $f : E \to \mathbb{R}$ be a Gâteaux differentiable and totally convex function. If $\bar{x} \in E$ and the sequence $\{D_f(x_n, \bar{x})\}$ is bounded, then the sequence $\{x_n\}$ is bounded too.*

Lemma 2.9 ([39]). *If $f : E \to (-\infty,+\infty]$ is a proper, lower semicontinuous and convex function, then $f^* : E^* \to (-\infty,+\infty]$ is a proper, weak* lower semicontinuous and convex function. Thus for all $y \in E$, we have*

$$D_f\left(y, \nabla f^*\left(\sum_{i=1}^N \lambda_i \nabla f(x_i)\right)\right) \leq \sum_{i=1}^N \lambda_i D_f(y, x_i),$$

where $\{x_i\}_i^N \subset E$ and $\{\lambda_i\}_i^N \subset (0,1)$ with $\sum_{i=1}^N \lambda_i = 1$.

Let K be a nonempty, closed and convex subset of a real Banach space E, and let $S : K \to K$ be a mapping. A point $x \in K$ is said to be a fixed point of S if $x = Sx$. A point x is called an asymptotic fixed point of S [42] if K contains a sequence $\{x_n\}_{n\geq 1}$ such that $x_n \rightharpoonup x$ and $\|x_n - Sx_n\| \to 0$ as $n \to \infty$. We denote by $F(S)$ and by $\hat{F}(S)$ the set of fixed points and the set of asymptotic fixed points of S, respectively. If $S : K \to int\ dom(f)$ is a mapping, then S is said to be

(a) Bregman firmly nonexpansive (BFNE) if

$$\langle \nabla f(Sx) - \nabla f(Sy), Sx - Sy\rangle \leq \langle \nabla f(x) - \nabla f(y), Sx - Sy\rangle \ \ \forall\, x, y \in K;$$

(b) Quasi-Bregman firmly nonexpansive (QFNE) if $F(S) \neq \emptyset$ and

$$\langle \nabla f(x) - \nabla f(Sx), x - y\rangle \geq 0 \ \ \forall\, x \in K,\ y \in F(S);$$

(c) Bregman strongly nonexpansive (BSNE) with respect to $\hat{F}(S)$ if

$$D_f(y, Sx) \leq D_f(y, x) \ \ \forall\, x \in K,\ y \in \hat{F}(S)$$

and for any bounded sequence $\{x_n\}_{n\geq 1} \subset K$,

$$\lim_{n\to\infty} D_f(y, x_n) - D_f(y, Sx_n) = 0$$

implies that

$$\lim_{n\to\infty} D_f(x_n, Sx_n) = 0$$

(see [29]). It is not difficult to see that if $F(S) = \hat{F}(S) \neq \emptyset$, then

$$BFNE \subset QFNE \subset BSNE.$$

If $h : E \to \mathbb{R} \cup \{\infty\}$ is a proper, convex and lower semicontinuous function, then the proximal map of h of order $\lambda > 0$ is given by

$$prox_{\lambda h}(x) = \arg\min_{y\in E} \left\{ h(x) + \frac{1}{2\lambda} D_f(x, y) \right\}.$$

A mapping $A : E \to 2^{E^*}$ is called monotone if for any $x, y \in domA$, we have

$$u \in Ax \text{ and } v \in Ay \Rightarrow \langle u - v, x - y\rangle \geq 0. \tag{2.5}$$

A is said to be maximal monotone if A is monotone and the graph of A is not properly contained in the graph of any other monotone mapping. The resolvent associated with the mapping A of order λ for any $\lambda > 0$ is the operator $J^f_{\lambda A} : E \to 2^E$ defined by

$$J^f_{\lambda A} := (\nabla f + \lambda A)^{-1} \circ \nabla f.$$

Let K be a nonempty, closed and nonempty subset of a reflexive Banach space E. The mapping $A : E \to 2^{E^*}$ is called Bregman inverse strongly monotone (BISM) on the set K if

$$K \cap dom(f) \cap int\ dom(f) \neq \emptyset$$

and for any $x, y \in K \cap int\ dom(f)$, $u \in Ax$ and $v \in Ay$, we have

$$\langle u - v, \nabla f^*(\nabla f(x) - u) - \nabla f^*(\nabla f(y) - v)\rangle \geq 0. \tag{2.6}$$

It is known (see [29]) that the class of BISM mappings is more general than the class of firmly nonexpansive operators in Hilbert spaces. To see this, let $f(\cdot) = \frac{1}{2}\|\cdot\|^2$. Then $\nabla f = \nabla f^* = I$, where I is the identity operator and (2.6) becomes

$$\langle u - v, x - y\rangle \geq \|u - v\|^2.$$

For more information on this class of operators, see [12, 44]. Let $A : E \to 2^{E^*}$ be a mapping. Then the anti-resolvent associated with A of order $\lambda > 0$ is the mapping $A^f_\lambda : E \to 2^E$ given by

$$A^f_\lambda := \nabla f^* \circ (\nabla f - \lambda A).$$

Proposition 2.10 ([38]). *Let $B : E \to 2^{E^*}$ be a maximal monotone operator and $A : E \to E^*$ be a Bregman inverse strongly monotone operator such that $(A + B)^{-1}(0) \neq \emptyset$. Let $f : E \to \mathbb{R}$ be a Legendre function which is uniformly Fréchet differentiable and bounded on bounded subsets of E. Then we have*

(i) $(A + B)^{-1}(0) = F(J^f_{\lambda B} \circ A^f_\lambda)$;
(ii) $J^f_{\lambda B} \circ A^f_\lambda$ *is a BSNE operator such that* $F(J^f_{\lambda B} \circ A^f_\lambda) = \hat{F}(J^f_{\lambda B} \circ A^f_\lambda)$;
(iii) $D_f(p, J^f_{\lambda B} \circ A^f_\lambda x) + D_f(J^f_{\lambda B} \circ A^f_\lambda x, x) \leq D_f(p, x)$ *for any* $p \in (A + B)^{-1}(0)$, $x \in X$ *and* $\lambda > 0$.

Lemma 2.11 ([47]). *Let $\{a_n\}$ be a sequence of nonnegative real numbers, $\{\alpha_n\}$ be a sequence of real numbers in $(0, 1)$ such that $\sum_{n=1}^{\infty} \alpha_n = \infty$ and $\{b_n\}$ be a sequence of real numbers. Assume that*

$$a_{n+1} \leq (1 - \alpha_n)a_n + \alpha_n b_n \ \forall\ n \geq 1.$$

If $\limsup_{k\to\infty} b_{n_k} \leq 0$ for every subsequence $\{a_{n_k}\}$ of $\{a_n\}$ satisfying the condition

$$\liminf_{k\to\infty}(a_{n_k+1} - a_{n_k}) \geq 0,$$

then $\lim_{n\to\infty} a_n = 0$.

3. Main result

In this section we first propose a strongly convergent algorithm for approximating a solution to the VIP (1.1) and then present its convergence analysis. Recall that the solution set Γ of the problem is closed and convex [38, 52].

Algorithm 3.1. Modified Halpern-Inertial Iterative Method (MHIM)

Initialization: Choose $v, x_0, x_1 \in E$ and $\theta > 0$. For all $n \in \mathbb{N}$, let $\{\alpha_n\}, \{\beta_n\}, \{\gamma_n\}$ be sequences in (0,1) such that $\alpha_n + \beta_n + \gamma_n = 1$.
Step 1: Given x_n, x_{n-1}, choose θ_n such that $\theta_n \in [0, \bar{\theta}_n]$, where

$$\bar{\theta}_n = \begin{cases} \min\left\{\theta, \dfrac{\epsilon_n}{\|\nabla f(x_n) - \nabla f(x_{n-1})\|}\right\}, & \text{if } \nabla f(x_n) \neq \nabla f(x_{n-1}), \\ \theta & \text{otherwise} \end{cases} \tag{3.1}$$

and $\{\epsilon_n\}$ is a sequence of nonnegative numbers such that $\epsilon_n = \circ(\alpha_n)$, that is, $\lim_{n\to\infty} \frac{\epsilon_n}{\alpha_n} = 0$.

Compute

$$\begin{cases} w_n = \nabla f^*(\nabla f(x_n) + \theta_n(\nabla f(x_n) - \nabla f(x_{n-1}))), \\ y_n = J^f_{\lambda_n B} \circ A^f_{\lambda_n} w_n. \end{cases} \tag{3.2}$$

If $y_n = w_n$, then stop. Otherwise, go to the next step.

Step 2: Compute

$$x_{n+1} = \nabla f^*(\alpha_n \nabla f(v) + \beta_n \nabla f(w_n) + \gamma_n \nabla f(y_n)) \ \ \forall\ n \geq 0. \tag{3.3}$$

Set $n := n+1$ and return to **Step 1**.

Lemma 3.2. *The sequence $\{x_n\}$ generated by Algorithm 3.1 is bounded. Consequently, the sequences $\{w_n\}$ and $\{y_n\}$ are bounded too.*

Proof. Let $p \in \Gamma$ and $v_n = \nabla f(x_n) + \theta_n(\nabla f(x_n) - \nabla f(x_{n-1}))$. Then $w_n = \nabla f^*(v_n)$. Using (3.2), we see that

(3.4)
$$\begin{aligned}
&D_f(p, w_n)\\
&= D_f(p, \nabla f^*(v_n)) = f(p) - f^*(v_n) + \langle p, v_n\rangle\\
&= f(p) - f^*(\nabla f(x_n) + \theta_n(\nabla f(x_n) - \nabla f(x_{n-1})))\\
&+ \langle p, \nabla f(x_n) + \theta_n(\nabla f(x_n) - \nabla f(x_{n-1}))\rangle\\
&\leq f(p) + \langle p, \nabla f(x_n)\rangle + \langle p, \theta_n(\nabla f(x_n) - \nabla f(x_{n-1}))\rangle\\
&- [(1+\theta_n) f^*(\nabla f(x_n)) - \theta_n f^*(\nabla f(x_{n-1}))\\
&+ \theta_n(1+\theta_n)\rho_r(\|\nabla f(x_n) - \nabla f(x_{n-1}\|))]\\
&= f(p) + \langle p, \nabla f(x_n)\rangle - f^*(\nabla f(x_n) + \langle p, \theta_n(\nabla f(x_n) - \nabla f(x_{n-1}))\rangle\\
&- [\theta_n f^*(\nabla f(x_n)) - \theta_n f^*(\nabla f(x_{n-1})) + \theta_n(1+\theta_n)\rho_r(\|\nabla f(x_n) - \nabla f(x_{n-1}\|))]\\
&= D_f(p, x_n) + \langle p, \theta_n(\nabla f(x_n) - \nabla f(x_{n-1}))\rangle\\
&- [\theta_n f^*(\nabla f(x_n)) - \theta_n f^*(\nabla f(x_{n-1})) + \theta_n(1+\theta_n)\rho_r(\|\nabla f(x_n) - \nabla f(x_{n-1}\|))]\\
&= D_f(p, x_n) - \theta_n f^*(\nabla f(x_n)) + \theta_n f^*(\nabla f(x_{n-1})) + \langle p, \theta_n(\nabla f(x_n) - \nabla f(x_{n-1}))\rangle\\
&- \theta_n(1+\theta_n)\rho_r(\|\nabla f(x_n) - \nabla f(x_{n-1}\|))\\
&\leq D_f(p, x_n) - \theta_n[f^*(\nabla f(x_n)) - f^*(\nabla f(x_{n-1}))] + \langle p, \theta_n(\nabla f(x_n) - \nabla f(x_{n-1}))\rangle\\
&\leq D_f(p, x_n) - \theta_n\langle x_n, \nabla f(x_n) - \nabla f(x_{n-1})\rangle + \theta_n\langle p, (\nabla f(x_n) - \nabla f(x_{n-1}))\rangle\\
&= D_f(p, x_n) + \theta_n\langle p - x_n, \nabla f(x_n) - \nabla f(x_{n-1})\rangle\\
&\leq D_f(p, x_n) + \theta_n\|x_n - p\|\|\nabla f(x_n) - \nabla f(x_{n-1})\|.
\end{aligned}$$

It follows from (3.1) that $\theta_n\|\nabla f(x_n) - \nabla f(x_{n-1})\| \leq \epsilon_n$. Combining this with $\lim_{n\to\infty} \frac{\epsilon_n}{\alpha_n} = 0$, we obtain

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

Thus, there exists a number $M_1 > 0$ such that

$$\frac{\theta_n}{\alpha_n}\|\nabla f(x_n) - \nabla f(x_{n-1})\| \leq M_1 \ \ \forall\ n \geq 1.$$

Hence

$$\begin{aligned} D_f(p, w_n) &\leq D_f(p, x_n) + \alpha_n M_1 M_2 \\ &= D_f(p, x_n) + \alpha_n M, \end{aligned} \tag{3.5}$$

where $M_2 = \sup_{n\in\mathbb{N}} \|x_n - p\|$.

Using (3.2) once again, we have

$$\begin{aligned} D_f(p, y_n) &= D_f(p, J^f_{\lambda_n B} \circ A^f_{\lambda_n} w_n) \\ &\leq D_f(p, w_n). \end{aligned} \tag{3.6}$$

It now follows from (3.3) that

$$\begin{aligned} D_f(p, x_{n+1}) &= D_f(p, \nabla f^*(\alpha_n \nabla f(v) + \beta_n \nabla f(w_n) + \gamma_n f(y_n))) \\ &\leq \alpha_n D_f(p, v) + \beta_n D_f(p, w_n) + \gamma_n D_f(p, y_n) \\ &\leq \alpha_n D_f(p, v) + (\beta_n + \gamma_n) D_f(p, w_n) \\ &= \alpha_n D_f(p, v) + (1 - \alpha_n) D_f(p, w_n) \\ &\leq (1 - \alpha_n) D_f(p, x_n) + \alpha_n (D_f(p, v) + M) \\ &\leq \max\{D_f(p, x_n), D_f(p, v) + M\} \\ &\vdots \\ &\leq \max\{D_f(p, x_0), D_f(p, v) + M\} \ \forall\ n \geq 0. \end{aligned} \tag{3.7}$$

Thus, we find that $\{D_f(p, x_n)\}$ is bounded. It now follows from Proposition 2.8 that the sequence $\{x_n\}$ is bounded. Consequently, the sequences $\{w_n\}$ and $\{y_n\}$ are bounded too, as asserted. □

Theorem 3.3. *Let E be a real reflexive Banach space and let E^* be its dual. Let $B : E \to 2^{E^*}$ be a maximal monotone operator and let $A : E \to E^*$ be a BISM mapping such that $\Gamma \neq \emptyset$. Let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a strongly coercive Legendre function, which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of E. If the sequences $\{\alpha_n\}$and $\{\gamma_n\} \subset (0, 1)$ are selected such that*

(C1) $\lim_{n\to\infty} \alpha_n = 0$ *and* $\sum_{n=1}^{\infty} \alpha_n = \infty$,

(C2) $0 < \liminf_{n\to\infty} \gamma_n \leq \limsup_{n\to\infty} \gamma_n < 1$,

then the sequence $\{x_n\}$ generated by Algorithm 3.1 converges strongly to a point $p \in \Gamma$, where $p = \Pi_\Gamma v$.

Proof. Let $p = \Pi_\Gamma v$. Since $p \in \Gamma$, it follows from (3.2), (3.3) and Proposition 2.10 (iii) that

$$\begin{aligned}
D_f(p, x_{n+1}) &= D_f(p, \nabla f^*(\alpha_n \nabla f(v) + \beta_n \nabla f(w_n) + \gamma_n f(y_n))) \\
&= V_f(p, \alpha_n \nabla f(v) + \beta_n \nabla f(w_n) + \gamma_n f(y_n)) \\
&\leq V_f(p, \alpha_n \nabla f(v) + \beta_n \nabla f(w_n) + \gamma_n f(y_n) - \alpha_n(\nabla f(v) - \nabla f(p))) \\
&+ 2\alpha_n \langle x_{n+1} - p, \nabla f(v) - \nabla f(p)\rangle \\
&= D_f(p, \nabla f^*(\alpha_n \nabla f(p) + \beta_n \nabla f(w_n) + \gamma_n f(y_n))) \\
&+ 2\alpha_n \langle x_{n+1} - p, \nabla f(v) - \nabla f(p)\rangle \\
&\leq \alpha_n D_f(p, p) + \beta_n D_f(p, w_n) + \gamma_n D_f(p, y_n) \\
&+ 2\alpha_n \langle x_{n+1} - p, \nabla f(v) - \nabla f(p)\rangle \\
&\leq \beta_n D_f(p, w_n) + \gamma_n (D_f(p, w_n) - D_f(y_n, w_n)) \\
&+ 2\alpha_n \langle x_{n+1} - p, \nabla f(v) - \nabla f(p)\rangle \\
&= (\beta_n + \gamma_n) D_f(p, w_n) - \gamma_n D_f(y_n, w_n) + 2\alpha_n \langle x_{n+1} - p, \nabla f(v) - \nabla f(p)\rangle \\
&\leq (1 - \alpha_n) D_f(p, x_n) - \gamma_n D_f(y_n, w_n) \\
&+ \alpha_n \left(2\langle x_{n+1} - p, \nabla f(v) - \nabla f(p)\rangle + \frac{\theta_n}{\alpha_n} \|\nabla f(x_n) - \nabla f(x_{n-1})\| M_2 \right) \\
&\leq (1 - \alpha_n) D_f(p, x_n) + \alpha_n b_n, \qquad (3.8)
\end{aligned}$$

where

$$b_n = \left(2\langle x_{n+1} - p, \nabla f(v) - \nabla f(p)\rangle + \frac{\theta_n}{\alpha_n} \|\nabla f(x_n) - \nabla f(x_{n-1})\| M_2 \right).$$

Therefore we have

$$\gamma_n D_f(y_n, w_n) \leq D_f(p, x_n) - D_f(p, x_{n+1}) + \alpha_n M', \qquad (3.9)$$

where $M' = \sup_{n \in \mathbb{N}} b_n$.

We next show that the sequence $\{x_n\}$ converges strongly to p. To this end, set $a_n = D_f(p, x_n)$. Then it follows from (3.8) that

$$a_{n+1} \leq (1 - \alpha_n) a_n + \alpha_n b_n.$$

To conclude the argument, we apply Lemma 2.11. Indeed, it suffices to show that $\limsup_{k\to\infty} b_{n_k} \leq 0$ whenever there exists a subsequence $\{a_{n_k}\}$ of $\{a_n\}$ verifying

$$\liminf_{k\to\infty} (a_{n_k+1} - a_{n_k}) \geq 0. \qquad (3.10)$$

Suppose such a subsequence exists. Then it follows from (C1) and (3.9) that

$$\begin{aligned}
\limsup_{k\to\infty} \gamma_{n_k} D_f(y_{n_k}, w_{n_k}) &\leq \limsup_{k\to\infty} (a_{n_k} - a_{n_k+1}) + M' \lim_{k\to\infty} \alpha_{n_k} \\
&= -\liminf_{k\to\infty} (a_{n_k+1} - a_{n_k}) \\
&\leq 0. \qquad (3.11)
\end{aligned}$$

Thus,

$$\lim_{k\to\infty} D_f(y_{n_k}, w_{n_k}) = 0. \qquad (3.12)$$

By Proposition 2.9, we have

$$\|y_{n_k} - w_{n_k}\| \to 0 \text{ as } k \to \infty. \tag{3.13}$$

Observe that

$$\begin{aligned}\|\nabla f(w_{n_k}) - \nabla f(x_{n_k})\| &= \theta_{n_k}\|\nabla f(x_{n_k}) - \nabla f(x_{n_{k-1}})\|\\ &\leq \alpha_{n_k}\frac{\theta_{n_k}}{\alpha_{n_k}}\|\nabla f(x_{n_k}) - \nabla f(x_{n_k-1})\| \to 0 \text{ as } k \to \infty. \end{aligned} \tag{3.14}$$

Therefore, since ∇f^* is continuous on bounded subsets of E^*, we obtain

$$\lim_{k\to\infty} \|w_{n_k} - x_{n_k}\| = 0. \tag{3.15}$$

Observe also from (3.3) that

$$\begin{aligned}\|\nabla f(x_{n_k+1}) - \nabla f(x_{n_k})\| &\leq \alpha_{n_k}\|\nabla f(v) - \nabla f(x_{n_k})\| + \beta_{n_k}\|\nabla f(w_{n_k}) - \nabla f(x_{n_k})\|\\ &\quad + \gamma_{n_k}\|\nabla f(y_{n_k}) - \nabla f(x_{n_k})\|\\ &\leq \alpha_{n_k}\|\nabla f(v) - \nabla f(x_{n_k})\| + \beta_{n_k}\|\nabla f(w_{n_k}) - \nabla f(x_{n_k})\|\\ &\quad + \gamma_{n_k}\|\nabla f(y_{n_k}) - \nabla f(w_{n_k})\| + \gamma_{n_k}\|\nabla f(w_{n_k}) - \nabla f(x_{n_k})\|\end{aligned}$$

which implies by condition $(C1)$, (3.13) and (3.14), that

$$\|\nabla f(x_{n_k+1}) - \nabla f(x_{n_k})\| \to 0 \text{ as } k \to \infty.$$

It now follows that

$$\lim_{k\to\infty} \|x_{n_k+1} - x_{n_k}\| = 0. \tag{3.16}$$

Since $\{x_{n_k}\}$ is bounded, there exists a subsequence, say $\{x_{n_{k_j}}\}$, of $\{x_{n_k}\}$ such that $x_{n_{k_j}} \rightharpoonup q \in K$ and (see (3.16)) such that

$$\begin{aligned}\limsup_{k\to\infty}\langle x_{n_k+1} - p, \nabla f(v) - \nabla f(p)\rangle &= \lim_{j\to\infty}\langle x_{n_{k_j}+1} - p, \nabla f(v) - \nabla f(p)\rangle\\ &= \langle q - p, \nabla f(v) - \nabla f(p)\rangle.\end{aligned}$$

It also follows from the fact that $x_{n_{k_j}} \rightharpoonup q$ and (3.15) that the sequence $\{w_{n_{k_j}}\}$ converges weakly to q. Thus, using (3.13), and Proposition 2.10 (i) and (ii), we see that $q \in (A+B)^{-1}(0)$. Since $p = \Pi_\Gamma v$, we infer that

$$\begin{aligned}\limsup_{k\to\infty}\langle x_{n_k+1} - p, \nabla f(v) - \nabla f(p)\rangle &= \langle q - p, \nabla f(v) - \nabla f(p)\rangle\\ &\leq 0.\end{aligned}$$

Now it follows from Lemma 2.11 and (3.8) that $D_f(p, x_n) \to 0$ as $n \to \infty$. It also follows from Proposition 2.6 that the sequence $\{x_n\}$ converges strongly to $p \in \Gamma$, as asserted. □

4. Applications

4.1. **Application to Minimization Problems (MP).** Let C be a nonempty, closed and convex subset of a real reflexive Banach space E. Consider the following Minimization Problem: Find $x \in C$ such that

$$\min_{y\in E} h_1(x) + h_2(x), \tag{4.1}$$

where $h_1 : E \to \mathbb{R} \cup \{+\infty\}$ is a proper, convex and lower semicontinuous function, and $h_2 : E \to \mathbb{R}$ is a smooth, convex and differentiable function with a Lipschitz continuous gradient ∇h_2. It is well known that many optimization problems such as image processing, machine learning, statistical regression and signal processing can be cast into the form (4.1) (see [51] and references therein). This formulation is premised on the fact that ∇h_2 is monotone and ∂h_1 is maximal monotone (see [46]). Note that the MP (4.1) is equivalent to the VIP

$$0 \in \partial h_1(x) + \nabla h_2(x). \tag{4.2}$$

It follows then that (4.2) is a special case of (1.1) with $A = \nabla h_2$ and $B = \partial h_1$.

By setting $A = \nabla h_2$ and $B = \partial h_1$ in Algorithm 3.1, we see that $y_n = prox_{\lambda_n h_1}(w_n - \lambda_n \nabla h_2(w_n))$. We then obtain the following strong convergence theorem for approximating a solution to (4.2), that is, a minimizer of (4.1). We denote the solution set of (4.2) by Ω.

Theorem 4.1. *Let C be a nonempty, closed and convex subset of a real reflexive Banach space E and let E^* be the dual space of E. Let $h_1 : E \to \mathbb{R} \cup \{+\infty\}$ be a proper, convex and lower semicontinuous function and let h_2 be a smooth, convex and differentiable function with L-Lipschitz continuous gradient. Assume that $\Omega \neq \emptyset$ and that $\lambda_n \in \left(0, \frac{2}{L}\right)$. Let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a strongly coercive Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of E. If the sequences $\{\alpha_n\}$and $\{\gamma_n\} \subset (0, 1)$ are selected so that*

(C1) $\lim_{n\to\infty} \alpha_n = 0$ *and* $\sum_{n=1}^{\infty} \alpha_n = \infty$,

(C2) $0 < \liminf_{n\to\infty} \gamma_n \leq \limsup_{n\to\infty} \gamma_n < 1$,

then the sequence $\{x_n\}$ generated by Algorithm 3.1 converges strongly to a point $p \in \Omega$, where $p = \Pi_\Omega v$.

4.2. **Appication to split feasibility problems.** Let E_1 and E_2 be real Banach spaces and let $T : E_1 \to E_2$ be a bounded linear operator. Let C and Q be nonempty, closed and convex subsets of E_1 and E_2, respectively. The split feasibility problem (SFP) is the problem of finding a point $x \in C$ such that

$$Tx \in Q.$$

We denote the solution set of the SFP by $\Upsilon = C \cap T^{-1}(Q) = \{y \in C : Ty \in Q\}$. The SFP was introduced by Censor and Elfving [18] for solving inverse problems arising from phase retrievals, medical image processing, and machine learning.

We recall that the indicator function of C is the function i_C, defined by

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

It is well known that the proximal mapping of i_C is the projection on C; that is,

$$\begin{aligned} prox_{i_C}(x) &= \arg\min_{y\in C} D_f(y,x) \\ &= \Pi_C x. \end{aligned}$$

Let C and Q be nonempty, closed and convex subsets of real reflexive Banach spaces E_1 and E_2, respectively. Let $T : E_1 \to E_2$ be a bounded linear operator and denote its adjoint by T^*. Take $h_2(x) = \frac{1}{2}\|Tx - \Pi_Q(Tx)\|^2$ and $h_1(x) = i_C$. The following theorem follows from Theorem 4.1 with $y_n = \Pi_C(w_n - \lambda_n T^*(I - \Pi_Q)Tw_n)$ in Algorithm 3.1.

Theorem 4.2. *Let C and Q be nonempty, closed and convex subsets of the real reflexive Banach spaces E_1 with E_2, respectively. Let $T : E_1 \to E_2$ be a bounded linear operator and let T^* denote the adjoint of T. Assume that $\Upsilon \neq \emptyset$ and that $\lambda_n \in \left(0, \frac{2}{\|T\|^2}\right)$. Let $f : E_1 \to \mathbb{R} \cup \{+\infty\}$ be a strongly coercive Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of E_1. If the sequences $\{\alpha_n\}$and $\{\gamma_n\} \subset (0, 1)$ are selected so that*

(C1) $\lim_{n\to\infty} \alpha_n = 0$ *and* $\sum_{n=1}^{\infty} \alpha_n = \infty$,

(C2) $0 < \liminf_{n\to\infty} \gamma_n \leq \limsup_{n\to\infty} \gamma_n < 1$,

then the sequence $\{x_n\}$ generated by Algorithm 3.1 converges strongly to a point $p \in \Upsilon$, where $p = \Pi_\Upsilon v$.

4.3. **Application to LASSO Problems.** The following application is a special case of the one given in Section 4.1.

Recall that the ℓ_1-norm regularized least squares model is formulated as follows:

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Dx - b\|^2 + \lambda\|x\|_1, \tag{4.3}$$

where $D \in \mathbb{R}^{m\times n}$ is a given matrix, b is a given vector and λ is a positive constant. Denote the solution set of (4.3) by Ω. The concept of ℓ_1 regularization has been studied for many years. The least squares problem with ℓ_1 penalty was presented and popularized independently with the name Least Absolute Shrinkage and Selection Operator (LASSO, see [49]) and Basis Pursuit Denoising (BPDN) [19]. Interest in compressed sensing is in recovering a solution x to an underdetermined system of linear equations $Dx = b$ in the case where $n > m$. We know from linear algebra that solutions to this system may not exist or be unique when the number of unknowns is greater than the number of equations. In order to solve a system of this formulation, the ℓ_1 norm regularized least squares model is used. If x is sparse, as is often the case in applications, then x can be recovered by solving the above ℓ_1-norm regularized least squares model (4.3). This model is often referred to as a LASSO problem. Two notable iterative methods that take advantage of the special structure of LASSO problems are the Iterative Shrinkage Thresholding Algorithm (ISTA) and the Fast Iterative Shrinkage Thresholding Algorithm (FISTA), which is an accelerated version of ISTA. The ISTA, which is also said to be a proximal gradient method, involves matrix and vector multiplication and possesses great advantage over standard convex algorithms by avoiding matrix factorization. To improve the speed of convergence, Beck and Teboulle [9] introduced the accelerated ISTA which follows an earlier method by Nesterov [36,37]. The ISTA and FISTA have been used for solving problems containing convex differentiable objectives combined with an ℓ_1 regularization terms like the problem

$$\min_{y\in E} h_1(x) + h_2(x), \tag{4.4}$$

where h_1 is a smooth convex function and h_2 is a continuous function not necessarily smooth. It is obvious that the LASSO problem (4.3) is a special case of (4.4) with $h_1(x) = \frac{1}{2}\|Dx-b\|^2$ and $h_2(x) = \lambda\|x\|_1$. The gradient of h_1 is $\nabla h_1(x) = D^*D - D^*b$ with Lipschitz constant $L = \|D^*D\|$. The proximal map with $g(x) = \lambda\|x\|_1$ is given by

$$\begin{aligned} prox_g(x) = prox_{\lambda\|\cdot\|}(x) &= \big(prox_{\lambda\|\cdot\|}(x_1), \dots, prox_{\lambda\|\cdot\|}(x_n)\big) \\ &= (\gamma_1, \dots, \gamma_n), \end{aligned}$$

where $\gamma_j = sgn(x_j)\max\{|x_j| - \lambda, 0\}$ for $j = 1, 2, \dots, n$. Using this, we obtain from Theorem 4.1 the following theorem for solving LASSO problems.

Theorem 4.3. *Let E be a real reflexive Banach space and let E^* denote its dual space. Let $h_1 : E \to \mathbb{R}$ and $h_2 : E \to \mathbb{R}$ be such that $h_1(x) = \frac{1}{2}\|Dx - b\|^2$ and $h_2(x) = \lambda\|x\|_1$. Assume Ω is nonempty and let $\lambda_n \in \big(0, \frac{2}{\|L\|}\big)$, where $L = \|D^*D\|$. Let $f : E \to \mathbb{R} \cup \{+\infty\}$ be a strongly coercive Legendre function which is bounded, uniformly Fréchet differentiable and totally convex on bounded subsets of E. If the sequences $\{\alpha_n\}$ and $\{\gamma_n\} \subset (0, 1)$ are selected so that*

(C1) $\lim_{n\to\infty} \alpha_n = 0$ *and* $\sum_{n=1}^{\infty} \alpha_n = \infty$,
(C2) $0 < \liminf_{n\to\infty} \gamma_n \le \limsup_{n\to\infty} \gamma_n < 1$,

then the sequence $\{x_n\}$ generated by Algorithm 3.1 converges strongly to a point $p \in \Omega$, where $p = \Pi_\Omega v$.

5. Numerical examples

In this section we present numerical examples regarding the performance of our algorithm.

Example 5.1. Let $E_1 = E_2 = \mathbb{R}$, $C = [-1, 1]$ and let $f : \mathbb{R} \to \mathbb{R}$ be defined by $f(x) = \frac{x^2}{2}$. Then f satisfies the assumptions of Theorem 3.3 (see [10]). We have $\nabla f(x) = x$, $f^*(x^*) = \frac{x^{*2}}{2}$ and $\nabla f^*(x^*) = x^*$. Let $A : \mathbb{R} \to \mathbb{R}$ be defined by $A(x) = 5x$ and let $B : \mathbb{R} \to \mathbb{R}$ be defined $B(x) = 7x - 1$. Then A and B are BISM and maximal monotone, respectively. Let y_n be as given in Algorithm 3.1, and let $x \in E$ and $\lambda_n > 0$. Then

$$\begin{aligned} J^f_{\lambda_n B} \circ A^f_{\lambda_n} x &= (\nabla f + \lambda_n B)^{-1} \nabla f \circ \nabla f^*(\nabla f - \lambda_n)A)(x) \\ &= (\nabla f + \lambda_n B)^{-1}(x - 5\lambda_n x) \\ &= \frac{x - 5\lambda_n x + \lambda_n}{1 + 7\lambda_n}. \end{aligned}$$

Hence,

$$y_n = \frac{(1 - 5\lambda_n)w_n + \lambda_n}{1 + 7\lambda_n}.$$

Let $\alpha_n = \frac{1}{5n+2}$, $\beta_n = \frac{3n}{5n+2}$, $\gamma_n = \frac{2n+1}{5n+2}$, $\epsilon_n = \frac{1}{n^2}$, $\theta = \frac{1}{3.3}$ and $\lambda_n = 1.1$. Choosing a stopping criterion given by $E_n = \|x_{n+1} - x_n\| = 10^{-4}$, we consider the following cases for this numerical experiment.

Case (i) $v = 0.8$, $x_0 = 1.75$ and $x_1 = 1.2$;
Case (ii) $v = 0.8$, $x_0 = 1.896$ and $x_1 = 1.896$;
Case (iii) $v = 1.2$, $x_0 = 1.25$ and $x_1 = 1.896$;
Case (iv) $v = 2.5$, $x_0 = 5.25$ and $x_1 = 7.69$.

The results we have obtained are displayed in Figure 1.

Table 1. Computational results for Example 5.1

Case	Test Parameter	Algorithm 3.1	Unaccelerated	Altered
(i)	No of Iter.	7	16	26
	CPU time (sec)	0.0013	2.9229	4.561
(ii)	No of Iter.	14	16	26
	CPU time (sec)	0.0130	3.6771	6.124
(iii)	No of Iter.	6	20	29
	CPU time (sec)	0.0050	5.8712	8.001
(iv)	No of Iter.	20	29	35
	CPU time (sec)	0.0079	5.8712	7.556

Example 5.2. Next, we give an example illustrating our Theorem 4.2. Let $f(x) = \frac{1}{2}\|x\|^2$ and $E_1 = E_2 = \mathcal{L}_2([\alpha, \beta])$ equipped with the inner product and norm given by $\langle x, y\rangle = \int_\alpha^\beta x(s)y(s)ds$ and $\|x\| = \sqrt{\int_\alpha^\beta \|x(s)|^2 ds}$, respectively, for all $x, y \in \mathcal{L}_2([\alpha, \beta])$ and $s \in [\alpha, \beta]$. Suppose

$$C := \{u \in \mathcal{L}_2([\alpha, \beta]) : \langle a, u\rangle \leq b\},$$

where $0 \neq a \in \mathcal{L}_2([\alpha, \beta])$ and $b \in \mathbb{R}$. Then we know (see [17]) that

$$\Pi_C(u) = P_C(u) = \begin{cases} u + \frac{b - \langle a, u\rangle}{\|a\|^2}, & \text{if } u \notin C \\ u, & \text{otherwise.} \end{cases}$$

Let Q be given by

$$Q = \{v \in \mathcal{L}_2([\alpha, \beta]) : \langle v, d\rangle = e\}$$

which is a closed hyperplane. Also,

$$\Pi_Q(v) = P_Q(v) = \max\left\{0, \frac{e - \langle v, d\rangle}{\|d\|^2}\right\} d + v. \tag{5.1}$$

For this example, we let $C := \{x \in \mathcal{L}_2([0, 1]) : \langle x, a\rangle \leq b\}$, where $a = \frac{t}{5}$ and $b = -1$. The set Q is given by $Q := \{y \in \mathcal{L}_2([0, 1]) : \langle c, y\rangle = e\}$, where $c = \frac{t}{3}$ and $e = 0$. Let the operator $T : \mathcal{L}_2([0, 1]) \to \mathcal{L}_2([0, 1])$ be given by $T(x(s)) = \frac{x(s)}{2}$ $\forall x \in \mathcal{L}_2([0, 1])$.

Let $\alpha_n = \frac{1}{200n+3}$, $\beta_n = \gamma_n = \frac{100n+1}{200n+3}$, $\epsilon_n = \frac{1}{n^{1.1}}$, $\theta = \frac{1}{3}$ and $\lambda_n = 2.2$. Choosing a stopping criterion given by $E_n = \|x_{n+1} - x_n\| = 10^{-2}$, we consider the following cases for the initial values x_0 and x_1:

Case (I) $v = 2t$, $x_0 = -2t + 1$ and $x_1 = \sin(t + 1)$;

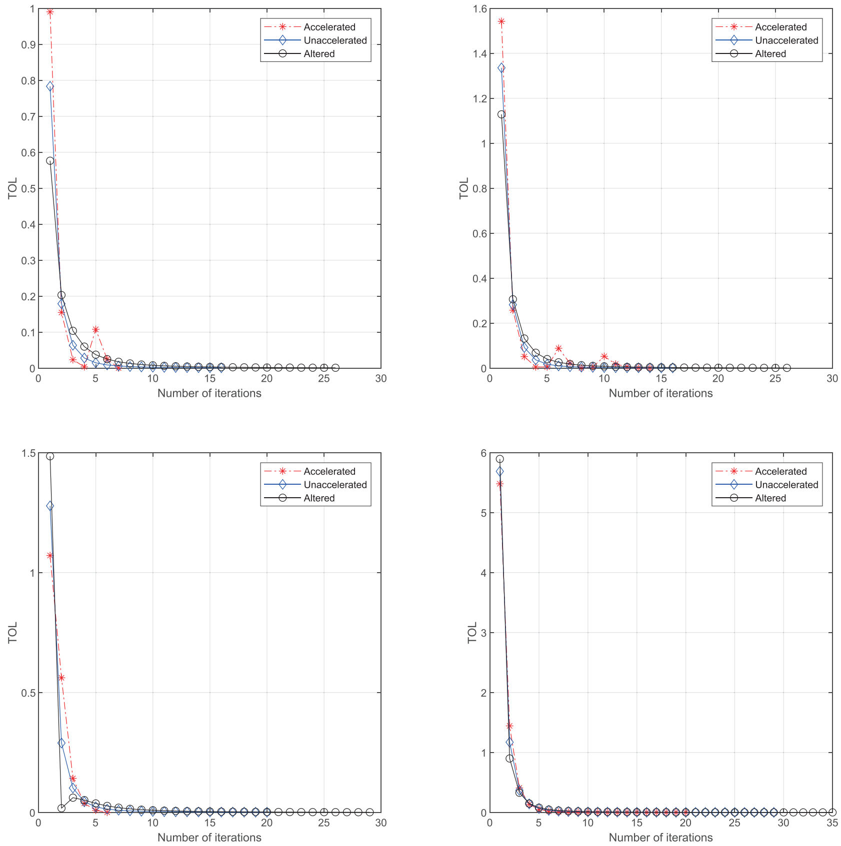


FIGURE 1. Example 5.1, Top left: Case (i); Top right: Case (ii); Bottom left: Case (iii); Bottom right: Case (iv)

Case (II) $v = 2t$, $x_0 = 2t^2 + 1$ and $x_1 = \sin(2t - 1)$;
Case (III) $v = \frac{t}{5}$, $x_0 = 11t^2 + 10t + 7$ and $x_1 = e^{-t}$;
Case (IV) $v = \frac{t}{5}$, $x_0 = \log(2t)$ and $x_1 = e^{-3t}$.

The numerical results we have obtained are displayed in Figure 2.

Example 5.3. We now give an application of our method to image restoration. We first recall that the general image recovery problem is formulated by the inversion of the following linear equation:

$$y = Dx + b, \tag{5.2}$$

where $x \in \mathbb{R}^n$, x, b and y are the original image, additive random noise, which is unknown, and the known degraded observation, respectively. The operator D is a linear operator which depends on the image recovery problem under consideration. This model has been applied to several optimization problems. For solving problems arising from this model, authors have used the ℓ_1 norm. The ℓ_1 regularization problem is given by

$$\min_x \left\{ \frac{1}{2}\|Dx - b\|_2^2 + \lambda_n \|x\|_1 \right\}, \tag{5.3}$$

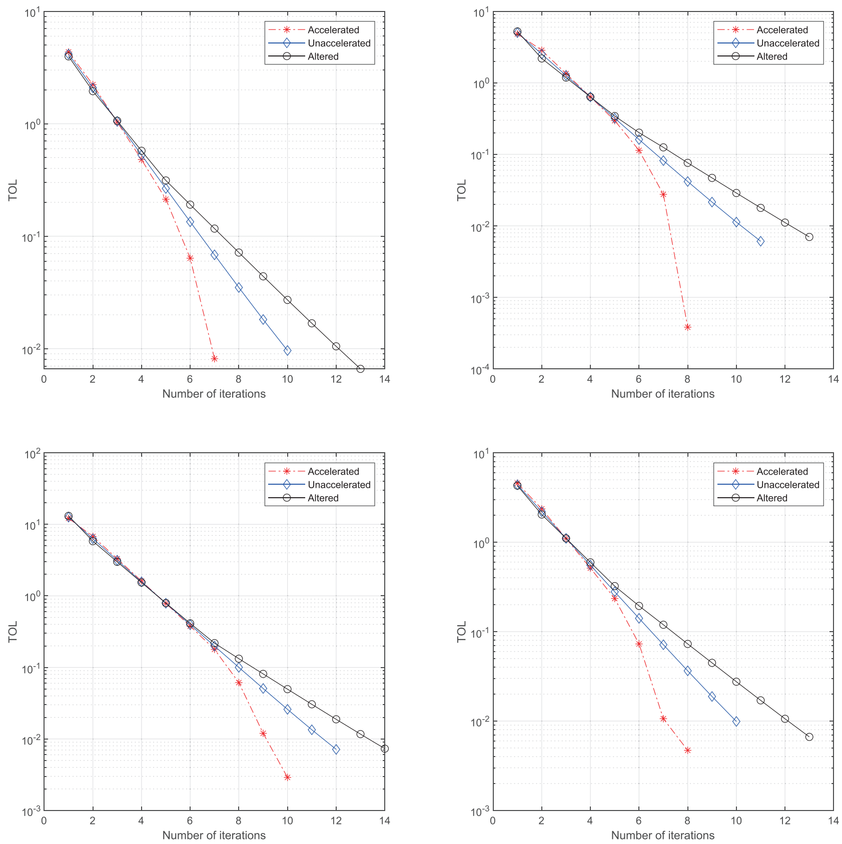


FIGURE 2. Example 5.2, Top left: Case (I); Top right: Case (ii); Bottom left : Case (III); Bottom right: Case (IV).

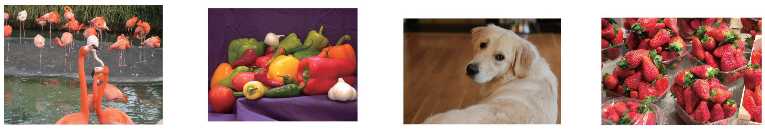
where $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$, D is an $m \times n$ matrix and λ_n is a nonnegative parameter. Thus, finding solutions to (5.2) amounts to approximating solutions of the convex minimization problem

$$\text{Find } x \in \arg\min_{x\in\mathbb{R}^n} \left\{ \frac{1}{2}\|Dx - b\|_2 + \lambda_n \|x\|_1 \right\}, \tag{5.4}$$

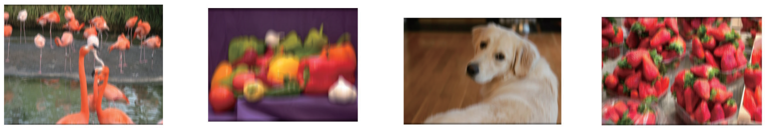
where b is the degraded image and D is an operator representing the mask. Setting $Ax := \nabla(\frac{1}{2}\|Dx - b\|_2^2) = D^T(Dx - b)$ and $Bx := \partial(\lambda\|x\|_1)$, we apply Algorithm 3.1 to solving (5.4).

In our experiments we have used the MATLAB blur function "P = special ('motion', 20,30)" and added the random noise $0.001 \times randn(size(x))$. We initialize x_0 and u to be zeroes in $\mathbb{R}^n$ and set $x_1 = Dx + b$. We also choose $\alpha_n = 0.5$, $\beta = 0.75$, $\gamma = 0.25$, $\theta = 0.95$ and $\lambda = 0.000001$. We have chosen flamingo, pepper, sherlock and strawberries from the MATLAB toolbox as our test images. The stopping criterion for the algorithms is $\frac{\|x_{n+1} - x_n\|}{\|x_{n+1}\|} < 10^{-4}$ and a maximum iteration number of $n = 100$.

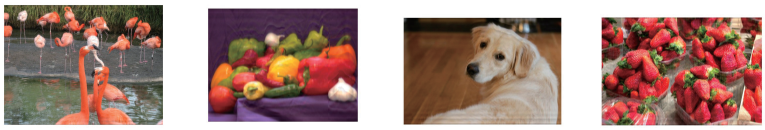
The signal-to-noise ratio (SNR) is a measure used in science and engineering to compare the level of a desired signal to the level of the background noise. In this



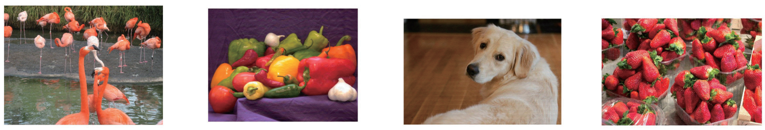
(a) Original test images



(b) Degraded test images with blur function and random noise



(c) Restored images with altered inertial term



(d) Restored images with Algorithm 3.1

FIGURE 3. Test images and restoration via Algorithm 3.1 and one of its variations

case, it is used to measure the performance of the algorithms and is defined by

$$SNR = 10 \log \frac{\|x_n\|}{\|x - x_n\|},$$

where x and x_n are the original image and estimated image at iteration n, respectively. By using this measuring technique, the best restoration process via the algorithm is determined by the higher SNR (see Table 2 and 3) for comparing two methods.

Remark 5.4. It is evident from Table 2, 3 and Figure 3 that our method performs better in applications than the method where the direction of the momentum $x_n - x_{n-1}$ is changed.

6. Conclusions

We have introduced an inertial iterative method for approximating a zero of the sum of two monotone operators in a reflexive Banach space. By using the forward-backward and the Halpern techniques, we have presented and proved a strong convergence theorem assuming standard conditions on the control parameters. We note that in choosing our inertial step the change of direction of the momentum which

Table 2. Numerical results of SNR in Figure 3.

	Algorithm 3.1		Altered direction	
n	Flamingo	Pepper	Flamingo	Pepper
1	32.58	32.12	30.59	28.50
10	36.96	36.95	34.53	34.56
20	39.74	40.20	35.82	36.27
30	43.48	44.37	36.79	37.54
40	45.70	46.62	37.58	38.55
50	47.28	48.21	38.22	39.38
60	48.69	49.47	38.78	40.07

Table 3. Continuation of Table 2.

	Algorithm 3.1		Altered direction	
n	Sherlock	Strawberries	Sherlock	Strawberries
1	47.08	28.07	39.72	26.12
10	52.50	32.19	50.23	29.82
20	55.63	34.94	52.17	31.01
30	58.12	38.09	53.49	31.92
40	59.19	41.62	54.49	32.65
50	59.71	43.55	55.30	33.27
60	59.99	45.43	55.97	33.80

has been mostly used when the framework is a Banach space ([1,5,30]) has been removed. Some applications of our main result have been reported. Furthermore, we have displayed several numerical examples which show the differences in the number of iterations and the CPU time for convergence for our method and its variations in terms of the inertial technique.

References

[1] H. A. Abass, G. C. Ugwnnadi and O. K. Narain, *A modifed inertial Halpern method for solving split monotone variational inclusion problems in Banach Spaces,* Rend. Circ. Mat. Palermo, II. Ser (2022), https://doi.org/10.1007/s12215-022-00795-y.

[2] A. Adamu, D. Kitkuan, P. Kumam, A. Padcharoen and T. Seanqwattana, *Approximation method for monotone inclusion problems in real Banach spaces with applications,* J. Ineq. Appl. **70** (2022), https://doi.org/10.1186/s13660-022-02805-0.

[3] 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** (2001), 3–11.

[4] H. Attouch and A. Cabot, *Convergence of a relaxed inertial forward-backward algorithm for structured monotone inclusions*, Available at https://hal.archives-ouvertes.fr/ hal-01782016 (2018), Hal ID: 01782016.

[5] B. Ali, G. C. Ugwunnadi, M. S. Lawan and A. R. Khan, *Modified inertial subgradient extragradient method in reflexive Banach spaces*, Bol. Soc. Mat. Mex. **27** (2021): 30.

[6] H. H. Bauschke, J. M. Borwein and P. L. Combettes, *Essential smoothness, essential strict convexity and Legendre functions in Banach spaces*, Commun. Contemp. Math. **3** (2001), 615–647.

[7] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev. **38 (3)** (1996), 367–426.

[8] A. Beck, *First-Order Methods in Optimization, Society for Industrial and Applied Mathematics*, Philadelphia, 2017.

[9] A. Beck and M Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems,* SIAM J. Imaging Sci. **2(1)** (2009), 183–202.

[10] J. M. Borwein, S. Reich and S. Sabach, *A characterization of Bregman firmly nonexpansive operators using a new monotonicity concept*, J. Nonlinear Convex Anal. **12** (2011), 161–184.

[11] L. M. Bregman, *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. **7** (1967), 200–217.

[12] D. Butnariu and G. Kassay, *A proximal-projection methods for finding zeroes of set-valued operators*, SIAM J. Control Optim. **47 (4)** (2008), 2096–2136.

[13] 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, 1–39.

[14] D. Butnariu, Y. Censor and S. Reich, *Iterative averaging of entropic projections for solving stochastic convex feasibility problems*, Comput. Optim. Appl. **8** (1997), 21–39.

[15] D. Butnariu, A. N. Iusem and C. Zălinescu, *On uniform convexity, total convexity and convergence of the proximal point and outer Bregman projection algorithms in Banach spaces*, J. Convex Anal. **10** (2003), 35–61.

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

[17] A. Cegielski, *Iterative Methods for Fixed Point Problems in Hilbert Spaces*, Lecture Notes in Mathematics, vol. 2057, Springer, Berlin, 2012.

[18] Y. Censor and T. Elfving, *A multiprojection algorithm using Bregman projections in a product space*, Numer. Algorithms. **8** (1994), 221–239.

[19] S. Chen, D. L. Donoho and M. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput. **20** (1998), 33–61.

[20] C. E. Chidume, P. Kumam and A. Adamu, *A hybrid inertial algorithm for approximating solution of convex feasibility problems with applications*, Fixed Point Theory Appl 2020, 12 (2020).

[21] C. E. Chidume, L. Okereke and A. Adamu, *A hybrid algorithm for approximating solutions of a variational inequality problem and a convex feasibility problem*, Adv. Nonlinear. Var. Ineq. **21 (1)** (2018), 46–64.

[22] W. Cholamjiak, P. Cholamjiak and S. Suantai, *An inertial forward-backward splitting method for solving inclusion problems in Hilbert spaces*, J. Fixed Point Theory Appl. **20** (2018): Article ID 42, 17 pages.

[23] I. Cioranescu, *Geometry of Banach spaces, Duality Mappings and Nonlinear*, Kluwer, Dordrecht, 1990.

[24] P.L. Combettes, V.R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul. **4** (2005), 1168–1200.

[25] Q. Dong, D. Jiang, P. Cholamjiak and Y. Shehu, *A strong convergence result involving an inertial forward-backward algorithm for monotone inclusions*, J. Fixed Point Theory Appl. **19** (2017), 3097-3118.

[26] M. Eslamian, *General algorithms for split common fixed point problem of demicontractive mappings*, Optimization **65** (2015), 443–465.

[27] Y.-P. Fang and N.-J. Huang, *H-accretive operators and resolvent operator technique for solving variational inclusions in Banach spaces*, J. Math. Anal. Lett. **17** (2004), 647–653.

[28] Y. Y. Huang, J. C. Jeng, T. Y. Kuo and C. C. Hong, *Fixed point and weak convergence theorems for point dependent λ-hybrid mappings in Banach spaces*, Fixed Point Theory and Appl. 2011:105 (2011).

[29] G. Kassay, S. Reich and S. Sabach, *Iterative methods for solving systems of variational inequalities in reflexive Banach spaces*, SIAM J. Optim. **21** (2011), 1319–1344.

[30] D. S. Kim and B.V. Dinh, *Parallel extragradient algorithms for multiple set split equilibrium problems in Hilbert spaces*, Numer Algor. **77** (2018), 741–761.
[31] F. Kohsaka and W. Takahashi, *Proximal point algorithm with Bregman functions in Banach spaces*, J. Nonlinear Convex Anal. **6** (2005), 505–523.
[32] P.-L. Lions and B. Mercier *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal. **16** (1979), 964–979.
[33] D. A. Lorenz and T. Pock, *An inertial forward-backward algorithm for monotone inclusions*, J. Math. Imaging Vis. **51** (2015), 311–325.
[34] A. A. Mebawondu, H. A. Abass, O. K. Oyewole, K. O. Aremu and O. K. Narain, *Generalized split null point of sum of monotone operators in Hilbert spaces*, Demonstratio Mathematica. **54 (1)** 359–376.
[35] A. Moudafi and M. Oliny, *Convergence of a splitting inertial proximal method for monotone operators*, J. Comput. Appl. Math. **155** (2003), 447–454.
[36] Y. Nesterov, *A method for solving the convex programming problem with convergence rate* $\mathcal{O}(\frac{1}{k^2})$, Dokl. Akad. Nauk SSSR. **269(3)** (1983), 543–547.
[37] Y. Nesterov, *Quelques proprietes des operateurs angle-bornes et n-cycliquement monotones*, Isreal J. Math. **26** (1977), 137–150.
[38] F.U. Ogbuisi and C. Izuchukwu, *Approximating a zero of sum of two monotone operators which solves a fixed point problem in reflexive Banach spaces*, Numer. Funct. Anal. Optim. **41 (3)** (2020), 322–343.
[39] R. R. Phelps, *Convex Functions. Monotone Operators and Differentiability*, 2nd edn. Lecture Notes in Mathematics, vol 1364, Berlin, 1993.
[40] B. T. Polyak, *Some methods of speeding up the convergence of iteration methods*, U.S.S.R Comput. Math. Math. Phys. **4 (5)** (1964), 1–17.
[41] D. Reem, S. Reich and A. De Pierro, *Re-examination of Bregman functions and new properties of their divergences*, Optimization **68** (2019), 279–348.
[42] 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, 1996, pp. 313–318.
[43] S. Reich and S. Sabach, *A strong convergence theorem for a proximal-type algorithm in reflexive Banach spaces*, J. Nonlinear Convex Anal. **10** (2009), 471–485.
[44] S. Reich and S. Sabach, *Two strong convergence theorems for Bregman strongly nonexpansive operators in reflexive banach spaces*, Nonlinear Anal. **73** (2010), 122–135.
[45] S. Reich and S. Sabach, *Two strong convergence theorems for a proximal method in reflexive Banach spaces*, Numer. Funct. Anal. Optim. **31** (2010), 24–44.
[46] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim. **14** (1976), 877–898.
[47] S. Saejung and P. Yotkaew, *Approximation of zeros of inverse strongly monotone operator in Banach spaces*, Nonlinear Anal. **75** (2012), 742–750.
[48] A. Taiwo, L. O. Jolaoso and O. T. Mewomo, *Inertial-type algorithm for solving split common fixed point problems in Banach spaces*, J. Sci. Comput. **86 (1)** https://doi.org/10.1007/s10915-020-01385-9
[49] R. Tibshirami, *Regression shrinkage and selection via lasso*, J. R. Statist. Soc. Ser. B. *58* (1996), 267–288.
[50] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim. **38** (2000), 431–446.
[51] Y. Wang and H.-K Xu, *Strong convergence for the proximal-gradient method*, J. Nonlinear Convex Anal. **15 (3)** (2014), 581–593.
[52] F. Q. Xia and N.-J. Huang, *Variational inclusions with a general H-monotone operator in Banach spaces*, Comput. Math. Appl. **54 (1)** (2007), 24–30.
[53] J. Yang and H. Liu, *The subgradient extragradient method extended to pseudomonotone equilibrium problems and fixed point problems in Hilbert space*, Optim. Lett. https://doi.org/10.1007/s11590-019-01474-1

[54] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing, Singapore, 2002.

Manuscript received April 12 2023
revised April 27 2023

O. K. Oyewole
Department of Mathematics, The Technion -– Israel Institute of Technology, 32000 Haifa, Israel
E-mail address: oyewoleolawalekazeem@gmail.com, oyewoleok@campus.technion.ac.il

S. Reich
Department of Mathematics, The Technion -– Israel Institute of Technology, 32000 Haifa, Israel
E-mail address: sreich@technion.ac.il