APPLICABLE ANALYSIS
2022, VOL. 101, NO. 15, 5386–5414
https://doi.org/10.1080/00036811.2021.1892080

Taylor & Francis
Taylor & Francis Group

Check for updates

Strong convergence of inertial forward–backward methods for solving monotone inclusions

Bing Tan [a] and Sun Young Cho [b]

[a]Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Chengdu, People's Republic of China; [b]Department of Liberal Arts, Gyeongnam National University of Science and Technology, Jinju-Si, South Korea

ABSTRACT
The paper presents four modifications of the inertial forward–backward splitting method for monotone inclusion problems in the framework of real Hilbert spaces. The advantages of our iterative schemes are that the single-valued operator is Lipschitz continuous monotone rather than cocoercive and the Lipschitz constant does not require to be known. The strong convergence of the suggested approaches is obtained under some standard and mild conditions. Finally, several numerical experiments in finite- and infinite-dimensional spaces are proposed to demonstrate the advantages of our algorithms over the existing related ones.

ARTICLE HISTORY
Received 2 October 2020
Accepted 12 February 2021

COMMUNICATED BY
J.-C. Yao

KEYWORDS
Inclusion problem; inertial forward–backward method; projection and contraction method; Tseng's splitting method; viscosity method

MATHEMATICS SUBJECT CLASSIFICATIONS
47H05; 47J22; 47J25; 68W10; 65K15

1. Introduction

In this paper, our interest is to devise fast iterative algorithms to solve the monotone inclusion problem in real Hilbert spaces. Our problem is described as follows:

$$\text{find } x^* \in \mathscr{H} \text{ such that } 0 \in (A+B)x^*, \tag{MIP}$$

where $\mathscr{H}$ is a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$, $A : \mathscr{H} \to \mathscr{H}$ is a monotone mapping and $B : \mathscr{H} \to 2^{\mathscr{H}}$ is a maximal monotone mapping. The solution set of (MIP) is denoted by Ω. It is known that many problems can be converted into the model of (MIP), such as image processing problems, convex minimization problems, split feasibility problems, equilibrium problems, variational inequalities and DC programming problems; see, e.g. [1–8]. Therefore, a large number of researchers are very interested in this problem and have developed many methods to solve such problems. One of the most famous of these approaches is the forward–backward algorithm (FBA), which generates an iterative sequence $\{x_n\}$ in the following way:

$$x_{n+1} = (I + \lambda_n B)^{-1}(I - \lambda_n A)x_n, \tag{1}$$

where stepsize $\lambda_n > 0$, I stands for the identity mapping on $\mathscr{H}$, the operator $(I - \lambda_n A)$ is referred to as forward operator and the operator $(I + \lambda_n B)^{-1}$ is the so-called backward operator (also referred to as resolvent operator). The FBA for monotone inclusion problems were first introduced by Lions

CONTACT Sun Young Cho sycho@gntech.ac.kr

and Mercier [9] (also by Passty [10] independently). In the past few decades, the convergence properties and the modified versions of this method have been extensively studied in the literature; see, e.g. [11–16] and the references therein. It should be mentioned that the FBA defined by (1) requires mapping A to be inverse strongly monotone (see the definition in Section 2). This assumption is very strict and it is difficult to meet the practical problems. In order to avoid this restriction, many scholars have made a lot of efforts and achieved some important results. Next, we introduce two methods to overcome this difficulty in the literature.

The first is the Tseng splitting algorithm (also known as forward–backward–forward method) proposed by Tseng [17] in 2000, which is a two-step iterative scheme. More precisely, the form of the algorithm is as follows:

$$\begin{cases} y_n = (I + \lambda_n B)^{-1}(I - \lambda_n A)x_n, \\ x_{n+1} = y_n - \lambda_n(Ay_n - Ax_n), \end{cases} \tag{2}$$

where the step size $\{\lambda_n\}$ can be automatically updated by Armijo-type search methods. When the mapping A is Lipschitz continuous monotone and the mapping B is maximal monotone, the sequence $\{x_n\}$ formed by iterative process (2) converges weakly to a solution of (MIP) in real Hilbert spaces. In 2018, Zhang and Wang [18] combined the projection and contraction method and (1), and proposed another iterative scheme to overcome the strong assumption on mapping A. To be more precise, the method is described as follows:

$$\begin{cases} y_n = (I + \lambda_n B)^{-1}(I - \lambda_n A)x_n, \\ x_{n+1} = x_n - \gamma\eta_n d_n, \end{cases} \tag{3}$$

where $d_n = x_n - y_n - \lambda_n(Ax_n - Ay_n)$, $\eta_n = \frac{\langle x_n - y_n, d_n\rangle}{\|d_n\|^2}$, $\gamma \in (0,2)$, $\{\lambda_n\}$ is a control sequence, operator A is assumed to be Lipschitz continuous monotone and operator B is assumed to be maximal monotone. They established the weak convergence of the iterative method (3) under some suitable conditions.

It is worth noting that the Tseng splitting method (2) and the Algorithm (3) are only weakly convergent in infinite-dimensional spaces. Examples in CT reconstruction and machine learning tell us that strong convergence is preferable to weak convergence in an infinite-dimensional space. Therefore, a natural question is how to modify method (1) such that it can achieve strong convergence in infinite-dimensional spaces. In fact, in the past few decades, researchers have proposed many modified forward–backward methods to achieve strong convergence in real Hilbert spaces; see, e.g. [19–22] and the references therein. It should be pointed out that the algorithms mentioned in the above literatures also require operator A to be inverse strongly monotone. Let us review some recent results to overcome this shortcoming. In 2018, Gibali and Thong [23] proposed two modifications of (2) based on Mann and viscosity ideas. They established two strong convergence theorems of the suggested algorithms in an infinite-dimensional Hilbert space. Moreover, Thong and Cholamjiak [24] and Gibali et al. [25] presented several new algorithms by means of the viscosity-type method and iterative method (3), and established the strong convergence theorems of the proposed algorithms in Hilbert spaces.

In recent years, the development of fast iterative algorithms has attracted enormous interest, especially for the inertial method, which is based on discrete versions of a second-order dissipative dynamic system. Many researchers have constructed various fast iterative algorithms by using inertial technology; see, e.g. [26–31] and the references therein. One of the common features of these algorithms is that the next iteration depends on the combination of the previous two iterations. Note that this minor change greatly improves the performance of the algorithms. In 2015, Lorenz and Pock [26] introduced the following inertial forward–backward algorithm (iFBA) for monotone inclusions:

$$\begin{cases} w_n = x_n + \theta_n(x_n - x_{n-1}), \\ x_{n+1} = (I + \lambda_n B)^{-1}(I - \lambda_n A)w_n. \end{cases} \tag{4}$$

Note that the iFBA (4) still achieves weak convergence in real Hilbert spaces. Their numerical experiments on image restoration show that iFBA converges faster than some existing algorithms.

Motivated and stimulated by the above work, in this paper, we propose four accelerated forward–backward splitting algorithms to solve the monotone inclusion problem (MIP) in real Hilbert spaces. The advantages of our iterative schemes are: (1) operator A is assumed to be Lipschitz continuous monotone and operator B is assumed to be maximal monotone; (2) the prior information of the Lipschitz constant of the operator is not required; (3) the strong convergence theorems of the suggested algorithms are established under some weaker conditions; (4) the inertial term is embedded to accelerate the convergence speed of the algorithms. Furthermore, we also give several theoretical applications of the proposed methods. Finally, some numerical experiments are provided to show the advantages of our stated algorithms over the previously existing algorithms. Our approaches obtained in this paper improve and summarize some results in the literature [18,20–25].

The organizational structure of our paper is built up as follows. Some essential definitions and technical lemmas that need to be used are given in Section 2. In Section 3, we propose several algorithms and analyze their convergence. Section 4 introduces four theoretical applications of the proposed methods. Some numerical experiments to verify our theoretical results are presented in Section 5. Finally, the paper ends with a brief summary in Section 6, the last section.

2. Preliminaries

Let C be a closed and convex nonempty subset of a real Hilbert space $\mathscr{H}$. The weak convergence and strong convergence of $\{x_n\}_{n=1}^{\infty}$ to x are represented by $x_n \rightharpoonup x$ and $x_n \to x$, respectively. For each $x, y, z \in \mathscr{H}$, we have the following facts:

(1) $\|x+y\|^2 \leq \|x\|^2 + 2\langle y, x+y\rangle$;
(2) $\|\alpha x + (1-\alpha)y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2 - \alpha(1-\alpha)\|x-y\|^2, \alpha \in \mathbb{R}$;
(3) $\|\alpha x + \beta y + \gamma z\|^2 = \alpha\|x\|^2 + \beta\|y\|^2 + \gamma\|z\|^2 - \alpha\beta\|x-y\|^2 - \alpha\gamma\|x-z\|^2 - \beta\gamma\|y-z\|^2$, where $\alpha, \beta, \gamma \in [0,1]$ with $\alpha+\beta+\gamma = 1$.

For every point $x \in \mathscr{H}$, there exists a unique nearest point in C, denoted by $P_C(x)$ such that $P_C(x) := \operatorname{argmin}\{\|x-y\|, y \in C\}$. P_C is called the metric projection of $\mathscr{H}$ onto C. It is known that P_C is nonexpansive and $P_C(x)$ has the following basic properties:

- $\langle x - P_C(x), y - P_C(x)\rangle \leq 0, \forall x \in \mathscr{H}, y \in C$;
- $\|P_C(x) - P_C(y)\|^2 \leq \langle P_C(x) - P_C(y), x - y\rangle, \forall x \in \mathscr{H}, y \in \mathscr{H}$.

The two projection formulas given next will be used in the sequel (see Section 5).

(1) The projection of x onto a half-space $H_{u,v} = \{x : \langle u, x\rangle \leq v\}$ is computed by

$$P_{H_{u,v}}(x) = x - \max\left\{[\langle u, x\rangle - v]/\|u\|^2, 0\right\}u.$$

(2) The projection of x onto a ball $B[p,q] = \{x : \|x-p\| \leq q\}$ is computed by

$$P_{B[p,q]}(x) = p + \frac{q}{\max\{\|x-p\|, q\}}(x-p).$$

For any $x, y \in \mathscr{H}$, the mapping $A : \mathscr{H} \to \mathscr{H}$ is said to be:

(1) *L-Lipschitz continuous* with $L > 0$ if

$$\|Ax - Ay\| \leq L\|x-y\|.$$

If $L \in (0,1)$, then mapping A is called a *contraction*. In particular, when $L = 1$, mapping A is called *nonexpansive*.

(2) *Monotone* if

$$\langle Ax - Ay, x - y\rangle \geq 0.$$

(3) *k-inverse strongly monotone* (also called *k-cocoercive*) if there exists a $k > 0$ such that

$$\langle Ax - Ay, x - y\rangle \geq k\|Ax - Ay\|^2.$$

Clearly, every k-cocoercive mapping is $1/k$-Lipschitz continuous and monotone.

Recall that a multi-valued mapping $B : \mathscr{H} \to 2^{\mathscr{H}}$ with domain $\mathrm{Dom}(B) := \{x \in \mathscr{H} : Bx \neq \emptyset\}$ is said to be (i) monotone if, for all $x, y \in \mathscr{H}$, $u \in Bx$ and $v \in By$ indicates that $\langle u - v, x - y\rangle \geq 0$; (ii) maximal monotone if it is monotone and if, for any $(x, u) \in \mathscr{H} \times \mathscr{H}$, $\langle u - v, x - y\rangle \geq 0$ for every $(y, v) \in \mathrm{Graph}(B)$ (the graph of mapping B) indicates that $u \in Bx$. Let mapping $B : \mathscr{H} \to 2^{\mathscr{H}}$ be set-valued maximal monotone. Then, for $\forall x \in \mathscr{H}$ and $\lambda > 0$, the resolvent mapping $J_{\lambda B} : \mathscr{H} \to \mathscr{H}$ associated with B is represented as $J_{\lambda B}(x) = (I + \lambda B)^{-1}(x)$, where I stands for the identity operator on $\mathscr{H}$.

We give the following definitions which will be used in the sequel.

(i) Let $h : \mathscr{H} \to \mathbb{R}$ be a proper convex and lower semicontinuous function. Its subdifferential is defined as

$$\partial h(x) = \left\{z \in \mathscr{H} : h(y) - h(x) \geq \langle z, y - x\rangle\right\}, \quad \forall x, y \in \mathrm{Dom}(h).$$

It is known that ∂h is maximal monotone.

(ii) Let $N_C(x)$ denote the normal cone of C at x, which is given by

$$N_C(x) := \begin{cases} \emptyset & \text{if } x \notin C; \\ \{z \in \mathscr{H} : \langle z, y - x\rangle \leq 0,\ \forall y \in C\} & \text{if } x \in C. \end{cases}$$

(iii) Let $\delta_C(x)$ be the indicator function of C at x, that is,

$$\delta_C(x) := \begin{cases} 0, & \text{if } x \in C; \\ \infty, & \text{if } x \notin C. \end{cases}$$

Then $\delta_C : \mathscr{H} \to \mathbb{R}$ is a proper convex and lower semicontinuous function. Set $B = \partial\delta_C$, we claim that $J_{\lambda B}(x) = P_C(x)$. Indeed, it is easy to see that $\partial\delta_C(x) = N_C(x)$. In addition, for all $x \in C$, we have

$$u = J_{\lambda B}(x) \Leftrightarrow x \in u + \lambda N_C(u) \Leftrightarrow \langle x - u, y - u\rangle \leq 0 \Leftrightarrow u = P_C(x).$$

Therefore, we deduce that $J_{\lambda B}(x) = P_C(x)$.

(iv) Let $\mathrm{prox}_{\lambda\varphi}$ represent the proximal mapping of a proper convex and lower semicontinuous function φ of parameter $\lambda > 0$, which is defined as follows:

$$\mathrm{prox}_{\lambda\varphi}(x) := \arg\min_{y \in \mathscr{H}} \left\{\lambda\varphi(y) + \frac{1}{2}\|y - x\|^2\right\}.$$

Note that it has closed-form expressions in some special situations. For example, if $\varphi = \delta_C$, we get $\mathrm{prox}_{\lambda\varphi}(x) = P_C(x) = \arg\min_{z \in C} \|x - z\|$.

The following lemmas are very helpful for the convergence analysis of the algorithms.

Lemma 2.1: *Assume that $\mathscr{H}$ is a real Hilbert space, $A : \mathscr{H} \to \mathscr{H}$ is a mapping and $B : \mathscr{H} \to 2^{\mathscr{H}}$ is a maximal monotone mapping. Define the fixed point set of the mapping T as* $\text{Fix}(T) = \{x : x = Tx\}$, *and* $T_\lambda = (I + \lambda B)^{-1}(I - \lambda A), \lambda > 0$. *Then,*

$$\text{Fix}(T_\lambda) = (A + B)^{-1}(0), \quad \forall \lambda > 0.$$

Proof: By the definition of T_λ, one has

$$\begin{aligned} x = T_\lambda x &\Leftrightarrow x = (I + \lambda B)^{-1}(I - \lambda A)x \\ &\Leftrightarrow x - \lambda Ax \in (I + \lambda B)x \Leftrightarrow x \in (A + B)^{-1}(0). \end{aligned}$$

Thus, $\text{Fix}(T_\lambda) = (A + B)^{-1}(0)$. ■

Lemma 2.2 ([32]): *Assume that $\mathscr{H}$ is a real Hilbert space. Let mapping $A : \mathscr{H} \to \mathscr{H}$ be Lipschitz continuous monotone and mapping $B : \mathscr{H} \to 2^{\mathscr{H}}$ be maximal monotone. Then the mapping $(A + B)$ is maximal monotone.*

Lemma 2.3 ([33]): *Let $\{a_n\}$ be a sequence of nonnegative real numbers, $\{\sigma_n\}$ be a sequence of real numbers in $(0, 1)$ with $\sum_{n=1}^{\infty} \sigma_n = \infty$, and $\{b_n\}$ be a sequence of real numbers. Assume that*

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

If $\limsup_{k\to\infty} b_{n_k} \leq 0$ *for every subsequence $\{a_{n_k}\}$ of $\{a_n\}$ satisfying* $\liminf_{k\to\infty} (a_{n_k+1} - a_{n_k}) \geq 0$, *then* $\lim_{n\to\infty} a_n = 0$.

3. Main results

In this section, we propose four inertial forward–backward algorithms to solve the inclusion problem (MIP) in real Hilbert spaces. The advantage of our algorithms is that they do not require the prior information of the Lipschitz constant of the operator. Before introducing our methods, assume that the following conditions are met.

(C1) The solution set of the inclusion problem (MIP) is nonempty, i.e. $\Omega := (A + B)^{-1}(0) \neq \emptyset$.
(C2) The mappings $A : \mathscr{H} \to \mathscr{H}$ is L-Lipschitz continuous monotone and $B : \mathscr{H} \to 2^{\mathscr{H}}$ is maximal monotone.
(C3) Let $\{\epsilon_n\}$ be a positive sequence such that $\lim_{n\to\infty} \frac{\epsilon_n}{\alpha_n} = 0$, where $\{\alpha_n\} \subset (0, 1)$ satisfies $\lim_{n\to\infty} \alpha_n = 0$ and $\sum_{n=1}^{\infty} \alpha_n = \infty$.
(C4) The mapping $f : \mathscr{H} \to \mathscr{H}$ is ρ-contraction with constant $\rho \in [0, 1)$.

3.1. The inertial viscosity-type projection algorithm

In this subsection, in order to solve the inclusion problem (MIP), we introduce an inertial projection method that is inspired by the inertial method, the forward–backward method, the projection and contraction method and the viscosity-type method. Details of the iterative scheme are described in Algorithm 3.1.

Algorithm 3.1 The inertial viscosity-type projection algorithm for solving (MIP).

Initialization: Set $\delta > 0, \theta > 0, l \in (0,1), \mu \in (0,1), \gamma \in (0,2)$ and let $x_0, x_1 \in \mathscr{H}$ be arbitrary.
Iterative Steps: Calculate x_{n+1} as follows:
Step 1. Given the iterates x_{n-1} and x_n $(n \geq 1)$. Set $w_n = x_n + \theta_n(x_n - x_{n-1})$, where

$$\theta_n = \begin{cases} \min\left\{\frac{\epsilon_n}{\|x_n - x_{n-1}\|}, \theta\right\}, & \text{if } x_n \neq x_{n-1}; \\ \theta, & \text{otherwise.} \end{cases} \tag{5}$$

Step 2. Compute $y_n = (I + \lambda_n B)^{-1}(I - \lambda_n A)w_n$, where λ_n is chosen to be the largest $\lambda \in \{\delta, \delta l, \delta l^2, \ldots\}$ satisfying the following:

$$\lambda \langle Aw_n - Ay_n, w_n - y_n \rangle \leq \mu \|w_n - y_n\|^2 . \tag{6}$$

If $w_n = y_n$, then stop and y_n is a solution of (MIP). Otherwise, go to **Step 3**.
Step 3. Compute $z_n = w_n - \gamma \eta_n d_n$, where

$$d_n := w_n - y_n - \lambda_n(Aw_n - Ay_n), \quad \eta_n := (1-\mu)\frac{\|w_n - y_n\|^2}{\|d_n\|^2}. \tag{7}$$

Step 4. Compute $x_{n+1} = \alpha_n f(x_n) + (1 - \alpha_n) z_n$.
Set $n := n + 1$ and go to **Step 1**.

Remark 3.1: We have the following observations for Algorithm 3.1.

(i) It follows from (5) that

$$\lim_{n\to\infty} \frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| = 0.$$

Indeed, we have $\theta_n\|x_n - x_{n-1}\| \leq \epsilon_n$ for all n, which together with $\lim_{n\to\infty} \frac{\epsilon_n}{\alpha_n} = 0$ implies that

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

(ii) If $w_n = y_n$ or $d_n = 0$, then $y_n \in \Omega$. Indeed, from the definition of d_n, one obtains

$$\langle w_n - y_n, d_n \rangle = \|w_n - y_n\|^2 - \lambda_n\langle w_n - y_n, Aw_n - Ay_n\rangle \geq (1-\mu)\|w_n - y_n\|^2.$$

This means that if $d_n = 0$, then $w_n = y_n$. We get $y_n \in \Omega$ by means of Lemma 2.1.

The following lemmas are quite helpful for the convergence analysis of our algorithms.

Lemma 3.1: *The Armijo-like search rule* (6) *is well defined and* $\min\{\delta, \frac{\mu l}{L}\} \leq \lambda_n \leq \delta$.

Proof: From the fact that A is L-Lipschitz continuous, one has $\langle Aw_n - Ay_n, w_n - y_n\rangle \leq L\|w_n - y_n\|^2$. Obviously, (6) holds for all $\lambda \leq \mu L^{-1}$. On the other hand, it is easy to see that $\lambda_n \leq \delta$. If $\lambda_n = \delta$, then this lemma is proved. Otherwise, if $\lambda_n < \delta$, then inequality (6) will be violated when $\lambda = \lambda_n l^{-1}$, which indicates that $\lambda_n l^{-1} > \mu L^{-1}$. Hence $\lambda_n \geq \min\{\delta, \frac{\mu l}{L}\}$. ■

Lemma 3.2: *Let $\{w_n\}$, $\{y_n\}$ and $\{z_n\}$ be three sequences formed by Algorithm* 3.1. *Then*

$$\|z_n - p\|^2 \le \|w_n - p\|^2 - \frac{2-\gamma}{\gamma}\|z_n - w_n\|^2, \quad \forall p \in \Omega$$

and

$$\|w_n - y_n\|^2 \le \frac{1 + L^2\delta^2}{[(1-\mu)\gamma]^2}\|z_n - w_n\|^2.$$

Proof: From the definition of z_n, one sees that

$$\begin{aligned}\|z_n - p\|^2 &= \left\|w_n - \gamma\eta_n d_n - p\right\|^2 \\ &= \|w_n - p\|^2 - 2\gamma\eta_n\langle w_n - p, d_n\rangle + \gamma^2\eta_n^2\|d_n\|^2. \end{aligned} \tag{8}$$

By (6) and the definition of d_n, we obtain

$$\begin{aligned}\langle w_n - p, d_n\rangle &= \langle w_n - y_n, d_n\rangle + \langle y_n - p, d_n\rangle \\ &= \big\langle w_n - y_n, w_n - y_n - \lambda_n(Aw_n - Ay_n)\big\rangle + \langle y_n - p, d_n\rangle \\ &= \|w_n - y_n\|^2 - \big\langle w_n - y_n, \lambda_n(Aw_n - Ay_n)\big\rangle + \langle y_n - p, d_n\rangle \\ &\ge (1-\mu)\|w_n - y_n\|^2 + \big\langle y_n - p, w_n - y_n - \lambda_n(Aw_n - Ay_n)\big\rangle. \end{aligned} \tag{9}$$

According to the definition of y_n, one has $(I - \lambda_n A)w_n \in (I + \lambda_n B)y_n$. Since B is maximal monotone, there exists $v_n \in By_n$ satisfying $(I - \lambda_n A)w_n = y_n + \lambda_n v_n$, which yields that

$$v_n = \lambda_n^{-1}(w_n - y_n - \lambda_n Aw_n). \tag{10}$$

Hence, $(A + B)$ is maximal monotone by means of Lemma 2.2. From $Ay_n + v_n \in (A + B)y_n$ and $0 \in (A + B)p$, we have $\langle Ay_n + v_n, y_n - p\rangle \ge 0$, which together with (10) yields

$$\langle w_n - y_n - \lambda_n(Aw_n - Ay_n), y_n - p\rangle \ge 0. \tag{11}$$

By use of (8), (9), (11), and the definitions of η_n and z_n, we get

$$\begin{aligned}\|z_n - p\|^2 &\le \|w_n - p\|^2 - 2\gamma\eta_n(1-\mu)\|w_n - y_n\|^2 + \gamma^2\eta_n^2\|d_n\|^2 \\ &\le \|w_n - p\|^2 - 2\gamma\eta_n^2\|d_n\|^2 + \gamma^2\eta_n^2\|d_n\|^2 \\ &= \|w_n - p\|^2 - \frac{2-\gamma}{\gamma}\|\gamma\eta_n d_n\|^2 \\ &= \|w_n - p\|^2 - \frac{2-\gamma}{\gamma}\|z_n - w_n\|^2. \end{aligned} \tag{12}$$

On the other hand, from the definition of η_n and z_n, we have

$$\begin{aligned}\|w_n - y_n\|^2 &= \frac{\eta_n}{1-\mu}\|d_n\|^2 = \frac{1}{1-\mu}\|\gamma\eta_n d_n\|^2 \cdot \frac{1}{\gamma^2} \cdot \frac{1}{\eta_n} \\ &= \frac{1}{1-\mu}\|z_n - w_n\|^2 \cdot \frac{1}{\gamma^2} \cdot \frac{1}{\eta_n}. \end{aligned} \tag{13}$$

From the definition of d_n and the fact that A is L-Lipschitz continuous monotone, we obtain

$$\|d_n\|^2 = \|w_n - y_n - \lambda_n(Aw_n - Ay_n)\|^2$$

$$= \|w_n - y_n\|^2 + \lambda_n^2\|Aw_n - Ay_n\|^2 - 2\lambda_n\langle w_n - y_n, Aw_n - Ay_n\rangle$$
$$\leq (1 + L^2\delta^2)\|w_n - y_n\|^2.$$

It follows from the definition of η_n that

$$\frac{1}{\eta_n} = \frac{\|d_n\|^2}{(1-\mu)\|w_n - y_n\|^2} \leq \frac{1 + L^2\delta^2}{1-\mu},$$

which, together with (13), deduces that

$$\|w_n - y_n\|^2 \leq \frac{1 + L^2\delta^2}{[(1-\mu)\gamma]^2}\|z_n - w_n\|^2.$$

This completes the proof of the lemma. ■

Lemma 3.3: *Assume that the sequences* $\{w_n\}$ *and* $\{y_n\}$ *are created by Algorithm* 3.1. *If* $\lim_{n\to\infty}\|w_n - y_n\| = 0$ *and* $\{w_{n_k}\}$ *converges weakly to some* $z \in \mathscr{H}$, *then* $z \in \Omega$.

Proof: Let $(v, u) \in \text{Graph}(A + B)$, i.e. $u \in (A + B)v$. From the definition of y_n, one sees that $(I - \lambda_{n_k}A)w_{n_k} \in (I + \lambda_{n_k}B)y_{n_k}$. This means that

$$\lambda_{n_k}^{-1}(w_{n_k} - y_{n_k} - \lambda_{n_k}Aw_{n_k}) \in By_{n_k}.$$

Since mapping B is maximal monotone, we obtain

$$\langle u - Av - \lambda_{n_k}^{-1}(w_{n_k} - y_{n_k} - \lambda_{n_k}Aw_{n_k}), v - y_{n_k}\rangle \geq 0.$$

This combining with the monotonicity of A finds that

$$\begin{aligned}\langle v - y_{n_k}, u\rangle &\geq \left\langle v - y_{n_k}, Av + \lambda_{n_k}^{-1}\left(w_{n_k} - y_{n_k} - \lambda_{n_k}Aw_{n_k}\right)\right\rangle \\ &= \langle v - y_{n_k}, Av - Ay_{n_k}\rangle + \langle v - y_{n_k}, Ay_{n_k} - Aw_{n_k}\rangle \\ &\quad + \left\langle v - y_{n_k}, \lambda_{n_k}^{-1}(w_{n_k} - y_{n_k})\right\rangle \\ &\geq \langle v - y_{n_k}, Ay_{n_k} - Aw_{n_k}\rangle + \left\langle v - y_{n_k}, \lambda_{n_k}^{-1}(w_{n_k} - y_{n_k})\right\rangle.\end{aligned}$$

Moreover, with the help of $\lim_{n\to\infty}\|w_n - y_n\| = 0$ and the fact that A is Lipschitz continuous, we find that $\lim_{k\to\infty}\|Ay_{n_k} - Aw_{n_k}\| = 0$. Since $\lambda_n > 0$, one infers that

$$\lim_{k\to\infty}\langle v - y_{n_k}, u\rangle = \langle v - z, u\rangle \geq 0,$$

which together with the maximal monotonicity of $(A + B)$ yields that $0 \in (A + B)z$, i.e. $z \in \Omega$. ■

Remark 3.2: It is worth noting that the proof of Lemma 3.3 does not use the definition of Armijo stepsize (6).

Theorem 3.1: *Suppose that Assumptions* (C1)–(C4) *hold. Then the sequence* $\{x_n\}$ *formed by Algorithm* 3.1 *converges to* $p \in \Omega$ *in norm, where* $p = P_\Omega \circ f(p)$.

Proof: First, we show that the sequence $\{x_n\}$ is bounded. Indeed, it follows from Lemma 3.2 that

$$\|z_n - p\| \le \|w_n - p\|, \quad \forall n \ge 1. \tag{14}$$

By the definition of w_n, we can write

$$\|w_n - p\| \le \|x_n - p\| + \alpha_n \cdot \frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\|. \tag{15}$$

According to Remark 3.1 (i), one has $\frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| \to 0$. Therefore, there exists a constant $M_1 > 0$ such that

$$\frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| \le M_1, \quad \forall n \ge 1. \tag{16}$$

Combining (14), (15) and (16), we obtain

$$\|z_n - p\| \le \|w_n - p\| \le \|x_n - p\| + \alpha_n M_1. \tag{17}$$

Thus, we obtain

$$\begin{aligned}
\|x_{n+1} - p\| &\le \alpha_n\|f(x_n) - f(p)\| + \alpha_n\|f(p) - p\| + (1-\alpha_n)\|z_n - p\| \\
&\le [1 - \alpha_n(1-\rho)]\|x_n - p\| + \alpha_n(1-\rho)\frac{\|f(p) - p\| + M_1}{1-\rho} \\
&\le \max\left\{\|x_n - p\|, \frac{\|f(p) - p\| + M_1}{1-\rho}\right\} \\
&\le \cdots \le \max\left\{\|x_0 - p\|, \frac{\|f(p) - p\| + M_1}{1-\rho}\right\}.
\end{aligned}$$

This implies that the sequence $\{x_n\}$ is bounded. So the sequences $\{f(x_n)\}$, $\{w_n\}$, $\{y_n\}$ and $\{z_n\}$ are also bounded. From (17), one sees that

$$\begin{aligned}
\|w_n - p\|^2 &\le \big(\|x_n - p\| + \alpha_n M_1\big)^2 \\
&= \|x_n - p\|^2 + \alpha_n\big(2M_1\|x_n - p\| + \alpha_n M_1^2\big) \\
&\le \|x_n - p\|^2 + \alpha_n M_2
\end{aligned} \tag{18}$$

for some $M_2 > 0$. Combining Lemma 3.2 and (18), we obtain

$$\begin{aligned}
\|x_{n+1} - p\|^2 &\le \alpha_n\|f(x_n) - p\|^2 + (1-\alpha_n)\|z_n - p\|^2 \\
&\le \alpha_n\big(\|f(x_n) - f(p)\| + \|f(p) - p\|\big)^2 + (1-\alpha_n)\|z_n - p\|^2 \\
&\le \alpha_n\big(\|x_n - p\| + \|f(p) - p\|\big)^2 + (1-\alpha_n)\|z_n - p\|^2 \\
&= \alpha_n\|x_n - p\|^2 + (1-\alpha_n)\|z_n - p\|^2 \\
&\quad + \alpha_n\big(2\|x_n - p\| \cdot \|f(p) - p\| + \|f(p) - p\|^2\big) \\
&\le \alpha_n\|x_n - p\|^2 + (1-\alpha_n)\|z_n - p\|^2 + \alpha_n M_3 \\
&\le \|x_n - p\|^2 - (1-\alpha_n)\frac{2-\gamma}{\gamma}\|w_n - z_n\|^2 + \alpha_n M_4,
\end{aligned}$$

where $M_4 := M_2 + M_3$. It follows that

$$(1-\alpha_n)\frac{2-\gamma}{\gamma}\|w_n - z_n\|^2 \le \|x_n - p\|^2 - \|x_{n+1} - p\|^2 + \alpha_n M_4. \tag{TC1}$$

In view of the definition of w_n, one asserts that

$$\begin{aligned}\|w_n - p\|^2 &\le \|x_n - p\|^2 + 2\theta_n\|x_n - p\|\|x_n - x_{n-1}\| + \theta_n^2\|x_n - x_{n-1}\|^2\\ &\le \|x_n - p\|^2 + 3M\theta_n\|x_n - x_{n-1}\|,\end{aligned} \tag{19}$$

where $M := \sup_{n\in\mathbb{N}}\{\|x_n - p\|, \theta\|x_n - x_{n-1}\|\} > 0$. Using (17) and (19), we have

$$\begin{aligned}\|x_{n+1} - p\|^2 &= \|\alpha_n(f(x_n) - f(p)) + (1-\alpha_n)(z_n - p) + \alpha_n(f(p) - p)\|^2\\ &\le \|\alpha_n(f(x_n) - f(p)) + (1-\alpha_n)(z_n - p)\|^2 + 2\alpha_n\langle f(p) - p, x_{n+1} - p\rangle\\ &\le \alpha_n\|f(x_n) - f(p)\|^2 + (1-\alpha_n)\|z_n - p\|^2 + 2\alpha_n\langle f(p) - p, x_{n+1} - p\rangle\\ &\le \alpha_n\rho\|x_n - p\|^2 + (1-\alpha_n)\|w_n - p\|^2 + 2\alpha_n\langle f(p) - p, x_{n+1} - p\rangle\\ &\le (1-(1-\rho)\alpha_n)\|x_n - p\|^2 + (1-\rho)\alpha_n\cdot\left[\frac{3M}{1-\rho}\cdot\frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\|\right.\\ &\quad\left. + \frac{2}{1-\rho}\langle f(p) - p, x_{n+1} - p\rangle\right].\end{aligned} \tag{TC2}$$

Next, one shows that $\{\|x_n - p\|^2\}$ converges to zero. Indeed, by Lemma 2.3, it suffices to show that $\limsup_{k\to\infty}\langle f(p) - p, x_{n_k+1} - p\rangle \le 0$ for every subsequence $\{\|x_{n_k} - p\|\}$ of $\{\|x_n - p\|\}$ satisfying $\liminf_{k\to\infty}(\|x_{n_k+1} - p\| - \|x_{n_k} - p\|) \ge 0$. For this purpose, we assume that $\{\|x_{n_k} - p\|\}$ is a subsequence of $\{\|x_n - p\|\}$ such that $\liminf_{k\to\infty}(\|x_{n_k+1} - p\| - \|x_{n_k} - p\|) \ge 0$. Then,

$$\begin{aligned}&\liminf_{k\to\infty}\left(\|x_{n_k+1} - p\|^2 - \|x_{n_k} - p\|^2\right)\\ &\quad= \liminf_{k\to\infty}\left[(\|x_{n_k+1} - p\| - \|x_{n_k} - p\|)(\|x_{n_k+1} - p\| + \|x_{n_k} - p\|)\right] \ge 0.\end{aligned}$$

From (TC1), the assumption on $\{\alpha_n\}$ and $\gamma \in (0,2)$, one finds that

$$\begin{aligned}(1-\alpha_{n_k})\frac{2-\gamma}{\gamma}\|w_{n_k} - z_{n_k}\|^2 &\le \limsup_{k\to\infty}\left[\|x_{n_k} - p\|^2 - \|x_{n_k+1} - p\|^2 + \alpha_{n_k}M_4\right]\\ &\le 0,\end{aligned}$$

which implies that

$$\lim_{k\to\infty}\|z_{n_k} - w_{n_k}\| = 0. \tag{20}$$

This together with Lemma 3.2 finds that $\lim_{k\to\infty}\|y_{n_k} - w_{n_k}\| = 0$. Moreover, using Remark 3.1(i) and Condition (C3), we have

$$\|x_{n_k+1} - z_{n_k}\| = \alpha_{n_k}\|z_{n_k} - f(x_{n_k})\| \to 0 \tag{21}$$

and

$$\|x_{n_k} - w_{n_k}\| = \alpha_{n_k}\cdot\frac{\theta_{n_k}}{\alpha_{n_k}}\|x_{n_k} - x_{n_k-1}\| \to 0. \tag{22}$$

From (20)–(22), we conclude that

$$\|x_{n_k+1} - x_{n_k}\| \le \|x_{n_k+1} - z_{n_k}\| + \|z_{n_k} - w_{n_k}\| + \|w_{n_k} - x_{n_k}\| \to 0. \tag{23}$$

Since the sequence $\{x_{n_k}\}$ is bounded, there exists a subsequence $\{x_{n_{k_j}}\}$ of $\{x_{n_k}\}$ such that $x_{n_{k_j}} \rightharpoonup z$. Furthermore,

$$\limsup_{k\to\infty}\langle f(p)-p, x_{n_k}-p\rangle = \lim_{j\to\infty}\langle f(p)-p, x_{n_{k_j}}-p\rangle = \langle f(p)-p, z-p\rangle. \tag{24}$$

We get $w_{n_k} \rightharpoonup z$ since $\|x_{n_k}-w_{n_k}\| \to 0$. This together with $\lim_{k\to\infty}\|w_{n_k}-y_{n_k}\| = 0$ and Lemma 3.3 obtains $z \in \Omega$. From the definition of p and (24), we obtain

$$\limsup_{k\to\infty}\langle f(p)-p, x_{n_k}-p\rangle = \langle f(p)-p, z-p\rangle \le 0. \tag{25}$$

Combining (23) and (25), we obtain

$$\limsup_{k\to\infty}\langle f(p)-p, x_{n_k+1}-p\rangle \le \limsup_{k\to\infty}\langle f(p)-p, x_{n_k}-p\rangle \le 0. \tag{26}$$

Thus, from Remark 3.1(i), (26), (TC2) and Lemma 2.3, we conclude that $x_n \to p$. That is the desired result. ■

3.2. The inertial Mann-type projection algorithm

In this subsection, we propose an inertial Mann-type projection algorithm to solve (MIP) and assume that our algorithm satisfies the conditions (C1)–(C3) and (C5).

(C5) Assume that the real sequence $\{\beta_n\} \subset (0,1)$ such that $\{\beta_n\} \subset (a,b) \subset (0, 1-\alpha_n)$ for some $a > 0, b > 0$.

The Algorithm (3.2) is of the form:

Algorithm 3.2 The inertial Mann-type projection algorithm for solving (MIP).

Initialization: Set $\delta > 0, \theta > 0, l \in (0,1), \mu \in (0,1), \gamma \in (0,2)$ and let $x_0, x_1 \in \mathscr{H}$ be arbitrary.
Iterative Steps: Calculate the next iteration point x_{n+1} as follows:

$$\begin{cases} w_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = (I+\lambda_n B)^{-1}(I-\lambda_n A)w_n, \\ z_n = w_n - \gamma\eta_n d_n, \\ x_{n+1} = (1-\alpha_n-\beta_n)w_n + \beta_n z_n, \end{cases}$$

where $\{\theta_n\}$, $\{\lambda_n\}$ and $\{d_n\}$ are defined in (5)- -(7), respectively.

Theorem 3.2: *Suppose that Assumptions* (C1)–(C3) *and* (C5) *hold. Then the sequence* $\{x_n\}$ *created by Algorithm* 3.2 *converges to* $p \in \Omega$ *in norm, where* $\|p\| = \min\{\|z\| : z \in \Omega\}$.

Proof: Thanks to Lemma 3.2, we have

$$\|z_n - p\| \le \|w_n - p\|, \quad \forall n \ge 1. \tag{27}$$

By use of the definition of x_{n+1}, one has

$$\|x_{n+1}-p\| = \|(1-\alpha_n-\beta_n)(w_n-p) + \beta_n(z_n-p) - \alpha_n p\|$$

$$\leq \|(1-\alpha_n-\beta_n)(w_n-p)+\beta_n(z_n-p)\|+\alpha_n\|p\|. \tag{28}$$

It follows from (27) that

$$\begin{aligned}
&\|(1-\alpha_n-\beta_n)(w_n-p)+\beta_n(z_n-p)\|^2\\
&\quad\leq (1-\alpha_n-\beta_n)^2\|w_n-p\|^2+2(1-\alpha_n-\beta_n)\beta_n\|z_n-p\|\|w_n-p\|+\beta_n^2\|z_n-p\|^2\\
&\quad\leq (1-\alpha_n-\beta_n)^2\|w_n-p\|^2+2(1-\alpha_n-\beta_n)\beta_n\|w_n-p\|^2+\beta_n^2\|w_n-p\|^2\\
&\quad= (1-\alpha_n)^2\|w_n-p\|^2,
\end{aligned}$$

which yields

$$\|(1-\alpha_n-\beta_n)(w_n-p)+\beta_n(z_n-p)\|\leq (1-\alpha_n)\|w_n-p\|. \tag{29}$$

Combining (17), (28) and (29), we deduce that

$$\begin{aligned}
\|x_{n+1}-p\| &\leq (1-\alpha_n)\|w_n-p\|+\alpha_n\|p\|\\
&\leq (1-\alpha_n)\|x_n-p\|+\alpha_n(\|p\|+M_1)\\
&\leq \max\{\|x_n-p\|,\|p\|+M_1\}\\
&\leq\cdots\leq \max\{\|x_0-p\|,\|p\|+M_1\}.
\end{aligned}$$

That is, the sequence $\{x_n\}$ is bounded, so are $\{z_n\}$ and $\{w_n\}$. From Lemma 3.2 and (18), we obtain

$$\begin{aligned}
\|x_{n+1}-p\|^2 &= \|(1-\alpha_n-\beta_n)(w_n-p)+\beta_n(z_n-p)+\alpha_n(-p)\|^2\\
&\leq (1-\alpha_n-\beta_n)\|w_n-p\|^2+\beta_n\|z_n-p\|^2+\alpha_n\|p\|^2\\
&\leq (1-\alpha_n-\beta_n)\|w_n-p\|^2+\beta_n\|w_n-p\|^2-\beta_n\frac{2-\gamma}{\gamma}\|w_n-z_n\|^2+\alpha_n\|p\|^2\\
&\leq \|x_n-p\|^2-\beta_n\frac{2-\gamma}{\gamma}\|w_n-z_n\|^2+\alpha_n(\|p\|^2+M_2).
\end{aligned}$$

Hence,

$$\beta_n\frac{2-\gamma}{\gamma}\|w_n-z_n\|^2\leq \|x_n-p\|^2-\|x_{n+1}-p\|^2+\alpha_n(\|p\|^2+M_2). \tag{TC3}$$

Setting $t_n=(1-\beta_n)w_n+\beta_n z_n$, one has

$$\|t_n-w_n\|=\beta_n\|w_n-z_n\|. \tag{30}$$

It follows from (27) that

$$\begin{aligned}
\|t_n-p\| &= \|(1-\beta_n)(w_n-p)+\beta_n(z_n-p)\|\\
&\leq (1-\beta_n)\|w_n-p\|+\beta_n\|w_n-p\|\\
&= \|w_n-p\|.
\end{aligned} \tag{31}$$

From (19), (30) and (31), we have

$$\begin{aligned}
\|x_{n+1}-p\|^2 &= \|(1-\beta_n)w_n+\beta_n z_n-\alpha_n w_n-p\|^2\\
&= \|(1-\alpha_n)(t_n-p)-\alpha_n(w_n-t_n)-\alpha_n p\|^2\\
&\leq (1-\alpha_n)^2\|t_n-p\|^2-2\alpha_n\langle w_n-t_n+p,x_{n+1}-p\rangle
\end{aligned}$$

$$
\begin{aligned}
&= (1-\alpha_n)^2\|t_n - p\|^2 + 2\alpha_n\langle w_n - t_n, p - x_{n+1}\rangle + 2\alpha_n\langle p, p - x_{n+1}\rangle \\
&\le (1-\alpha_n)\|t_n - p\|^2 + 2\alpha_n\|w_n - t_n\|\|x_{n+1} - p\| + 2\alpha_n\langle p, p - x_{n+1}\rangle \\
&\le (1-\alpha_n)\|x_n - p\|^2 + \alpha_n\Bigg[2\beta_n\|w_n - z_n\|\|x_{n+1} - p\| \\
&\quad +2\langle p, p - x_{n+1}\rangle + \frac{3M\theta_n}{\alpha_n}\|x_n - x_{n-1}\|\Bigg]. \qquad \text{(TC4)}
\end{aligned}
$$

Finally, one shows that $\{\|x_n - p\|^2\}$ converges to zero. Assume that $\{\|x_{n_k} - p\|\}$ is a subsequence of $\{\|x_n - p\|\}$ such that $\liminf_{k\to\infty}(\|x_{n_k+1} - p\| - \|x_{n_k} - p\|) \ge 0$. By use of (TC3) and Condition (C5), we have

$$
\begin{aligned}
\beta_{n_k}\frac{2-\gamma}{\gamma}\|w_{n_k} - z_{n_k}\|^2 &\le \limsup_{k\to\infty}\left[\|x_{n_k} - p\|^2 - \|x_{n_k+1} - p\|^2\right] + \limsup_{k\to\infty}\alpha_{n_k}\left(\|p\|^2 + M_2\right) \\
&\le 0,
\end{aligned}
$$

which indicates that

$$
\lim_{k\to\infty}\|z_{n_k} - w_{n_k}\| = 0. \qquad (32)
$$

In view of Lemma 3.2, one observes that $\lim_{k\to\infty}\|y_{n_k} - w_{n_k}\| = 0$. From (32) and the boundedness of $\{x_n\}$, we can further obtain

$$
\lim_{k\to\infty}\beta_{n_k}\|w_{n_k} - z_{n_k}\|\|x_{n_k+1} - p\| = 0. \qquad (33)
$$

Moreover, using (32), Condition (C5) and Remark 3.1 (i), we have

$$
\|x_{n_k+1} - w_{n_k}\| = \alpha_{n_k}\|w_{n_k}\| + \beta_{n_k}\|z_{n_k} - w_{n_k}\| \to 0
$$

and

$$
\|x_{n_k} - w_{n_k}\| = \alpha_{n_k}\cdot\frac{\theta_{n_k}}{\alpha_{n_k}}\|x_{n_k} - x_{n_k-1}\| \to 0.
$$

Thus, we conclude that

$$
\|x_{n_k+1} - x_{n_k}\| \le \|x_{n_k+1} - w_{n_k}\| + \|w_{n_k} - x_{n_k}\| \to 0. \qquad (34)
$$

Since the sequence $\{x_{n_k}\}$ is bounded, there exists a subsequence $\{x_{n_{k_j}}\}$ of $\{x_{n_k}\}$ such that $x_{n_{k_j}} \rightharpoonup z$. Moreover,

$$
\limsup_{k\to\infty}\langle p, p - x_{n_k}\rangle = \lim_{j\to\infty}\langle p, p - x_{n_{k_j}}\rangle = \langle p, p - z\rangle. \qquad (35)
$$

Since $\|x_{n_k} - w_{n_k}\| \to 0$, one has $w_{n_k} \rightharpoonup z$, which together with $\lim_{k\to\infty}\|w_{n_k} - y_{n_k}\| = 0$ and Lemma 3.3, gets $z \in \Omega$. From the definition of p and (35), we obtain

$$
\limsup_{k\to\infty}\langle p, p - x_{n_k}\rangle = \langle p, p - z\rangle \le 0. \qquad (36)
$$

Combining (34) and (36), we have

$$
\limsup_{k\to\infty}\langle p, p - x_{n_k+1}\rangle \le \limsup_{k\to\infty}\langle p, p - x_{n_k}\rangle \le 0. \qquad (37)
$$

Thus, from Remark 3.1(i), (33), (37), (TC4) and Lemma 2.3, we conclude that $x_n \to p$. The proof is completed. ■

3.3. The inertial Mann-type Tseng algorithm

In this subsection, an inertial Mann-type Tseng algorithm will be given. It is worth noting that this method uses a new step size update criterion that does not require any line search process. More precisely, the Algorithm 3.3 is described as follows.

Algorithm 3.3 The inertial Mann-type Tseng algorithm for solving (MIP).

Initialization: Set $\lambda_0 > 0, \theta > 0, \mu \in (0,1)$ and let $x_0, x_1 \in \mathcal{H}$ be arbitrary.
Iterative Steps: Calculate the next iteration point x_{n+1} as follows:

$$\begin{cases} w_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = (I + \lambda_n B)^{-1}(I - \lambda_n A)w_n, \\ z_n = y_n - \lambda_n(Ay_n - Ax_n), \\ x_{n+1} = (1 - \alpha_n - \beta_n)w_n + \beta_n z_n, \end{cases}$$

where $\{\theta_n\}$ is defined in (5) and the stepsize λ_n is updated by the following:

$$\lambda_{n+1} = \begin{cases} \min\left\{\dfrac{\mu\|w_n - y_n\|}{\|Aw_n - Ay_n\|}, \lambda_n\right\}, & \text{if } Aw_n - Ay_n \neq 0; \\ \lambda_n, & \text{otherwise.} \end{cases} \tag{38}$$

The following two lemmas are very important for the convergence analysis of the algorithms.

Lemma 3.4: *The sequence $\{\lambda_n\}$ formed by (38) is nonincreasing and $\lim_{n\to\infty} \lambda_n = \lambda \geq \min\{\lambda_0, \frac{\mu}{L}\}$.*

Proof: It is easy to see that the sequence $\{\lambda_n\}$ is nonincreasing by the definition of $\{\lambda_n\}$. Moreover, from the fact that operator A is L-Lipschitz continuous, one obtains,

$$\frac{\mu\|w_n - y_n\|}{\|Aw_n - Ay_n\|} \geq \frac{\mu}{L}, \quad \text{if } Aw_n - Ay_n \neq 0.$$

Therefore, we deduce that $\lim_{n\to\infty} \lambda_n = \lambda \geq \min\{\lambda_0, \frac{\mu}{L}\}$. ■

Lemma 3.5: *Suppose that Conditions (C1)–(C3) hold. Let the sequences $\{w_n\}$ and $\{z_n\}$ be made by Algorithm 3.3. Then*

$$\|z_n - p\|^2 \leq \|w_n - p\|^2 - \left(1 - \mu^2\frac{\lambda_n^2}{\lambda_{n+1}^2}\right)\|w_n - y_n\|^2, \quad \forall p \in \Omega.$$

Proof: From the definition of z_n, one has

$$\begin{aligned} \|z_n - p\|^2 &= \|y_n - p\|^2 + \lambda_n^2\|Ay_n - Aw_n\|^2 - 2\lambda_n\langle y_n - p, Ay_n - Aw_n\rangle \\ &= \|w_n - p\|^2 + \|y_n - w_n\|^2 - 2\langle y_n - w_n, y_n - w_n\rangle + 2\langle y_n - w_n, y_n - p\rangle \\ &\quad + \lambda_n^2\|Ay_n - Aw_n\|^2 - 2\lambda_n\langle y_n - p, Ay_n - Aw_n\rangle \\ &= \|w_n - p\|^2 - \|w_n - y_n\|^2 - 2\langle w_n - y_n - \lambda_n(Aw_n - Ay_n), y_n - p\rangle \\ &\quad + \lambda_n^2\|Ay_n - Aw_n\|^2. \end{aligned} \tag{39}$$

It follows from the definition of λ_n that

$$\|Aw_n - Ay_n\| \leq \frac{\mu}{\lambda_{n+1}}\|w_n - y_n\|, \quad \forall n. \tag{40}$$

Indeed, if $Aw_n = Ay_n$, then inequality (40) holds. Otherwise, we obtain

$$\lambda_{n+1} \leq \frac{\mu\|w_n - y_n\|}{\|Aw_n - Ay_n\|},$$

which yields that $\|Aw_n - Ay_n\| \leq \frac{\mu}{\lambda_{n+1}}\|w_n - y_n\|$. Therefore, inequality (40) holds for all n.

Next, we will show that

$$\left\langle w_n - y_n - \lambda_n(Aw_n - Ay_n), y_n - p\right\rangle \geq 0. \tag{41}$$

From the definition of y_n, one sees that $(I - \lambda_n A)w_n \in (I + \lambda_n B)y_n$. By use of the maximal monotonicity of B, there exists $v_n \in By_n$ such that $(I - \lambda_n A)w_n = y_n + \lambda_n v_n$. This indicates that

$$v_n = \frac{1}{\lambda_n}(w_n - y_n - \lambda_n Aw_n). \tag{42}$$

Using the definition of p, one obtains $0 \in (A + B)p$. From $Ay_n + v_n \in (A + B)y_n$ and the fact that $(A + B)$ is maximal monotone, we get $\langle Ay_n + v_n, y_n - p\rangle \geq 0$. This together with (42) yields

$$\lambda_n^{-1}\langle w_n - y_n - \lambda_n Aw_n + \lambda_n Ay_n, y_n - p\rangle \geq 0,$$

which further infers that (41) holds. Combining (39)–(41), we obtain

$$\|z_n - p\|^2 \leq \|w_n - p\|^2 - \left(1 - \mu^2\frac{\lambda_n^2}{\lambda_{n+1}^2}\right)\|w_n - y_n\|^2.$$

The proof of the lemma is now complete. ■

Next, the convergence theorem of Algorithm 3.3 is given. This proof has some common characteristics with the proof of Theorem 3.2, but there are still some differences. Therefore, we decide to introduce it in a compact form.

Theorem 3.3: *Suppose that Assumptions* (C1)–(C3) *and* (C5) *hold. Then the sequence* $\{x_n\}$ *formed by Algorithm* 3.3 *converges to* $p \in \Omega$ *in norm, where* $\|p\| = \min\{\|z\| : z \in \Omega\}$.

Proof: From $\lim_{n\to\infty}(1 - \mu^2\frac{\lambda_n^2}{\lambda_{n+1}^2}) = 1 - \mu^2 > 0$, one sees that there exists $n_0 \in \mathbb{N}$ such that $1 - \mu^2\frac{\lambda_n^2}{\lambda_{n+1}^2} > 0, \forall n \geq n_0$, which together with Lemma 3.5 yields that

$$\|z_n - p\| \leq \|w_n - p\|, \quad \forall n \geq n_0. \tag{43}$$

From (17), (28) and (29), we obtain

$$\begin{aligned}\|x_{n+1} - p\| &\leq (1 - \alpha_n)\|w_n - p\| + \alpha_n\|p\| \\ &\leq (1 - \alpha_n)\|x_n - p\| + \alpha_n(\|p\| + M_1) \\ &\leq \max\{\|x_n - p\|, \|p\| + M_1\} \\ &\leq \cdots \leq \max\{\|x_{n_0} - p\|, \|p\| + M_1\},\end{aligned}$$

where M_1 is defined as in Theorem 3.1. Thus, the sequence $\{x_n\}$ is bounded. Consequently, the sequences $\{w_n\}$ and $\{z_n\}$ are also bounded. From Lemma 3.5 and (18), we have

$$\begin{aligned}\|x_{n+1}-p\|^2 &= \left\|(1-\alpha_n-\beta_n)(w_n-p)+\beta_n(z_n-p)+\alpha_n(-p)\right\|^2\\ &\le (1-\alpha_n-\beta_n)\|w_n-p\|^2+\beta_n\|z_n-p\|^2+\alpha_n\|p\|^2\\ &\quad -\beta_n(1-\alpha_n-\beta_n)\|w_n-z_n\|^2\\ &\le \|x_n-p\|^2-\beta_n\left(1-\mu^2\frac{\lambda_n^2}{\lambda_{n+1}^2}\right)\|w_n-y_n\|^2\\ &\quad +\alpha_n(\|p\|^2+M_2)-\beta_n(1-\alpha_n-\beta_n)\|w_n-z_n\|^2,\end{aligned}$$

where M_2 is defined as in Theorem 3.1. So,

$$\begin{aligned}&\beta_n\left(1-\mu^2\frac{\lambda_n^2}{\lambda_{n+1}^2}\right)\|w_n-y_n\|^2+\beta_n(1-\alpha_n-\beta_n)\|w_n-z_n\|^2\\ &\quad\le \|x_n-p\|^2-\|x_{n+1}-p\|^2+\alpha_n\big(\|p\|^2+M_2\big).\end{aligned}\tag{TC5}$$

Using the same facts as (TC4) of Theorem 3.2, we find that

$$\begin{aligned}\|x_{n+1}-p\|^2 \le (1-\alpha_n)\|x_n-p\|^2+\alpha_n\Big[&2\beta_n\|w_n-z_n\|\|x_{n+1}-p\|\\ &+2\langle p,p-x_{n+1}\rangle+\frac{3M\theta_n}{\alpha_n}\|x_n-x_{n-1}\|\Big].\end{aligned}\tag{TC6}$$

Finally, we show that $\{\|x_n-p\|^2\}$ converges to zero. Assume that $\{\|x_{n_k}-p\|\}$ is a subsequence of $\{\|x_n-p\|\}$ such that $\liminf_{k\to\infty}(\|x_{n_k+1}-p\|-\|x_{n_k}-p\|)\ge 0$. From (TC5) and Condition (C5), we have

$$\begin{aligned}&\beta_{n_k}\left(1-\mu^2\frac{\lambda_{n_k}^2}{\lambda_{n_k+1}^2}\right)\|w_{n_k}-y_{n_k}\|^2+\beta_{n_k}\left(1-\alpha_{n_k}-\beta_{n_k}\right)\|w_{n_k}-z_{n_k}\|^2\\ &\le \limsup_{k\to\infty}\left[\|x_{n_k}-p\|^2-\|x_{n_k+1}-p\|^2\right]+\limsup_{k\to\infty}\alpha_{n_k}\big(\|p\|^2+M_2\big)\le 0,\end{aligned}$$

which implies that $\lim_{k\to\infty}\|z_{n_k}-w_{n_k}\|=0$ and $\lim_{k\to\infty}\|y_{n_k}-w_{n_k}\|=0$. From (33)–(37), we can show that

$$\lim_{k\to\infty}\beta_{n_k}\|w_{n_k}-z_{n_k}\|\|x_{n_k+1}-p\|=0$$

and

$$\limsup_{k\to\infty}\langle p,p-x_{n_k+1}\rangle\le 0.$$

Combining these with Remark 3.1(i), (TC6) and Lemma 2.3, we deduce that $x_n\to p$. This completes the proof. ∎

3.4. The inertial viscosity-type Tseng algorithm

Finally, we introduce a modified version of Algorithm 3.3, which uses the viscosity-type method to ensure the strong convergence of the suggested iterative scheme. The Algorithm 3.4 is stated as follows.

Algorithm 3.4 The inertial viscosity-type Tseng algorithm for solving (MIP).

Initialization: Set $\lambda_0 > 0, \theta > 0, \mu \in (0,1)$ and let $x_0, x_1 \in \mathscr{H}$ be arbitrary.
Iterative Steps: Calculate the next iteration point x_{n+1} as follows:

$$\begin{cases} w_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = (I + \lambda_n B)^{-1}(I - \lambda_n A)w_n, \\ z_n = y_n - \lambda_n(Ay_n - Ax_n), \\ x_{n+1} = \alpha_n f(x_n) + (1-\alpha_n)z_n, \end{cases}$$

where $\{\theta_n\}$ and $\{\lambda_n\}$ are defined in (5) and (38), respectively.

Based on the proofs of Theorems 3.1 and 3.3, we will give the convergence analysis of Algorithm 3.4 in a compact way.

Theorem 3.4: *Suppose that Assumptions* (C1)–(C3) *and* (C5) *hold. Then the sequence* $\{x_n\}$ *created by Algorithm* 3.4 *converges to* $p \in \Omega$ *in norm, where* $p = P_\Omega \circ f(p)$.

Proof: Using (15)–(17) and (43), we have

$$\begin{aligned} \|x_{n+1} - p\| &= \left\|\alpha_n(f(x_n) - p) + (1-\alpha_n)(z_n - p)\right\| \\ &\le \max\left\{\|x_{n_0} - p\|, \frac{\|f(p) - p\| + M_1}{1-\rho}\right\}. \end{aligned}$$

This means that the sequence $\{x_n\}$ is bounded. Hence, the sequences $\{f(x_n)\}$, $\{w_n\}$, $\{y_n\}$ and $\{z_n\}$ are also bounded. From (18) and Lemma 3.5, we obtain

$$\begin{aligned} \|x_{n+1} - p\|^2 &\le \alpha_n\|f(x_n) - p\|^2 + (1-\alpha_n)\|z_n - p\|^2 \\ &\le \alpha_n(\|f(x_n) - f(p)\| + \|f(p) - p\|)^2 + (1-\alpha_n)\|z_n - p\|^2 \\ &\le \alpha_n\|x_n - p\|^2 + (1-\alpha_n)\|z_n - p\|^2 \\ &\quad + \alpha_n(2\|x_n - p\| \cdot \|f(p) - p\| + \|f(p) - p\|^2) \\ &\le \alpha_n\|x_n - p\|^2 + (1-\alpha_n)\|z_n - p\|^2 + \alpha_n M_3 \\ &\le \|x_n - p\|^2 - (1-\alpha_n)\left(1 - \mu^2\frac{\lambda_n^2}{\lambda_{n+1}^2}\right)\|w_n - y_n\|^2 + \alpha_n M_4, \end{aligned}$$

where $M_4 := M_2 + M_3$. That is,

$$(1-\alpha_n)\left(1 - \mu^2\frac{\lambda_n^2}{\lambda_{n+1}^2}\right)\|w_n - y_n\|^2 \le \|x_n - p\|^2 - \|x_{n+1} - p\|^2 + \alpha_n M_4. \tag{TC7}$$

From the same line in Theorem 3.1, one arrives at

$$\begin{aligned} \|x_{n+1} - p\|^2 \le (1-(1-\rho)\alpha_n)\|x_n - p\|^2 + (1-\rho)\alpha_n \cdot \bigg[&\frac{3M}{1-\rho} \cdot \frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| \\ &+ \frac{2}{1-\rho}\langle f(p) - p, x_{n+1} - p\rangle\bigg]. \end{aligned} \tag{TC8}$$

Finally, one shows $\{\|x_n - p\|^2\}$ converges to zero. We assume that $\{\|x_{n_k} - p\|\}$ is a subsequence of $\{\|x_n - p\|\}$ such that $\liminf_{k\to\infty}(\|x_{n_k+1} - p\| - \|x_{n_k} - p\|) \geq 0$. By use of Condition (C3) and (TC7), we have

$$(1-\alpha_{n_k})\left(1-\mu^2\frac{\lambda_{n_k}^2}{\lambda_{n_k+1}^2}\right)\|w_{n_k} - y_{n_k}\|^2 \leq \limsup_{k\to\infty}\left[\|x_{n_k} - p\|^2 - \|x_{n_k+1} - p\|^2 + \alpha_{n_k}M_4\right] \leq 0,$$

which implies that $\lim_{k\to\infty}\|w_{n_k} - y_{n_k}\| = 0$. From the definition of z_n and (40), we have

$$\|z_n - y_n\| \leq \mu\frac{\lambda_n}{\lambda_{n+1}}\|w_n - y_n\|,$$

which yields $\lim_{k\to\infty}\|y_{n_k} - z_{n_k}\| = 0$. Therefore, we obtain $\lim_{k\to\infty}\|z_{n_k} - w_{n_k}\| = 0$. From (21)–(26), we observe that

$$\limsup_{k\to\infty}\langle f(p) - p, x_{n_k+1} - p\rangle \leq 0.$$

This together with Remark 3.1(i), (TC8) and Lemma 2.3, we conclude that $x_n \to p$. We have thus proved the theorem. ■

Remark 3.3: We note here that the proposed algorithms directly improve some known results in the literature. The details are as follows:

(i) In the algorithms proposed by Dong et al. [20] and Cholamjiak et al. [21], the operator A is assumed to be inverse strongly monotone, while operator A in our algorithms is Lipschitz continuous monotone. It is known that inverse strongly monotone mappings are monotone and thus the suggested methods are applicable to a wider class of mappings.

(ii) The selection of the step size in the algorithms provided by [20,21,25] requires the knowledge of the Lipschitz constant of the mapping, while our algorithms can adaptively update the step size of each iteration. On the one hand, it is not easy to estimate the Lipschitz constant of the mapping in practical applications. On the other hand, it should be pointed out that Armijo-type search methods need to evaluate the value of the iterative sequences w_n, y_n at operator A multiple times in each iteration. The proposed Algorithms 3.3 and 3.4 use a new step size update criterion that does not involve any line search process. The criterion only needs to use known information for a simple calculation in each iteration to complete the step size update. Therefore, our self-adaptive iterative schemes (especially for Algorithms 3.3 and 3.4) are more preferable than the fixed stepsize methods and the Armijo-type methods [24].

(iii) In the methods introduced by [20,21], they combined FBA with the projection methods to ensure the strong convergence of the proposed algorithms. It is known that calculating projection requires additional cost and is not easy to implement. However, our suggested algorithms do not involve any projection calculations, so they are easier to implement.

(iv) Our presented methods are strongly convergent in real Hilbert spaces, which is more preferable than the weak convergence results of Tseng [17] and Zhang and Wang [18].

(v) If the inertial parameter $\theta_n = 0$, then the suggested Algorithms 3.1, 3.3 and Algorithm 3.4 degenerate into [24, Algorithm 3.1], [23, Algorithm 1] and [23, Algorithm 2], respectively. Our algorithms embed inertial terms so that they converge faster than algorithms without inertial terms (see Section 5).

4. Applications

In this section, we apply our proposed Algorithms 3.1–3.4 to some problems, including convex minimization problems, variational inequality problems, split feasibility problems and image processing problems.

4.1. Application to convex minimization problems

Recall that the convex minimization problem is stated as follows:

$$\text{find } x^* \in \mathscr{H}, \text{ such that } h(x^*) + g(x^*) = \min_{x \in \mathscr{H}} \{h(x) + g(x)\},$$

where $h : \mathscr{H} \to \mathbb{R}$ is a convex differentiable function with L-Lipschitz continuous gradient and monotone, and $g : \mathscr{H} \to \mathbb{R}$ is a proper convex and lower semicontinuous function. In fact, the above problem is a special form of the inclusion problem (MIP), that is, it is equivalent to the following problem: $0 \in \nabla h(x^*) + \partial g(x^*)$, where ∇h and ∂g represent the gradient of function h and the subdifferential of function g, respectively. Set $A = \nabla h$ and $B = \partial g$ in Theorem 3.1, it is known that B is maximal monotone, then we have the following result.

Corollary 4.1: *Assume that $\mathscr{H}$ is a real Hilbert space. Let mapping $\nabla h : \mathscr{H} \to \mathscr{H}$ be L-Lipschitz continuous monotone and let $\partial g : \mathscr{H} \to 2^{\mathscr{H}}$ be maximal monotone. Suppose that $\Omega = (\nabla h + \partial g)^{-1}(0) \neq \emptyset$ and Conditions* (C3)–(C4) *hold. Set $A = \nabla h$ and $B = \partial g$. Let $x_0, x_1 \in \mathscr{H}$ and $\{x_n\}$ be a sequence defined by*

$$\begin{cases} w_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = (I + \lambda_n \partial g)^{-1}(I - \lambda_n \nabla h) w_n, \\ z_n = w_n - \gamma \eta_n d_n, \\ x_{n+1} = \alpha_n f(x_n) + (1 - \alpha_n) z_n, \end{cases}$$

where $\{\theta_n\}$, $\{\lambda_n\}$ and $\{d_n\}$ are defined in (5)–(7), *respectively. Then the iterative sequence $\{x_n\}$ provided above converges to $p \in \Omega$ in norm, where $p = P_\Omega \circ f(p)$.*

Theorems 3.2–3.4 can also have a similar sub-results, which we omit here.

4.2. Application to variational inequality problems

Let C be a nonempty convex subset of Hilbert space $\mathscr{H}$. Set $B = \partial\varphi : \mathscr{H} \to 2^{\mathscr{H}}$, where $\partial\varphi$ is the subdifferential of the proper convex and lower semicontinuous function φ. Then problem (MIP) is equivalent to the following mixed quasi-variational inequality problem (MQVIP):

$$\text{find } x^* \in \mathscr{H}, \text{ such that } \langle Ax^*, y - x^* \rangle + \varphi(y) - \varphi(x^*) \geq 0, \quad \forall y \in \mathscr{H}.$$

On the other hand, if $\varphi = \delta_C$ is the indicator function of C, then the above (MQVIP) can be written as the following variational inequality problem (VIP):

$$\text{find } x^* \in C, \text{ such that } \langle Ax^*, y - x^* \rangle \geq 0, \quad \forall y \in C. \tag{VIP}$$

We shall denote Ω the solution set of (VIP) and assume $\Omega \neq \emptyset$. It is easy to check that the above (VIP) is a special case of (MIP), that is, $x^* \in (A + \partial\varphi)^{-1}(0) \Leftrightarrow x^* \in \Omega$. Moreover, we know that $B = \partial\varphi$ is maximal monotone and $(I + \lambda_n B)^{-1}(x) = P_C(x)$. Thus, the following corollary follows directly from Theorem 3.2.

Corollary 4.2: *Assume that mapping $A : \mathscr{H} \to \mathscr{H}$ is L-Lipschitz continuous monotone and mapping $P_C : \mathscr{H} \to 2^{\mathscr{H}}$ is maximal monotone. Suppose that $\Omega \neq \emptyset$, and Conditions* (C3) *and* (C5) *hold. Let*

$x_0, x_1 \in \mathscr{H}$ and $\{x_n\}$ be a sequence generated by

$$\begin{cases} w_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = P_C(w_n - \lambda_n A w_n), \\ z_n = w_n - \gamma \eta_n d_n, \\ x_{n+1} = (1 - \alpha_n - \beta_n) w_n + \beta_n z_n, \end{cases}$$

where $\{\theta_n\}$, $\{\lambda_n\}$ and $\{d_n\}$ are defined in (5)–(7), respectively. Then the iterative sequence $\{x_n\}$ created above converges to $p \in \Omega$ in norm, where $\|p\| = \min\{\|z\| : z \in \Omega\}$.

4.3. Application to split feasibility problems

Suppose that C and Q are nonempty closed convex subsets of real Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$, respectively. The split feasibility problem (SFP) is described as follows:

$$\text{find } x^* \in C \text{ such that } Tx^* \in Q, \tag{SFP}$$

where $T : \mathscr{H}_1 \to \mathscr{H}_2$ is a bounded linear operator. We also use Ω to represent the solution set of (SFP). Problem (SFP) appears in image reconstruction and signal processing. From an optimization point of view, $x^* \in \Omega$ iff x^* is a solution of the following minimization problem with zero optimal value:

$$\min_{x \in C} h(x) := \frac{1}{2}\|Tx - P_Q Tx\|^2.$$

It should be noted that h is convex differentiable. Moreover, note that $\nabla h(x) = T^*(I - P_Q)Tx$ and it is $\|T\|^2$-Lipschitz continuous monotone. Thus, x^* solves (SFP) iff x^* solves the following variational inclusion problem:

$$\text{find } x \in \mathscr{H}_1, \text{ such that } 0 \in \nabla h(x) + \partial \delta_C(x),$$

where δ_C is the indicator function of C. In Theorem 3.3, choosing $A = \nabla h$ and $B = \partial \delta_C$, then we obtain the following result.

Corollary 4.3: *Let the mappings A and B be defined above. Suppose that $\Omega \neq \emptyset$, and Conditions (C3) and (C5) hold. Let $x_0, x_1 \in \mathscr{H}$ and $\{x_n\}$ be a sequence formed by*

$$\begin{cases} w_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = P_C(I - \lambda_n \nabla h) w_n, \\ z_n = y_n - \lambda_n(\nabla h(y_n) - \nabla h(x_n)), \\ x_{n+1} = (1 - \alpha_n - \beta_n) w_n + \beta_n z_n, \end{cases}$$

where $\{\theta_n\}$ and $\{\lambda_n\}$ are defined in (5) and (38), respectively. Then the iterative sequence $\{x_n\}$ constructed above converges to $p \in \Omega$ in norm, where $\|p\| = \min\{\|z\| : z \in \Omega\}$.

4.4. Application to image processing problems

Using known information from the contaminated signal/image to estimate the original and clean signal/image is called the signal processing/image restoration problem. This kind of problem can usually be expressed as the following linear inverse problem:

$$\mathbf{b} = \mathbf{C}\mathbf{x} + \mathbf{w},$$

where $\mathbf{C}$, $\mathbf{x}$, $\mathbf{b}$ and $\mathbf{w}$ represent degradation operator, unknown real image, contaminated image and noise function, respectively. Regularization methods have aroused considerable interest in many researchers for dealing with such problems. In particular, the l_1 regularization method considers finding the solution to the following problem:

$$\min_{\mathbf{x}} \left\{ \frac{1}{2}\|\mathbf{C}\mathbf{x}-\mathbf{b}\|^2 + \gamma\|\mathbf{x}\|_1 \right\},$$

where γ stands for the regularization parameter, and $\|\mathbf{x}\|_1$ represents the sum of the absolute values of the components of $\mathbf{x}$. Set $h(\mathbf{x}) = \frac{1}{2}\|\mathbf{C}\mathbf{x}-\mathbf{b}\|^2$ and $g(\mathbf{x}) = \|\mathbf{x}\|_1$, then $\nabla h(\mathbf{x}) = \mathbf{C}^*(\mathbf{C}\mathbf{x}-\mathbf{b})$ and thus it is Lipschitz continuous with constant $L(h) = \|\mathbf{C}^*\mathbf{C}\|$. The proximal map of $g(\mathbf{x}) = \gamma\|\mathbf{x}\|_1$ is expressed as $\operatorname{prox}_{\lambda g}(\mathbf{x}) = (I+\lambda\partial g)^{-1}$ and it can be calculated by the following:

$$\begin{aligned}\operatorname{prox}_{\lambda g}(\mathbf{x}) &= \operatorname{prox}_{\lambda\gamma\|\cdot\|_1}(\mathbf{x}) = (\operatorname{prox}_{\lambda\gamma|\cdot|_1}(x_1), \dots, \operatorname{prox}_{\lambda\gamma|\cdot|_1}(x_n)) \\ &= (p_1, \dots, p_n),\end{aligned}$$

where $p_k = \operatorname{sgn}(x_k)\max\{|x_k| - \lambda\gamma, 0\}$ for $k = 1, 2, \dots, n$. Set $A = \nabla h$ and $B = \partial g$, then we immediately get the following result by using Theorem 3.4.

Corollary 4.4: *Let the mappings A and B be defined above. Suppose that $\Omega \neq \emptyset$, and Conditions* (C3)–(C4) *hold. Let $x_0, x_1 \in \mathscr{H}$ and $\{x_n\}$ be a sequence produced by*

$$\begin{cases} w_n = x_n + \theta_n(x_n - x_{n-1}), \\ y_n = \operatorname{prox}_{\lambda_n g}(I - \lambda_n \nabla h) w_n, \\ z_n = y_n - \lambda_n(\nabla h(y_n) - \nabla h(x_n)), \\ x_{n+1} = \alpha_n f(x_n) + (1-\alpha_n) z_n, \end{cases}$$

where $\{\theta_n\}$ and $\{\lambda_n\}$ are defined in (5) *and* (38), *respectively. Then the iterative sequence $\{x_n\}$ presented above converges to $p \in \Omega$ in norm, where $p = P_\Omega \circ f(p)$.*

5. Numerical experiments

In this section, we provide some numerical examples occurring in finite- and infinite-dimensional spaces to show the advantages of our algorithms and compare them with some known strongly convergent algorithms, including Gibali and Thong's Algorithm 1 (GT Alg. 1) and Algorithm 2 (GT Alg. 2) [23], Thong and Cholamjiak's Algorithm 3.1 (TC Alg. 3.1) [24], and Gibali et al.'s Algorithm 3 (GTV Alg. 3) [25]. All the programs were implemented in MATLAB 2018a on a Intel(R) Core(TM) i5-8250U CPU @ 1.60 GHz computer with RAM 8.00 GB.

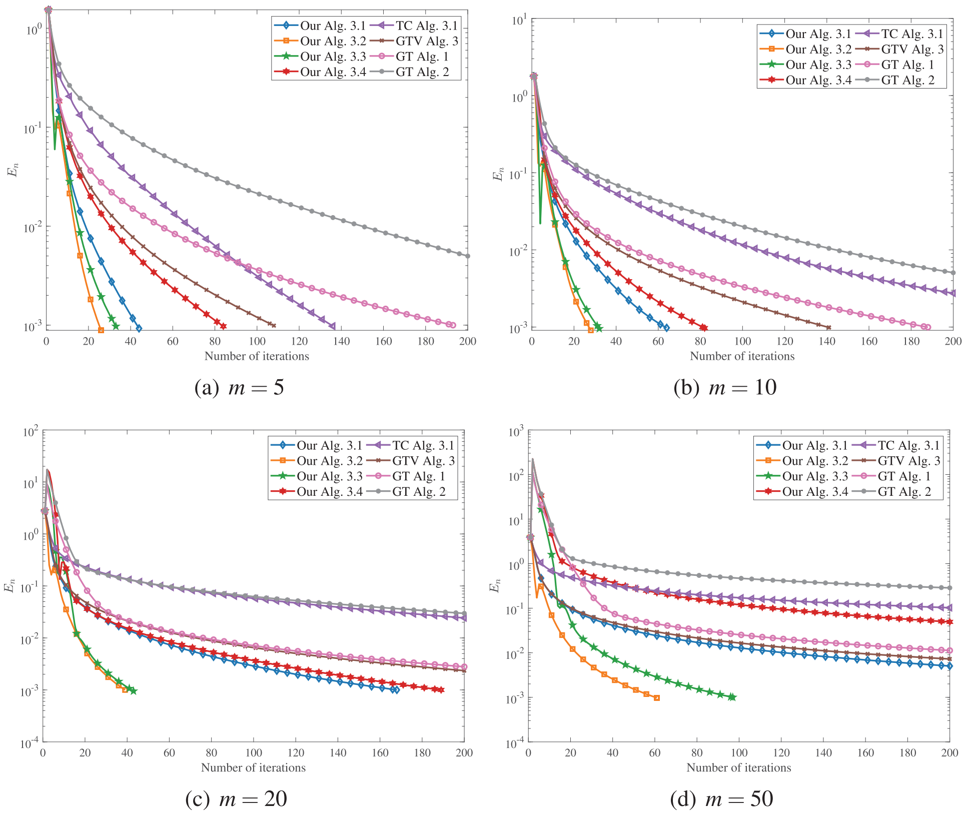
In the following numerical experiments, the parameters of all algorithms are set as follows:

- In our Algorithms 3.1–3.2 and TC Alg. 3.1, set $\delta = 2$, $l = 0.5$, $\mu = 0.5$, $\gamma = 1$, $\alpha_n = 1/(n+1)$, $\beta_n = 0.5(1-\alpha_n)$ and $f(x) = 0.5x$.
- In our Algorithms 3.3–3.4, GT Alg. 1 and GT Alg. 2, choose $\lambda_0 = 1$, $\mu = 0.5$, $\alpha_n = 1/(n+1)$, $\beta_n = 0.5(1-\alpha_n)$ and $f(x) = 0.5x$.
- In GTV Alg. 3, update the inertia parameter α_n through (5), adopt stepsize $\lambda_n = 0.2/L$, $\gamma = 1$, $\theta_n = 1/(n+1)$ and $f(x) = 0.5$. In GTV Alg. 3 and our Algorithms 3.1–3.4, take $\epsilon_n = 100/(n+1)^2$ and $\theta = 0.5$.

Example 5.1: In the first example, we study the proposed algorithms to solve the variational inequality problem (VIP). Consider the form of linear operator $A : \mathbb{R}^m \to \mathbb{R}^m$ ($m = 5, 10, 20, 50$) as follows:

Table 1. The number of termination iterations and execution time of all algorithms under different dimensions (Example 5.1).

	$m=5$		$m=10$		$m=20$		$m=50$	
Algorithms	Iter.	Time (s)	Iter.	Time (s)	Iter.	Time (s)	Iter.	Time (s)
Our Alg. 3.1	43	0.0107	63	0.0117	167	0.0176	199	0.0255
Our Alg. 3.2	25	0.0097	27	0.0100	38	0.0128	60	0.0146
Our Alg. 3.3	32	0.0057	31	0.0051	42	0.0078	96	0.0059
Our Alg. 3.4	83	0.0057	81	0.0056	188	0.0073	199	0.0074
TC Alg. 3.1	135	0.0174	199	0.0221	199	0.027	199	0.0307
GTV Alg. 3	107	0.0068	140	0.0089	199	0.0119	199	0.0093
GT Alg. 1	192	0.0077	187	0.0088	199	0.0069	199	0.0082
GT Alg. 2	199	0.0062	199	0.0067	199	0.0067	199	0.0068



(a) $m=5$ (b) $m=10$

(c) $m=20$ (d) $m=50$

Figure 1. Numerical behavior of all algorithms under different dimensions (Example 5.1).

$A(x) = Gx + g$, where $g \in \mathbb{R}^m$ and $G = DD^{\mathsf{T}} + S + E$, matrix $D \in \mathbb{R}^{m\times m}$, matrix $S \in \mathbb{R}^{m\times m}$ is skew-symmetric, and matrix $E \in \mathbb{R}^{m\times m}$ is diagonal matrix whose diagonal terms are non-negative (hence G is positive symmetric definite). We choose the feasible set C is a box constraint with the form $C = [-2, 5]^m$. It is easy to see that A is Lipschitz continuous monotone and its Lipschitz constant $L = \|G\|$. In this numerical example, all entries of D, E are generated randomly in $[0, 2]$ and S is generated randomly in $[-2, 2]$. Let $q = \mathbf{0}$, then the solution set is $x^* = \{\mathbf{0}\}$. Note that $1/L$ is much smaller than 0.01 in this example, so we adjust $\lambda_0 = 1$ to $\lambda_0 = 0.01$. We use $E_n = \|x_n - x^*\|$ to measure the n-th iteration error of all algorithms. The stopping condition is $E_n < 10^{-3}$ or the maximum number of iterations is 199 times. The numerical results of all algorithms in different dimensions are shown in Table 1 and Figure 1.

Remark 5.1: From the numerical results of Example 5.1, we have the following observations.

(1) The four iterative schemes proposed in this paper are efficient and easy to implement. The most important thing is that they converge quickly.
(2) Our methods converge faster than some known algorithms in the literature in terms of the number of iterations and execution time, and these observations have no significant relationship with the dimensions of the problem and the selection of initial values (cf. Table 1 and Figure 1).
(3) Note that the execution time of self-adaptive methods (our Algorithms. 3.3–3.4, GT Alg. 1 and GT Alg. 2) is less than that of Armijo-type methods (our Algorithms 3.1–3.2 and TC Alg. 3.1) on average. The reason for this result is that Armijo-type methods take extra time to find the appropriate step size, while adaptive-type methods only need some previous iteration information to calculate the next iteration step size. In addition, it should be pointed out that GTV Alg. 3 has great limitations because it needs to know the prior information of the Lipschitz constant of the variational inequality mapping before it can work.
(4) It should be noted that the numerical behavior of the algorithms (our Algorithms 3.3–3.4, GT Alg. 1 and GT Alg. 2) in Figure 1(c) and 1(d) has a sudden change when the dimensions of the problem are equal to 20 and 50. Since $1/L$ is much smaller than $\lambda_0 = 0.01$ in these cases, the step size of the algorithms mentioned above is much larger than $1/L$ in the previous dozens of iterations. It is known that a small step size means slow convergence and a large step size will cause the algorithms to oscillate.

Example 5.2: In the second example, we explore the proposed methods to solve the split feasibility problem (SFP) in infinite-dimensional Hilbert spaces. For any $x, y \in L^2([0,1])$, we consider $\mathscr{H}_1 = \mathscr{H}_2 = L^2([0,1])$ embedded with the inner product $\langle x, y\rangle := \int_0^1 x(t)y(t)\,\mathrm{d}t$ and the induced norm $\|x\| := (\int_0^1 |x(t)|^2\,\mathrm{d}t)^{\frac{1}{2}}$. Consider the following nonempty closed and convex subsets C and Q in $L^2([0,1])$:

$$C = \left\{ x \in L_2([0,1]) \,\middle|\, \int_0^1 x(t)\,\mathrm{d}t \le 1 \right\},$$

$$Q = \left\{ x \in L_2([0,1]) \,\middle|\, \int_0^1 |x(t) - \sin(t)|^2\,\mathrm{d}t \le 16 \right\}.$$

Let $T : L^2([0,1]) \to L^2([0,1])$ be the Volterra integration operator, which is given by $(Tx)(t) = \int_0^t x(s)\,\mathrm{d}s,\ \forall t \in [0,1], x \in \mathscr{H}$. Then T is a bounded linear operator (see [34, Exercise 20.16]) and its operator norm is $\|T\| = \frac{2}{\pi}$. Moreover, the adjoint T^* of T is defined by $(T^*x)(t) = \int_t^1 x(s)\,\mathrm{d}s$. Note that $x(t) = 0$ is a solution of (SFP) and thus the solution set of the problem is nonempty. On the other hand, it is known that projections on sets C and Q have display formulas, that is,

$$P_C(x) = \begin{cases} 1 - a + x, & a > 1; \\ x, & a \le 1 \end{cases} \quad \text{and} \quad P_Q(x) = \begin{cases} \sin(\cdot) + \dfrac{4(x - \sin(\cdot))}{\sqrt{b}}, & b > 16; \\ x, & b \le 16, \end{cases}$$

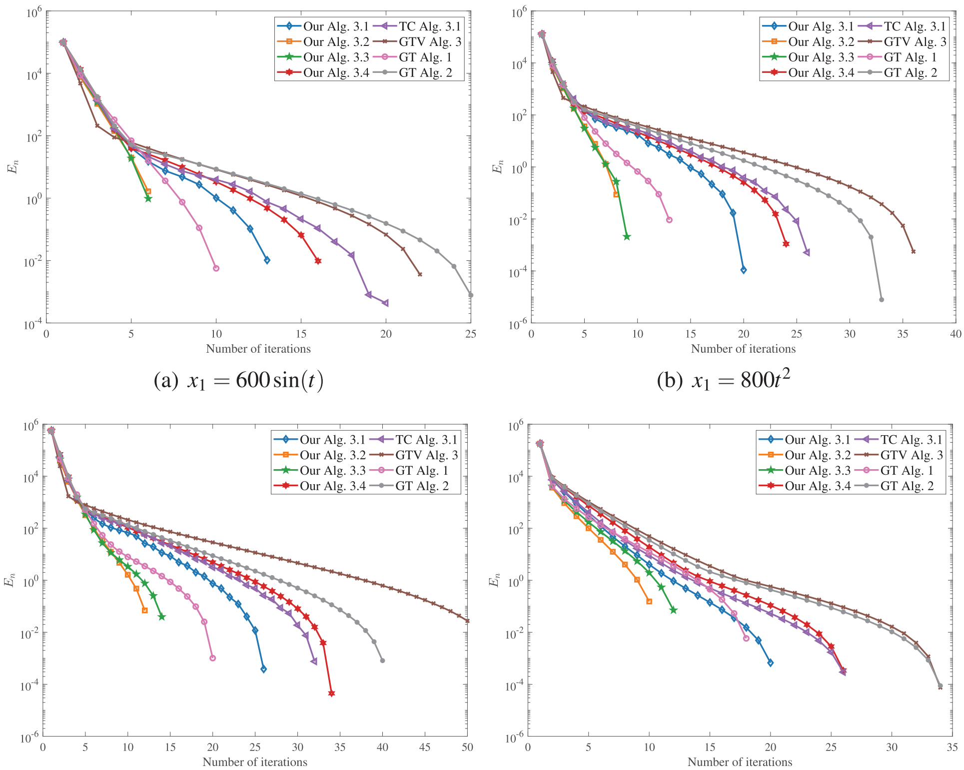
where $a := \int_0^1 x(t)\,\mathrm{d}t$ and $b := \int_0^1 |x(t) - \sin(t)|^2\,\mathrm{d}t$.

We use symbolic computation in MATLAB to implement these algorithms for generating the sequences of iterates and use $E_n = \|(I - P_C)x_n\|^2 + \|T^*(I - P_Q)Tx_n\|^2 < 10^{-5}$ or the maximum iteration 49 times as the stopping criterion. Table 2 and Figure 2 show the numerical behavior of all algorithms under four different initial values $x_0 = x_1$.

Remark 5.2: It can be seen from Table 2 and Figure 2 that the proposed approaches are easy to implement and efficient. In addition, our suggested methods (especially the Mann-type Algorithms 3.2

Table 2. The number of termination iterations and execution time of all algorithms under different initial values (Example 5.2).

	$x_1 = 600\sin(t)$		$x_1 = 800t^2$		$x_1 = 500(t^3+2t)$		$x_1 = 300\log(t)$	
Algorithms	Iter.	Time (s)	Iter.	Time (s)	Iter.	Time (s)	Iter.	Time (s)
Our Alg. 3.1	13	33.717	20	26.997	26	74.9293	20	124.447
Our Alg. 3.2	6	11.503	8	7.7113	12	27.9015	10	51.1319
Our Alg. 3.3	6	12.545	9	8.2073	14	32.1083	12	46.0236
Our Alg. 3.4	16	43.161	24	28.891	34	111.492	26	197.977
TC Alg. 3.1	20	27.571	26	27.144	32	40.9894	26	118.734
GTV Alg. 3	22	45.361	36	51.580	49	120.534	34	284.084
GT Alg. 1	10	9.8095	13	8.7694	20	19.8141	18	49.3417
GT Alg. 2	25	33.668	32	24.672	40	57.3985	34	151.074



(a) $x_1 = 600\sin(t)$ (b) $x_1 = 800t^2$

(c) $x_1 = 500(t^3+2t)$ (d) $x_1 = 300\log(t)$

Figure 2. Numerical behavior of all algorithms under different initial values (Example 5.2).

and 3.3) require fewer iterations than some algorithms in the literature to achieve the same error accuracy, and these results are independent of the selection of initial values. It is worth noting that our Algorithms 3.1 and 3.4 enjoy fewer iterations while accompanied by more execution time. Moreover, it should be pointed out that the operator norm $\|T\|$ of this problem is not easy to obtain, which means that the fixed step size algorithm (GTV Alg. 3) will fail. However, the self-adaptive algorithms presented in this paper can work well.

Example 5.3: Compressed sensing is an effective method to recover a clean signal from a polluted signal. This requires us to solve the following underdetermined system problems:

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \epsilon,$$

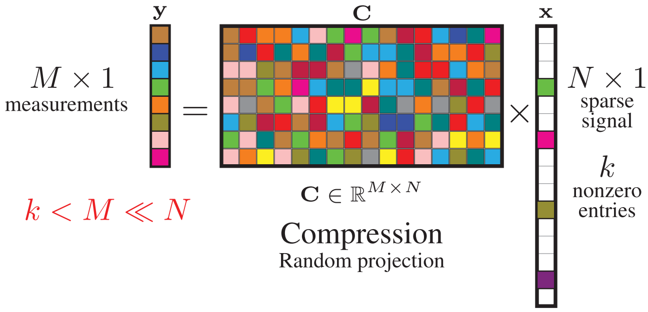
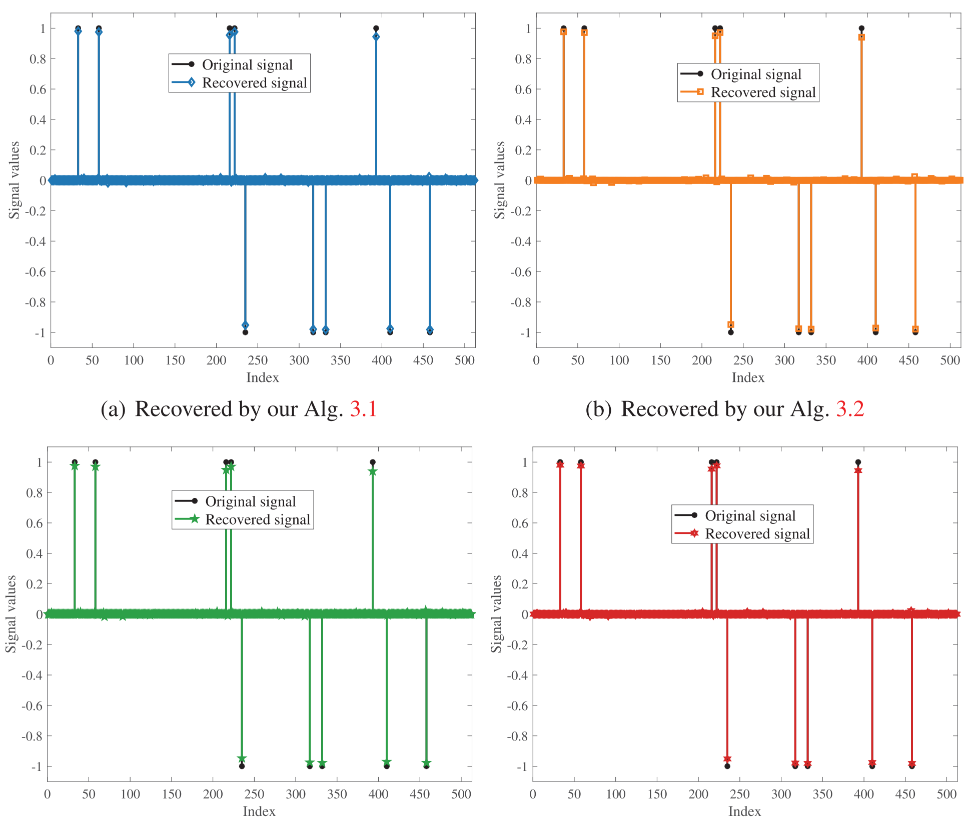


Figure 3. Structure of compressive sensing matrices.

where $\mathbf{y} \in \mathbb{R}^M$ is the observed noise data, $\mathbf{C} : \mathbb{R}^{M \times N}$ is a bounded linear observation operator, $\mathbf{x} \in \mathbb{R}^N$ with k ($k \ll N$) non-zero elements is the original and clean data that needs to be restored, and ε is the noise observation encountered during data transmission. An important consideration of this problem is that the signal $\mathbf{x}$ is sparse, that is, the number of non-zero elements in the signal $\mathbf{x}$ is much smaller than the dimension of the signal $\mathbf{x}$. Figure 3 visually shows the matrix structure expression of compressed sensing.

A successful model used to solve the above problem can be translated into the following convex



(a) Recovered by our Alg. 3.1

(b) Recovered by our Alg. 3.2

(c) Recovered by our Alg. 3.3

(d) Recovered by our Alg. 3.4

Figure 4. The original signal and the signal recovered by our algorithms.

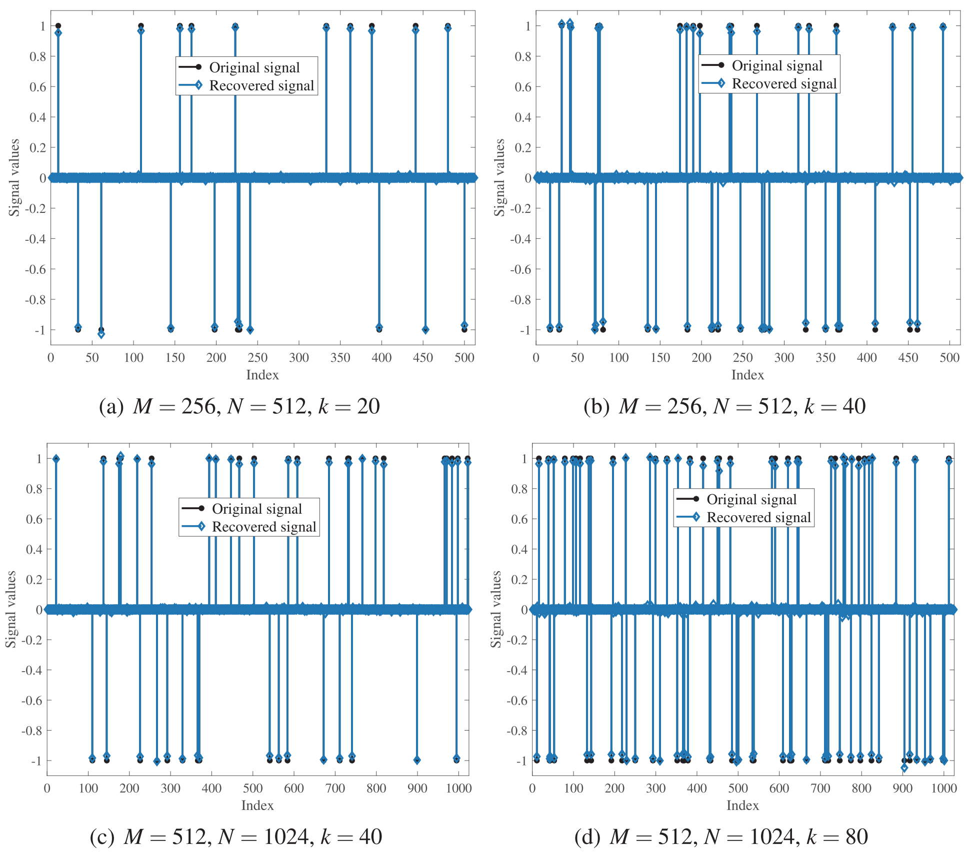
constraint minimization problem:

$$\min_{\mathbf{x}\in\mathbb{R}^N} \frac{1}{2}\|\mathbf{y}-\mathbf{C}\mathbf{x}\|^2 \text{ subject to } \|\mathbf{x}\|_1 \le t, \tag{LASSO}$$

where t is a positive constant. It should be pointed out that this problem is related to the least absolute shrinkage and selection operator (LASSO) problem. Note that the (LASSO) problem described above can be regarded as a special case of (SFP) when $C = \{\mathbf{x} \in \mathbb{R}^N : \|\mathbf{x}\|_1 \le t\}$ and $Q = \{\mathbf{y}\}$. In this situation, we can use the projection formulas described in Section 2 to calculate P_C and P_Q.

Table 3. The numerical results of all algorithms for solving (LASSO) under different situations (Example 5.3).

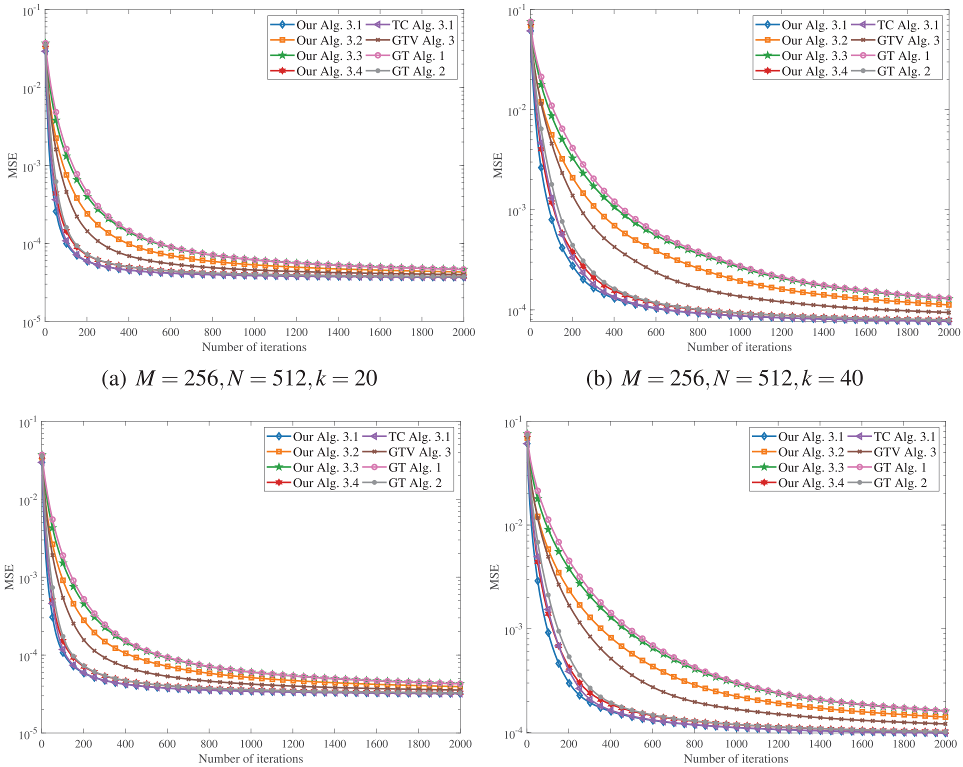
	$M = 256, N = 512, k = 20$		$M = 256, N = 512, k = 40$		$M = 512, N = 1024$ k = 40		$M = 512, N = 1024$ k = 80	
Algorithms	MSE ($\times 10^{-4}$)	Time (s)	MSE ($\times 10^{-3}$)	Time (s)	MSE ($\times 10^{-4}$)	Time (s)	MSE ($\times 10^{-3}$)	Time (s)
Our Alg. 3.1	0.3675	0.4426	0.0769	0.4657	0.3205	1.5904	0.0989	1.4212
Our Alg. 3.2	0.4294	0.4573	0.1120	0.4543	0.3934	1.5970	0.1408	1.4111
Our Alg. 3.3	0.4627	0.1862	0.1290	0.1978	0.4280	0.7325	0.1603	0.6410
Our Alg. 3.4	0.3742	0.1875	0.0793	0.1823	0.3274	0.7264	0.1020	0.6264
TC Alg. 3.1	0.3675	0.5043	0.0769	0.4781	0.3205	1.7834	0.0990	1.5341
GTV Alg. 3	0.3975	0.1927	0.0941	0.1944	0.3558	0.7795	0.1211	0.6446
GT Alg. 1	0.4632	0.1897	0.1303	0.1909	0.4285	0.7627	0.1612	0.6366
GT Alg. 2	0.3742	0.1819	0.0794	0.1818	0.3275	0.7700	0.1021	0.6166



(a) $M = 256, N = 512, k = 20$ (b) $M = 256, N = 512, k = 40$

(c) $M = 512, N = 1024, k = 40$ (d) $M = 512, N = 1024, k = 80$

Figure 5. The original signal and the signal recovered by our Algorithm 3.1.



(a) $M = 256, N = 512, k = 20$ (b) $M = 256, N = 512, k = 40$

(c) $M = 512, N = 1024, k = 40$ (d) $M = 512, N = 1024, k = 80$

Figure 6. The discrepancy of mean squared error (MSE) of all algorithms.

We now consider using our proposed iterative schemes to solve (LASSO) and compare them with some known algorithms in the literature. In our numerical experiments, the matrix $\mathbf{C} : \mathbb{R}^{M\times N}$ is created from a standard normal distribution with zero mean and unit variance and then orthonormalizing the rows. The clean signal $\mathbf{x} \in \mathbb{R}^N$ contains k ($k \ll N$) randomly generated ± 1 spikes. The observation $\mathbf{y}$ is formed by $\mathbf{y} = \mathbf{Cx} + \epsilon$ with white Gaussian noise ϵ of variance 10^{-4}. The recovery process starts with the initial signals $\mathbf{x}_0 = \mathbf{x}_1 = \mathbf{0}$ and ends after 2000 iterations. We use the mean squared error $\text{MSE} = (1/N)\|\mathbf{x}^* - \mathbf{x}\|^2$ ($\mathbf{x}^*$ is an estimated signal of $\mathbf{x}$) to measure the restoration accuracy of all algorithms. In our first test, we set $M = 256, N = 512$ and $k = 10$. The recovery results of our suggested algorithms are shown in Figure 4.

Next, in order to demonstrate the robustness of the proposed algorithms, we conduct signal recovery tests with different dimensions and different sparsity. The numerical results are reported in Table 3, Figures 5 and 6.

Remark 5.3: As can be seen from the numerical results of Example 5.3, the proposed algorithms can be applied to signal processing problems in compressed sensing, and they can work well (see Figures 4 and 5). Under the same number of iterations, the presented algorithms have smaller mean squared error and cpu time than the compared algorithms (cf. Table 3), which implies that our proposed algorithms perform better and converge faster in the signal recovery tests (cf. Figure 6). Moreover, it is worth noting that the Armijo-type methods (our Algorithms 3.1–3.2 and TC Alg. 3.1) also take more time than the others. Furthermore, as shown in the previous two examples, our algorithms are still robust in this example, because the dimension and sparsity of the signal have no significant influence on our results.

6. The conclusion

In this paper, we introduced four modified inertial forward–backward splitting methods for finding a zero of the sum of two monotone operators. The proposed algorithms are constructed based on several existing methods, including the inertial method, the forward–backward splitting method, the Tseng method, the projection and contraction method, the Mann-type method and the viscosity-type method. Strong convergence theorems of the suggested algorithms are built without knowing the prior information of the Lipschitz constant of the single-valued operator. As applications, our methods are applied to convex minimization problems, variational inequality problems, split feasibility problems and image processing problems. Finally, we give some numerical experiments to illustrate the computational efficiency of our proposed algorithms compared with other ones. The approaches obtained in this paper improved and summarized some relevant results in the literature.

Acknowledgements

The authors are very grateful to the anonymous referees for their constructive comments, which significantly improved the original manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Bing Tan http://orcid.org/0000-0003-1509-1809

References

[1] Combettes PL, Wajs VR. Signal recovery by proximal forward–backward splitting. Multiscale Model Simul. 2005;4:1168–1200.
[2] Sahu DR, Yao JC, Verma M, et al. Convergence rate analysis of proximal gradient methods with applications to composite minimization problems. Optimization. 2021;70:75–100.
[3] Qin X, An NT. Smoothing algorithms for computing the projection onto a Minkowski sum of convex sets. Comput Optim Appl. 2019;74:821–850.
[4] Cuong TH, Yao JC, Yen ND. Qualitative properties of the minimum sum-of-squares clustering problem. Optimization. 2020;69:2131–2154.
[5] Wang Y, Zhang H. Strong convergence of the viscosity Douglas-Rachford algorithm for inclusion problems. Appl Set-Valued Anal Optim. 2020;2:339–349.
[6] Qin X, Yao JC. A viscosity iterative method for a split feasibility problem. J Nonlinear Convex Anal. 2019;20:1497–1506.
[7] Tan B, Xu S, Li S. Inertial shrinking projection algorithms for solving hierarchical variational inequality problems. J Nonlinear Convex Anal. 2020;21:871–884.
[8] Ansari QH, Islam M, Yao JC. Nonsmooth variational inequalities on Hadamard manifolds. Appl Anal. 2020;99:340–358.
[9] Lions PL, Mercier B. Splitting algorithms for the sum of two nonlinear operators. SIAM J Numer Anal. 1979;16:964–979.
[10] Passty GB. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. J Math Anal Appl. 1979;72:383–390.
[11] Attouch H, Peypouquet J. The rate of convergence of Nesterov's accelerated forward–backward method is actually faster than $1/k^2$. SIAM J Optim. 2016;26:1824–1834.
[12] Boț RI, Csetnek ER. Convergence rates for forward–backward dynamical systems associated with strongly monotone inclusions. J Math Anal Appl. 2018;457:1135–1152.
[13] Malitsky Y, Tam MK. A forward–backward splitting method for monotone inclusions without cocoercivity. SIAM J Optim. 2020;30:1451–1472.
[14] Moudafi A, Shehu Y. Convergence of the forward–backward method for split null-point problems beyond coerciveness. J Nonlinear Convex Anal. 2019;20:1659–1672.
[15] Shehu Y, Yao JC. Rate of convergence for inertial iterative method for countable family of certain quasi-nonexpansive mappings. J Nonlinear Convex Anal. 2020;21:533–541.

[16] Tan B, Xu S. Strong convergence of two inertial projection algorithms in Hilbert spaces. J Appl Numer Optim. 2020;2:171–186.
[17] Tseng P. A modified forward–backward splitting method for maximal monotone mappings. SIAM J Control Optim. 2000;38:431–446.
[18] Zhang C, Wang Y. Proximal algorithm for solving monotone variational inclusion. Optimization. 2018;67:1197–1209.
[19] Wang Y, Wang F. Strong convergence of the forward–backward splitting method with multiple parameters in Hilbert spaces. Optimization. 2018;67:493–505.
[20] Dong QL, Jiang D, Cholamjiak P, et al. A strong convergence result involving an inertial forward–backward algorithm for monotone inclusions. J Fixed Point Theory Appl. 2017;19:3097–3118.
[21] Cholamjiak W, Cholamjiak P, Suantai S. An inertial forward–backward splitting method for solving inclusion problems in Hilbert spaces. J Fixed Point Theory Appl. 2018;20:42.
[22] Shehu Y, Cai G. Strong convergence result of forward–backward splitting methods for accretive operators in Banach spaces with applications. Rev R Acad Cienc Exactas Fís Nat Ser A Mat. 2018;112:71–87.
[23] Gibali A, Thong DV. Tseng type methods for solving inclusion problems and its applications. Calcolo. 2018;55:49.
[24] Thong DV, Cholamjiak P. Strong convergence of a forward–backward splitting method with a new step size for solving monotone inclusions. Comput Appl Math. 2019;38:94.
[25] Gibali A, Thong DV, Vinh NT. Three new iterative methods for solving inclusion problems and related problems. Comput Appl Math. 2020;39:187.
[26] Lorenz DA, Pock T. An inertial forward–backward algorithm for monotone inclusions. J Math Imaging Vision. 2015;51:311–325.
[27] Qin X, Wang L, Yao JC. Inertial splitting method for maximal monotone mappings. J Nonlinear Convex Anal. 2020;21:2325–2333.
[28] Gibali A, Hieu DV. A new inertial double-projection method for solving variational inequalities. J Fixed Point Theory Appl. 2019;21:97.
[29] Cholamjiak P, Shehu Y. Inertial forward–backward splitting method in Banach spaces with application to compressed sensing. Appl Math. 2019;64:409–435.
[30] Shehu Y, Dong QL, Liu LL, et al. New strong convergence method for the sum of two maximal monotone operators. Optim Eng. 2020. doi:10.1007/s11081-020-09544-5.
[31] Tan B, Li S. Strong convergence of inertial Mann algorithms for solving hierarchical fixed point problems. J Nonlinear Var Anal. 2020;4:337–355.
[32] Brézis H, Chapitre II. Operateurs maximaux monotones; 1973. North-Holland Math Stud. 1973;5:19–51.
[33] Saejung S, Yotkaew P. Approximation of zeros of inverse strongly monotone operators in Banach spaces. Nonlinear Anal. 2012;75:742–750.
[34] Bauschke HH, Combettes PL. Convex analysis and monotone operator theory in Hilbert spaces. 2nd ed. New York: Springer; 2017.