CARPATHIAN J. MATH. Online version at https://www.carpathian.cunbm.utcluj.ro
36 (2020), No. 1, 159 - 177 Print Edition: ISSN 1584 - 2851 Online Edition: ISSN 1843 - 4401

Dedicated to Prof. Hong-Kun Xu on the occasion of his 60^{th} anniversary

Multi-step inertial proximal contraction algorithms for monotone variational inclusion problems

CUIJIE ZHANG, QIAO-LI DONG and JIAJIA CHEN

ABSTRACT. In this article, we introduce the multi-step inertial proximal contraction algorithms (MiPCA) to approximate a zero of the sum of two monotone operators, with one of the two operators being monotone and Lipschitz continuous. The weak convergence of the MiPCA is shown under the summability condition formulated in terms of the iterative sequence in a Hilbert space setting. We also investigate the unconditional convergence of the one-step inertial proximal contraction algorithm. Finally, numerical experiments are given to illustrate the advantage of the multi-step inertial proximal contraction algorithms.

1. INTRODUCTION

Let H be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. We focus on the following monotone variational inclusion:

$$\text{Find } x^* \in H \quad \text{such that} \quad 0 \in A(x^*) + f(x^*), \tag{1.1}$$

where $A : H \to 2^H$ is a maximally monotone operator and $f : H \to H$ is a single-valued monotone operator. When $f \equiv 0$, the problem (1.1) reduces to the inclusion problem in [27], which plays an important role in minimization problems and other fields of mathematics. In recent years, the problem (1.1) attracts the increasing attention due to its applicability in machine learning and like fields, where data sets of unprecedented size are processed, often in real time.

A great deal of iterative algorithms have been introduced and developed for solving this problem (1.1), most of which are splitting algorithms by involving the operators individually (see, for example, [5, 12, 14, 16, 17, 30, 33, 34, 35, 36]). One of the most popular splitting methods is the well-known forward-backward algorithm (see, e.g., [20]).

It is easy to verify that $x^* \in H$ is a solution of the problem (1.1) if and only if x^* is a fixed point of the operator $J_\lambda^A(I - \lambda f)$, i.e., $x^* = J_\lambda^A(I - \lambda f)x^*$, where J_λ^A is the resolvent operator of A defined by

$$J_\lambda^A = (I + \lambda A)^{-1}, \quad \lambda > 0.$$

By using this relation, the forward-backward splitting algorithm

$$x_{k+1} = J_{\lambda_k}^A (I - \lambda_k f)x_k, \quad \lambda_k > 0, \quad k \geq 0, \tag{1.2}$$

was naturally introduced (see, [2]).

Received: 09.11.2019. In revised form: 31.01.2020. Accepted: 07.02.2020
2010 *Mathematics Subject Classification.* 90C47, 49J35.
Key words and phrases. *monotone variational inclusion problem, resolvent operator, multi-step inertial proximal contraction algorithm, one-step inertial proximal contraction algorithm.*
Corresponding author: Qiao-Li Dong; dongql@lsec.cc.ac.cn

However, the forward-backward splitting algorithm (1.2) may take a lot of iterations: indeed, when f is the gradient of a convex and differential function, the forward-backward splitting algorithm (1.2) becomes the proximal gradient algorithm for convex optimization problems, which can be slow in practice [24, Section 5].

To speed up the convergence of iterative algorithms, the inertial extrapolation technique was introduced in 1964, which originates from the heavy ball method [26]. The main features of inertial type algorithms is that the next iterate is defined by making use of the previous two iterates. Nesterov [23] in 1983 introduced and developed an inertial type method for minimizing a smooth convex function, which was proven to be an "optimal" first order (gradient) method in the sense of complexity analysis.

By combining the inertial extrapolation technique and the forward-backward algorithm (1.2), the Moudafi and Oliny [22] introduced the following inertial proximal algorithm:

$$\begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}), \\ x_{k+1} = J^A_{\lambda_k}(y_k - \lambda_k f(y_k)), \quad k \geq 0, \end{cases} \tag{1.3}$$

where f was cocoercive and α_k is the inertial parameter. The convergence of the iterative sequence $\{x_k\}$ was established under the summability condition formulated in terms of $\{x_k\}$ and α_k. An open problem proposed in [22] is "*to investigate, theoretically as well as numerically, which are the best choices for the inertial parameter α_k in order to accelerate the convergence*".

Since the open problem was proposed, there has been little progress. Until 2009, Beck and Teboulle [3] introduced the well-known fast iterative shrinkage-thresholding algorithm (FISTA) to solve the linear inverse problems, which is an inertial version of the iterative shrinkage-thresholding algorithm (ISTA). The inertial parameters $\{\alpha_k\}$ in FISTA is chosen as follows:

$$\alpha_k = \frac{t_k - 1}{t_{k+1}},$$

where $t_1 = 1$, and

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \quad k \geq 1.$$

They proved that FISTA has global rate $O(1/k^2)$ of convergence, while the global rate of convergence of ISTA is $O(1/k)$. Since its inception, FISTA has received a great deal of attention due to its excellent computational effect and important applications (see [1] and the references therein). Many authors improved it in various ways (see, e.g., [1, 6, 18, 24]). Note that the linear inverse problems studied in [3] can be seen as a special case of the monotone variational inclusion (1.1), for which f is the gradient of a convex continuously differentiable function. So, the open problem of Moudafi and Oliny has not been fully resolved until now.

Lorenz and Pock [21] introduced the following inertial forward-backward splitting method

$$\begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}), \\ x_{k+1} = J^A_{\lambda_k M^{-1}}(y_k - \lambda_k M^{-1} f(y_k)), \end{cases} \tag{1.4}$$

where M is a linear selfadjoint and positive definite map and f is cocoercive. Two kinds of the conditions on inertial parameters $\{\alpha_k\}$ were considered, one of which involves the iterative sequence $\{x_k\}$, and the other can be chosen a-priori. In subsequent work [9], by combining Nesterov's acceleration scheme and Haugazeau's algorithm, an inertial forward-backward method was proposed, the sequence generated by which has the strong convergence.

Polyak [25] suggested to accelerate the convergence of the iterative algorithms by using more previous iterates. Following his idea, Liang [19] proposed the variable metric multi-step inertial operator splitting method (MUSTARD). Let $S = \{0, \cdots, s-1\}$, $s \in \mathbb{N}_+$. MUSTARD is defined as follows:

$$\begin{cases} y_{a,k} = x_k + \sum_{i\in S} a_{i,k}(x_{k-i} - x_{k-i-1}), \\ y_{b,k} = x_k + \sum_{i\in S} b_{i,k}(x_{k-i} - x_{k-i-1}), \\ x_{k+1} = J_{\gamma_k M_k^{-1} A}(y_{a,k} - \gamma_k M_k^{-1} f(y_{b,k})), \end{cases} \tag{1.5}$$

where $\{M_k\}$ is a sequence of symmetric positive definite operators and f is cocoercive. The weak convergence of MUSTARD was established under the summability conditions of the inertial parameters $\{a_{i,k}\}$, $\{b_{i,k}\}$ and the iterative sequence $\{x_k\}$. Furthermore, numerical examples illustrated that MUSTARD behaves better than FISTA. Very recently, motivated by Liang's work, the authors [8] introduced the multi–step inertial Krasnosel'skiĭ–Mann algorithm (MiKM), which was extended to the monotone variational inclusion (1.1) with the cocoercive f.

As far as we know, the inertial algorithms and their convergence for the problem (1.1) when A is maximal monotone and f is cocoercive has been studied extensively in the literature. However, there are still few results on the inertial algorithms concerning more general case of the problem (1.1) when A is maximal monotone and f is monotone and Lipschitz continuous. This is the gap that this paper aims to achieve.

Our aim in this paper is to propose a multi-step inertial proximal contraction algorithm for solving the monotone variational inclusion (1.1) by combining the proximal contraction method in [37] and the inertial technique. We also introduce one-step inertial proximal contraction algorithm and two convergent results are given. The application of the multi-step inertial proximal contraction algorithm is presented in the variational inequality problems. In summary,

- We introduce the multi-step inertial proximal contraction algorithm for the monotone variational inclusion (1.1) and its weak convergence is established under the summability condition formulated in terms of the iterative sequence and inertial parameters when A is maximal monotone and f is monotone and Lipschitz continuous.
- We present the conditional convergence and unconditional convergence for the one-step inertial proximal contraction algorithm. The inertial parameters in the former involve the iterative sequence, while the inertial parameters in the latter can be chosen a-priori.
- We extend the multi-step inertial proximal contraction algorithm to the variational inequality problems and first propose the multi-step inertial projection contraction algorithm for the variational inequality problems
- Two numerical examples are given to confirm the importance of the presence of multi-step inertial terms in our method, especially for two-step inertial terms.

The paper is organized as follows: We first recall some definitions and preliminary results which will be used in main results in Section 2. The multi-step inertial proximal contraction algorithms is introduced and the analysis of the convergence is investigated in Section 3. Two convergence results for the one-step inertial proximal contraction algorithm are presented. An application of the proposed algorithms is presented in the variational inequality problems in Section 4. We give numerical implementations in Section 5.

2. PRELIMINARIES

Throughout this paper, we use $x_k \rightharpoonup x$ and $x_k \to x$ to indicate that $\{x_k\}$ converges weakly to x and converges strongly to x, respectively. Assume that T is an operator, we denote the fixed point set of T by $Fix(T)$.

In order to prove our results, we collect some necessary conceptions and lemmas in this section.

Definition 2.1. A mapping $T : H \to H$ is said to be
• L-Lipshcitz continuous iff for all $x, y \in H$,

$$\|Tx - Ty\| \le L\|x - y\|, \tag{2.6}$$

where $L > 0$ is Lipschitz constant;
• nonexpansive iff $L \equiv 1$ in (2.6);
• firmly nonexpansive iff $2T - I$ is nonexpansive, or equivalently for all $x, y \in H$,

$$\langle Tx - Ty, x - y\rangle \ge \|Tx - Ty\|^2. \tag{2.7}$$

If T is firmly nonexpansive, then $I - T$ is firmly nonexpansive. If $Fix(T) \neq \emptyset$, we have that

$$\langle x - Tx, Tx - y\rangle \ge 0 \text{ for all } x \in H,\ y \in Fix(T). \tag{2.8}$$

Definition 2.2. A single-valued mapping $T : H \to H$ is said to be
• monotone iff for all $x, y \in H$,

$$\langle Tx - Ty, x - y\rangle \ge 0. \tag{2.9}$$

A multi-valued mapping $A : H \to 2^H$ is called
•monotone iff

$$\langle u - v, x - y\rangle \ge 0 \text{ whenever } u \in A(x),\ v \in A(y), \tag{2.10}$$

• maximal monotone iff, in addition, its graph

$$G(A) := \{(x, y) \in H \times H : y \in A(x)\} \tag{2.11}$$

is not properly contained in the graph of any other monotone operator.

It is well-known that a monotone mapping A is maximal if and only if for $(x, y) \in H \times H$,$\langle x - v, y - w\rangle \ge 0$ for every $(v, w) \in G(A)$ implies $y \in A(x)$.

In the proof of the main theorem, we will use Pythagoras relation:

$$2\langle c_1 - c_2, c_1 - c_3\rangle = \|c_1 - c_2\|^2 + \|c_1 - c_3\|^2 - \|c_2 - c_3\|^2, \tag{2.12}$$

where $c_1, c_2, c_3 \in H$.

Lemma 2.1. ([4]) *Let $A : H \to 2^H$ be a maximal monotone mapping and let $f : H \to H$ be a Lipschitz continuous mapping. Then the mapping $B = A + f$ is a maximal monotone mapping.*

Lemma 2.2. ([2]) *In a Hilbert space H, there holds the equality*

$$\|tx + (1 - t)y\|^2 = t\|x\|^2 + (1 - t)\|y\|^2 - t(1 - t)\|x - y\|^2$$

for all $x, y \in H$ and $t \in \mathbb{R}$.

Lemma 2.3. ([2]) *Let C be a nonempty set of H, and $\{x_k\}$ be a sequence in H. If the following conditions hold:*

(i) *for every $x \in C$, $\lim_{k\to\infty} \|x_k - x\|$ exists;*

(ii) *every weak sequential cluster point of* $\{x_k\}$ *belongs to* C.

Then the sequence $\{x_k\}$ *converges weakly to a point in* C.

Lemma 2.4. ([19]) *Let* $s \in \mathbb{N}_+$ *and* $S = \{0, 1, \cdots, s-1\}$. *Let* $\{d_k\}$, $\{\delta_k\}$ *be two non-negative sequences, and* $\omega = (\omega_i)_s \in \mathbb{R}^s$ *such that*

$$d_{k+1} \leq \sum_{i \in S} \omega_i d_{k-i} + \delta_k,$$

for all $k \geq s$. *If* $\sum_{i \in S} \omega_i \in [0, 1)$ *and* $\sum_{k=1}^{\infty} \delta_k < +\infty$, *then*

$$\sum_{k=1}^{\infty} d_k < +\infty.$$

3. Algorithms and convergence analysis

In this section we introduce the multi-step inertial proximal contraction algorithm and establish its weak convergence. Two convergence results of the one-step inertial proximal contraction algorithm are also presented.

3.1. The multi-step inertial proximal contraction algorithm.

In this subsection, we give a precise statement of our multi-step inertial proximal contraction algorithm. Its convergence analysis is postponed to the next subsection. We first state the assumptions that we will assume to hold through the rest of this paper.

Condition (i) The solution set of (1.1), denoted by Γ, is nonempty.

Condition (ii) The mapping f is monotone on H.

Condition (iii) The mapping f is Lipschitz continuous on H with constant $L > 0$.

We next give a precise statement of the multi-step inertial proximal contraction algorithm (MiPCA):

Algorithm 3.1. Let $s \in \mathbb{N}_+$ and $S := \{0, \cdots, s-1\}$. Let $\{\alpha_{i,k}\}_{i \in S} \in (-1, 1)^s$. Choose $x_0 \in H$, $x_{-i-1} = x_0$, $i \in S$.

(S.1) Compute

$$\omega_k = x_k + \sum_{i \in S} \alpha_{i,k}(x_{k-i} - x_{k-i-1}),$$

and

$$y_k = J^A_{\lambda_k}(\omega_k - \lambda_k f(\omega_k)). \tag{3.13}$$

If $w_k - y_k = 0$: STOP.

(S.2) Calculate

$$d(\omega_k, y_k) = (\omega_k - y_k) - \lambda_k(f(\omega_k) - f(y_k)),$$

and

$$x_{k+1} = \omega_k - \gamma\beta_k d(\omega_k, y_k),$$

where $\gamma \in (0, 2)$,

$$\beta_k := \begin{cases} \dfrac{\phi(\omega_k, y_k)}{\|d(\omega_k, y_k)\|^2} & \|d(\omega_k, y_k)\| \neq 0; \\ c & \|d(\omega_k, y_k)\| = 0, \end{cases} \tag{3.14}$$

where $c > 1$ is an arbitrary constant and $\phi(\omega_k, y_k) := \langle \omega_k - y_k, d(\omega_k, y_k) \rangle$.

(S.3) Set $k \leftarrow k + 1$, and go to (S.1).

Remark 3.1. Based on the choice of the inertial parameters $\alpha_{i,k}$, the relations between Algorithm 3.1 with the other work are as following:

(i) Let $s = 1$ and $\alpha_{0,k} = \alpha_k$, then Algorithm 3.1 reduces to the following one-step inertial proximal contraction scheme (1-MiPCA):

$$
\begin{cases}
\omega_k = x_k + \alpha_k(x_k - x_{k-1}), \\
y_k = J^A_{\lambda_k}(\omega_k - \lambda_k f(\omega_k)), \\
d(\omega_k, y_k) = (\omega_k - y_k) - \lambda_k(f(\omega_k) - f(y_k)), \\
x_{k+1} = \omega_k - \gamma\beta_k d(\omega_k, y_k).
\end{cases}
\tag{3.15}
$$

(ii) Let $\alpha_{i,k} \equiv 0$ for $i \in S$ and $k \in \mathbb{N}$, then Algorithm 3.1 becomes the original proximal contraction algorithm (PCA) in [37] as follows:

$$
\begin{cases}
y_k = J^A_{\lambda_k}(x_k - \lambda_k f(x_k)), \\
d(x_k, y_k) = (x_k - y_k) - \lambda_k(f(x_k) - f(y_k)), \\
x_{k+1} = x_k - \gamma\beta_k d(x_k, y_k).
\end{cases}
\tag{3.16}
$$

Definition 3.3. Let $c_1, c_2 > 0$ be given constants in $(0, 1)$. λ_k is said to satisfy the `stepsize conditions` in Algorithm 3.1, if λ_k satisfies

$$\phi(\omega_k, y_k) \geq c_1\|\omega_k - y_k\|^2, \tag{3.17}$$

$$\beta_k \geq c_2, \tag{3.18}$$

and

$$\inf_{k\geq 0}\{\lambda_k\} \geq \underline{\lambda} > 0. \tag{3.19}$$

Using [29, Lemma 5.2] and [38, Lemma 3.4], we obtain the following result.

Lemma 3.5. *Assume that the sequence $\{x_k\}$ is generated by Algorithm 3.1. Then λ_k satisfies the stepsize conditions when we take $\lambda_k \in [a, b] \subset (0, 1/L)$ or λ_k is given self-adaptively, i.e., $\lambda_k = \sigma\eta^{m_k}$, $\sigma > 0$, $\eta \in (0, 1)$, where m_k is the smallest nonnegative integer such that*

$$\lambda_k\|f(x_k) - f(y_k)\| \leq \nu\|x_k - y_k\|, \tag{3.20}$$

where $\nu \in (0, 1)$ is given.

From the next lemma, it follows that stopping condition in Algorithm 3.1 is valid.

Lemma 3.6. ([37, Lemma 3.1]) *If $\omega_k = y_k$ in* (3.13)*, then $\omega_k \in \Gamma$.*

3.2. **Convergence analysis.**

Before proceeding with the main theorem we establish a key result (see Lemma 3.7 below) that is crucial for the convergence analysis.

Lemma 3.7. *Let λ_k satisfy* (3.17) *and* (3.18)*, and let $u \in \Gamma$. Then under Conditions (i), (ii) and (iii), we have*

$$\|x_{k+1} - u\|^2 \leq \|\omega_k - u\|^2 - \frac{2-\gamma}{\gamma}\|\omega_k - x_{k+1}\|^2, \tag{3.21}$$

and

$$\|\omega_k - y_k\|^2 \leq \frac{1}{c_1 c_2 \gamma^2}\|\omega_k - x_{k+1}\|^2. \tag{3.22}$$

Proof. We prove the inequality (3.21) firstly. By the definition of x_{k+1}, we have

$$\begin{aligned} \|x_{k+1}-u\|^2 &= \|\omega_k - u - \gamma\beta_k d(\omega_k, y_k)\|^2 \\ &= \|\omega_k - u\|^2 - 2\gamma\beta_k\langle \omega_k - u, d(\omega_k, y_k)\rangle + \gamma^2\beta_k^2\|d(\omega_k, y_k)\|^2. \end{aligned} \tag{3.23}$$

Since $J^A_{\lambda_k}$ is firmly-nonexpansive, it follows that

$$\begin{aligned} &\langle J^A_{\lambda_k}(I-\lambda_k f)\omega_k - J^A_{\lambda_k}(I-\lambda_k f)u, (I-\lambda_k f)\omega_k - (I-\lambda_k f)u\rangle \\ \geq &\|J^A_{\lambda_k}(I-\lambda_k f)\omega_k - J^A_{\lambda_k}(I-\lambda_k f)u\|^2 = \|y_k - u\|^2. \end{aligned} \tag{3.24}$$

Using (3.24), we have

$$\begin{aligned} &\langle y_k - u, \omega_k - y_k - \lambda_k f(\omega_k)\rangle \\ =&\langle J^A_{\lambda_k}(I-\lambda_k f)\omega_k - J^A_{\lambda_k}(I-\lambda_k f)u, (I-\lambda_k f)\omega_k - (I-\lambda_k f)u + (I-\lambda_k f)u - y_k\rangle \\ \geq&\|y_k - u\|^2 + \langle y_k - u, u - y_k\rangle + \langle y_k - u, -\lambda_k f(u)\rangle \\ =&-\langle y_k - u, \lambda_k f(u)\rangle. \end{aligned}$$

Hence,

$$\langle y_k - u, \omega_k - y_k - \lambda_k(f(\omega_k) - f(u))\rangle \geq 0. \tag{3.25}$$

From the monotonicity of f and $\lambda_k > 0$, we have

$$\langle y_k - u, \lambda_k f(y_k) - \lambda_k f(u)\rangle \geq 0. \tag{3.26}$$

Adding (3.25) and (3.26), we obtain that

$$\langle y_k - u, d(\omega_k, y_k)\rangle = \langle y_k - u, \omega_k - y_k - \lambda_k f(\omega_k) + \lambda_k f(y_k)\rangle \geq 0.$$

Thus,

$$\begin{aligned} \langle \omega_k - u, d(\omega_k, y_k)\rangle =&\langle \omega_k - y_k, d(\omega_k, y_k)\rangle + \langle y_k - u, d(\omega_k, y_k)\rangle \\ \geq&\langle \omega_k - y_k, d(\omega_k, y_k)\rangle \\ =&\phi(\omega_k, y_k). \end{aligned} \tag{3.27}$$

Substituting (3.27) into (3.23), it follows that

$$\begin{aligned} \|x_{k+1} - u\|^2 &\leq \|\omega_k - u\|^2 - 2\gamma\beta_k\phi(\omega_k, y_k) + \gamma^2\beta_k\phi(\omega_k, y_k) \\ &= \|\omega_k - u\|^2 - \gamma(2-\gamma)\beta_k\phi(\omega_k, y_k). \end{aligned} \tag{3.28}$$

From the definition of x_{k+1}, it follows that

$$\beta_k\phi(\omega_k, y_k) = \|\beta_k d(\omega_k, y_k)\|^2 = \frac{1}{\gamma^2}\|\omega_k - x_{k+1}\|^2. \tag{3.29}$$

Combining (3.28) and (3.29), we obtain (3.21). Then we prove the inequality (3.22).

From (3.29) and (3.18), we get

$$\phi(\omega_k, y_k) = \frac{1}{\beta_k\gamma^2}\|\omega_k - x_{k+1}\|^2 \leq \frac{1}{c_2\gamma^2}\|\omega_k - x_{k+1}\|^2. \tag{3.30}$$

Combining (3.30) and (3.17), we know

$$\|\omega_k - y_k\|^2 \leq \frac{1}{c_1}\phi(\omega_k, y_k) \leq \frac{1}{c_1 c_2\gamma^2}\|\omega_k - x_{k+1}\|^2.$$

The proof is completed. □

Theorem 3.1. *Assume that Conditions* (i), (ii) *and* (iii) *hold and* λ_k *satisfies the stepsize conditions. Support that* $\sum_{i\in S}\overline{\alpha}_i < 1$, *where* $\overline{\alpha}_i := \sup_{k\in\mathbb{N}}|\alpha_{i,k}|$. *Then the sequence* $\{x_k\}$ *generated by Algorithm 3.1 is bounded. If moreover the following summability condition holds*

$$\sum_{k=1}^{+\infty}\max_{i\in S}|\alpha_{i,k}|\sum_{i\in S}\|x_{k-i}-x_{k-i-1}\|^2 < +\infty, \tag{3.31}$$

then $\{x_k\}$, $\{y_k\}$ *and* $\{\omega_k\}$ *weakly converge to the same solution of the variational inclusion problem* (1.1).

Proof. Fix $u\in\Gamma$. From the definition of ω_k, we have

$$\begin{aligned}
&\|x_{k+1}-u\|^2-\|x_k-u\|^2\\
=&-\|x_{k+1}-x_k\|^2-2\langle x_{k+1}-u,x_k-x_{k+1}\rangle\\
=&-\|x_{k+1}-x_k\|^2-2\langle x_{k+1}-u,\omega_k-x_{k+1}\rangle+2\langle x_{k+1}-u,\omega_k-x_k\rangle\\
=&-\|x_{k+1}-x_k\|^2-2\langle x_{k+1}-u,\omega_k-x_{k+1}\rangle+2\sum_{i\in S}\alpha_{i,k}\langle x_{k+1}-u,x_{k-i}-x_{k-i-1}\rangle.
\end{aligned}\tag{3.32}$$

By using (2.12), we have

$$\begin{aligned}
&\langle x_{k-i}-x_{k-i-1},x_{k+1}-u\rangle\\
=&\langle x_{k-i}-x_{k-i-1},x_{k+1}-x_k+x_k-x_{k-i}\rangle+\langle x_{k-i}-x_{k-i-1},x_{k-i}-u\rangle\\
=&\langle x_{k-i}-x_{k-i-1},x_{k+1}-x_k\rangle+\frac{1}{2}(\|x_k-x_{k-i-1}\|^2-\|x_k-x_{k-i}\|^2)\\
&+\frac{1}{2}(\|x_{k-i}-u\|^2-\|x_{k-i-1}-u\|^2),
\end{aligned}\tag{3.33}$$

and

$$\langle x_{k+1}-u,\omega_k-x_{k+1}\rangle=\frac{1}{2}(\|\omega_k-u\|^2-\|\omega_k-x_{k+1}\|^2-\|x_{k+1}-u\|^2). \tag{3.34}$$

Hence, from (3.21) and (3.34), it follows that

$$\langle x_{k+1}-u,\omega_k-x_{k+1}\rangle\geq\frac{1-\gamma}{\gamma}\|\omega_k-x_{k+1}\|^2. \tag{3.35}$$

Combining (3.32), (3.33) and (3.35), we have

$$\begin{aligned}
&\|x_{k+1}-u\|^2-\|x_k-u\|^2\\
\leq&-\|x_{k+1}-x_k\|^2-\frac{2(1-\gamma)}{\gamma}\|\omega_k-x_{k+1}\|^2+2\langle x_{k+1}-x_k,\sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1})\rangle\\
&+\sum_{i\in S}\alpha_{i,k}(\|x_k-x_{k-i-1}\|^2-\|x_k-x_{k-i}\|^2)+\sum_{i\in S}\alpha_{i,k}(\|x_{k-i}-u\|^2-\|x_{k-i-1}-u\|^2),
\end{aligned}\tag{3.36}$$

which follows that

$$\begin{aligned}
&\|x_{k+1}-u\|^2-\|x_k-u\|^2-\sum_{i\in S}\alpha_{i,k}(\|x_{k-i}-u\|^2-\|x_{k-i-1}-u\|^2)\\
\leq&-\|x_{k+1}-x_k\|^2-\frac{2(1-\gamma)}{\gamma}\|\omega_k-x_{k+1}\|^2+2\langle x_{k+1}-x_k,\sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1})\rangle\\
&+\sum_{i\in S}\alpha_{i,k}(\|x_k-x_{k-i-1}\|^2-\|x_k-x_{k-i}\|^2).
\end{aligned}\tag{3.37}$$

Since

$$\|x_{k+1}-\omega_k\|^2 = \|x_{k+1}-x_k-\sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1})\|^2$$
$$= \|x_{k+1}-x_k\|^2 + \|\sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1})\|^2 - 2\sum_{i\in S}\alpha_{i,k}\langle x_{k+1}-x_k, x_{k-i}-x_{k-i-1}\rangle. \tag{3.38}$$

Combining (3.37) and (3.38), we have

$$\begin{aligned}
&\|x_{k+1}-u\|^2 - \|x_k-u\|^2 - \sum_{i\in S}\alpha_{i,k}(\|x_{k-i}-u\|^2 - \|x_{k-i-1}-u\|^2)\\
\le & -\|x_{k+1}-x_k\|^2 - \frac{2(1-\gamma)}{\gamma}\Big[\|x_{k+1}-x_k\|^2 + \|\sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1})\|^2\\
& - 2\sum_{i\in S}\alpha_{i,k}\langle x_{k+1}-x_k, x_{k-i}-x_{k-i-1}\rangle\Big] + 2\langle x_{k+1}-x_k, \sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1})\rangle\\
& + \sum_{i\in S}\alpha_{i,k}(\|x_k-x_{k-i-1}\|^2 - \|x_k-x_{k-i}\|^2)\\
= & -\frac{2-\gamma}{\gamma}\|x_k-x_{k+1}\|^2 - \frac{2(1-\gamma)}{\gamma}\|\sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1})\|^2\\
& + \sum_{i\in S}\alpha_{i,k}(\|x_k-x_{k-i-1}\|^2 - \|x_k-x_{k-i}\|^2)\\
& + \frac{2(2-\gamma)}{\gamma}\langle x_{k+1}-x_k, \sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1})\rangle.
\end{aligned} \tag{3.39}$$

In the following, we apply some techniques from [19]. Define

$$\beta = \frac{2-\gamma}{\gamma} \quad \text{and} \quad \nu_k = x_{k+1}-x_k-\sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1}).$$

Then, we obtain

$$\begin{aligned}
&\|x_{k+1}-u\|^2 - \|x_k-u\|^2 - \sum_{i\in S}\alpha_{i,k}(\|x_{k-i}-u\|^2 - \|x_{k-i-1}-u\|^2)\\
\le & -\beta\|x_{k+1}-x_k\|^2 + \sum_{i\in S}\alpha_{i,k}(\|x_k-x_{k-i-1}\|^2 - \|x_k-x_{k-i}\|^2)\\
& + (1-\beta)\|\sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1})\|^2 + 2\beta\langle x_{k+1}-x_k, \sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1})\rangle.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\|x_{k+1}-u\|^2 - \|x_k-u\|^2 - \sum_{i\in S}\alpha_{i,k}(\|x_{k-i}-u\|^2 - \|x_{k-i-1}-u\|^2)\\
\le & -\beta\|\nu_k\|^2 + \|\sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1})\|^2\\
& + \sum_{i\in S}|\alpha_{i,k}|(\|x_k-x_{k-i-1}\|^2 + \|x_k-x_{k-i}\|^2).
\end{aligned} \tag{3.40}$$

Then, we define the sequence $\{\delta_k\}$ by

$$\delta_k = \|\sum_{i\in S}\alpha_{i,k}(x_{k-i}-x_{k-i-1})\|^2 + \sum_{i\in S}|\alpha_{i,k}|(\|x_k-x_{k-i-1}\|^2 + \|x_k-x_{k-i}\|^2).$$

Hence, apply the Jensen's inequality, we have

$$\begin{aligned}\delta_k &\le \Big\{\sum_{i\in S}|\alpha_{i,k}|\|x_{k-i}-x_{k-i-1}\|\Big\}^2+\sum_{i\in S}|\alpha_{i,k}|(\|x_k-x_{k-i-1}\|^2+\|x_k-x_{k-i}\|^2)\\ &\le \sum_{i\in S}|\alpha_{i,k}|\sum_{i\in S}|\alpha_{i,k}|\|x_{k-i}-x_{k-i-1}\|^2+2s\sum_{i\in S}|\alpha_{i,k}|\sum_{j\in S}\|x_{k-j}-x_{k-j-1}\|^2\\ &\le s\max_{i\in S}|\alpha_{i,k}|\sum_{i\in S}\|x_{k-i}-x_{k-i-1}\|^2+2s^2\max_{i\in S}|\alpha_{i,k}|\sum_{i\in S}\|x_{k-i}-x_{k-i-1}\|^2\\ &\le (s+2s^2)\max_{i\in S}|\alpha_{i,k}|\sum_{i\in S}\|x_{k-i}-x_{k-i-1}\|^2.\end{aligned}\tag{3.41}$$

Since $s+2s^2$ is a constant, and the sequence $\{\delta_k\}$ is summable if condition (3.31) holds. Next we define $\theta_k=\|x_k-u\|^2-\|x_{k-1}-u\|^2$, then from (3.40), we get

$$\theta_{k+1}\le -\beta\|\nu_k\|^2+\sum_{i\in S}\alpha_{i,k}\theta_{k-i}+\delta_k.\tag{3.42}$$

Set $[\theta]_+=\max\{\theta,0\}$, then from (3.42), we obtain

$$\begin{aligned}[\theta_{k+1}]_+ &\le \sum_{i\in S}|\alpha_{i,k}|[\theta_{k-i}]_++\delta_k\\ &\le \sum_{i\in S}\overline{\alpha}_i[\theta_{k-i}]_++\delta_k.\end{aligned}$$

From Lemma 2.4, since $\sum_{i\in S}\overline{\alpha}_i<1$, $\{[\theta_k]_+\}$ is summable. In turn,

$$\begin{aligned}\|x_{k+1}-u\|^2-\sum_{j=1}^{k+1}[\theta_j]_+ &\le \|x_{k+1}-u\|^2-\theta_{k+1}-\sum_{j=1}^{k}[\theta_j]_+\\ &=\|x_k-u\|^2-\sum_{j=1}^{k}[\theta_j]_+.\end{aligned}$$

So the sequence $\{\|x_k-u\|^2-\sum_{j=1}^k[\theta_j]_+\}$ is decreasing and bounded from below, therefore convergent, and we can also obtain that $\{\|x_k-u\|\}$ is convergent and $\{x_k\}$ is bounded.

From assumption (3.31), we have, for any $i\in S$,

$$\lim_{k\to\infty}|\alpha_{i,k}|\|x_{k-i}-x_{k-i-1}\|=0.\tag{3.43}$$

Moreover, from (3.42), we obtain

$$\sum_{k=1}^{\infty}\|\nu_k\|^2\le\frac{1}{\beta}(\|x_1-u\|^2+\sum_{k=1}^{\infty}(\sum_{i\in S}\overline{\alpha}_i[\theta_{k-i}]_++\delta_k))<+\infty,$$

and $\nu_k\to0$. From (3.43), we have $x_{k+1}-x_k\to0$. Then we get

$$\|x_{k+1}-\omega_k\|\le\|x_{k+1}-x_k\|+\sum_{i\in S}|\alpha_{i,k}|\|x_{k-i}-x_{k-i-1}\|.$$

So we have $\lim_{k\to\infty}\|x_{k+1}-\omega_k\|=0$. From (3.22), it follows that $\lim_{k\to\infty}\|y_k-\omega_k\|=0$.

Since the sequence $\{x_k\}$ is bounded, there exists a subsequence $\{x_{k_j}\}\subset\{x_k\}$ such that $x_{k_j}\rightharpoonup\hat{x}$ as $j\to\infty$. Subsequently, we get $\omega_{k_j}\rightharpoonup\hat{x}$ as $j\to\infty$ and $y_{k_j}\rightharpoonup\hat{x}$ as $j\to\infty$. Now, we will show that $\hat{x}$ is a solution of (1.1), that is, $\hat{x}\in\Gamma$.

We observe that the mapping f is Lipschitz continuous. From Lemma 2.1, we know that $A+f$ is maximal monotone. Let $(v,w)\in G(A+f)$, that is, $w-f(v)\in A(v)$. Since $y_{k_j}=J^A_{\lambda_{k_j}}(\omega_{k_j}-\lambda_{kj}f(\omega_{k_j}))$, we have

$$\omega_{k_j}-\lambda_{k_j}f(\omega_{k_j})\in(I+\lambda_{k_j}A)(y_{k_j}),$$

that is,

$$\frac{\omega_{k_j}-y_{k_j}-\lambda_{k_j}f(\omega_{k_j})}{\lambda_{k_j}}\in Ay_{k_j}. \tag{3.44}$$

By virtue of the maximal monotonicity of A, we have

$$\langle v-y_{k_j}, w-f(v)-\frac{\omega_{k_j}-y_{k_j}-\lambda_{k_j}f(\omega_{k_j})}{\lambda_{k_j}}\rangle\geq 0.$$

Hence,

$$\begin{aligned}\langle v-y_{k_j},w\rangle&\geq\langle v-y_{k_j},f(v)+\frac{\omega_{k_j}-y_{k_j}-\lambda_{k_j}f(\omega_{k_j})}{\lambda_{k_j}}\rangle\\&=\langle v-y_{k_j},f(v)-f(y_{k_j})+f(y_{k_j})-f(\omega_{k_j})+\frac{\omega_{k_j}-y_{k_j}}{\lambda_{k_j}}\rangle\\&\geq\langle v-y_{k_j},f(y_{k_j})-f(\omega_{k_j})\rangle+\langle v-y_{k_j},\frac{\omega_{k_j}-y_{k_j}}{\lambda_{k_j}}\rangle.\end{aligned}$$

Since $\lim_{k\to\infty}\|y_k-\omega_k\|=0$, and f is Lipschitz continuous, we obtain that $\lim_{j\to\infty}\|f(y_{k_j})-f(\omega_{k_j})\|=0$. And since $\inf_{k\geq0}\{\lambda_k\}\geq\underline{\lambda}>0$, it follows that

$$\lim_{j\to\infty}\langle v-y_{k_j},w\rangle=\langle v-\hat{x},w\rangle\geq 0.$$

It follows from the maximal monotonicity of $A+f$ that $0\in(A+f)(\hat{x})$, that is, $\hat{x}\in\Gamma$.

Then from the Lemma 2.3, we can know that the sequences $\{x_k\}$, $\{y_k\}$ and $\{\omega_k\}$ converge weakly to a point in Γ. □

Remark 3.2. If the inertial parameters $\{\alpha_{i,k}\}$ are chosen in $[0,1)$, then the condition (3.31) simplifies to

$$\sum_{k=1}^{+\infty}\max_{i\in S}\alpha_{i,k}\sum_{i\in S}\|x_{k-i}-x_{k-i-1}\|^2<+\infty. \tag{3.45}$$

The condition (3.45) can be enforced by a simple online updating rule such as, for each $i\in S$ and $\alpha_i\in[0,1)$,

$$\alpha_{i,k}=\min\{\alpha_i,c_{i,k}\}, \tag{3.46}$$

where $c_{i,k}$ and $\max\{c_{i,k}\}\sum_{i\in S}\|x_{k-i}-x_{k-i-1}\|^2$ is summable. For instance, one can choose

$$c_{i,k}=\frac{c_i}{k^{1+\delta}\sum_{i\in S}\|x_{k-i}-x_{k-i-1}\|^2},\quad c_i>0,\quad \delta>0.$$

Remark 3.3. Using Liang's way in [19], it is easy to generalize Algorithm 3.1 and Theorem 3.1 and give the corresponding variable metric versions.

From Theorem 3.1, we get the convergence of the one-step inertial proximal contraction algorithm (3.15).

Corollary 3.1. (Conditional convergence for $s = 1$) *Assume that Conditions (i), (ii) and (iii) hold and λ_k satisfies the stepsize conditions. Suppose that $\alpha_k \in (-1,1)$, and $\sup_{k\in\mathbb{N}} |\alpha_k| < 1$. Then the sequence $\{x_k\}$ generated by the algorithm* (3.15) *is bounded. If moreover the following summability condition holds*

$$\sum_{k=1}^{+\infty} |\alpha_k| \|x_k - x_{k-1}\|^2 < +\infty, \tag{3.47}$$

then $\{x_k\}$, $\{y_k\}$ and $\{\omega_k\}$ weakly converge to the same solution of the variational inclusion problem (1.1).

In Corollary 3.1, the inertial parameters $\{\alpha_k\}$ depend on the iterative sequence $\{x_k\}$ which is the origin of its name. Next we address the question whether inertial parameters $\{\alpha_k\}$ can be chosen a-priori such that the algorithm (3.15) is guaranteed to converge.

Theorem 3.2. (Unconditional convergence for $s = 1$) *Assume that Conditions* (i), (ii), *and* (iii) *hold and λ_k satisfies the stepsize conditions. Assume that $\{\alpha_k\}$ is nondecreasing with $\alpha_1 = 0$ and $0 \le \alpha_k \le \alpha < 1$, and $\sigma, \delta > 0$ are such that*

$$\delta > \frac{\alpha^2(1+\alpha) + \alpha\sigma}{1-\alpha^2}, \quad 0 < \gamma \le \frac{2[\delta - \alpha(\alpha(1+\alpha) + \alpha\delta + \sigma)]}{\delta[1 + \alpha(1+\alpha) + \alpha\delta + \sigma]}. \tag{3.48}$$

Then the sequences $\{x_k\}$, $\{y_k\}$ and $\{\omega_k\}$ generated by the algorihtm (3.15) *weakly converge to the same solution of the variational inclusion problem* (1.1).

Proof. Fix $u \in \Gamma$, from lemma 2.2, we have

$$\begin{aligned} \|\omega_k - u\|^2 &= \|(1+\alpha_k)(x_k - u) - \alpha_k(x_{k-1} - u)\|^2 \\ &= (1+\alpha_k)\|x_k - u\|^2 - \alpha_k\|x_{k-1} - u\|^2 + \alpha_k(1+\alpha_k)\|x_k - x_{k-1}\|^2. \end{aligned} \tag{3.49}$$

From (3.21), it follows that

$$\begin{aligned} &\|x_{k+1} - u\|^2 - (1+\alpha_k)\|x_k - u\|^2 + \alpha_k\|x_{k-1} - u\|^2 \\ \le& -\frac{2-\gamma}{\gamma}\|\omega_k - x_{k+1}\|^2 + \alpha_k(\alpha_k+1)\|x_k - x_{k-1}\|^2. \end{aligned} \tag{3.50}$$

Then,

$$\begin{aligned} &\|x_{k+1} - \omega_k\|^2 \\ =&\|(x_{k+1} - x_k) - \alpha_k(x_k - x_{k-1})\|^2 \\ =&\|x_{k+1} - x_k\|^2 + \alpha_k^2\|x_k - x_{k-1}\|^2 - 2\alpha_k\langle x_{k+1} - x_k, x_k - x_{k-1}\rangle \\ \ge&\|x_{k+1} - x_k\|^2 + \alpha_k^2\|x_k - x_{k-1}\|^2 + \alpha_k(-\rho_k\|x_{k+1} - x_k\|^2 - \frac{1}{\rho_k}\|x_k - x_{k-1}\|^2) \\ \ge&(1 - \alpha_k\rho_k)\|x_{k+1} - x_k\|^2 + \frac{\alpha_k(\alpha_k\rho_k - 1)}{\rho_k}\|x_k - x_{k-1}\|^2, \end{aligned} \tag{3.51}$$

where $\rho_k = \frac{2}{2\alpha_k + \delta\gamma}$. Combining (3.50) and (3.51), we obtain

$$\begin{aligned} &\|x_{k+1} - u\|^2 - (1+\alpha_k)\|x_k - u\|^2 + \alpha_k\|x_{k-1} - u\|^2 \\ \le& -\frac{2-\gamma}{\gamma}(1 - \alpha_k\rho_k)\|x_{k+1} - x_k\|^2 + \tau_k\|x_k - x_{k-1}\|^2, \end{aligned} \tag{3.52}$$

where

$$\tau_k = \alpha_k(1+\alpha_k) + \frac{(2-\gamma)\alpha_k(1-\alpha_k\rho_k)}{\gamma\rho_k}. \tag{3.53}$$

Since $\alpha_k\rho_k < 1$ and $\gamma \in (0,2)$, $\tau_k \geq 0$. From the choice of ρ_k, we have

$$\delta = \frac{2(1-\rho_k\alpha_k)}{\rho_k\gamma}.$$

It follows that

$$\tau_k = \alpha_k(1+\alpha_k) + (1-\frac{\gamma}{2})\alpha_k\delta \leq \alpha(1+\alpha) + \alpha\delta.$$

Define the sequence $\{\phi_k\}$ and $\{\xi_k\}$ for all $k \geq 1$ by

$$\phi_k = \|x_k - u\|^2, \xi_k = \phi_k - \alpha_k\phi_{k-1} + \tau_k\|x_k - x_{k-1}\|^2.$$

Using the facts that $\phi_k \geq 0$ and $\{\alpha_k\}$ is nondecreasing, we have

$$\xi_{k+1} - \xi_k \leq \phi_{k+1} - (1+\alpha_k)\phi_k + \alpha_k\phi_{k-1} + \tau_{k+1}\|x_{k+1} - x_k\|^2 - \tau_k\|x_k - x_{k-1}\|^2.$$

From (3.52), we have

$$\xi_{k+1} - \xi_k \leq (\frac{(2-\gamma)(\alpha_k\rho_k - 1)}{\gamma} + \tau_{k+1})\|x_{k+1} - x_k\|^2. \tag{3.54}$$

Now we claim that

$$\frac{(2-\gamma)(\alpha_k\rho_k - 1)}{\gamma} + \tau_{k+1} \leq -\sigma. \tag{3.55}$$

If this assertion does not hold, then

$$\begin{aligned} &\frac{(2-\gamma)(\alpha_k\rho_k - 1)}{\gamma} + \tau_{k+1} > -\sigma \\ \Longleftrightarrow &\gamma(\tau_{k+1} + \sigma) + (2-\gamma)(\alpha_k\rho_k - 1) > 0 \\ \Longleftrightarrow &\gamma(\tau_{k+1} + \sigma) - \frac{\delta\gamma(2-\gamma)}{2\alpha_k + \delta\gamma} > 0 \\ \Longleftrightarrow &(2\alpha_k + \delta\gamma)(\tau_{k+1} + \sigma) + \delta\gamma > 2\delta. \end{aligned} \tag{3.56}$$

However, from (3.48), it follows

$$(2\alpha_k + \delta\gamma)(\tau_{k+1} + \sigma) + \delta\gamma \leq (2\alpha + \delta\gamma)(\alpha(1+\alpha) + \alpha\delta + \sigma) + \delta\gamma \leq 2\delta.$$

It is obvious that they are contradictory. Hence (3.55) is true.

Then from (3.54) and (3.55), it follows that

$$\xi_{k+1} - \xi_k \leq -\sigma\|x_{k+1} - x_k\|^2 \leq 0, \tag{3.57}$$

which implies that $\{\xi_k\}$ is nonincreasing. From $\phi_k \geq 0$, $\alpha_k \leq \alpha$, $\tau_k \geq 0$ and the definition of $\{\xi_k\}$, we have for all $k \geq 1$,

$$-\alpha\phi_{k-1} \leq \phi_k - \alpha\phi_{k-1} \leq \phi_k - \alpha_k\phi_{k-1} + \tau_k\|x_k - x_{k-1}\|^2 = \xi_k \leq \xi_1. \tag{3.58}$$

It follows that

$$\phi_k \leq \alpha^k\phi_0 + \xi_1\sum_{n=0}^{k-1}\alpha^n \leq \alpha^k\phi_0 + \frac{\xi_1}{1-\alpha}.$$

Since $\alpha_1 = 0$, we notice that $\xi_1 = \phi_1 \geq 0$. Combining (3.57) and (3.58), we have

$$\begin{aligned} \sigma\sum_{n=1}^{k}\|x_{n+1} - x_n\|^2 &\leq \xi_1 - \xi_{k+1} \leq \xi_1 + \alpha\phi_k \\ &\leq \alpha^{k+1}\phi_0 + \frac{\xi_1}{1-\alpha} \leq \phi_0 + \frac{\phi_1}{1-\alpha}. \end{aligned} \tag{3.59}$$

It follows that

$$\sum_{k=1}^{\infty} \|x_{k+1} - x_k\|^2 < +\infty, \tag{3.60}$$

and $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0$. Using similar arguments in the proof of Theorem 3.1 in [7], we obtain that $\lim_{k\to\infty} \|x_k - u\|$ exits. Hence $\{x_k\}$ is bounded. From (3.15), we have

$$\|x_{k+1} - \omega_k\| \le \|x_{k+1} - x_k\| + \alpha_k\|x_k - x_{k-1}\| \le \|x_{k+1} - x_k\| + \alpha\|x_k - x_{k-1}\|. \tag{3.61}$$

So $\lim_{k\to\infty} \|x_{k+1} - \omega_k\| = 0$. And from (3.22), we obtain

$$\lim_{k\to\infty} \|y_k - \omega_k\| = 0.$$

Next, using similar arguments in the proof of Theorem 3.1, we get that $\{x_k\}$, $\{y_k\}$ and $\{\omega_k\}$ converge weakly to the same solution of problem (1.1).

4. Applications to variational inequality problems

Let C be a nonempty closed convex subset of a real Hilbert space H. Let $A = N_C$, where $N_C(v)$ is the normal cone of C at $v \in C$, i.e.,

$$N_C(v) := \{d \in H | \langle d, y - v\rangle \le 0, \forall y \in C\}.$$

Let $f : C \to H$ be a singe-valued mapping.Then the problem (1.1) becomes

$$\text{Find } x^* \in H \quad \text{such that} \quad 0 \in N_C(x^*) + f(x^*). \tag{4.62}$$

It is easy to show that the problem (4.62) equals to the classical variational inequality problem, which is to find a point $x^* \in C$ such that

$$\langle f(x^*), x - x^*\rangle \ge 0, \quad \text{for all } x \in C. \tag{4.63}$$

This problem captures various applications arising in many areas, such as partial differential equations, and optimization problems. Recently, there are increasing research on the inertial-type projection algorithms [7, 10, 11, 31, 28]. However, there is no results on the multi-step inertial projection algorithms.

Now we extend Algorithm 3.1 to the variational inequality problems (4.63). Since $J_{\lambda_k}^{N_C} = P_C$, Algorithm 3.1 reduces to the following:

Algorithm 4.1. Let $s \in \mathbb{N}_+$ and $S := \{0, \cdots, s-1\}$. Let $\{\alpha_{i,k}\}_{i\in S} \in (-1, 1)^s$. Choose $x_0 \in H$, $x_{-i-1} = x_0$, $i \in S$.

(S.1) Compute

$$\omega_k = x_k + \sum_{i\in S} \alpha_{i,k}(x_{k-i} - x_{k-i-1}),$$

and

$$y_k = P_C(\omega_k - \lambda_k f(\omega_k)).$$

If $w_k - y_k = 0$: STOP.

(S.2) Calculate

$$d(\omega_k, y_k) = (\omega_k - y_k) - \lambda_k(f(\omega_k) - f(y_k)),$$

and

$$x_{k+1} = \omega_k - \gamma\beta_k d(\omega_k, y_k),$$

where $\gamma \in (0, 2)$ and β_k is given as in (3.14).

(S.3) Set $k \leftarrow k + 1$, and go to (S.1).

By Theorem 3.1, we obtain the following convergence result for the problem (4.63).

Corollary 4.2. *Assume that Conditions* (ii) *and* (iii) *hold and λ_k satisfies the stepsize conditions. Assume that the solution set of the problem* (4.63) *is nonempty. Suppose that $\sum_{i\in S}\overline{\alpha}_i < 1$, where $\overline{\alpha}_i := \sup_{k\in\mathbb{N}}|\alpha_{i,k}|$. Then the sequence $\{x_k\}$ generated by Algorithm 4.1 is bounded. If moreover the following summability condition holds*

$$\sum_{k=1}^{+\infty}\max_{i\in S}|\alpha_{i,k}|\sum_{i\in S}\|x_{k-i}-x_{k-i-1}\|^2 < +\infty, \tag{4.64}$$

then $\{x_k\}$, $\{y_k\}$ and $\{\omega_k\}$ weakly converge to the same solution of the variational inequality problem (4.63).

Remark 4.4. When $s = 1$, Algorithm 4.1 and Theorem 3.2 becomes Algorithm 3.1 and Theorem 3.1 in [7], respectively.

5. NUMERICAL RESULTS

In this section, we present two numerical examples to compare the performance of the original proximal contraction algorithm (PCA), one-step inertial PCA with the unconditional convergence (1-uMiPCA), one-step inertial PCA with the conditional convergence (1-MiPCA) and two-step inertial PCA (2-MiPCA).

All programs are written in Matlab version 7.0. and performed on a PC Desktop Intel(R) Core(TM) i5-4200U CPU @ 1.60GHz 2.30 GHz, RAM 4.00 GB.

In the numerical results listed in the following tables, 'Iter.' and 'CPU time' denote the number of iterations and the execution time in seconds, respectively.

Example 5.1. Consider the variational inequality (4.63) with the linear operator $f(x) := Mx + q$, which is taken from [13] and has been considered by many authors for numerical experiments, see, for example [15], where

$$M = BB^T + S + D,$$

and B is an $m \times m$ matrix, S is an $m \times m$ skew-symmetric matrix, and D is an $m \times m$ diagonal matrix, whose diagonal entries are nonnegative (so M is positive semidefinite), q is a vector in $\mathbb{R}^m$. The feasible set $C \subseteq \mathbb{R}^m$ is closed and convex and defined as

$$C := \{x \in \mathbb{R}^m \mid Qx \le b\},$$

where Q is an $l \times m$ matrix and b is a nonnegative vector. It is clear that f is monotone and L-Lipschitz continuous with $L = \|M\|$. For $q = 0$, the solution set of the variational inequality problem (4.63) $\{0\}$.

TABLE 1. Comparison of four algorithms for $l = 60$ and different m

	Iter.				CPU time			
m	PCA	1-uMiPCA	1-MiPCA	2-MiPCA	PCA	1-uMiPCA	1-MiPCA	2-MiPCA
20	14122	2157	2001	1756	0.2344	0.0938	0.0625	0.0469
30	20611	4140	4401	3669	0.3438	0.1563	0.0781	0.0625
40	30385	6737	6488	5559	0.4063	0.2813	0.1406	0.0938
50	27778	7866	5635	4606	0.4844	0.1875	0.1719	0.1563
60	38615	10607	5255	4957	0.7500	0.2031	0.1719	0.1406

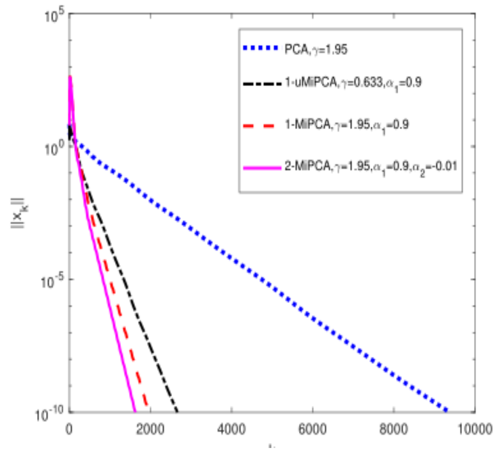

FIGURE 1. Comparison of four algorithms for $m = 50$ and $l = 60$.

Just as in [7], we randomly choose the starting points $x_0 \in [0, 1]^m$ in PCA, 1-uMiPCA, 1-MiPCA and 2-MiPCA and let $x_{-1} = x_0$ in 1-uMiPCA and 1-MiPCA. Take the stopping criterion as $\|x_k\| \leq \epsilon$ and the size $l = 60$ and $m = 20, 30, 40, 50, 60$. The matrices B, S, D and the vector b are generated randomly.

The numerical results in Figure 1 and Table 1 illustrate that the three inertial-type PCA behave far better that the original PCA from the number of iterations and CPU time. 1-MiPCA is the better than 1-uMiPCA and the performance of 2-MiPCA is best among all inertial-type algorithms.

Example 5.2. Let $x^* \in \mathbb{R}^n$ be a K-sparse signal, $K \ll n$. The sampling matrix $G \in \mathbb{R}^{m\times n}(m \ll n)$ is stimulated by standard Gaussian distribution and vector $b = Gx^* + e$, where e is additive noise. When $e = 0$, it means that there is no noise to the observed data. Our task is to recover the signal x^* from the data b.

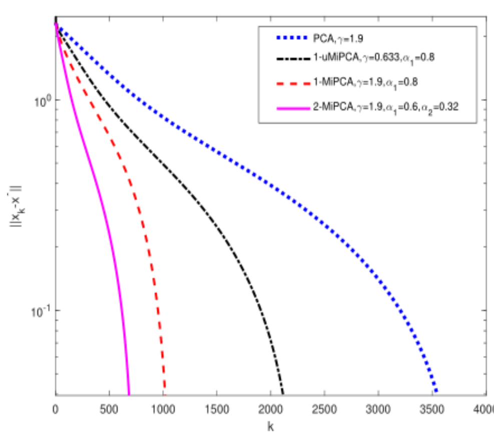

FIGURE 2. Comparison of four algorithms for $m = 600$, $n = 2560$ and $K = 100$.

It's well-known that the sparse signal x^* can be recovered by solving the following LASSO problem [32],

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Gx-b\|_2^2+\gamma\|x\|_1, \tag{5.65}$$

which equals to

$$\text{Find } x^* \in H \text{ such that } 0 \in \gamma\partial\|x^*\|_1 + G^T(Gx^*-b). \tag{5.66}$$

We have that (5.66) is a special case of the problem (1.1). The resolvent $J_\lambda^{\partial\|\cdot\|_1}$ is given by the Moreau decomposition

$$\begin{aligned} J_\lambda^{\partial\|\cdot\|_1}(x) &= (I+\lambda\partial\|.\|_1)^{-1}(x) \\ &= \text{Prox}_{\lambda\|\cdot\|_1}(x) = \text{sgn}(x)\max\{|x|-\lambda, 0\}. \end{aligned}$$

Take the stopping criteria $\|x_k - x^*\| \leq \epsilon$, where $\epsilon > 0$ is a given small constant. We randomly choose the starting points $x_0 \in [0,1]^n$.

TABLE 2. Computational results of four algorithms for different problem sizes.

Problem size			Iter.				CPU time			
n	m	K	PCA	1-uMiPCA	1-MiPCA	2-MiPCA	PCA	1-uMiPCA	1-MiPCA	2-MiPCA
120	512	20	1170	541	289	250	6.2813	3.1719	1.7656	1.3594
120	512	25	1462	681	346	272	8.3438	3.4375	2.2344	1.5469
240	1024	40	2306	1079	570	496	26.125	12.1563	7.6719	6.4219
240	1024	50	2694	1262	625	462	28.4531	14.7344	6.9375	5.3906
360	1536	60	2911	1365	725	628	52.125	27.1563	13.5469	12.1563
360	1536	70	3934	1852	934	752	70.4688	34.0938	17.0156	14.4531
480	2048	80	4333	2034	1044	898	102.4844	46.8438	27.0781	23.3125
480	2048	90	5153	2428	1213	952	126.7969	60.0313	28.7031	24.5313
600	2560	100	5277	2486	1285	1163	183.5	86.4219	46.3438	42.5781
600	2560	110	6528	3077	1538	1198	224.1406	101.0469	48.2031	44.4063

From Figure 2 and Table 2, we get the similar results with Example 5.1. By these two examples, it is necessary to investigate the multi-step inertial algorithms. Note that we don't give comparison with three-step inertial PCA since it improves 2-MiPCA very little.

6. CONCLUSION

In this article we introduce the multi-step inertial proximal contraction algorithms (MiPCA) for the monotone variational inclusion problems and present the weak convergence analysis of the MiPCA. We establish the conditional and unconditional convergence of the one-step inertial proximal contraction algorithm. The numerical experiments are given to illustrate the advantage of the multi-step inertial proximal contraction algorithms.

Note that it needs further research on the convergence rate of the MiPCA and the inertial parameters which do not involve the iterative sequence $\{x_k\}$.

Acknowledgements. The authors would like to thank the anonymous reviewer for the comments on the manuscript which helped us very much in improving and presenting the original version of this paper.

The first author is supported by Scientific Research Project of Tianjin Municipal Education Commission (No: 2018KJ253).

REFERENCES

[1] Bauschke, H. H., Bui, M. N. and Wang, X., *Applying FISTA to optimization problems (with or) without minimizers*, Math. Program., https://doi.org/10.1007/s10107-019-01415-x, (2019)

[2] Bauschke, H. H. and Combettes, P. L., *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Second Edition, Springer, 2017

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

[4] Brézis, H., *Opérateur maximaux monotones,* In Mathematics Studies, 1973

[5] Ceng, L. C., Lin, Y. C. and Wen, C. F., *Iterative methods for triple hierarchical variational inequalities with mixed equilibrium problems, variational inclusions, and variational inequalities constraints*, J. Inequal. Appl., **2015** 2015:16, 62 pp.

[6] Chambolle, A. and Dossal, Ch., *On the Convergence of the Iterates of the Fast Iterative Shrinkage/ Thresholding Algorithm*, J. Optim. Theory Appl., **166** (2015), 968–982

[7] Dong Q. L., Cho Y. J., Zhong L. L. and Rassias T. M., *Inertial projection and contraction algorithms for variational inequalities*, J. Global Optim., **70** (2018), 687–704

[8] Dong, Q. L., Huang, J., Li, X. H., Cho, Y. J. and Rassias, Th. M., *MiKM: Multi–step inertial Krasnosel'skiĭ–Mann algorithm and its applications*, J. Global Optim., **73** (2019), No. 4, 801–824

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

[10] Dong, Q. L., Lu, Y. Y. and Yang, J., *The extragradient algorithm with inertial effects for solving the variational inequality*, Optim., **65** (2016), No. 12, 2217–2226

[11] Fan, J., Liu, L. and Qin, X., *A subgradient extragradient algorithm with inertial effects for solving strongly pseudo-monotone variational inequalities*, Optimization, https://doi.org/10.1080/02331934.2019.1625355, (2020)

[12] Fang, Y. P. and Huang, N. J., *H-monotone operator and resolvent operator technique for variational inclusion*, Appl. Math. Comput., **145** (2006), 795–803

[13] Harker, P. T. and Pang, J. S., *A damped-newton method for the linear complementarity problem*, Lect. Appl. Math., **26** (1990), 265–284

[14] He, B. S., *A class of projection and contraction methods for monotone variational inequalities*, Appl. Math. Optim., **35** (1997), 69–76

[15] Hieu, D. V., Anh, P. K. and Muu, L. D., *Modified hybrid projection methods for finding common solutions to variational inequality problems*, Comput. Optim. Appl., **66** (2017), 75–96

[16] Huang, N. J., *A new completely general class of variational inclusions with noncompact valued mappings*, Comput. Math. Appl., **35** (1998), No. 10, 9–14

[17] Jitpeera, T. and Kumam, P., *A new hybrid algorithm for a system of mixed equilibrium problems, fixed point problems for nonexpansive Semigroup, and variational inclusion problem*, Fixed Point Theory Appl., **2011** (2011), 217407

[18] Liang, J. and Schönlieb, C. B., *Improving FISTA: Faster, Smarter and Greedier*, https://arxiv.org/abs/1811.01430v2, (2019)

[19] Liang, J., *Convergence Rates of First-Order Operator Splitting Methods,* Optim. Cont [math.OC]. Normandie Université. Greyc Cnrs Umr 6072, 2016

[20] Lions, P. L. and Mercier, B., *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal., **16** (1979), 964–979

[21] Lorenz, D. A. and Pock, T., *An inertial forward-backward algorithm for monotone inclusions*, J. Math. Imaging Vis., **51** (2015), 311–325

[22] Moudafi, A. and Oliny, M., *Convergence of a splitting inertial proximal method for monotone operators*, J. Comp. Appl. Math., **155** (2003), 447–454

[23] Nesterov, Y. E., *A method for solving the convex programming problem with convergence rate $O(1/k^2)$*, Dokl. Akad. Nauk SSSR., **269** (1983), 543–547 (in Russian)

[24] O'Donoghue, B. and Candès, E. J., *Adaptive restart for accelerated gradient schemes*, Found. Comput. Math., **15** (2015), No. 3, 715–732

[25] Polyak, B. T., *Introduction to Optimization,* Optimization Software 1987

[26] Polyak, B. T., *Some methods of speeding up the convergence of iteration methods*, U.S.S.R. Comput. Math. Math. Phys., **4** (1964), No. 5, 1–17

[27] Rockafellar, R. T., *Monotone operators and the proximal point algorithms*, SIAM J. Control Optim., **14** (1976), No. 5, 877–898

[28] Shehu, Y. and Cholamjiak, P., *Iterative method with inertial for variational inequalities in Hilbert spaces*, CALCOLO., **56** (2019), No. 1, Art. 4, 21 pp.

[29] Shehu, Y., Dong, Q. L. and Jiang, D., *Single projection method for pseudo-monotone variational inequality in Hilbert spaces*, Optimization, **68** (2019), 385–409

[30] Tang, Y. C., Wu, G. R. and Zhu, C. X., *A first-order splitting method for solving a large-scale composite convex optimization Problem*, J. Comp. Math., **37** (2019), No. 5, 668–690
[31] Thong, D. V. and Hieu, D. V., *Inertial subgradient extragradient algorithms with line-search process for solving variational inequality problems and fixed point problems*, Numer. Algorithms, **80** (2019), 1283–1307
[32] Tibshirani, R., *Regression Shrinkage and Selection Via the Lasso*, J. Royal Stat. Soc., **58** (1996), 267–288
[33] Verma, R. U., *A-monotone nonlinear relaxed cocoercive variational inclusions*, Cent. Eur. J. Math., **5** (2007), 386–396
[34] Verma, R. U., *Approximation-solvability of a class of A-monotone variational inclusion problems*, J. Korea Soc. Indus. Appl Math., **8** (2004), No. 1, 55–66
[35] Zeng, L. C., Guu, S. M. and Yao, J. C., *Characterization of H-monotone operators with applications to variational inclusions*, Comput. Math. Appl., **50** (2005), No. 3-4, 329–337
[36] Zhang, S. S., Lee, J. H. W. and Chan, C. K., *Algorithms of common solutions to quasi variational inclusion and fixed point problems*, Appl. Math. Mech., **29** (2008), 571–581
[37] Zhang, C. and Wang, Y., *Proximal algorithm for solving monotone variational inclusion*, Optimization, **67** (2018), No. 8, 1197–1209
[38] Zhao, J. and Yang, Q., *Self-adaptive projection methods for the multiple-sets split feasibility problem*, Inverse Probl., **27** (2011), No. 3, 035009, 13 pp.

CIVIL AVIATION UNIVERSITY OF CHINA
COLLEGE OF SCIENCE
TIANJIN, 300300, CHINA
E-mail address: cjzhang@cauc.edu.cn
E-mail address: dongql@lsec.cc.ac.cn
E-mail address: 2428743880@qq.com