J. Appl. Numer. Optim. 2 (2020), No. 3, pp. 373–385
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.2.2020.3.07

STRONG CONVERGENCE OF A MODIFIED INERTIAL FORWARD-BACKWARD SPLITTING ALGORITHM FOR A INCLUSION PROBLEM

LIYA LIU

Department of Mathematics, Hangzhou Normal University, Zhejiang, China

Abstract. The aim of this paper is to introduce a new inertial-like forward-backward splitting algorithm for solving a maximal monotone inclusion problem. A strong convergence theorem is established in the setting of real Hilbert spaces. A numerical example is also provided to support our main result.

Keywords. Forward-backward splitting method; Inertial extrapolation; Maximal monotone mapping; Strong convergence; Tseng's algorithm.

1. Introduction

A fundamental and classical problem in nonlinear functional analysis and optimization is to find a zero point of a maximal monotone mapping G, that is,

$$\text{find } x \in H \text{ such that } 0 \in G(x). \tag{1.1}$$

This problem finds a number of important real-world applications in scientific fields, such as, image reconstructions, computer visions, the machine learning and the signal processing; see, e.g., [3, 29, 31] and the references therein. The celebrated Proximal Point Algorithm (PPA), which was first studied by Martinet [14, 15] and Rockafellar [21, 22], is one of classical methods to solve the above inclusion problem. Problem (1.1) includes, as special cases, equilibrium problems, convex-concave saddle-point problems, non-smooth variational inequalities, split feasibility problems and so on [4, 6, 23].

The framework of this paper is a real Hilbert space. From now on, we always assume that H is a real Hilbert space, and its inner product and norm are denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. Let C be a convex and closed nonempty set in space H. Let $G : H \rightrightarrows H$ be a set-valued mapping with $\operatorname{dom} G = \{x \in H : G(x) \neq \emptyset\}$. G is said to be maximal monotone if and only if G is monotone, i.e., $\langle x - x^{'}, y - y^{'} \rangle \geq 0$, for all $y \in G(x), y^{'} \in G(x^{'})$, and its graph, denoted by $\operatorname{gph} G := \{(x, y) \in H \times H \mid y \in G(x)\}$, is not properly contained in the graph of any other monotone mapping on H. On the other hand, one knows that every monotone continuous mapping is maximal. The theory of maximal monotone operators has been well studied by Minty, Moreau and Rockafellar [16, 17, 21].

One of celebrated examples of set-valued maximal monotone mappings is the subdifferential of proper, convex, and closed functions. The subdifferential, ∂g, of a proper, convex, and closed function $g : H \mapsto (-\infty, \infty]$ is defined by

$$\partial g(x) := \{x^* \in H : \langle x' - x, x^* \rangle \leq g(x') - g(x), \forall x' \in H\}, \quad \forall x \in H.$$

E-mail address: liya42@qq.com.
Received June 12, 2020; Accepted October 2, 2020.

It is known that the inclusion problem of operator $\partial g(x)$ is equivalent to the minimization problem of function $g(x)$, that is, $0 \in \partial g(v)$ if and only if $g(v) = \min_{x\in H} g(x)$. Let N_C be the normal cone operator defined by

$$N_C(x) = \{y \in H : \langle y, x' - x\rangle \leq 0 \text{ for all } x' \in C\}$$

for each $x \in C$ and is empty if otherwise. Let G be a continuous single-valued maximal monotone mapping defined on C. It follows that $x \in C$ satisfies the variational inequality of $\langle Gx, x' - x\rangle \geq 0$ for all $x' \in C$ if and only if $0 \in N_C(x) + Gx$.

The interdisciplinary nature of problem (1.1) is evident through a vast literature which includes a large body of mathematical and algorithmic developments; see, for instance, [1, 2] and the references therein.

(i) When $G = \nabla g$ is the gradient of a differentiable convex function g. The most simple method to solve (1.1) is to apply the following iterative scheme. For any point $x_0 \in C$, the sequence $(x_n)_{n\geq 0}$ can be defined as $x_{n+1} = (I - \gamma_n G)x_n$, $\forall n \in \mathbb{N}$, where the step size parameter $\{\gamma_n\}$ is chosen according to a rule that guarantees the weak convergence of the algorithm. $I - \gamma_n G$ is the so-called forward operator. The above scheme is nothing else than the classical steepest descend method.

(ii) When G is a general maximal monotone mapping. A classical method for solving (1.1) is the proximal point algorithm, which can be traced back to the early works in [19, 22]. Starting with any point $x_0 \in C$, a sequence $(x_n)_{n\geq 0} \subset H$ can be defined as $x_{n+1} = (I + \gamma_n G)^{-1}(x_n)$, $\forall n \in \mathbb{N}$, where $\gamma_n > 0$ is a regularization parameter, see [18]. $(I + \gamma_n G)^{-1}$ is the so-called resolvent operator. In the context of algorithms, the resolvent operator is often referred as the backward operator. In addition, Rockafellar [22] has shown that the sequence $(x_n)_{n\geq 0}$ generated by the proximal point algorithm weakly converges to a point x^* such that $0 \in G(x^*)$. Quite often, the evaluation of the resolvent operator, which is an inverse problem, is too difficult to solve problem (1.1). This limits the practicability of the proximal point algorithm. Aiming at this, Rockafellar [22] proved that this algorithm still has a weakly convergence by using the inexact evaluation of the resolvent operator. The evaluation error has to satisfy a certain summability condition which essentially implies that the resolvent operator has to be computed with the increasing accuracy. In fact, this is still somewhat limiting due to the error of the resolvent operator is often hard to control in practice.

A favorable situation occurs when the operator G can be written as the sum of two maximal monotone operators, i.e., $G = A + B$, such that $(I + \alpha A)^{-1}$ or $(I + \alpha B)^{-1}$ is much easier to compute than the full resolvent $(I + \alpha G)^{-1}$. In general, by combining the resolvent operator with respect to A or B in a certain way, one might be able to mimic the effect of the full proximal step based on G. Motivated by the proximal point method and the steepest decent method, Lions and Mercier [11] proposed the forward-backward splitting method in the case of $G = A + B$. It is based on the recursive application of an explicit forward step with respect to B, followed by an implicit backward step with respect to A. The algorithm takes the following form:

Algorithm 1.1

Input: initial data: x_0;iterative stepsize: $(\gamma_i)_{i\geq 0}$.
Output: Output x

1: Set $n \leftarrow 1$.
2: Initialize the data $x_0, x_1 \in H$.
3: **while** not converged **do**
$$x_{n+1} = (I + \gamma_n A)^{-1}(x_n - \gamma_n B x_n).$$
4: **end while**
5: **return** $x = x_n$

Basically, A is maximal monotone and B is single valued and L-Lipschitz continuous, i.e., $\|Bx^{'} - Bx\| \leq L\|x^{'} - x\|, (x^{'}, x \in H)$. If, in addition, B is the gradient of a smooth convex function, the situation becomes much more beneficial. This method was recently extensively analyzed and further studied, see [7, 10, 24, 28] and references therein. Of course, the problem decomposition is not the only consideration, the convergence rate is another.

In 1964, Polyak [19] introduced a heavy-ball technique in the context of optimization problems. In the context of algorithms, it is often referred as the inertial extrapolation. It involves two iterative steps and the second one is defined based on the previous two iterates. Recently, the inertial extrapolation, which has been intensively investigated, is viewed as an acceleration procedure of speeding up the rate of convergence; see, e.g., [13, 20, 25, 27, 32]. Some authors constructed fast iterative algorithms by using the inertial extrapolation, including inertial proximal algorithms, inertial forward-backward algorithms, inertial forward-backward splitting algorithms, and so on; see [9, 26] and the references therein. On the other hand, the heavy-ball method can also be interpreted as an explicit finite difference discretization of the two-order time dynamical system. To motivate the so-called the inertial forward backward splitting algorithm, we consider the following regularized dynamic system

$$M_{A,B,\gamma(t)}(x(t)) + \ddot{x}(t) + \iota(t)\dot{x}(t) = 0. \tag{1.2}$$

where $A : H \rightrightarrows H$ and $B : H \rightrightarrows H$ are maximal monotone operators, and $M_{A,B,\gamma(t)} : H \rightrightarrows H$ is the operator defined by

$$M_{A,B,\gamma}(x) = \frac{(x - (x + \gamma A(x))(x - \gamma B(x)))}{\gamma}.$$

There is classical trick to replace the inclusion $0 \in (A+B)(x)$ with a regular equation $M_{A,B,\gamma(t)}(x) = 0$. Indeed, for any $\gamma > 0$, we have the equivalences

$$0 \in (A+B)(x) \Leftrightarrow x - (x + \gamma A(x))^{-1}(x - \gamma B(x)) = 0 \Leftrightarrow M_{A,B,\gamma(t)}(x) = 0.$$

It comes naturally by discretizing (1.2) explicitly with a time step $h_k > 0$. Denote $t_k = \sum_{i=1}^{k} h_i$, $\iota_k = \iota(t_k)$, $\gamma_k = \gamma(t_k)$ and $x_k = x(t_k)$. An explicit finite-difference scheme for (1.2) with centered second-order variations gives that

$$\frac{1}{h_k^2}(x_{k+1} - 2x_k + x_{k-1}) + \frac{\iota_k}{h_k}(x_k - x_{k-1}) + M_{A,B,\gamma_k}(y_k) = 0, \tag{1.3}$$

where y_k is a point belonging to the line passing through x_{k-1} and x_k. The above equality can be rewritten as

$$x_{k+1} = \frac{\gamma_k - h_k^2}{\gamma_k}((1 - \iota_k h_k)(x_k - x_{k-1}) + x_k) + \frac{h_k^2}{\gamma_k}(I + \gamma_t A)^{-1}(y_k - \gamma_k B(y_k)). \tag{1.4}$$

Set $\gamma_k = h_k^2$ and $\alpha_k = 1 - \iota_k h_k$. Using the classical Nesterov extrapolation choice for y_k, one has

$$y_k = \alpha_k(x_k - x_{k-1}) + x_k$$

which, together with (1.4), yields that

$$x_{k+1} = (I + \gamma_k A)^{-1}(y_k - \gamma_k B(y_k)).$$

where $\{\gamma_t\}$ is a step-size parameter. The extrapolation term $\alpha_k(x_k - x_{k-1})$ is intended to speed up the rate of convergence. Therefore, this dynamical approach leads to a special case of the forward-backward algorithm of inertial type. An alternative modification to the inertial forward-backward splitting algorithm is the following algorithm proposed by Lorzenz and Pock [12]. The algorithm weakly converges to a zero of the sum of two maximal monotone operators with one of two operators being L-Lipschitz continuous and single valued. The scheme is as follows:

Algorithm 1.2

Input: initial data: x_0; iterative stepsizes: $(\alpha_i)_{i\geq 0}, (\gamma_i)_{i\geq 0}$.
Output: Output x

1: Set $k \leftarrow 1$.
2: Initialize the data $x_0, x_1 \in H$.
3: **while** not converged **do**
   $y_k = x_k + \alpha_k(x_k - x_{k-1})$,
   $x_{k+1} = (M + \gamma_t A)^{-1}(M - \gamma_t B)(y_t)$.
4: **end while**
5: **return** $x = x_k$

where M is a linear self-adjoint positive definite mapping, γ_t is a step-size parameter and $\alpha_k \in [0,1)$ is an extrapolation factor.

For all $x \in H$, there exists a unique nearest point in C, denoted by $P_C x$, such that $\|x - P_C x\| = \inf\{\|x - y\| : y \in C\}$. Then P_C is called the metric projection of H onto C. It is known that the projection operator can be characterized by

(i) $\|P_C x - P_C y\| \leq \|x - y\|,\ \forall x, y \in H$;
(ii) $\langle x - P_C x, y - P_C x\rangle \leq 0,\ \forall x \in H, y \in C$.

Among first-order methods, there have always been some trade-off between methods with variable stepsizes and ones with fixed stepsizes. These methods with fixed stepsizes require less computation per iteration, however, very often they require to know much more information, for example, the Lipschitz constant of mappings, see [12]. Usually, we can estimate the Lipschitz constant from the above. However, this estimation is often quite conservative and the method will use tiny steps.

In [28], Tseng described the modified forward-backward splitting method with the Lipschitz constant unknown. Let $X \subset \operatorname{dom} A$ be a convex closed set with the restriction that $X \cap (A+B)^{-1}(0) \neq \emptyset$. The scheme is stated as follows:

Algorithm 1.3

Input: initial data: x_0; iterative stepsize: $(\gamma_i)_{i\geq 0}$.
Output: Output x

1: Set $n \leftarrow 1$.
2: Initialize the data $x_0, x_1 \in H$.
3: **while** not converged **do**
   $\bar{x}_n = (I + \gamma_n A)^{-1}(x_n - \gamma_n B x_n)$,
   $x_{n+1} = P_X(\bar{x}_n - \gamma_n(A(\bar{x}_n) - A(x_n)))$.
4: **end while**
5: **return** $x = x_n$

where B is single-valued, A is a set-valued and $\{\gamma_n\}$ satisfies an Armijo-Goldstein-type stepsize rule. Specifically, $\{\gamma_n\}$ is chosen to be the largest $\gamma \in \{\rho, \rho\sigma, \rho\sigma^2, \cdots\}$ such that $\gamma\|Az_n - Ay_n\| \leq \kappa\|z_n - y_n\|$. It was shown that, under certain conditions on κ, ρ and σ, the algorithm converges weakly to an element of $(B+A)^{-1}(0)$. This method seems to be simple and the search procedure is inexpensive and flexible. We note that variable stepsizes run some search procedures by using the Armijo-Goldstein-type stepsize rule, which aims to find an appropriate stepsize in each iteration. This method often allows to use a larger step than what is predicted by the Lipschitz constant. It uses variable and nonmonotone stepsizes to avoid the modulus of A or B or the Lipschitz constant of B.

Unfortunately, those methods are only known to be weakly convergent in the setting of Hilbert spaces. In many disciplines, problems always arises in infinite dimensional spaces. For these problems, strong convergence, that is, norm convergence, is often much more desirable than weak convergence (convergence in weak topology). The natural question that arises is how to construct an algorithm which guarantees the norm convergence in the setting of infinite dimensional real Hilbert spaces. To answer this question, Takahashi, Takeuchi and Kubota [33] introduced the following shrinking projection algorithm

Algorithm 1.4

Input: initial data: x_0; iterative stepsize: $(\lambda_i)_{i\geq 0}$.
Output: Output x

1: Set $n \leftarrow 1$.
2: Initialize the data $x_0, x_1 \in H$.
3: **while** not converged **do**
   $z_n = \lambda_n x_n + (1-\lambda_n)Tx_n$,
   $C_{n+1} = \{w \in C_n : \|z_n - w\| \leq \|x_n - w\|\}$,
   $x_{n+1} = P_{C_{n+1}}(x_0)$.
4: **end while**
5: **return** $x = x_{n+1}$

where $T : H \to H$ is a nonexpansive mapping, i.e., 1-Lipschitz continuous. They proved that the sequence $\{x_n\}$ generated above converges strongly to an element of the variational inequality solution set. Note that the shrinking projection method is essentially important for proving the strongly convergence in many cases.

In this paper, motivated by the mentioned works in literature, and the ongoing research in these directions, we propose a modified inertial forward-backward splitting algorithm with the Armijo-like step size rule for finding a common zero point of the sum of two families of maximal monotone mappings. A norm convergence theorem is established by using the shrinking projection method. This algorithm is based on the inertial method and the Tseng's splitting method. Finally, we give a numerical example to illustrate the performance of the proposed algorithm within the framework of the signal processing problem.

In order to prove our main result, we need the following lemmas.

Lemma 1.1. [28] *Let H be a real Hilbert space and let $S : H \rightrightarrows H$ be a maximal monotone mapping. Let $\{x_n\}$ be a bounded sequence in H converging weakly to some x, and let $\{w_n\}$ be a sequence in H converging strongly to some w. If $w_n \in S(x_n)$ for all n, then $w \in S(x)$.*

Lemma 1.2. [8] *Let H be a real Hilbert space and let $\{x_n\}$ be a vector sequence in H. $\|x_n\| \to \|x\|$ and $x_n \rightharpoonup x$ imply that $\{x_n\}$ converges to x in norm as $n \to \infty$.*

Lemma 1.3. [30] *Let $\{a_n\}$ be a sequence of nonnegative real numbers and let $\{b_n\}$ be a sequence of real numbers. Let $\{\alpha_n\}$ be a sequence of real numbers in $(0,1)$ such that $\sum_{n=1}^{\infty} \alpha_n = \infty$. Assume that*

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

If $\limsup_{n\to\infty} b_n \leq 0$, then $\lim_{n\to\infty} a_n = 0$.

2. Main Results

First, we give our algorithm for our inclusion problem involving two family of maximal monotone operators $\{A_1, A_2, \cdots, A_M\}$ and $\{B_1, B_2, \cdots, B_M\}$, where M is some positive integer.

Algorithm 2.1 Shrinking inertial splitting algorithm with Armijo-like step (SISA)

Input: initial data: $x_0, u \in H, x_1 = P_{C_1}u$; algorithm parameters: $(\alpha_i)_{i\geq 0}, \rho, \sigma, \kappa$.
Output: Output x

1: Set $n \leftarrow 1$.
2: Initialize the data $x_0, x_1 \in H$.
3: **while** not converged **do**

$$y_n = x_n + \alpha_n(x_n - x_{n-1}),$$

4: **for** $m = 1, 2, \cdots, M$ **do**

$$z_{n,m} = (I + \gamma_{n,m}A_m)^{-1}(y_n - \gamma_{n,m}B_m y_n), \tag{2.1}$$

where $\gamma_{n,m}$ is chosen to be the largest $\gamma \in \{\rho, \rho\sigma, \rho\sigma^2, \cdots\}$ such that

$$\gamma\|B_m y_n - B_m z_{n,m}\| \leq \kappa\|y_n - z_{n,m}\|, \tag{2.2}$$

$$w_{n,m} = z_{n,m} - \gamma_n(B_m(z_{n,m}) - B_m(y_n)), \tag{2.3}$$

5: **end for**

$$w_n = \frac{1}{M}\sum_{m=1}^{M} w_{n,m},$$

$$C_{n+1} = \{w \in C_n : \|w_n - w\|^2 \leq \|x_n - w\|^2 + 2\alpha_n\langle x_n - w, x_n - x_{n-1}\rangle, \\ +\alpha_n^2\|x_n - x_{n-1}\|^2 - (1-\kappa^2)\frac{1}{M}\sum_{m=1}^{M}\|y_n - z_{n,m}\|^2\} \tag{2.4}$$

$$x_{n+1} = P_{C_{n+1}}u. \tag{2.5}$$

6: **end while**
7: **return** $x = x_{n+1}$

The following assumptions will be used for the convergence analysis of the above algorithm.

(a) The solution set $\Lambda = \bigcap_{m=1}^{M}(A_m + B_m)^{-1}(0)$ is nonempty.
(b) The mappings $A_m : H \rightrightarrows H$ and $B_m : H \rightrightarrows H$ are maximal monotone with B_m being Lipschitz continuous and single-valued on dom $B_m \supset \mathrm{dom} A_m$, where $m = 1, 2, \cdots, M$.

Remark 2.1. From (2.4), one has

$$\|w_n - x\|^2 \leq \|x_n - x\|^2 + 2\alpha_n\langle x_n - x, x_n - x_{n-1}\rangle + \alpha_n^2\|x_n - x_{n-1}\|^2 - (1-\kappa^2)\|y_n - z_n\|^2, \quad x \in C_n,$$

which is equivalent to

$$2\langle x, y_n - w_n\rangle \leq \|y_n\|^2 - \|w_n\|^2 - 2(1-\kappa^2)\|y_n - z_n\|^2.$$

This shows that the set C_{n+1} defined in algorithm (SISA) is a half space. As a result, the nearest point projection $P_{C_{n+1}}$ has a closed form expression, which can be easily computed.

Lemma 2.1. *Let $m \in \{1, 2, \cdots, r\}$. If B_m is Lipschitz continuous on $\mathrm{dom} B_m \supset \mathrm{dom} A_m$ with constant $L_m > 0$, then $\gamma_{n,m}$ is well defined for all n, and the Armijo-like search rule (2.2) always terminates and*

$$\min\left\{\rho, \frac{\kappa\sigma}{L_m}\right\} \leq \gamma_{n,m} \leq \rho.$$

Proof. It follows from the Lipschitz continuity of B_m, we have that $\|B_m(y_n) - B_m(z_{n,m})\| \leq L_m\|y_n - z_{n,m}\|$, that is,

$$\frac{\kappa}{L_m}\|B_m(y_n) - B_m(z_{n,m})\| \leq \kappa\|y_n - z_{n,m}\|.$$

For all $\gamma \leq \frac{\kappa}{L_m}$, this shows that (2.2) holds. Hence, sequences $\{\gamma_{n,m}\}$ are well defined. Observe $\rho \geq \gamma_{n,m}$. If $\gamma_{n,m} = \rho$, this is done. Otherwise, if $\gamma_{n,m} < \rho$, one can easily reach a contradiction. Since B_m is L_m-Lipschitz continuous on domA_m, one sees that

$$\begin{aligned}\kappa\|y_n - z_{n,m}\| &\leq \frac{\gamma_{n,m}}{\sigma}\|B_m y_n - B_m z_{n,m}\| \\ &\leq \frac{L_m \gamma_{n,m}}{\sigma}\|y_n - z_{n,m}\|.\end{aligned}$$

This implies that $\frac{\kappa\sigma}{L_m} < \gamma_{n,m}$. This completes the proof. □

Now we are in a position to prove our main convergence theorem.

Theorem 2.1. *Let $\{x_n\}$ be a sequence generated by Algorithm (SISA) and let $G_m = A_m + B_m$, where $m \in \{1,2,\cdots,r\}$. Let one of the following assumptions be satisfied (i) $\liminf_{n\to\infty} \gamma_{n,m} > 0$; (ii) B_m is continuous on dom$B_m \supset$ domA_m and the function $x \mapsto \min_{\Phi \in G_m(x)} \|\Phi\|$ is locally bounded on domA_m. Then $\{x_n\}, \{y_n\}$ and $\{w_n\}$ generated by Algorithm (SISA) converge to $z \in P_\Lambda u$ in norm.*

Proof. For any $n \in N$, $m \in \{1,2,\cdots,M\}$, we denote $a_{n,m} = B_m y_n$ and $p_{n,m} = B_m z_{n,m}$. Furthermore, there exists $q_{n,m}$ in $A_m z_{n,m}$. Invoking (2.1) and (2.5), we find that

$$z_{n,m} + \gamma_{n,m} q_{n,m} = y_n - \gamma_{n,m} a_n,$$

and

$$w_{n,m} = z_{n,m} - \gamma_{n,m}(p_{n,m} - a_{n,m}).$$

This sends us to

$$y_n - w_{n,m} = \gamma_{n,m} p_{n,m} + \gamma_{n,m} q_{n,m}. \tag{2.6}$$

For each $e \in \Lambda$, that there exists $p, q \in H$ such that $p = B_m(e)$ and $q = A_m(e)$. Hence, $p + q = 0$. It follows from (2.3) and (2.6) that

$$\begin{aligned}\|y_n - e\|^2 &= \|y_n - z_{n,m}\|^2 + \|z_{n,m} - w_{n,m}\|^2 + \|w_{n,m} - e\|^2 + 2\langle y_n - z_{n,m}, z_{n,m} - e\rangle \\ &\quad + 2\langle z_{n,m} - w_{n,m}, w_{n,m} - e\rangle \\ &= \|y_n - z_{n,m}\|^2 - \gamma_{n,m}^2\|B_m z_{n,m} - B_m y_n\|^2 + \|w_{n,m} - e\|^2 + 2\gamma_{n,m}\langle p_{n,m} + q_{n,m}, z_{n,m} - e\rangle \\ &= \|y_n - z_{n,m}\|^2 - \gamma_{n,m}^2\|B_m z_{n,m} - B_m y_n\|^2 + \|w_{n,m} - e\|^2 + 2\gamma_{n,m}\langle p_{n,m} - p, z_{n,m} - e\rangle \\ &\quad + 2\gamma_{n,m}\langle q_{n,m} - q, z_{n,m} - e\rangle.\end{aligned} \tag{2.7}$$

The monotonicity of A_m and B_m gives that $\langle p_{n,m} - p, z_{n,m} - e\rangle \geq 0$ and $\langle q_{n,m} - q, z_{n,m} - e\rangle \geq 0$. Borrowing (2.7) to yeild

$$\|w_{n,m} - e\|^2 \leq \|y_n - e\|^2 - \|y_n - z_{n,m}\|^2 + \gamma_{n,m}^2\|B_m y_n - B_m z_{n,m}\|^2. \tag{2.8}$$

Employing (2.8), one arrives at

$$\begin{aligned}\|w_n-e\|^2\le&\frac{1}{M}\sum_{m=1}^{M}\|w_{n,m}-e\|^2\\ \le&\|\alpha_n(x_n-x_{n-1})+x_n-e\|^2-\|y_n-z_{n,m}\|^2+\frac{1}{M}\sum_{m=1}^{M}\gamma_{n,m}^2\|B_my_n-B_mz_{n,m}\|^2\\ \le&\|x_n-e\|^2+2\alpha_n\langle x_n-e,x_n-x_{n-1}\rangle+\alpha_n^2\|x_n-x_{n-1}\|^2-\frac{1}{M}\sum_{m=1}^{M}\|y_n-z_{n,m}\|^2\\ &+\frac{1}{M}\sum_{m=1}^{M}\gamma_{n,m}^2\|B_my_n-B_mz_{n,m}\|^2\\ \le&\|x_n-e\|^2+2\alpha_n\langle x_n-e,x_n-x_{n-1}\rangle+\alpha_n^2\|x_n-x_{n-1}\|^2-(1-\kappa^2)\frac{1}{M}\sum_{m=1}^{M}\|y_n-z_{n,m}\|^2.\end{aligned}\tag{2.9}$$

It is evident that set C_{n+1} is convex and closed. In light of (2.9), we have that $\Lambda\subseteq C_{n+1}$, $\forall n\in\mathbb{N}$. For this, sequence $\{x_n\}$ is well defined. Denote $z=P_\Lambda u$. From $x_n=P_{C_n}u$ and $z\in\Lambda\subseteq C_n$, one arrives at $\|x_n-u\|\le\|z-u\|$, which further implies that $\{x_n\}$ is a bounded vector sequence. Since $x_n=P_{C_n}u$, we observe that $\langle x_n-u,x-x_n\rangle\ge 0$, $\forall x\in C_n$. This together with the fact that $x_{n+1}=P_{C_{n+1}}u\in C_{n+1}\subset C_n$, leads us to $\langle x_{n+1}-x_n,x_n-u\rangle\ge 0$. By using the above inequality, we find that

$$\begin{aligned}\|x_n-x_{n+1}\|^2=&\|x_{n+1}-x_u\|^2-\|x_n-u\|^2-2\langle x_n-u,x_{n+1}-x_n\rangle\\ \le&\|x_{n+1}-u\|^2-\|x_n-u\|^2.\end{aligned}\tag{2.10}$$

This yields that $\|x_{n+1}-u\|^2\ge\|x_n-u\|^2$, Since $\{x_n\}$ is bounded, we conclude that $\lim_{n\to\infty}\|x_n-u\|$ exists. From this, we obtain from (2.10) that

$$\lim_{n\to\infty}\|x_n-x_{n+1}\|=0.\tag{2.11}$$

Coming back to (2.4) and substituting $p=x_{n+1}$ into set C_{n+1}, one gets that

$$\begin{aligned}\|w_n-x_{n+1}\|^2\le&\|x_n-x_{n+1}\|^2+2\alpha_n\langle x_n-x_{n+1},x_n-x_{n-1}\rangle+\alpha_n^2\|x_n-x_{n-1}\|^2\\ &-(1-\kappa^2)\frac{1}{M}\sum_{m=1}^{M}\|y_n-z_{n,m}\|^2.\end{aligned}\tag{2.12}$$

By letting n tend to infinity in (2.12), one sees from (2.11) that

$$\begin{aligned}&\lim_{n\to\infty}\|w_n-x_{n+1}\|^2+(1-\kappa^2)\lim_{n\to\infty}\frac{1}{M}\sum_{m=1}^{M}\|y_n-z_{n,m}\|^2\\ &\le\lim_{n\to\infty}\|x_n-x_{n+1}\|^2+2\lim_{n\to\infty}\alpha_n\langle x_n-x_{n+1},x_n-x_{n-1}\rangle+\lim_{n\to\infty}\alpha_n^2\|x_n-x_{n-1}\|^2=0.\end{aligned}\tag{2.13}$$

Taking account of condition $\kappa\in(0,1)$, one infers that $1-\kappa^2$ is also in $(0,1)$. Recall that the above inequality and (2.13) lead us to

$$\lim_{n\to\infty}\|w_n-x_{n+1}\|=0,\tag{2.14}$$

and

$$\lim_{n\to\infty}\|y_n-z_{n,m}\|^2=0\ (m=1,2,\cdots,M).\tag{2.15}$$

Invoking (2.11), which together with the definition of y_n finds that

$$\lim_{n\to\infty}\|x_n-y_n\|=\lim_{n\to\infty}\alpha_n\|x_n-x_{n-1}\|=0.\tag{2.16}$$

By combining (2.15) with (2.16), one concludes that

$$\lim_{n\to\infty}\|x_n - z_{n,m}\| \leq \lim_{n\to\infty}(\|x_n - y_n\| + \|y_n - z_{n,m}\|) = 0. \tag{2.17}$$

Using (2.11), which together with (2.14) derives that

$$\lim_{n\to\infty}\|w_n - x_n\| \leq \lim_{n\to\infty}(\|w_n - x_{n+1}\| + \|x_{n+1} - x_n\|) = 0. \tag{2.18}$$

It directly follows from (2.11) that $\{x_n\}$ is a bounded vector sequence. So there exists a sequence $\{x_{n_i}\}$, which is indeed a subsequence of $\{x_n\}$, such that $\{x_{n_i}\}$ converges weakly to some $\hat{x} \in H$. Thanks to (2.16), (2.17), (2.18), one concludes that $\{y_{n_i}\}, \{z_{n_i,m}\}$ and $\{w_{n_i}\}$ also weakly converge to $\hat{x}$.

Next, one focus on $\hat{x} \in \Lambda$. Invoking (2.1), one derives that

$$\frac{y_n - z_{n,m}}{\gamma_{n,m}} \in A_m(z_{n,m}) + B_m(y_n), \tag{2.19}$$

which further implies that

$$\frac{y_n - z_{n,m}}{\gamma_{n,m}} + B_m(z_{n,m}) - B_m(y_n) \in A_m(z_{n,m}) + B_m(z_{n,m}) = G_m(z_{n,m}). \tag{2.20}$$

We consider the following two cases.

Case 1. Let $\liminf_{n\to\infty}\gamma_{n,m} > 0$. It follows from (2.2) and (2.15) that

$$\|B_m(y_n) - B_m(z_{n,m})\| \to 0, \text{ as } n \to \infty. \tag{2.21}$$

Hence, (2.15), (2.20) and (2.21) guarantees that the left-hand side of (2.20) converges strongly to 0. In view of the fact that $z_{n_i,m}$ converges weakly to $\hat{x}$, which together with Lemma 1.1 deduces that $0 \in G_m(\hat{x})$.

Case 2. Let B_m be continuous on $\operatorname{dom} B_m \supset \operatorname{dom} A_m$ and function $x \mapsto \min_\chi \|\chi\|$ be locally uniformly continuous on $\operatorname{dom} A_m$.

If there is a subsequence of $\{\gamma_{n,m}\}$ and the subsequence is bounded below by a positive value, one, with a similar discussion as above, deduces that $\hat{x} \in \Lambda$. Otherwise, one assumes $\gamma_{n,m} \to 0, n \to \infty$. For all $n \in \mathbb{N}$ sufficiently large, one finds that $\gamma_{n,m} < \rho$. The choice of $\gamma_{n,m}$ shows that (2.2) fails to hold if $\gamma = \gamma_{n,m}/\sigma$, i.e.,

$$\kappa\|y_n - \hat{z}_{n,m}\| < \hat{\gamma}_{n,m}\|B_m y_n - B_m\hat{z}_{n,m}\|. \tag{2.22}$$

where $\hat{\gamma}_{n,m} = \gamma_{n,m}/\sigma$ and

$$\hat{z}_{n,m} = (I + \hat{\gamma}_{n,m}A_m)^{-1}(y_n - \hat{\gamma}_{n,m}B_m y_n).$$

It follows that

$$\frac{y_n - \hat{z}_{n,m}}{\hat{\gamma}_{n,m}} \in A_m(\hat{z}_{n,m}) + B_m(y_n). \tag{2.23}$$

Hence,

$$\left\langle y_n - \hat{z}_{n,m}, \chi - \frac{y_n - \hat{z}_{n,m}}{\hat{\gamma}_{n,m}} \right\rangle \geq 0, \quad \forall \chi \in A_m(y_n) + B_m(y_n)$$

for each $m \in (1, 2, \cdots, r)$ and

$$\|y_n - \hat{z}_{n,m}\|^2 \leq \hat{\gamma}_{n,m}\langle \chi, y_n - \hat{z}_{n,m}\rangle \leq \hat{\gamma}_{n,m}\|\chi\|\|y_n - \hat{z}_{n,m}\|.$$

This indicates that

$$\frac{\|y_n - \hat{z}_{n,m}\|}{\hat{\gamma}_{n,m}} \leq \min_{\chi \in A_m y_n + B_m y_n}\|\chi\|. \tag{2.24}$$

Note that $x \mapsto \min_{\Phi \in G_m(x)}\|\Phi\|$ is locally bounded on $\operatorname{dom} A_m$. From the assumption on $\gamma_{n,m}$, one has $\hat{\gamma}_{n,m} = \gamma_{n,m}/\sigma \to 0$ $(n \to \infty)$. Accordingly,

$$\lim_{n\to\infty}\|y_n - \hat{z}_{n,m}\| = 0. \tag{2.25}$$

Since $\{y_{n_i}\}$ converges weakly to $\hat{x}$, we concludes that $\{\hat{z}_{n_i,m}\}$ converges weakly to $\hat{x}$, too. Borrowing (2.23), one finds

$$\frac{y_n - \hat{z}_{n,m}}{\hat{\gamma}_{n,m}} + B_m(\hat{z}_{n,m}) - B_m(y_n) \in A_m(\hat{z}_{n,m}) + B_m(\hat{z}_{n,m}) = G_m(\hat{z}_{n,m}). \tag{2.26}$$

Denote $\delta_{n,m} = \frac{y_n - \hat{z}_{n,m}}{\hat{\gamma}_{n,m}} + B_m(\hat{z}_{n,m}) - B_m(y_n)$. It immediately follows from (2.26) that $\delta_{n,m} \in G_m(\hat{z}_{n,m})$, $\forall n \in \mathbb{N}$.

(i) Assume $\liminf_{i\to\infty} \|\delta_{n_i,m}\| > 0$. Since B_m is continuous on $\operatorname{dom} B_m \supset \operatorname{dom} A_m$, one deduces that $\lim_{i\to\infty} \|B_m y_{n_i} - B_m \hat{z}_{n_i,m}\| = 0$, for each $m \in \{1, \cdots, r\}$. Using (2.25), one finds

$$\liminf_{i\to\infty} \frac{y_{n_i} - \hat{z}_{n_i,m}}{\hat{\gamma}_{n_i,m}} = \liminf_{i\to\infty} \|\delta_{n_i,m}\| > 0.$$

This contradicts the fact that (2.22) holds for all $n \in \mathbb{N}$ sufficiently large.

(ii) Otherwise, assume $\liminf_{i\to\infty} \|\delta_{n_i,m}\| = 0$. In view of $\hat{z}_{n_i,m}$ converges weakly to $\hat{x}$, we assert from Lemma 1.1 that $0 \in G_m \hat{x}$ $(m = 1, \cdots, r)$. Hence, $\hat{x} \in \Lambda$.

By putting together $\Lambda \subseteq C_{n_i}$ and $x_{n_i} = P_{C_{n_i}}(u)$, we conclude that

$$\|u - z\| \le \|u - \hat{x}\| \le \liminf_{i\to\infty} \|u - x_{n_i}\| \le \limsup_{i\to\infty} \|u - x_{n_i}\| \le \|u - z\|.$$

where $z = P_\Lambda u$. Thus, $\lim_{i\to\infty} \|u - x_{n_i}\| = \|u - \hat{x}\|$. From this and $u - x_{n_i} \rightharpoonup u - \hat{x}$ as $i \to \infty$, one concludes from Lemma 1.2 that $u - x_{n_i} \to u - \hat{x}$ as $i \to \infty$. Hence, $x_{n_i} \to \hat{x}$ as $i \to \infty$. In view of $x_{n_i} = P_{C_{n_i}}(u)$ and $z \in \Lambda \subseteq C_{n_i}$, $\forall n_i \in \mathbb{N}$, one has

$$\|z - x_{n_i}\|^2 = \langle u - x_{n_i}, z - x_{n_i} \rangle + \langle z - u, z - x_{n_i} \rangle \le \langle z - u, z - x_{n_i} \rangle. \tag{2.27}$$

Letting i tend to infinity in (2.27), and combining $z = P_\Lambda$ with $\hat{x} \in \Lambda$, one finds that

$$\|z - \hat{x}\|^2 \le \langle z - u, z - \hat{x} \rangle \le 0. \tag{2.28}$$

Therefore, $z = \hat{x}$. Since subsequence $\{x_{n_i}\}$ was arbitrarily chosen in $\{x_n\}$, one finds that $x_n \to z$. Coming back to (2.16) and (2.18), one concludes that $y_n \to z$ and $w_n \to z$ as $n \to \infty$. □

3. Numerical examples

In this section, we consider a computational experiment in support of the convergence of our proposed algorithm. Since set C_n in Algorithm (SISA) is either a half-space or the whole space H. All programs are written in Matlab version 5.0. and performed on a PC Desktop Intel(R) Core (TM) i5-8250U CPU @1.60GHz. We apply our proposed algorithm to solve the following convex feasibility problem.

Many problems in signal processing can be formulated as inverting a linear system of the form

$$\mu = \nu + Qx, \tag{3.1}$$

where $x \in \mathbb{R}^n$ is the 'true' and unknown signal vector to be estimated, $\mu \in \mathbb{R}^m$ is the vector of measurements or noisy observations, $\nu \in \mathbb{R}^m$ is the noise/perturbation vector of unknown, and the data matrix $Q \in \mathbb{R}^{m\times n}$ describes the blur operator, which is often ill conditioned. The problem of recovering x from the observed blurred and noisy signal μ is called an signal deblurring problem. A classical and efficient model for the linear problems of type (3.1) is the l_1-regularized least squares in which one seeks to find the sparse solutions of

$$\text{minmize}_{x\in\mathbb{R}^n} \frac{\|Qx - \mu\|_2^2 + 2\alpha\|x\|_1}{2}, \tag{3.2}$$

where $\alpha(\alpha \ge 0)$ provides a tradeoff between the fidelity to the measurements and the noise sensitivity. The solution set of problem (3.2) is always nonempty. The optimal solution tends to zero as $\alpha \to \infty$.

As $\alpha \to 0$, the limiting point has the minimum l_1 norm among all points that satisfy $Q^T(Qx-\mu)=0$, i.e., $x = \arg\min_{Q^T(Qx-\mu)=0} \|x\|_1$. The convergence occurs for a finite value of $\alpha(\alpha \geq \|Q^T\mu\|_\infty)$ for l_1-regularized least squares. One now uses a first-order optimality condition based on the subdifferential calculus. For $i=1,2,\cdots$, one can obtain the necessary and sufficient conditions for the optimal solution as follows

$$(Q^T(Qx-\mu))_i \in \begin{cases} \{-\alpha_i\}, & x_i>0; \\ \{+\alpha_i\}, & x_i<0; \\ [-\alpha_i,+\alpha_i], & x_i=0. \end{cases}$$

From above, one sees that the optimal solution of Problem (3.2) is 0, when $\alpha \geq \|Q^T\mu\|_\infty$. Thus one can now derive the formula

$$\alpha_{\max} = \|Q^T\mu\|_\infty. \tag{3.3}$$

By substituting $g(x)=\alpha\|x\|_1$ and $f(x)=\frac{\|Qx-\mu\|_2^2}{2}$, one can see that Problem (3.2) is a special case of problem

$$\text{find } x \in H \text{ such that } 0 \in (\partial g + \nabla f)x,$$

∂g and ∇f are maximal monotone and

$$\begin{cases} (I+\gamma_n\partial g)^{-1}x = (\operatorname{sgn}(x^1)\max\{|x^1|-\alpha\gamma_1,0\},\cdots,\operatorname{sgn}(x^n)\max\{|x^n|-\alpha\gamma_n,0\}), \\ (I-\gamma_n\nabla f)x = x-\gamma_n Q^T(Qx-\mu), \ \forall\, \gamma_n>0. \end{cases}$$

One now considers a general compressed sensing scenario, where the goal is to reconstruct a n-length sparse signal x with exactly m nonzero components from t observations. In this experiment, considering the storage limitation of the PC, one tests a small size signal with $n=2^8$ and the original signal contains $m=2^6$ randomly nonzero elements. One reconstructs this signal from $t=2^6$ observations.

More precisely, the observation $\mu = \nu + Qx$, where ν is the Gaussian noise distributed as $N(0,\sigma^2 I)$ with $\sigma^2 = 10^{-4}$ and Q is a $t\times n$ Gaussian matrix whose elements are randomly obtained from the standard normal distribution $N(0,1)$. Based on μ, the noisy measurement, our aim is to reconstruct x, the true signal. We restrict our attention to l_1-regularized least squares model (3.2). To avoid that the optimal sparse solution is a zero vector, the regularization parameter, in our experiment, is denoted by $\alpha = 0.01\alpha_{max}$, where the value of α_{max} is from (3.3). One randomly chooses $\{\alpha_i\}$ $(i=1,2,\cdots)$, the inertial parameter, in $(0,1)$. We take $\kappa=0.5$ $\rho=0.8$ and $\sigma=0.5$. One chooses the starting signals in the range of $(0,1)^{256}$ randomly. One takes the number of iterations 100 as the stopping criterion in the following experiment.

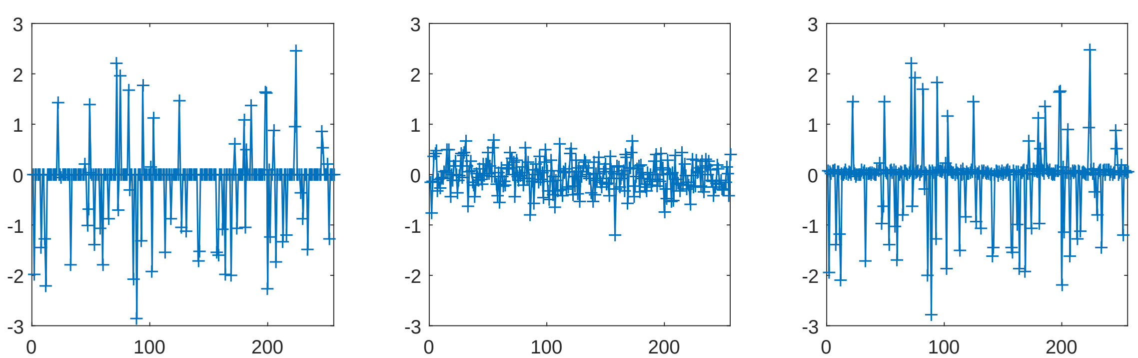

FIGURE 1. The original signal with length 2^8 and 2^6 nonzero elements (left). The minimum norm solution (middle). The reconstructed signal (right). The number of iterations is 100.

As depicted in Figure 1, one can find the result of this experiment for a signal sparse reconstruction. The minimum norm solution, in this experiment, is the point in $\{x \in \mathbb{R}^{256} \mid Q^T Qx = Q^T \mu\}$, which is the observed noisy signal. Comparing the left plot and the right plot in Figure 1, one finds that the original sparse signal is recovered almost exactly. That is, the reconstructed signal is closest to the original sparse signal. Our algorithm therefore provides an efficient and important approach to deal with the deblurring problem.

Acknowledgments
The author is grateful to the referees for useful suggestions, which improve the presentation of this paper.

REFERENCES

[1] F. Alvarez, Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space, SIAM J. Optim. 14 (2004), 773-782.
[2] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (2009), 183-202.
[3] P.L. Combettes, J. C. Pesquet, Stochastic forward-backward and primal-dual approximation algorithms with application to online image restoration, 24th European Signal Processing Conference (EUSIPCO), pp. 1813-1817, Budapest, 2016.
[4] S.S. Chang, C.F. Wen, J.C. Yao, Common zero point for a finite family of inclusion problems of accretive mappings in Banach spaces, Optimization 67 (2018), 1183-1196.
[5] H.G. Chen, R.T. Rockafellar, Convergence rates in forward-backward splitting, SIAM J. Optim. 7 (1997), 421-444.
[6] S.Y. Cho, X. Qin, L. Wang, Strong convergence of a splitting algorithm for treating monotone operators, Fixed Point Theory Appl. 2014 (2014), Article ID 94.
[7] S.Y. Cho, Implicit extragradient-like method for fixed point problems and variational inclusion problems in a Banach space, Symmetry, 12 (2020), 998.
[8] I. Cioranescu, Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems, Kluwer, Dordrecht, 1990.
[9] Q. Dong, Y.J. Cho, L.L. Zhong, T.M. Rassias, Inertial projection and contraction algorithms for variational inequalities, J. Global Optim. 70(2018), 687-704.
[10] A. Gibali, L.W. Liu, Y.C. Tang, Note on the modified relaxation CQ algorithm for the split feasibility problem, Optim. Lett. 12 (2018), 817-830.
[11] P. L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM J. Numer. Anal. 16 (1979), 964-979.
[12] D. A. Lorenz, T. Pock, An inertial forward-backward algorithm for monotone inclusions, J. Math. Imaging Vision 51 (2015), 311-325.
[13] Y. Luo, M. Shang, B. Tan, A general inertial viscosity type method for nonexpansive mappings and its applications in signal processing, Mathematics 8 (2020), Article ID 288.
[14] B. Martinet, Régularisation d'inéquations variationelles par approximations successives, Revue Française D'Informatique et de Recherche Opérationelle, 4 (1970), 154-158.
[15] B. Martinet, Détermination approchée d'un point fixe d'une application pseudo-contractante, C. R. Acad. Sci. Paris Ser. A-B, 274 (1972), 163-165.
[16] G.J. Minty, Monotone (nonlinear) operators in Hilbert space, Duke Math. J. 29 (1962), 341-345.
[17] J.J. Moreau, Proximité et dualité dans un espace Hilbertien, Bull. Soc. Math. France, 93 (1965), 273-299.
[18] J.J. Moreau, Proximité et dualité dans un espace hilbertien, Bull. Soc. Math. France, 93 (1965), 273-299.
[19] B.T. Polyak, Some methods of speeding up the convergence of iterative methods, Zh. Vychisl. Mat. Mat. Fiz. 4 (1964), 1-17.
[20] X. Qin, L. Wang, J.C. Yao, Inertial splitting method for maximal monotone mappings, J. Nonlinear Convex Anal. 21 (2020), 2325-2333.
[21] R.T. Rockafellar, Monotone operators and the proximal point algorithm, SIAM J. Control Optim. 14 (1976), 877-898.
[22] R.T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, Math. Oper. Res. 1 (1976), 97-116.
[23] D.R. Sahu, J.C. Yao, M. Verma, K.K. Shukla, Convergence rate analysis of proximal gradient methods with applications to composite minimization problems, Optimization, (2020), 10.1080/02331934.2019.1702040.

[24] D.R. Sahu, Q.H. Ansari, J.C. Yao, The prox-Tikhonov-like forward-backward method and applications, Taiwanese J. Math. 19 (2015), 481-503.

[25] Y. Shehu, P. Cholamjiak, Iterative method with inertial for variational inequalities in Hilbert spaces, Calcolo, 56 (2019), Article ID 4.

[26] Y. Shehu, O. S. Iyiola, Convergence analysis for the proximal split feasibility problem using an inertial extrapolation term method, J. Fixed Point Theory Appl. 19 (2017), 2483-2510.

[27] B. Tan, S. Xu, S. Li, Inertial hybrid and shrinking projection algorithms for solving variational inequality problems, J. Nonlinear Convex Anal. 21 (2020), 2193-2206.

[28] P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, SIAM J. Control Optim. 38 (2000), 431-446.

[29] M. Xu, C. Li, S. Zhang, P. Le Callet, State-of-the-art in 360 video/image processing: Perception, assessment and compression, IEEE J. Sel. Topics Signal Process. 14 (2020), 5-26.

[30] H.K. Xu, Iterative algorithms for nonlinear operators, J. Lond. Math. Soc. 66 (2002), 240-256.

[31] X. Zenggang, T. Zhiwen, C. Xiaowen, Z. Xue-min, Z. Kaibin, Y. Conghuan, Research on image retrieval algorithm based on combination of color and shape features, J. Sign. Process. Syst. (2019). https://doi.org/10.1007/s11265-019-01508-y.

[32] Z. Zhou, B. Tan, S. Li, An inertial shrinking projection algorithm for split common fixed point problems, J. Appl. Anal. Comput. 10 (2020), 2104-2120.

[33] W. Takahashi, Y. Takeuchi, R. Kubota, Strong convergence theorems by hybrid methods for families of nonexpansive mappings in Hilbert spaces, J. Math. Anal. Appl. 341 (2008), 276-286.