THE SUBGRADIENT EXTRAGRADIENT-TYPE ALGORITHMS FOR SOLVING A CLASS OF MONOTONE VARIATIONAL INCLUSION PROBLEMS

CUIJIE ZHANG*, JIAJIA CHEN

College of Science, Civil Aviation University of China, Tianjin 300300, China

Abstract. In this paper, we introduce two subgradient extragradient-type algorithms for finding a solution of a class of monotone variational inclusion problem. Weak and strong convergence theorems are established under some conditions. A numerical example is presented in order to illustrate the performance of our proposed algorithm. From the example, the effect of parameters on convergence speed is shown.

Keywords. Subgradient extragradient-type algorithm; Monotone variational inclusion problem; Maximally monotone operator; Resolvent.

1. INTRODUCTION

Let H be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. In this paper, we consider the monotone variational inclusion problem, which is to find a point $x^* \in H$ such that

$$0 \in A(x^*) + f(x^*), \tag{1.1}$$

where $A : H \to 2^H$ is a maximally monotone operator and $f : H \to H$ is a single-valued monotone operator. This problem plays an important role in many optimization-related areas, such as, mathematical programming, optimal control and many others; see, e.g., [1, 6, 17, 18] and the references therein. This problem was investigated by many researchers and a great deal of iterative algorithms were introduced and developed; see, for example, [4, 5, 19, 14, 25].

Let C be a nonempty closed convex subset of H. If A is the normal cone of C at $x \in C$ defined by

$$N_C(x) = \{d \in H | \langle d, y - x \rangle \leq 0, \forall y \in C\}. \tag{1.2}$$

Then, problem (1.1) is reduced to the problem, which consists of finding a point $x^* \in C$ such that

$$0 \in N_C(x^*) + f(x^*). \tag{1.3}$$

It is clear that problem (1.3) indeed is the classical variational inequality problem, which consists of finding a point $x^* \in C$ such that

$$\langle f(x^*), x - x^* \rangle \geq 0, \forall x \in C. \tag{1.4}$$

Among the algorithms for solving variational inequality (1.3), the extragradient method introduced by Korpelevich [16] is one of the simple and attractive methods. It reads as follows

$$\begin{cases} y_k = P_C(x_k - \lambda_k f(x_k)), \\ x_{k+1} = P_C(x_k - \lambda_k f(y_k)), \end{cases} \tag{1.5}$$

where $\lambda_k \in (0, \frac{1}{L})$, L is the Lipschitz constant of f, or λ_k is updated by some adaptive rule, for example, $\lambda_k \|f(x_k) - f(y_k)\| \leq v \|x_k - y_k\|$, where $v \in (0,1)$. In recent years, the extragradient method has received a great deal of attention from many authors, who improved this method in various ways (see, e.g., [8, 9, 10, 12, 13, 22] and references therein).

However, if set C is not simple enough to calculate, then it affects the computation effort and complexity of the method. The projection and contraction method of He [14] and Sun [20] was introduced to overcome this obstacle recently. Cai, Gu and He [3] gave the "optimal" step size in the contraction sense. On the other hand, Censor, Gibali and Reich [7] modified the extragradient algorithm and gave the subgradient extragradient method. In their algorithm, the second projection onto C is replaced by a projection onto a half-space which constraint C. The projection onto a half-space is easier to compute. The subgradient extragradient method is of the form:

$$\begin{cases} y_k = P_C(x_k - \lambda f(x_k)), \\ T_k = \{\omega \in H | \langle x_k - \lambda f(x_k) - y_k, \omega - y_k \rangle \leq 0\}, \\ x_{k+1} = P_{T_k}(x_k - \lambda f(y_k)), \end{cases} \tag{1.6}$$

where $\lambda \in (0, \frac{1}{L})$.

Recently, Dong, Jiang and Gibali [11] further introduced a modified subgradient extragradient method by improving the stepsize of the second step. Their algorithm reads as follows.

Algorithm 1 Dong-Jiang-Gibali modified subgradient extragradient algorithm
---

Initialization: Give $\tau > 0$, $\ell \in (0,1)$, $v \in (0,1)$ and $\gamma \in (0,2)$. Let $x_0 \in H$ be arbitrary.
Iterative Steps: Calculate x_{k+1} as follows:
    Step 1. Compute $y_k = P_C(x_k - \lambda_k f(x_k))$, where λ_k is chosen to be the largest $\lambda \in \{\tau, \tau\ell, \tau\ell^2, \ldots\}$ satisfying

$$\lambda \|f(x_k) - f(y_k)\| \leq v \|x_k - y_k\|.$$

If $x_k - y_k = 0$, then STOP.
    Step 2. Compute

$$x_{k+1} = P_{T_k}(x_k - \gamma\eta_k\lambda_k f(y_k)),$$

where

$$T_k := \{\omega \in H | \langle x_k - \lambda_k f(x_k) - y_k, \omega - y_k \rangle \leq 0\},$$

and

$$\eta_k := \frac{\langle x_k - y_k, d(x_k, y_k) \rangle}{\|d(x_k, y_k)\|^2},$$

where

$$d(x_k, y_k) := x_k - y_k - \lambda_k(f(x_k) - f(y_k)).$$

    Step 3. Set $k \leftarrow k+1$, and go to **Step 1**.

Motivated and inspired by the above works, two natural questions which raise are: can the subgradient extragradient method and the modified subgradient extragradient method be used to solve problem (1.1)? Can one devise their algorithms and obtain the convergence of the algorithms? In order to answer these questions, we introduce two subgradient extragradient-type algorithms for solving monotone variational inclusion problem (1.1). The weak convergence and the strong convergence are obtained, respectively.

The paper is organized as follows. We collect some necessary definitions and lemmas for the convergence analysis in Section 2. We introduce two subgradient extragradient-type algorithms and prove their convergence in Section 3. We give a numerical example in Section 4, the last section.

2. PRELIMINARIES

Throughout this section, 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. Let T be an operator. We denote the fixed point set of T by $Fix(T)$. For each $x,y,z \in H$ and $\alpha, \beta, \gamma \in [0,1]$ with $\alpha + \beta + \gamma = 1$, we have

$$\|x+y\|^2 \le \|x\|^2 + 2\langle y, x+y \rangle, \tag{2.1}$$

and

$$\|\alpha x + \beta y + \gamma z\|^2 = \alpha \|x\|^2 + \beta \|y\|^2 + \gamma \|z\|^2 - \alpha\beta \|x-y\|^2 - \alpha\gamma \|x-z\|^2 - \beta\gamma \|y-z\|^2. \tag{2.2}$$

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

$$\|Tx - Ty\| \le L\|x - y\|, \tag{2.3}$$

where $L > 0$ is Lipschitz constant;
• `nonexpansive` iff $L \equiv 1$ in (2.3);
• `firmly nonexpansive` iff $2T - I$ is nonexpansive, or equivalently, for all $x, y \in H$,

$$\langle Tx - Ty, x - y \rangle \ge \|Tx - Ty\|^2. \tag{2.4}$$

It is clear that firm nonexpansiveness implies nonexpansiveness. If T is firmly nonexpansive and $Fix(T) \ne \emptyset$, we have that

$$\langle x - Tx, Tx - y \rangle \ge 0 \text{ for all } x \in H, \ y \in Fix(T). \tag{2.5}$$

For every $x \in H$, there exists a unique nearest point in C, denoted by $P_C x$, such that, for every $y \in C$, $\|x - P_C x\| \le \|x - y\|$. P_C is called the metric projection of H onto C. It is well-known that P_C is firmly nonexpansive. It is easy to prove that, for every $y \in C$,

$$\|P_C x - y\|^2 \le \|x - y\|^2 - \|x - P_C x\|^2. \tag{2.6}$$

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

$$\langle Tx - Ty, x - y \rangle \ge 0. \tag{2.7}$$

• β-`inverse strongly monotone`(or β-`cocoercive`) iff for all $x, y \in H$,

$$\langle Tx - Ty, x - y \rangle \ge \beta \|Tx - Ty\|^2. \tag{2.8}$$

A multi-valued operator $A : H \to 2^H$ is said to be
- `monotone` iff
$$\langle u - v, x - y \rangle \geq 0 \text{ whenever } u \in A(x), \ v \in A(y). \tag{2.9}$$
- `maximally monotone` iff, in addition, its graph
$$G(A) := \{(x, y) \in H \times H : y \in A(x)\} \tag{2.10}$$
is not properly contained in the graph of any other monotone operator.

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

Lemma 2.1. [27] *Let $A : H \to 2^H$ be a maximally monotone operator and let $f : H \to H$ be a Lipschitz continuous operator. Then the operator $B = A + f$ is a maximally monotone operator.*

Lemma 2.2. [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.

From [21, Lemma 3.4], we have the following Lemma immediately.

Lemma 2.3. *Let $A : H \to 2^H$ be a maximally monotone operator and $J_\lambda^A = (I + \lambda A)^{-1}$ be the resolvent of A for $\lambda > 0$. Let $f : H \to H$ be a β-inverse strongly monotone with $\beta > 0$. Suppose that $A^{-1}(0) \cap f^{-1}(0) \neq \emptyset$. Let $\rho > 0$ and $u \in H$. Then the following are equivalent:*

    (i) $u = J_\lambda^A(I - \rho f)u$;

    (ii) $u \in A^{-1}(0) \cap f^{-1}(0)$;

    (iii) $0 \in A(u) + f(u)$.

Consequently, $\mathrm{Fix}(J_\lambda^A(I - \rho f)) = (A + f)^{-1}(0) = A^{-1}(0) \cap f^{-1}(0)$.

Proof. Since $A^{-1}(0) \cap f^{-1}(0) \neq \emptyset$, there exists u_0 such that $0 \in A(u_0)$ and $f(u_0) = 0$.

$(i) \Rightarrow (ii)$. Assume that $u = J_\lambda^A(I - \rho f)u$. Using the definition of the resolvent operator, we have
$$u - \rho f(u) \in (I + \lambda A)u,$$
which follows that
$$-\frac{\rho}{\lambda} f(u) \in A(u). \tag{2.11}$$
Since A is monotone and $0 \in A(u_0)$, we obtain
$$\left\langle -\frac{\rho}{\lambda} f(u), u - u_0 \right\rangle \geq 0.$$
Then
$$\langle f(u), u - u_0 \rangle \leq 0. \tag{2.12}$$
On the other hand, since f is β-inverse strongly monotone and $f(u_0) = 0$, we get
$$\langle f(u), u - u_0 \rangle = \langle f(u) - f(u_0), u - u_0 \rangle \geq \beta \|f(u) - f(u_0)\|^2 = \beta \|f(u)\|^2 \geq 0. \tag{2.13}$$
Combining (2.12) and (2.13), we have $f(u) = 0$. From (2.11), we have $0 \in A(u)$. Therefore, $u \in A^{-1}(0) \cap f^{-1}(0)$.

$(ii) \Rightarrow (iii)$. It is obvious.

$(iii) \Rightarrow (i)$. The assumption $0 \in A(u) + f(u)$ can be rewritten as $-f(u) \in A(u)$. Note that $0 \in A(u_0)$, the monotonicity of A implies that $\langle f(u), u - u_0 \rangle \leq 0$. From (2.13), we get $f(u) = 0$. It follows that $0 \in A(u)$. Then $u = J_\lambda^A u$ and

$$J_\lambda^A (I - \rho f)u = J_\lambda^A u = u.$$

The proof is complete. $\qquad\square$

Lemma 2.4. [15] *Let $\{s_k\}$ be a sequence of nonnegative real numbers such that*

$$s_{k+1} \leq (1 - \alpha_k)s_k + \alpha_k \delta_k, k \geq 0,$$

$$s_{k+1} \leq s_k - \tau_k + \zeta_k, k \geq 0,$$

where $\{\alpha_k\}$ is a sequence in $(0, 1)$, $\{\tau_k\}$ is a sequence of nonnegative real numbers, and $\{\delta_k\}$ and $\{\zeta_k\}$ are two sequences in $\mathbb{R}$ such that

(i) $\Sigma_{k=0}^\infty \alpha_k = \infty$;
(ii) $\lim_{k \to \infty} \zeta_k = 0$;
(iii) $\lim_{j \to \infty} \tau_{k_j} = 0$ *implies* $\limsup_{j \to \infty} \delta_{k_j} \leq 0$.

Then $\lim_{k \to \infty} s_k = 0$.

3. ALGORITHMS AND THE CONVERGENCE ANALYSIS

In this section, we introduce two self-adaptive subgradient extragradient-type algorithms for solving the monotone variational inclusion problem. The convergence of two algorithms is analyzed in Hilbert spaces.

3.1. **Weak convergence algorithm and convergence analysis.** In this subsection, we give a subgradient extragradient-type algorithm and establish its weak convergence. We first state the assumptions that we will use through the rest of this paper.

Condition (i) The solution set of (1.1), denoted by Γ, and $A^{-1}(0) \cap f^{-1}(0)$ are nonempty.

Condition (ii) The operator f is β-inverse strongly monotone on H with $\beta > 0$.

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

First, we give a self-adaptive subgradient extragradient-type algorithm. The algorithm is of the form:

Definition 3.1. Let c_1, $c_2 > 0$ be given constants in $(0, 1)$. λ_k is said to satisfy the `stepsize` `conditions` in Algorithm 2 if λ_k satisfies

$$\langle x_k - y_k, d(x_k, y_k) \rangle \geq c_1 \|x_k - y_k\|^2, \tag{3.3}$$

$$\eta_k \geq c_2, \tag{3.4}$$

and

$$\inf_{k \geq 0} \{\lambda_k\} \geq \underline{\lambda} > 0. \tag{3.5}$$

Algorithm 2 The weakly convergent subgradient extragradient-type algorithm

---

Initialization: Give $\gamma \in (0,2)$. Let $x_0 \in H$ be arbitrary.

Iterative Steps: Calculate x_{k+1} as follows:

   Step 1: Compute

$$y_k = J_{\lambda_k}^A(x_k - \lambda_k f(x_k)). \tag{3.1}$$

If $x_k - y_k = 0$: STOP.

   Step 2: Compute

$$x_{k+1} = P_{T_k}(x_k - \gamma \eta_k \lambda_k f(y_k)), \tag{3.2}$$

where

$$T_k := \{\omega \in H \mid \langle x_k - \lambda_k f(x_k) - y_k, \omega - y_k \rangle \le 0\},$$

and

$$\eta_k := \frac{\langle x_k - y_k, d(x_k, y_k) \rangle}{\|d(x_k, y_k)\|^2},$$

where

$$d(x_k, y_k) := x_k - y_k - \lambda_k(f(x_k) - f(y_k)).$$

   Step 3: Set $k \leftarrow k+1$, and go to **Step 1**.

Lemma 3.1. [23] *(i) We know that $d(x_k, y_k) = 0$ if and only if $x_k = y_k$. Hence if $d(x_k, y_k) = 0$, then $x_k \in \Gamma$. (ii) Assume that the sequence $\{x_k\}$ is generated by Algorithm 2. 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*

$$\lambda_k \|f(x_k) - f(y_k)\| \le \nu \|x_k - y_k\|, \tag{3.6}$$

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

Lemma 3.2. *Assume that $\{x_k\}$ and $\{y_k\}$ are the sequences generated by Algorithm 2. Let $u \in \Gamma$ and $\{\lambda_k\}$ satisfy the stepsize conditions. Then, under Conditions (i), (ii) and (iii), we have*

$$\|x_{k+1} - u\|^2 \le \|x_k - u\|^2 - \|x_k - x_{k+1} - \gamma \eta_k d(x_k, y_k)\|^2 - c_1 c_2 \gamma(2-\gamma)\|x_k - y_k\|^2. \tag{3.7}$$

Proof. Let $u \in \Gamma$. From Lemma 2.3, we have $0 \in A(u)$ and $f(u) = 0$. From (3.1), we have

$$x_k - \lambda_k f(x_k) \in y_k + \lambda_k A(y_k),$$

which follows that

$$\frac{x_k - y_k}{\lambda_k} - f(x_k) \in A(y_k). \tag{3.8}$$

By virtue of the maximal monotonicity of A, we have

$$\langle \frac{x_k - y_k}{\lambda_k} - f(x_k), y_k - u \rangle \ge 0.$$

Thus, it follows from $\lambda_k > 0$ that

$$\langle x_k - y_k - \lambda_k f(x_k), u - y_k \rangle \le 0. \tag{3.9}$$

Hence $u \in T_k$. From the definition of x_{k+1} and (2.6), we get

$$
\begin{aligned}
\|x_{k+1} - u\|^2 &= \|P_{T_k}(x_k - \gamma\eta_k\lambda_k f(y_k)) - u\|^2 \\
&\leq \|x_k - \gamma\eta_k\lambda_k f(y_k) - u\|^2 - \|x_k - \gamma\eta_k\lambda_k f(y_k) - x_{k+1}\|^2. \\
&= \|x_k - u\|^2 - \|x_k - x_{k+1}\|^2 - 2\langle \gamma\eta_k\lambda_k f(y_k), x_{k+1} - u\rangle.
\end{aligned}
\tag{3.10}
$$

From $f(u) = 0$, the fact that f is monotone, we get $\langle f(y_k), y_k - u\rangle \geq 0$, which further implies that

$$
\begin{aligned}
\langle f(y_k), x_{k+1} - u\rangle &= \langle f(y_k), x_{k+1} - y_k\rangle + \langle f(y_k), y_k - u\rangle \\
&\geq \langle f(y_k), x_{k+1} - y_k\rangle.
\end{aligned}
\tag{3.11}
$$

Then

$$
-2\gamma\eta_k\lambda_k\langle f(y_k), x_{k+1} - u\rangle \leq -2\gamma\eta_k\lambda_k\langle f(y_k), x_{k+1} - y_k\rangle.
\tag{3.12}
$$

On the other hand, from (3.2), we have $x_{k+1} \in T_k$, which implies that

$$
\langle x_k - \lambda_k f(x_k) - y_k, x_{k+1} - y_k\rangle \leq 0.
\tag{3.13}
$$

It follows that

$$
\langle x_k - y_k - \lambda_k(f(x_k) - f(y_k)), x_{k+1} - y_k\rangle \leq \lambda_k\langle f(y_k), x_{k+1} - y_k\rangle,
$$

and thus

$$
\langle d(x_k, y_k), x_{k+1} - y_k\rangle \leq \lambda_k\langle f(y_k), x_{k+1} - y_k\rangle.
$$

Note that

$$
-2\gamma\eta_k\lambda_k\langle f(y_k), x_{k+1} - y_k\rangle \leq -2\gamma\eta_k\langle d(x_k, y_k), x_{k+1} - y_k\rangle.
\tag{3.14}
$$

Combining (3.12) and (3.14), we arrive at

$$
\begin{aligned}
&-2\gamma\eta_k\lambda_k\langle f(y_k), x_{k+1} - u\rangle \\
&\leq -2\gamma\eta_k\langle d(x_k, y_k), x_{k+1} - y_k\rangle \\
&= -2\gamma\eta_k\langle d(x_k, y_k), x_k - y_k\rangle + 2\gamma\eta_k\langle d(x_k, y_k), x_k - x_{k+1}\rangle \\
&= -2\gamma\eta_k^2\|d(x_k, y_k)\|^2 + 2\gamma\eta_k\langle d(x_k, y_k), x_k - x_{k+1}\rangle.
\end{aligned}
\tag{3.15}
$$

Now, we estimate $2\gamma\eta_k\langle d(x_k, y_k), x_k - x_{k+1}\rangle$. Observe that

$$
\begin{aligned}
&2\gamma\eta_k\langle d(x_k, y_k), x_k - x_{k+1}\rangle \\
&= -\|x_k - x_{k+1} - \gamma\eta_k d(x_k, y_k)\|^2 + \|x_k - x_{k+1}\|^2 + \gamma^2\eta_k^2\|d(x_k, y_k)\|^2.
\end{aligned}
\tag{3.16}
$$

It implies from (3.15) and (3.16) that

$$
\begin{aligned}
&-2\gamma\eta_k\lambda_k\langle f(y_k), x_{k+1} - u\rangle \\
&\leq -2\gamma\eta_k^2\|d(x_k, y_k)\|^2 - \|x_k - x_{k+1} - \gamma\eta_k d(x_k, y_k)\|^2 + \|x_k - x_{k+1}\|^2 + \gamma^2\eta_k^2\|d(x_k, y_k)\|^2 \\
&\leq -\|x_k - x_{k+1} - \gamma\eta_k d(x_k, y_k)\|^2 + \|x_k - x_{k+1}\|^2 - \gamma(2 - \gamma)\eta_k^2\|d(x_k, y_k)\|^2.
\end{aligned}
\tag{3.17}
$$

Substituting (3.17) into (3.10), we have

$$
\|x_{k+1} - u\|^2 \leq \|x_k - u\|^2 - \|x_k - x_{k+1} - \gamma\eta_k d(x_k, y_k)\|^2 - \gamma(2 - \gamma)\eta_k^2\|d(x_k, y_k)\|^2.
\tag{3.18}
$$

Combining (3.3) and (3.4), we obtain

$$
\eta_k^2\|d(x_k, y_k)\|^2 = \eta_k\langle d(x_k, y_k), x_k - y_k\rangle \geq c_1 c_2\|x_k - y_k\|^2.
\tag{3.19}
$$

From (3.18) and (3.19), we obtain (3.7). The proof is completed. $\square$

Theorem 3.1. *Assume that Conditions* (i)*,* (ii) *and* (iii) *hold and* $\{\lambda_k\}$ *satisfies the stepsize conditions. Then, the sequences* $\{x_k\}$ *and* $\{y_k\}$ *generated by Algorithm* 2 *weakly converge to the same solution of variational inclusion problem* (1.1).

Proof. Let $u \in \Gamma$. From (3.7), we have

$$\|x_{k+1} - u\| \leq \|x_k - u\|.$$

It implies that the sequence $\{\|x_k - u\|\}$ is monotonically decreasing and bounded from below. Hence $\lim_{k \to \infty} \|x_k - u\|$ exists and $\{x_k\}$ is bounded. From $0 < \gamma < 2$ and (3.7), we get

$$\|x_k - y_k\|^2 \leq \frac{1}{c_1 c_2 \gamma (2 - \gamma)} (\|x_k - u\|^2 - \|x_{k+1} - u\|^2). \tag{3.20}$$

It follows that $\lim_{k \to \infty} \|x_k - y_k\| = 0$ and $\{y_k\}$ is bounded. 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 $y_{k_j} \rightharpoonup \hat{x}$ as $j \to \infty$.

Now, we show that $\hat{x}$ is a solution of problem (1.1), that is, $\hat{x} \in \Gamma$. We observe that the operator f is Lipschitz continuous. From Lemma 2.1, we know that $A + f$ is maximally monotone. Let $(v, w) \in G(A + f)$, that is, $w - f(v) \in A(v)$. Since $y_{k_j} = J^A_{\lambda_{k_j}}(x_{k_j} - \lambda_{k_j} f(x_{k_j}))$, we have

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

that is,

$$\frac{x_{k_j} - y_{k_j}}{\lambda_{k_j}} - f(x_{k_j}) \in Ay_{k_j}. \tag{3.21}$$

By virtue of the maximal monotonicity of A, we have

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

Hence

$$\langle v - y_{k_j}, w \rangle \geq \langle v - y_{k_j}, f(v) + \frac{x_{k_j} - y_{k_j}}{\lambda_{k_j}} - f(x_{k_j}) \rangle$$

$$= \langle v - y_{k_j}, f(v) - f(y_{k_j}) + f(y_{k_j}) - f(x_{k_j}) + \frac{x_{k_j} - y_{k_j}}{\lambda_{k_j}} \rangle$$

$$\geq \langle v - y_{k_j}, f(y_{k_j}) - f(x_{k_j}) \rangle + \langle v - y_{k_j}, \frac{x_{k_j} - y_{k_j}}{\lambda_{k_j}} \rangle.$$

Since $\lim_{k \to \infty} \|y_k - x_k\| = 0$, and f is Lipschitz continuous, we obtain that $\lim_{j \to \infty} \|f(y_{k_j}) - f(x_{k_j})\| = 0$. Since $\lambda_k \geq \lambda > 0$, 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, $\hat{x} \in \Gamma$. Using Lemma 2.2, we find that the $\{x_k\}$ and $\{y_k\}$ converge weakly to a point in Γ. This completes the proof. $\qquad \square$

Remark 3.1. If $\gamma \eta_k \equiv 1$, Algorithm 2 is reduced to the follow algorithm:

$$\begin{cases} y_k = J^A_{\lambda_k}(x_k - \lambda_k f(x_k)), \\ x_{k+1} = P_{T_k}(x_k - \lambda_k f(y_k)). \end{cases} \tag{3.22}$$

where $T_k := \{\omega \in H | \langle x_k - \lambda_k f(x_k) - y_k, \omega - y_k \rangle \le 0\}$. By using the similar method, we can obtain the the weak convergence result with $c_2 \in (\frac{1}{2}, 1)$.

3.2. Strong convergence algorithm and convergence analysis.
In this subsection, we give the other subgradient extragradient-type algorithm. Different conditions is necessary to establish the strong convergence.

Condition(iv) Let $\{\alpha_k\}$ and $\{\beta_k\}$ be two real sequences in (0,1) such that $\{\beta_k\} \subset (a,b) \subset (0, 1 - \alpha_k)$ for some $a > 0, b > 0$ and

$$\lim_{k \to \infty} \alpha_k = 0, \sum_{k=1}^{\infty} \alpha_k = \infty. \tag{3.23}$$

Algorithm 3 The strongly convergent subgradient extragradient-type algorithm

Initialization: Give $\gamma \in (0,2)$. Let $x_0 \in H$ be arbitrary.
Iterative Steps: Calculate x_{k+1} as follows:
   Step 1: Compute

$$y_k = J_{\lambda_k}^A (x_k - \lambda_k f(x_k)), \tag{3.24}$$

If $x_k - y_k = 0$, then STOP.
   Step 2: Compute

$$z_k = P_{T_k}(x_k - \gamma \eta_k \lambda_k f(y_k)), \tag{3.25}$$

where $T_k := \{\omega \in H | \langle x_k - \lambda_k f(x_k) - y_k, \omega - y_k \rangle \le 0\}$.
   and

$$\eta_k := \frac{\langle x_k - y_k, d(x_k, y_k) \rangle}{\|d(x_k, y_k)\|^2},$$

where

$$d(x_k, y_k) := x_k - y_k - \lambda_k(f(x_k) - f(y_k)).$$

   Step 3: Compute

$$x_{k+1} = (1 - \alpha_k - \beta_k)x_k + \beta_k z_k. \tag{3.26}$$

Step 4: Set $k \leftarrow k + 1$, and go to **Step 1**.

Theorem 3.2. *Assume that Conditions (i), (ii), (iii) and (iv) hold and λ_k satisfies the stepsize conditions. The sequence $\{x_k\}$ generated by Algorithm 3 converges strongly to $\hat{u} \in \Gamma$, where $\|\hat{u}\| = min\{\|u\| \, |u \in \Gamma\}$.*

Proof. From Lemma 3.2, we have

$$\|z_k - \hat{u}\|^2 \le \|x_k - \hat{u}\|^2 - \|x_k - z_k - \gamma \eta_k d(x_k, y_k)\|^2 - c_1 c_2 \gamma(2 - \gamma)\|x_k - y_k\|^2, \tag{3.27}$$

and

$$\|z_k - \hat{u}\| \le \|x_k - \hat{u}\|. \tag{3.28}$$

First, we prove $\{x_k\}$ is bounded. Using (3.28), we get

$$
\begin{aligned}
\|x_{k+1} - \hat{u}\| &= \|(1 - \alpha_k - \beta_k)(x_k - \hat{u}) + \beta_k(z_k - \hat{u}) - \alpha_k \hat{u}\| \\
&\leq \|(1 - \alpha_k - \beta_k)(x_k - \hat{u}) + \beta_k(z_k - \hat{u})\| + \alpha_k \|\hat{u}\| \\
&\leq (1 - \alpha_k - \beta_k)\|x_k - \hat{u}\| + \beta_k \|z_k - \hat{u}\| + \alpha_k \|\hat{u}\| \\
&\leq (1 - \alpha_k)\|x_k - \hat{u}\| + \alpha_k \|\hat{u}\| \\
&\leq \max\{\|x_k - \hat{u}\|, \|\hat{u}\|\} \\
&\leq \ldots \leq \max\{\|x_0 - \hat{u}\|, \|\hat{u}\|\}.
\end{aligned}
\tag{3.29}
$$

This proves that $\{x_k\}$ is bounded, so is $\{z_k\}$.

Second, setting $t_k = (1 - \beta_k)x_k + \beta_k z_k$, we have

$$
\begin{aligned}
\|t_k - \hat{u}\| &\leq (1 - \beta_k)\|x_k - \hat{u}\| + \beta_k \|z_k - \hat{u}\| \\
&\leq (1 - \beta_k)\|x_k - \hat{u}\| + \beta_k \|x_k - \hat{u}\| \\
&= \|x_k - \hat{u}\|,
\end{aligned}
\tag{3.30}
$$

and

$$
\|t_k - x_k\| = \beta_k \|x_k - z_k\|.
\tag{3.31}
$$

From (3.30), (3.31) and (2.1), we have

$$
\begin{aligned}
\|x_{k+1} - \hat{u}\|^2 &= \|(1 - \alpha_k)(t_k - \hat{u}) - \alpha_k(x_k - t_k) - \alpha_k \hat{u}\|^2 \\
&\leq (1 - \alpha_k)^2 \|t_k - \hat{u}\|^2 - 2\langle \alpha_k(x_k - t_k) + \alpha_k \hat{u}, x_{k+1} - \hat{u}\rangle \\
&= (1 - \alpha_k)^2 \|t_k - \hat{u}\|^2 + 2\langle \alpha_k(x_k - t_k), \hat{u} - x_{k+1}\rangle + 2\langle \alpha_k \hat{u}, \hat{u} - x_{k+1}\rangle \\
&\leq (1 - \alpha_k)\|x_k - \hat{u}\|^2 + 2\alpha_k \|x_k - t_k\|\|\hat{u} - x_{k+1}\| + 2\langle \alpha_k \hat{u}, \hat{u} - x_{k+1}\rangle \\
&\leq (1 - \alpha_k)\|x_k - \hat{u}\|^2 + \alpha_k(2\beta_k \|x_k - z_k\|\|\hat{u} - x_{k+1}\| + 2\langle \hat{u}, \hat{u} - x_{k+1}\rangle).
\end{aligned}
\tag{3.32}
$$

On the other hand, from (3.2) and (2.2), we have

$$
\begin{aligned}
\|x_{k+1} - \hat{u}\|^2 &= \|(1 - \alpha_k - \beta_k)(x_k - \hat{u}) + \beta_k(z_k - \hat{u}) + \alpha_k(-\hat{u})\|^2 \\
&= (1 - \alpha_k - \beta_k)\|x_k - \hat{u}\|^2 + \beta_k \|z_k - \hat{u}\|^2 + \alpha_k \|\hat{u}\|^2 \\
&\quad - \beta_k(1 - \alpha_k - \beta_k)\|x_k - z_k\|^2 \\
&\quad - \alpha_k(1 - \alpha_k - \beta_k)\|x_k\|^2 - \alpha_k \beta_k \|z_k\|^2 \\
&\leq (1 - \alpha_k - \beta_k)\|x_k - \hat{u}\|^2 + \beta_k \|z_k - \hat{u}\|^2 + \alpha_k \|\hat{u}\|^2.
\end{aligned}
\tag{3.33}
$$

Combining (3.27) and (3.33), we obtain

$$
\begin{aligned}
\|x_{k+1} - \hat{u}\|^2 &\leq (1 - \alpha_k - \beta_k)\|x_k - \hat{u}\|^2 + \beta_k \|x_k - \hat{u}\|^2 - \beta_k \|x_k - z_k - \gamma \eta_k d(x_k, y_k)\|^2 \\
&\quad - \beta_k c_1 c_2 \gamma(2 - \gamma)\|x_k - y_k\|^2 + \alpha_k \|\hat{u}\|^2 \\
&\leq \|x_k - \hat{u}\|^2 - \beta_k \|x_k - z_k - \gamma \eta_k d(x_k, y_k)\|^2 - \beta_k c_1 c_2 \gamma(2 - \gamma)\|x_k - y_k\|^2 + \alpha_k \|\hat{u}\|^2.
\end{aligned}
\tag{3.34}
$$

From Condition (iv), we have

$$
\begin{aligned}
\|x_{k+1} - \hat{u}\|^2 &\leq \|x_k - \hat{u}\|^2 - a\|x_k - z_k - \gamma \eta_k d(x_k, y_k)\|^2 \\
&\quad - a c_1 c_2 \gamma(2 - \gamma)\|x_k - y_k\|^2 + \alpha_k \|\hat{u}\|^2.
\end{aligned}
\tag{3.35}
$$

Setting

$$s_k = \|x_k - \hat{u}\|^2, \quad \delta_k = 2\beta_k\|x_k - z_k\|\|\hat{u} - x_{k+1}\| + 2\langle \hat{u}, \hat{u} - x_{k+1}\rangle,$$

$$\tau_k = a\|x_k - z_k - \gamma\eta_k d(x_k, y_k)\|^2 + ac_1c_2\gamma(2 - \gamma)\|x_k - y_k\|^2, \zeta_k = \alpha_k\|\hat{u}\|^2,$$

we find that (3.32) and (3.35) can be rewritten as the following, respectively:

$$s_{k+1} \leq (1 - \alpha_k)s_k + \alpha_k\delta_k,$$

and

$$s_{k+1} \leq s_k - \tau_k + \zeta_k.$$

Finally, from conditions (iv), we know that $\lim_{k\to\infty}\zeta_k = 0$ and $\Sigma_{k=1}^{\infty}\alpha_k = \infty$. In order to complete the proof by using Lemma 2.4, it suffices to verify the fact that $\lim_{j\to\infty}\tau_{k_j} = 0$ implies $\limsup_{j\to\infty}\delta_{k_j} \leq 0$ for any subsequence $\{k_j\} \subset \{k\}$. In fact, if $\lim_{j\to\infty}\tau_{k_j} = 0$, then

$$\lim_{j\to\infty}\|x_{k_j} - y_{k_j}\| = 0 \quad \text{and} \quad \lim_{j\to\infty}\|x_{k_j} - z_{k_j} - \gamma\eta_{k_j}d(x_{k_j}, y_{k_j})\| = 0.$$

Since

$$\begin{aligned}
\|x_{k_j} - z_{k_j}\| &\leq \|x_{k_j} - z_{k_j} - \gamma\eta_{k_j}d(x_{k_j}, y_{k_j})\| + \gamma\eta_{k_j}\|d(x_{k_j}, y_{k_j})\| \\
&= \|x_{k_j} - z_{k_j} - \gamma\eta_{k_j}d(x_{k_j}, y_{k_j})\| + \gamma\frac{\langle x_{k_j} - y_{k_j}, d(x_{k_j}, y_{k_j})\rangle}{\|d(x_{k_j}, y_{k_j})\|} \\
&\leq \|x_{k_j} - z_{k_j} - \gamma\eta_{k_j}d(x_{k_j}, y_{k_j})\| + \gamma\frac{\|x_{k_j} - y_{k_j}\|\|d(x_{k_j}, y_{k_j})\|}{\|d(x_{k_j}, y_{k_j})\|} \\
&= \|x_{k_j} - z_{k_j} - \gamma\eta_{k_j}d(x_{k_j}, y_{k_j})\| + \gamma\|x_{k_j} - y_{k_j}\|,
\end{aligned} \tag{3.36}$$

it follows that $\lim_{j\to\infty}\|x_{k_j} - z_{k_j}\| = 0$. We also have

$$\|x_{k_j+1} - x_{k_j}\| \leq \alpha_{k_j}\|x_{k_j}\| + \beta_{k_j}\|x_{k_j} - z_{k_j}\|, \tag{3.37}$$

which implies that $\lim_{j\to\infty}\|x_{k_j+1} - x_{k_j}\| = 0$. Since $\{x_k\}$ is bounded, we obtain that $\omega_w(x_k)$ is nonempty. Using the similar method in Theorem 3.1, we can obtain $\omega_w(x_k) \subset \Gamma$. Since $\|\hat{u}\| = \min\{\|u\||u \in \Gamma\}$, it follows that

$$\limsup_{j\to\infty}\langle \hat{u}, \hat{u} - x_{k_j+1}\rangle = \limsup_{j\to\infty}\langle \hat{u}, \hat{u} - x_{k_j}\rangle = \sup_{\omega\in\omega_w(x_{k_j})}\langle \hat{u}, \hat{u} - \omega\rangle \leq 0.$$

Since $\|x_{k_j} - y_{k_j}\| \to 0$ and $\{x_{k_j}\}$ is bounded, it is easy to see that $\limsup_{j\to\infty}\delta_{k_j} \leq 0$. From Lemma 2.4, we conclude that $\lim_{k\to\infty}x_k = \hat{u}$ and the proof is completed. $\square$

4. THE NUMERICAL EXAMPLE

In this section, we present an example to demonstrate the performance of the proposed Algorithm 2 (denoted by SEA). With different initial values, we compare the SEA with the proximal and contraction algorithm (denoted by PCA) in [26] as follows:

$$\begin{cases}
y_k = J_{\lambda_k}^A(x_k - \lambda_k f(x_k)); \\
d(x_k, y_k) = (x_k - y_k - \lambda_k(f(x_k) - f(y_k))); \\
x_{k+1} = x_k - \gamma\beta_k d(x_k, y_k).
\end{cases} \tag{4.1}$$

We use figures to show their difference in the convergence speed of the algorithms. Besides, we also show the different convergence speed about SEA with the different λ values.

All experiments are carried out using MATLAB on a personal computer with 1.6GHz central processing unit (Intel(R) Celeron(R) CPU N3150), 4.00 GB memory, and Window operating system (Window 10).

Example 4.1. Let $x^* \in \mathbb{R}^n$ be a K-sparse signal, $K \ll n$. The sampling matrix $A \in \mathbb{R}^{m \times n}(m <<$ $n)$ is stimulated by standard Gaussian distribution and vector $b = Ax^* + e$, 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.
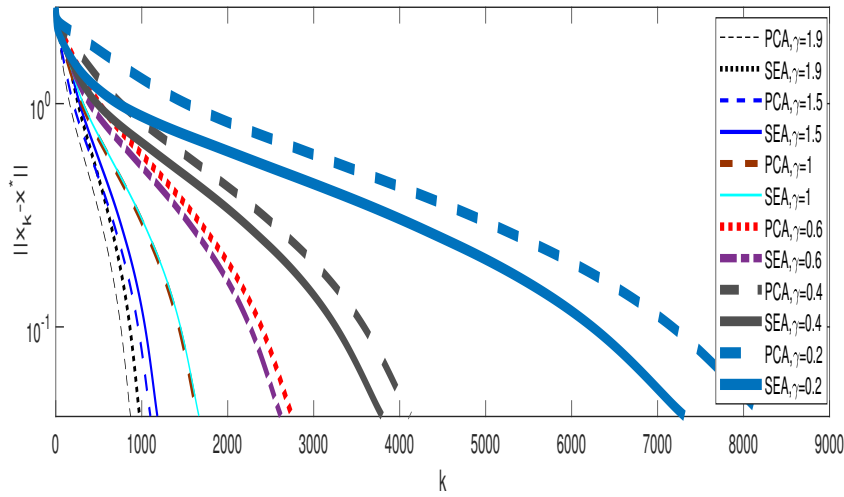

FIGURE 1. Comparision results of SEA and PCA with the different parameter γ selection.
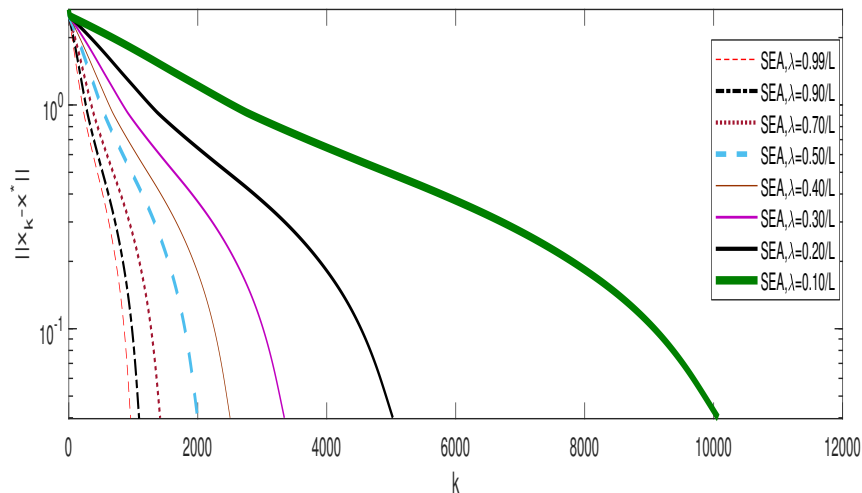

FIGURE 2. Comparision results of SEA with the different parameter λ selection.

It is well-known that the sparse signal x^* can be recovered by solving the following LASSO problem [24],

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2 + \kappa\|x\|_1, \tag{4.2}$$

which is to

Find $x^* \in H$ such that $0 \in \kappa\partial\|x^*\|_1 + A^T(Ax^*-b)$.

The resolvent $J_\lambda^{\partial\|.\|_1}$ is given by the Moreau decomposition

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

We take $m = 120$, $n = 512$ and $K = 60$, the stopping criteria $\|x_k - x^*\| \leq \varepsilon$ where $\varepsilon > 0$ is a given small constant. We randomly choose the starting points $x_0 \in [0,1]^n$.

We can obtain two figures under the different parameters of γ and λ. In FIGURE 1, we take $\lambda_k = \frac{0.99}{L}$, and SEA is compared with PCA. If γ is larger, then SEA is worse than PCA, and if γ is smaller, then SEA is better than PCA. In FIGURE 2, we obtain that the value of λ is the greater, and the convergence is the better. In conclusion, as we can see from the figures, our algorithm (SEA) is feasible.

Acknowledgements

REFERENCES

[1] N.T. An, N.M. Nam, X. Qin, Solving k-center problems involving sets based on optimization techniques, J. Global Optim. 76 (2020), 189–209.

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

[3] X. Cai, G. Gu, B. He, On the $O(1/t)$ convergence rate of the projection and contraction methods for variational inequalities with Lipschitz continuous monotone operators, Comput. Optim. Appl. 57 (2014), 339-363.

[4] S.Y. Cho, Implicit extragradient-like method for fixed point problems and variational inclusion problems in a Banach space, Symmetry, 12 (2020), 998.

[5] S.Y. Cho, A convergence theorem for generalized mixed equilibrium problems and multivalued asymptotically nonexpansive mappings, J. Nonlinear Convex Anal. 21 (2020), 1017-1026.

[6] T.H. Cuong, J.C. Yao, N.D. Yen, Qualitative properties of the minimum sum-of-squares clustering problem, Optimization, 69 (2020), 2131-2154.

[7] Y. Censor, A. Gibali, S. Reich, The subgradient extragradient method for solving variational inequalities in Hilbert space. J. Optim. Theory Appl. 148 (2011), 318-335.

[8] Y. Censor, A. Gibali, S. Reich, Strong convergence of subgradient extragradient methods for the variational inequality problem in Hilbert space, Optim. Meth. Softw. 26 (2011), 827-845.

[9] Y. Censor, A. Gibali, S. Reich, Extensions of Korpelevich's extragradient method for the variational inequality problem in Euclidean space, Optimization 61 (2011), 1119-1132.

[10] Y. Censor, A. Gibali, S. Reich, Algorithms for the split variational inequality problem. Numer. Algo. 56 (2012), 301-323.

[11] Q.L. Dong, D. Jiang, A. Gibali, A modified subgradient extragradient method for solving the variational inequality problem, Numer. Algo. 79 (2018), 927–940.

[12] Q.L. Dong, Y.Y. Lu, J. Yang, The extragradient algorithm with inertial effects for solving the variational inequality. Optim. 65 (2016), 2217-2226.

[13] F. Fan, L. Liu, X. Qin, A subgradient extragradient algorithm with inertial effects for solving strongly pseu-domonotone variational inequalities, Optim. 69 (2020), 2199-2215.

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

[15] S. He, C. Yang Solving the variational inequality problem defined on intersection of finite level sets, Abstr. Appl. Anal. 2013 (2013), 1-8.

[16] G.M. Korpelevich, The extragradient method for finding saddle points and other problems, Ekonomikai Mat. Metody. 12 (1976), 747-756.

[17] L.V. Nguyen, X. Qin, The minimal time function associated with a collection of sets, ESAIM Control Optim. Calc. Var. 26 (2020), 93.

[18] X. Qin, N.T. An, Smoothing algorithms for computing the projection onto a Minkowski sum of convex sets, Comput. Optim. Appl. 74 (2019), 821-850.

[19] X. Qin, L. Wang, J.C. Yao, Inertial splitting method for maximal monotone mappings, J. Nonlinear Convex Anal. 21 (2020), 2325-2333.

[20] D.F. Sun, A class of iterative methods for solving nonlinear projection equations, J. Optim. Theory Appl. 91 (1996), 123-140.

[21] W. Takahashi, H.K. Xu, J.C. Yao, Iterative methods for generalized split feasibility problems in Hilbert space, Set-Valued Var Anal. 23 (2015), 205-221.

[22] W. Takahashi, C.F. Wen, J.C. Yao, The shrinking projection method for a finite family of demimetric map-pings with variational inequality problems in a Hilbert space, Fixed Point Theory 19 (2018), 407-419.

[23] D.V. Thong, A. Gibali, Two strong convergence subgradient extragradient methods for solving variational inequalities in Hilbert spaces, Japan J. Ind. Appl. Math. 36 (2019), 299-321.

[24] R. Tibshirani, Regression Shrinkage and Selection Via the Lasso, J. Royal Stat. Soc. 58 (1996), 267-288.

[25] W. Takahahsi, J.C. Yao, The split common fixed point problem for two finite families of nonlinear mappings in Hilbert spaces, J. Nonlinear Convex Anal. 20 (2019), 173-195.

[26] C. Zhang, Y. Wang, Proximal algorithm for solving monotone variational inclusion, Optim. 67 (2018), 1197-1209.

[27] H. Zhou, X. Qin, Fixed Points of Nonlinear Operators. Iterative Methods, De Gruyter, Berlin, 2020.