



Research article

# An extrapolated fixed-point optimization method for strongly convex smooth optimizations

Duangdaw Rakjarungkiat and Nimit Nimana\*

Department of Mathematics, Faculty of Science, Khon Kaen University, Khon Kaen 40002, Thailand

\* **Correspondence:** Email: nimitni@kku.ac.th; Tel: +66818866798.

**Abstract:** In this work, we focused on minimizing a strongly convex smooth function over the common fixed-point constraints. We proposed an extrapolated fixed-point optimization method, which is a modified version of the extrapolated sequential constraint method with conjugate gradient direction. We proved the convergence of the generated sequence to the unique solution to the considered problem without boundedness assumption. We also investigated some numerical experiments to underline the effectiveness and performance of the proposed method.

**Keywords:** conjugate gradient method; cutter; extrapolated method; strong convergence; strong convexity

**Mathematics Subject Classification:** 47H09, 47J05, 65K10, 90C25

## 1. Introduction

Let  $\mathcal{H}$  be a real Hilbert space equipped with an inner product  $\langle \cdot, \cdot \rangle$  and its induced norm  $\| \cdot \|$ . In this work, we deal with the minimization of a strongly convex smooth function over the common fixed-point constraints, which is in the following form:

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{x} \in \bigcap_{i=1}^m \text{Fix } T_i, \end{aligned} \tag{1.1}$$

where  $\text{Fix } T_i := \{ \mathbf{x} \in \mathcal{H} : T_i(\mathbf{x}) = \mathbf{x} \}$  and the objective function and the constrained operators satisfy the following assumptions:

(A1) The function  $f : \mathcal{H} \rightarrow \mathbb{R}$  is  $\eta$ -strongly convex and  $L$ -smooth.

(A2) The operators  $T_i : \mathcal{H} \rightarrow \mathcal{H}, i = 1, \dots, m$ , are cutters with  $\bigcap_{i=1}^m \text{Fix } T_i \neq \emptyset$ .

Thanks to the closedness and the convexity of the common fixed-point sets  $\bigcap_{i=1}^m \text{Fix } T_i$  and the strong convexity of  $f$ , we can ensure that the solution set to the problem (1.1) is a singleton set. In this situation, we denote the solution point to the problem (1.1) by  $\mathbf{x}^*$  throughout this work.

### 1.1. Related works

The minimization problem and its generalization (such as variational inequality problem) over the common fixed-point constraints have been investigated by many authors, in which we briefly note some related works as follows. Bauschke [1] considered the problem (1.1) in the case when the objective function is given by  $f(\mathbf{x}) := \frac{1}{2}\|\mathbf{x} - \mathbf{a}\|^2$ , where the point  $\mathbf{a} \in \mathcal{H}$  is a given point and the operators  $T_i, i = 1, 2, \dots, m$  are nonexpansive operators. He proposed the following method: Given  $\mathbf{x}^1 \in \mathcal{H}$  and a nonnegative sequence  $\{\beta_k\}_{k=1}^\infty$ , for all  $k \in \mathbb{N}$ , compute

$$\mathbf{x}^{k+1} = T_{[k+1]}(\mathbf{x}^k) - \beta_{k+1}(T_{[k+1]}(\mathbf{x}^k) - \mathbf{a}), \quad (1.2)$$

where the function  $[\cdot]$  is the modulo  $m$  function taking values in  $\{1, \dots, m\}$ . Bauschke proved that the sequence  $\{\mathbf{x}^k\}_{k=1}^\infty$  generated by the proposed method (1.2) converges strongly to the unique solution of the considered problem, provided that the constrained set satisfies

$$\begin{aligned} \bigcap_{i=1}^m \text{Fix } T_i &= \text{Fix}(T_m \cdots T_1) \\ &= \text{Fix}(T_1 T_m \cdots T_3 T_2) \\ &= \cdots = \text{Fix}(T_{m-1} T_{m-2} \cdots T_1 T_m), \end{aligned} \quad (1.3)$$

and the sequence  $\{\beta_k\}_{k=1}^\infty \subset [0, 1)$  satisfies  $\lim_{k \rightarrow \infty} \beta_k = 0$ ,  $\sum_{k=1}^\infty \beta_k = \infty$ , and  $\sum_{k=1}^\infty |\beta_{k+m} - \beta_k| < \infty$ .

After that, Yamada [2] considered the solving of a more general setting of the problem (1.1) in the sense of variational inequality problem governed by the strongly monotone and Lipschitz continuous operator  $F : \mathcal{H} \rightarrow \mathcal{H}$ . Thanks to the optimality condition, it is well known that such a considered problem can be reduced to the problem (1.1) by setting  $F := \nabla f$ . He proposed the so-called hybrid steepest descent method, which is in the following form: Given  $\mathbf{x}^1 \in \mathcal{H}$  and a nonnegative sequence  $\{\beta_k\}_{k=1}^\infty$ , for all  $k \in \mathbb{N}$ , compute

$$\mathbf{x}^{k+1} = T_{[k+1]}(\mathbf{x}^k) - \beta_{k+1}F(T_{[k+1]}(\mathbf{x}^k)). \quad (1.4)$$

With the similar assumption on the constrained set (1.3) and the control sequence  $\{\beta_k\}_{k=1}^\infty$ , the strong convergence of the proposed method (1.4) to the unique solution of the considered problem was obtained.

Xu and Kim [3] also investigated the problem in the same setting as in Yamada [2] and considered the convergence result of the method (1.4) by proposing a variant condition of the control sequence  $\{\beta_k\}_{k=1}^\infty \subset (0, 1]$ , which satisfies  $\lim_{n \rightarrow \infty} \frac{\beta_n}{\beta_{n+1}} = 1$  in place of the condition  $\sum_{k=1}^\infty |\beta_{k+m} - \beta_k| < \infty$  given in Yamada [2]. Another condition of the control sequence  $\{\beta_k\}_{k=1}^\infty$  and the assumption on the constrained set (1.3) are also the same as above. After the works of Yamada [2] and Xu and Kim [3] were presented, many authors considered the variances and generalizations; see, for instance, [4–9].

Prangprakhon et al. [10] also considered the solving of the variational inequality problem over the intersection of fixed point sets of firmly nonexpansive operators. They presented an iterative algorithm, which can be viewed as a generalization of the hybrid steepest descent method in the allowance of adding appropriated information when computing operators values. The main feature of their proposed method is the presence of added information that can be viewed as the allowance of possible numerical errors on the computations of operators' values. They subsequently proved the strong convergence of the proposed method without the assumption on the constrained set (1.3).

Prangprakhon and Nimana [11] subsequently proposed the so-called extrapolated sequential constraint method with conjugate gradient direction (ESCoM-CGD) for solving the variational inequality problem over the intersection of the fixed-point sets of cutters. ESCoM-CGD is motivated by the ideas of the hybrid conjugate gradient method [12], which is an accelerated version of the hybrid steepest descent method together with the extrapolated cyclic cutter methods [13, 14]. This method consists of two interesting features, namely, it consists of the extrapolation stepsize function  $\sigma(\mathbf{y}^k)$  (see Step 2 of Algorithm 1) and the search direction  $\mathbf{d}^{k+1}$ , which is the combination of the current direction  $-\nabla f(\mathbf{x}^{k+1})$  together with a previous search direction  $\mathbf{d}^k$ . By assuming the boundedness of the generated sequence  $\{\mathbf{x}^k\}_{k=1}^{\infty}$  together with the assumptions on control sequences, they also proved the strong convergence of the proposed method.

Very recently, Petrot et al. [15] proposed the so-called dynamic distributed three-term conjugate gradient method for solving the strongly monotone variational inequality problem over the intersection of fixed-point sets of firmly nonexpansive operators. Unlike [11], this method has a simultaneous structure so that it allows the independent computation of each operator along with the dynamic weight, which is updated at every iteration. Moreover, the strong convergence of the method was obtained without assuming that the sequence  $\{\mathbf{x}^k\}_{k=1}^{\infty}$  is bounded.

## 1.2. Our contribution

In this work, we will present a modification version of ESCoM-CGD by using the search direction considered in [15]. We will prove the convergence of the proposed method without assuming that the generated sequence is bounded. It is important to underline that we will simplify the proving lines compared to the proof of ESCoM-CGD given in [11, Theorem 3] in a shortened way. We also show that the proposed algorithm and convergence result are applicable in the binary classification via support vector machine learning.

The remainder of this work is organized as follows. In the rest of this section, we collect some mathematical preliminaries consisting of some useful definitions and facts needed in the proving lines. In Section 2, we present the modified version of ESCoM-CGD for finding the solution to the problem (1.1). The main convergence theorem and some important remarks are also given in this section. In Section 3, we provide some usefulness of the proposed method and its convergence result in solving the minimum-norm problem to the system of homogeneous linear inequalities, as well as the binary classification via support vector machine learning. In Section 4, we provide some technical convergence properties of the generated sequences and, subsequently, prove the main convergence theorem. Lastly, we give a concluding remark.

### 1.3. Mathematical preliminaries

In this subsection, we will recall some definitions, properties, and key tools in proving our convergence results. The readers can consult the books [16, 17] for more details.

We denote by  $Id$  the identity operator on a Hilbert space  $\mathcal{H}$ . For a sequence  $\{\mathbf{x}^k\}_{k=1}^{\infty}$ , the strong and weak convergences of a sequence  $\{\mathbf{x}^k\}_{k=1}^{\infty}$  to a point  $\mathbf{x} \in \mathcal{H}$  are defined by the expression  $\mathbf{x}^k \rightarrow x$  and  $\mathbf{x}^k \rightharpoonup x$ , respectively.

We will recall the convexity and smoothness of a function. Now, let  $f : \mathcal{H} \rightarrow \mathbb{R}$  be a real-valued function. Let  $\alpha > 0$  be given. We say that the function  $f$  is convex if, for all  $\mathbf{x}, \mathbf{y} \in \mathcal{H}$  and for all  $\lambda \in [0, 1]$ , it holds that

$$f((1 - \lambda)\mathbf{x} + \lambda\mathbf{y}) \leq (1 - \lambda)f(\mathbf{x}) + \lambda f(\mathbf{y}).$$

In particular, we say that the function  $f$  is  $\alpha$ -strongly convex if, for all  $\mathbf{x}, \mathbf{y} \in \mathcal{H}$  and for all  $\lambda \in [0, 1]$ , it holds that

$$f((1 - \lambda)\mathbf{x} + \lambda\mathbf{y}) \leq (1 - \lambda)f(\mathbf{x}) + \lambda f(\mathbf{y}) - \frac{1}{2}\alpha\lambda(1 - \lambda)\|\mathbf{x} - \mathbf{y}\|^2.$$

The constant  $\alpha$  is called the strongly convex parameter.

Let  $\mathbf{x} \in \mathcal{H}$  be given. We say that the function  $f$  is Fréchet differentiable or, in short, differentiable at  $\mathbf{x}$  if there exists a vector  $\mathbf{g} \in \mathcal{H}$  such that

$$\lim_{\mathbf{h} \rightarrow \mathbf{0}} \frac{f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) - \langle \mathbf{g}, \mathbf{h} \rangle}{\|\mathbf{h}\|} = 0.$$

Moreover, we call this unique vector  $\mathbf{g}$  by the gradient of  $f$  at  $\mathbf{x}$  and denote it by  $\nabla f(\mathbf{x})$ .

It is well known that the convexity of  $f$  can be characterized in terms of differentiability property as the followings.

**Fact 1.1.** [16, Theorem 5.24] *Let  $f : \mathcal{H} \rightarrow \mathbb{R}$  be a real-valued differentiable function, then  $f$  is  $\alpha$ -strongly convex if, and only if, for all  $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ , we have*

$$\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq \alpha\|\mathbf{x} - \mathbf{y}\|^2.$$

The following fact is the cornerstone of our proving lines. It is the necessary and sufficient condition for the optimality of a convex function over a nonempty closed and convex set.

**Fact 1.2.** [16, Corollary 3.68] *Let  $f : \mathcal{H} \rightarrow \mathbb{R}$  be a continuously differentiable convex function and  $C \subset \mathcal{H}$  be a closed and convex set, then  $f$  attains its minimum at a point  $\mathbf{x}^* \in C$  if, and only if, for all  $\mathbf{y} \in C$ , it holds that*

$$\langle \nabla f(\mathbf{x}^*), \mathbf{y} - \mathbf{x}^* \rangle \geq 0.$$

Let  $L \geq 0$  be given. We say that the function  $f$  is  $L$ -smooth if it is differentiable and, for all  $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ , it satisfies that

$$\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\| \leq L\|\mathbf{x} - \mathbf{y}\|.$$

The constant  $L$  is called the smoothness parameter. It is clear that an  $L$ -smooth function is a continuously differentiable function.

The following fact is a very useful tool in obtaining the convergence result. The proof can be found in [2, Lemma 3.1(b)].

**Fact 1.3.** Let  $f : \mathcal{H} \rightarrow \mathbb{R}$  be an  $\eta$ -strongly convex and  $L$ -smooth. For any  $\mu \in (0, 2\eta/L^2)$  and  $\beta \in (0, 1]$ , if we define the operator  $F^\beta : \mathcal{H} \rightarrow \mathcal{H}$  by  $F^\beta(\mathbf{x}) := \mathbf{x} - \mu\beta\nabla f(\mathbf{x})$  for all  $\mathbf{x} \in \mathcal{H}$ , then

$$\|F^\beta(\mathbf{x}) - F^\beta(\mathbf{y})\| \leq (1 - \beta\tau)\|\mathbf{x} - \mathbf{y}\|,$$

for all  $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ , where  $\tau := 1 - \sqrt{1 + \mu^2 L^2 - 2\mu\eta} \in (0, 1]$ .

Next, we recall the definition of cutters and some useful properties.

Let  $T : \mathcal{H} \rightarrow \mathcal{H}$  be an operator with  $\text{Fix } T := \{\mathbf{x} \in \mathcal{H} : T(\mathbf{x}) = \mathbf{x}\} \neq \emptyset$ . We say the operator  $T$  is a cutter if, for all  $\mathbf{x} \in \mathcal{H}$  and  $\mathbf{z} \in \text{Fix } T$ , we have

$$\langle \mathbf{x} - T(\mathbf{x}), \mathbf{z} - T(\mathbf{x}) \rangle \leq 0.$$

The followings are the important properties of cutters.

**Fact 1.4.** [17, Remark 2.1.31 and Lemma 2.1.36.] Let  $T : \mathcal{H} \rightarrow \mathcal{H}$  be a cutter, then the following statements hold:

- (i)  $\text{Fix } T$  is closed and convex.
- (ii) For all  $\mathbf{x} \in \mathcal{H}$  and for all  $\mathbf{z} \in \text{Fix } T$ , we have  $\langle T\mathbf{x} - \mathbf{x}, \mathbf{z} - \mathbf{x} \rangle \geq \|T(\mathbf{x}) - \mathbf{x}\|^2$ .

Next, we recall the definitions of relaxations of an operator. Let  $T : \mathcal{H} \rightarrow \mathcal{H}$  be an operator and  $\lambda \in [0, 2]$  be given. The relaxation of the operator  $T$  is defined by  $T_\lambda(\mathbf{x}) := (1 - \lambda)\mathbf{x} + \lambda T(\mathbf{x})$ , for all  $\mathbf{x} \in \mathcal{H}$ . In this case, we call the parameter  $\lambda$  by a relaxation parameter. Moreover, let  $\sigma : \mathcal{H} \rightarrow (0, \infty)$  be a function. The generalized relaxation of the operator  $T$  is defined by

$$T_{\sigma,\lambda}(\mathbf{x}) := \mathbf{x} + \lambda\sigma(\mathbf{x})(T(\mathbf{x}) - \mathbf{x}),$$

for all  $\mathbf{x} \in \mathcal{H}$ . In this case, we call the function  $\sigma$  by a stepsize function. If the function  $\sigma(\mathbf{x}) \geq 1$  for all  $\mathbf{x} \in \mathcal{H}$ , then we call the operator  $T_{\sigma,\lambda}$  by an extrapolation of  $T_\lambda$ . We denote  $T_\sigma := T_{\sigma,1}$ . It is worth noting that, for all  $\mathbf{x} \in \mathcal{H}$ , the followings hold

$$T_{\sigma,\lambda}(\mathbf{x}) - \mathbf{x} = \lambda\sigma(\mathbf{x})(T(\mathbf{x}) - \mathbf{x}) = \lambda(T_\sigma(\mathbf{x}) - \mathbf{x}),$$

and for all  $\lambda \neq 0$ , we also have

$$\text{Fix } T_{\sigma,\lambda} = \text{Fix } T_\sigma = \text{Fix } T.$$

For the simplicity, we denote the compositions of operators as

$$T := T_m T_{m-1} \cdots T_1,$$

$$S_0 := Id$$

and

$$S_i := T_i T_{i-1} \cdots T_1, i = 1, 2, \dots, m.$$

We recall the important properties that will be useful for the convergence properties.

**Fact 1.5.** [17, Section 4.10] Let  $T_i : \mathcal{H} \rightarrow \mathcal{H}$ ,  $i = 1, 2, \dots, m$ , be cutters with  $\bigcap_{i=1}^m \text{Fix } T_i \neq \emptyset$ . Let  $\sigma : \mathcal{H} \rightarrow (0, \infty)$  be defined by

$$\sigma(\mathbf{x}) := \begin{cases} \frac{\sum_{i=1}^m \langle T(\mathbf{x}) - S_{i-1}(\mathbf{x}), S_i(\mathbf{x}) - S_{i-1}(\mathbf{x}) \rangle}{\|T(\mathbf{x}) - \mathbf{x}\|^2}, & \text{if } \mathbf{x} \notin \bigcap_{i=1}^m \text{Fix } T_i, \\ 1, & \text{otherwise,} \end{cases} \quad (1.5)$$

then the following properties hold:

(i) For all  $\mathbf{x} \notin \bigcap_{i=1}^m \text{Fix } T_i$ , we have

$$\sigma(\mathbf{x}) \geq \frac{\frac{1}{2} \sum_{i=1}^m \|S_i(\mathbf{x}) - S_{i-1}(\mathbf{x})\|^2}{\|T(\mathbf{x}) - \mathbf{x}\|^2} \geq \frac{1}{2m}.$$

(ii) The operator  $T_\sigma$  is a cutter.

Next, we will recall the notion of demi-closedness principle as follows.

Let  $T : \mathcal{H} \rightarrow \mathcal{H}$  be an operator having a fixed point. The operator  $T$  is said to satisfy the demi-closedness principle if, for every sequence  $\{\mathbf{x}^k\} \subset \mathcal{H}$  such that  $\mathbf{x}^k \rightarrow \mathbf{x} \in \mathcal{H}$  and  $\|T(\mathbf{x}^k) - \mathbf{x}^k\| \rightarrow 0$ , we have  $\mathbf{x} \in \text{Fix } T$ .

We close this section by recalling the well-known metric projection, which is defined in the following: Let  $C$  be a nonempty subset of  $\mathcal{H}$  and  $\mathbf{x} \in \mathcal{H}$ . If there is a point  $\mathbf{y} \in C$  such that  $\|\mathbf{x} - \mathbf{y}\| \leq \|\mathbf{x} - \mathbf{z}\|$ , for any  $\mathbf{z} \in C$ , then  $\mathbf{y}$  is said to be a metric projection of  $\mathbf{x}$  onto  $C$  and is denoted by  $P_C(\mathbf{x})$ . If  $P_C(\mathbf{x})$  exists and is uniquely determined for all  $\mathbf{x} \in \mathcal{H}$ , then the operator  $P_C : \mathcal{H} \rightarrow \mathcal{H}$  is said to be the metric projection onto  $C$ . Actually, we need some additional properties of the set  $C$  to ensure the existence and the uniqueness of a metric projection  $P_C(\mathbf{x})$  for a point  $\mathbf{x} \in \mathcal{H}$  as the following fact.

**Fact 1.6.** [17, Theorem 1.2.3.] Let  $C$  be a nonempty closed convex subset of  $\mathcal{H}$ , then for any  $\mathbf{x} \in \mathcal{H}$ , there exists a metric projection  $P_C(\mathbf{x})$  and it is uniquely determined.

Moreover, we finally note that the metric projection onto a nonempty closed convex set  $C$  is a cutter with  $\text{Fix } P_C = C$ ; see [17, Theorem 2.2.21.] for further details.

## 2. Algorithm and its convergence

We will proceed in this section by presenting a modified version of the extrapolated sequential constraint method with conjugate gradient direction (ESCoM-CGD) for solving the problem (1.1). Now, we are in a position to state a modified version of ESCoM-CGD as the following algorithm. We call the proposed method by the modified extrapolated sequential constraint method with conjugate gradient direction (in short, MESCoM-CGD).

---

**Algorithm 1: MESCoM-CGD**


---

**Initialization:** Put nonnegative sequences  $\{\beta_k\}_{k=1}^\infty$ ,  $\{\varphi_k\}_{k=1}^\infty$  and  $\{\lambda_k\}_{k=1}^\infty$ , and a parameter  $\mu \in (0, \frac{2\eta}{L^2})$ . Choose an initial point  $\mathbf{x}^1 \in \mathcal{H}$  arbitrarily and set  $\mathbf{d}^1 = -\nabla f(\mathbf{x}^1)$ .

**Iterative Step** ( $k \in \mathbb{N}$ ): For a current iterate  $\mathbf{x}^k \in \mathcal{H}$  and direction  $\mathbf{d}^k \in \mathcal{H}$ , repeat the following steps:

**Step 1.** Compute the iterate  $\mathbf{y}^k \in \mathcal{H}$  as

$$\mathbf{y}^k := \mathbf{x}^k + \mu\beta_k \frac{\mathbf{d}^k}{\max\{1, \|\mathbf{d}^k\|\}}.$$

**Step 2.** Compute the stepsize  $\sigma(\mathbf{y}^k)$  as

$$\sigma(\mathbf{y}^k) := \begin{cases} \frac{\sum_{i=1}^m \langle T(\mathbf{y}^k) - S_{i-1}(\mathbf{y}^k), S_i(\mathbf{y}^k) - S_{i-1}(\mathbf{y}^k) \rangle}{\|T(\mathbf{y}^k) - \mathbf{y}^k\|^2}, & \text{if } \mathbf{y}^k \notin \bigcap_{i=1}^m \text{Fix } T_i, \\ 1, & \text{otherwise.} \end{cases}$$

**Step 3.** Compute the recurrence  $\mathbf{x}^{k+1} \in \mathcal{H}$  and the search direction  $\mathbf{d}^{k+1} \in \mathcal{H}$  as

$$\mathbf{x}^{k+1} := T_{\sigma, \lambda_k}(\mathbf{y}^k)$$

and

$$\mathbf{d}^{k+1} := -\nabla f(\mathbf{x}^{k+1}) + \varphi_{k+1} \frac{\mathbf{d}^k}{\max\{1, \|\mathbf{d}^k\|\}}.$$

**Step 4.** Update  $k := k + 1$  and return to **Step 1**.

---

Some comments and particular situations of MESCoM-CGD are the following:

**Remark 2.1.** (i) As we have mentioned that the convergence of ESCoM-CGD (see, [11, Theorem 3]) relies on the assumption that the generated sequence  $\{\mathbf{x}^k\}_{k=1}^\infty$  is bounded, however, it is also pointed in [11, Remark 3] that the boundednesses of the search direction  $\{\mathbf{d}^k\}_{k=1}^\infty \in \mathcal{H}$  and the sequence  $\{\varphi_k\}_{k=1}^\infty$  together with the definition of the search direction yield the boundedness of the sequence  $\{\mathbf{x}^k\}_{k=1}^\infty$ . In this situation, if we construct the bounded search direction  $\left\{ \frac{\mathbf{d}^k}{\max\{1, \|\mathbf{d}^k\|\}} \right\}_{k=1}^\infty \in \mathcal{H}$  in place of the previous version, then the boundedness of the generated sequence  $\{\mathbf{x}^k\}_{k=1}^\infty$  will be provable as well (see, Lemma 3.3 below). One can notice that the key idea of this bounded strategy is nothing else but restricting the search direction in a unit ball centered at the origin.

(ii) If the search direction  $\{\mathbf{d}^k\}_{k=1}^\infty$  is guaranteed to stay within the unit ball centered at the origin, then it holds that  $\max\{1, \|\mathbf{d}^k\|\} = 1$  so that MESCoM-CGD is reduced to the ESCoM-CGD in the case when the operator  $T_m = Id$ .

(iii) If the number of  $m = 1$ , the relaxation parameter  $\lambda_k = 1$ , the stepsize  $\sigma(\mathbf{y}^k) = 1$ , and  $\max\{1, \|\mathbf{d}^k\|\} = 1$  for all  $n \in \mathbb{N}$ , then MESCoM-CGD is nothing else but the hybrid conjugate gradient method proposed in [12].

Now, let us begin this section by investigating the assumptions needed for the convergence result.

**Assumption 2.2.** Assume that

(H1) The sequence of stepsizes  $\{\beta_k\}_{k=1}^\infty \subset (0, 1]$  satisfies  $\sum_{k=1}^\infty \beta_k = \infty$  and  $\sum_{k=1}^\infty \beta_k^2 < \infty$ .

(H2) The sequence of parameters  $\{\varphi_k\}_{k=1}^\infty \subset [0, \infty)$  satisfies  $\lim_{k \rightarrow \infty} \varphi_k = 0$ .

(H3) There is a constant  $\varepsilon \in (0, 1)$  such that the relaxation parameters  $\{\lambda_k\}_{k=1}^\infty \subset [\varepsilon, 2 - \varepsilon]$ .

(H4) The operators  $T_i$ ,  $i = 1, \dots, m$ , satisfy the demi-closedness principle.

Let us briefly discuss some particular situations in which hypotheses (H1)–(H4) hold as follows:

**Remark 2.3.** (i) An example of the stepsizes  $\{\beta_k\}_{k=1}^\infty$  satisfying (H1) is, for instance,  $\beta_k = \frac{\beta_0}{(k+1)^b}$  with  $\beta_0 \in (0, 1]$  and  $b \in (0, 1]$  for all  $k \in \mathbb{N}$ .

(ii) An example of the parameters  $\{\varphi_k\}_{k=1}^\infty$  satisfying (H2) is, for instance,  $\varphi_k = \frac{\varphi_0}{(k+1)^c}$  with  $\varphi_0 \geq 0$  and  $c > 0$  for all  $k \in \mathbb{N}$ .

(iii) The role of the constant  $\varepsilon \in (0, 1)$  is to ensure that the term  $\lambda_k(2 - \lambda_k)$  stays away from zero for all  $k \in \mathbb{N}$ . This will yield that the cancellation of this term can be processed. One can take, for instance,  $\lambda_k = \lambda_0$  for some  $\lambda_0 \in (0, 2)$  as a trivial example of relaxation parameters  $\{\lambda_k\}_{k=1}^\infty$ , satisfying hypothesis (H3).

(iv) The demi-closedness principle of the operators  $T_i$ ,  $i = 1, \dots, m$ , in (H4) is a key property and it is typically assumed when dealing with the convergence result of the common fixed-point type problems. Actually, the demi-closedness principle is satisfied by a nonexpansive operator, i.e.,  $\|T_i(\mathbf{x}) - T_i(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|$  for all  $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ ; see [17, Lemma 3.2.5.] for the proving lines with some historical notes. This yields that the metric projection onto a nonempty closed and convex also satisfies the demi-closedness principle, see [17, Theorem 2.2.21]. Moreover, it is worth mentioning that the demi-closedness principle is also satisfied by a subgradient projection operator  $P_f$  for a convex continuous function  $f : \mathcal{H} \rightarrow \mathbb{R}$ , which is Lipschitz continuous relative to every bounded subset of  $\mathcal{H}$ ; see [17, Theorem 4.2.7] for more details.

The main result of this work is the following theorem:

**Theorem 2.4.** Let the sequence  $\{\mathbf{x}^k\}_{k=1}^\infty$  be generated by MESCOM-CGD and assume that assumptions (A1) and (A2) and hypotheses (H1)–(H4) hold, then the sequence  $\{\mathbf{x}^k\}_{k=1}^\infty$  converges strongly to the unique solution  $\mathbf{x}^*$  of the problem (1.1).

### 3. Proof of Theorem 2.4

In this section, we will provide some technical convergence properties and close this section by proving the convergence of the proposed method to the unique solution of the problem (1.1).

For simplicity of notations, we denote

$$\tau := 1 - \sqrt{1 + \mu^2 L^2 - 2\mu\eta} \in (0, 1]$$

and, for every  $n \in \mathbb{N}$  and  $\mathbf{x} \in \bigcap_{i=1}^m \text{Fix } T_i$ , we denote

$$\varepsilon_k := \mu^2 \beta_k^2 + \frac{2\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \langle \mathbf{x}^k - \mathbf{x}, \mathbf{d}^k \rangle$$



and

$$\alpha_k := \frac{\beta_k \tau}{\max\{1, \|\mathbf{d}^k\|\}}.$$

Moreover, for every  $k \geq 2$ , we denote

$$\delta_k := \frac{2\mu}{\tau} \left( \frac{\varphi_k}{\max\{1, \|\mathbf{d}^{k-1}\|\}} \langle \mathbf{d}^{k-1}, \mathbf{y}^k - \mathbf{x} \rangle - \langle \nabla f(\mathbf{x}), \mathbf{y}^k - \mathbf{x} \rangle \right).$$

In order to prove the convergence result in Theorem 2.4, we collect several facts that will be useful in what follows. We first state the lemma relating to the norms of iterate  $\mathbf{x}^k$  to a common fixed-point. The analysis is similar to those given in [11, Lemma 4], however, we state here again for the sake of completeness.

**Lemma 3.1.** For all  $k \in \mathbb{N}$  and  $\mathbf{x} \in \bigcap_{i=1}^m \text{Fix } T_i$ , we have

$$\|\mathbf{x}^{k+1} - \mathbf{x}\|^2 \leq \|\mathbf{x}^k - \mathbf{x}\|^2 - \frac{\lambda_k(2 - \lambda_k)}{4m} \sum_{i=1}^m \|S_i(\mathbf{y}^k) - S_{i-1}(\mathbf{y}^k)\|^2 + \varepsilon_k.$$

*Proof.* Let  $k \in \mathbb{N}$  and  $\mathbf{x} \in \bigcap_{i=1}^m \text{Fix } T_i$  be given. By using the properties of extrapolation and Facts 1.4 (ii) and 1.5(ii), we note that

$$\begin{aligned} \|\mathbf{x}^{k+1} - \mathbf{x}\|^2 &= \|T_{\sigma, \lambda_k}(\mathbf{y}^k) - \mathbf{x}\|^2 \\ &= \|\mathbf{y}^k + \lambda_k \sigma(\mathbf{y}^k)(T(\mathbf{y}^k) - \mathbf{y}^k) - \mathbf{x}\|^2 \\ &= \|\mathbf{y}^k - \mathbf{x}\|^2 + \lambda_k^2 \|\sigma(\mathbf{y}^k)(T(\mathbf{y}^k) - \mathbf{y}^k)\|^2 + 2\lambda_k \langle \mathbf{y}^k - \mathbf{x}, \sigma(\mathbf{y}^k)(T(\mathbf{y}^k) - \mathbf{y}^k) \rangle \\ &= \|\mathbf{y}^k - \mathbf{x}\|^2 + \lambda_k^2 \|T_{\sigma}(\mathbf{y}^k) - \mathbf{y}^k\|^2 + 2\lambda_k \langle \mathbf{y}^k - \mathbf{x}, T_{\sigma}(\mathbf{y}^k) - \mathbf{y}^k \rangle \\ &\leq \|\mathbf{y}^k - \mathbf{x}\|^2 + \lambda_k^2 \|T_{\sigma}(\mathbf{y}^k) - \mathbf{y}^k\|^2 - 2\lambda_k \|T_{\sigma}(\mathbf{y}^k) - \mathbf{y}^k\|^2 \\ &= \|\mathbf{y}^k - \mathbf{x}\|^2 - \lambda_k(2 - \lambda_k) \|T_{\sigma}(\mathbf{y}^k) - \mathbf{y}^k\|^2 \\ &= \|\mathbf{y}^k - \mathbf{x}\|^2 - \lambda_k(2 - \lambda_k) \sigma^2(\mathbf{y}^k) \|T(\mathbf{y}^k) - \mathbf{y}^k\|^2 \\ &\leq \|\mathbf{y}^k - \mathbf{x}\|^2 - \lambda_k(2 - \lambda_k) \frac{\left( \sum_{i=1}^m \|S_i(\mathbf{y}^k) - S_{i-1}(\mathbf{y}^k)\|^2 \right)^2}{\|T(\mathbf{y}^k) - \mathbf{y}^k\|^4} \|T\mathbf{y}^k - \mathbf{y}^k\|^2 \\ &= \|\mathbf{y}^k - \mathbf{x}\|^2 - \lambda_k(2 - \lambda_k) \frac{\left( \sum_{i=1}^m \|S_i(\mathbf{y}^k) - S_{i-1}(\mathbf{y}^k)\|^2 \right)^2}{\|T(\mathbf{y}^k) - \mathbf{y}^k\|^2} \\ &\leq \|\mathbf{y}^k - \mathbf{x}\|^2 - \frac{\lambda_k(2 - \lambda_k)}{4m} \sum_{i=1}^m \|S_i(\mathbf{y}^k) - S_{i-1}(\mathbf{y}^k)\|^2. \end{aligned} \tag{3.1}$$

We note from the definition of  $\{\mathbf{y}^k\}_{k=1}^{\infty}$  that

$$\begin{aligned}
\|\mathbf{y}^k - \mathbf{x}\|^2 &\leq \left\| \mathbf{x}^k + \mu\beta_k \frac{\mathbf{d}^k}{\max\{1, \|\mathbf{d}^k\|\}} - \mathbf{x} \right\|^2 \\
&= \|\mathbf{x}^k - \mathbf{x}\|^2 + \frac{\mu^2\beta_k^2}{\max\{1, \|\mathbf{d}^k\|\}^2} \|\mathbf{d}^k\|^2 + \frac{2\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \langle \mathbf{x}^k - \mathbf{x}, \mathbf{d}^k \rangle \\
&\leq \|\mathbf{x}^k - \mathbf{x}\|^2 + \mu^2\beta_k^2 + \frac{2\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \langle \mathbf{x}^k - \mathbf{x}, \mathbf{d}^k \rangle \\
&= \|\mathbf{x}^k - \mathbf{x}\|^2 + \varepsilon_k.
\end{aligned}$$

By invoking this relation in (3.1), we obtain

$$\|\mathbf{x}^{k+1} - \mathbf{x}\|^2 \leq \|\mathbf{x}^k - \mathbf{x}\|^2 - \frac{\lambda_k(2 - \lambda_k)}{4m} \sum_{i=1}^m \|S_i(\mathbf{y}^k) - S_{i-1}(\mathbf{y}^k)\|^2 + \varepsilon_k$$

as required.  $\square$

In order to prove the boundedness of the generated sequence  $\{\mathbf{x}^k\}_{k=1}^\infty$ , we need the following proposition [18, Lemma 3.1].

**Proposition 3.2.** *Let  $\{a_k\}_{k=1}^\infty$  be a sequence of nonnegative real numbers such that there exists a subsequence  $\{a_{k_j}\}_{j=1}^\infty$  of  $\{a_k\}_{k=1}^\infty$  with  $a_{k_j} < a_{k_{j+1}}$  for all  $j \in \mathbb{N}$ . If, for all  $k \geq k_0$ , we define*

$$v(k) = \max\{\bar{k} \in \mathbb{N} : k_0 \leq \bar{k} \leq k, a_{\bar{k}} < a_{\bar{k}+1}\},$$

*then the sequence  $\{v(k)\}_{k \geq k_0}$  is nondecreasing,  $a_{v(k)} \leq a_{v(k)+1}$ , and  $a_k \leq a_{v(k)+1}$  for every  $k \geq k_0$ .*

Now, we are in a position to prove the boundedness property of the generated sequence  $\{\mathbf{x}^k\}_{k=1}^\infty$  as the following lemma. The idea proof is based on [15, Lemma 5].

**Lemma 3.3.** *The sequence  $\{\mathbf{x}^k\}_{k=1}^\infty$  is a bounded sequence.*

*Proof.* Let  $k \in \mathbb{N}$  and  $\mathbf{x} \in \bigcap_{i=1}^m \text{Fix } T_i$  be given. Since  $\lambda_k \in [\varepsilon, 2 - \varepsilon]$ , we have  $\varepsilon^2 \leq \lambda_k(2 - \lambda_k)$ , and so

$$\frac{\lambda_k(2 - \lambda_k)}{4m} \sum_{i=1}^m \|S_i(\mathbf{y}^k) - S_{i-1}(\mathbf{y}^k)\|^2 \geq 0.$$

Thus, by using this property together with the inequality obtained in Lemma 3.1, we get

$$\|\mathbf{x}^{k+1} - \mathbf{x}\|^2 \leq \|\mathbf{x}^k - \mathbf{x}\|^2 + \mu^2\beta_k^2 + \frac{2\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \langle \mathbf{x}^k - \mathbf{x}, \mathbf{d}^k \rangle. \quad (3.2)$$

For the sake of simplicity, we set

$$\Gamma_k := \|\mathbf{x}^k - \mathbf{x}\|^2 - \mu^2 \sum_{j=1}^{k-1} \beta_j^2$$

for all  $k \in \mathbb{N}$ . In this case, the inequality (3.2) can be rewritten as

$$\Gamma_{k+1} - \Gamma_k + \frac{2\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \langle \mathbf{x}^k - \mathbf{x}, \mathbf{d}^k \rangle \leq 0. \quad (3.3)$$

Next, we divide the proof into two cases based on the behaviors of the sequence  $\{\Gamma_k\}_{k=1}^\infty$ :

Case I: Suppose that there exists  $k_0 \in \mathbb{N}$  such that  $\Gamma_{k+1} \leq \Gamma_k$  for all  $k \geq k_0$ . In this case, we have  $\Gamma_k \leq \Gamma_{k_0}$  for all  $k \geq k_0$ , which is

$$\|\mathbf{x}^k - \mathbf{x}\|^2 \leq \Gamma_{k_0} + \mu^2 \sum_{j=1}^{k-1} \beta_j^2$$

for all  $k \geq k_0$ . Since the series  $\sum_{k=1}^\infty \beta_k^2$  converges, we obtain that the sequence  $\{\|\mathbf{x}^k - \mathbf{x}\|^2\}_{k=1}^\infty$  is bounded and, subsequently, the sequence  $\{\mathbf{x}^k\}_{k=1}^\infty$  is also bounded.

Case II: Suppose that there exists a subsequence  $\{\Gamma_{k_j}\}_{j=1}^\infty$  of  $\{\Gamma_k\}_{k=1}^\infty$  such that  $\Gamma_{k_j} < \Gamma_{k_{j+1}}$  for all  $k \in \mathbb{N}$ , and let  $\{v(k)\}_{k=1}^\infty$  be defined as in Proposition 3.2. This yields, for all  $k \geq k_0$ , that

$$\Gamma_{v(k)} < \Gamma_{v(k)+1} \quad (3.4)$$

and

$$\Gamma_k < \Gamma_{v(k)+1}. \quad (3.5)$$

By using the relation (3.4) in the inequality (3.3) and the fact that the term  $\frac{2\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}}$  is positive, we have

$$\langle \mathbf{x}^{v(k)} - \mathbf{x}, -\mathbf{d}^{v(k)} \rangle \leq 0.$$

Now, let us note that

$$0 \geq \langle \mathbf{x}^{v(k)} - \mathbf{x}, -\mathbf{d}^{v(k)} \rangle = \langle \mathbf{x}^{v(k)} - \mathbf{x}, -\nabla f(\mathbf{x}^{v(k)}) \rangle - \frac{\varphi_{v(k)}}{\max\{1, \|\mathbf{d}^{v(k)-1}\|\}} \langle \mathbf{x}^{v(k)} - \mathbf{x}, -\mathbf{d}^{v(k)-1} \rangle$$

and, since  $0 \leq \varphi_k \leq 1$ , we have

$$\begin{aligned} \langle \mathbf{x}^{v(k)} - \mathbf{x}, -\nabla f(\mathbf{x}^{v(k)}) \rangle &\leq \frac{\varphi_{v(k)}}{\max\{1, \|\mathbf{d}^{v(k)-1}\|\}} \langle \mathbf{x}^{v(k)} - \mathbf{x}, -\mathbf{d}^{v(k)-1} \rangle \\ &\leq \varphi_{v(k)} \|\mathbf{x}^{v(k)} - \mathbf{x}^*\| \leq \|\mathbf{x}^{v(k)} - \mathbf{x}\|. \end{aligned} \quad (3.6)$$

Since  $f$  is  $\eta$ -strongly convex, we have from Fact 1.1 that

$$\begin{aligned} \eta \|\mathbf{x}^{v(k)} - \mathbf{x}\|^2 &\leq \langle \nabla f(\mathbf{x}^{v(k)}) - \nabla f(\mathbf{x}), \mathbf{x}^{v(k)} - \mathbf{x} \rangle \\ &\leq \langle \nabla f(\mathbf{x}^{v(k)}), \mathbf{x}^{v(k)} - \mathbf{x} \rangle + \langle -\nabla f(\mathbf{x}), \mathbf{x}^{v(k)} - \mathbf{x} \rangle \\ &\leq \|\mathbf{x}^{v(k)} - \mathbf{x}\| + \|\nabla f(\mathbf{x})\| \|\mathbf{x}^{v(k)} - \mathbf{x}\|, \end{aligned}$$

where the last inequality holds by the inequality (3.6). Thus, we obtain that

$$\|\mathbf{x}^{v(k)} - \mathbf{x}\| \leq \eta^{-1}(1 + \|\nabla f(\mathbf{x})\|),$$

which implies that the sequence  $\{\|\mathbf{x}^{v(k)} - \mathbf{x}^*\|\}_{k=1}^\infty$  is bounded. Now, let us observe that

$$\Gamma_{v(k)+1} = \|\mathbf{x}^{v(k)+1} - \mathbf{x}\|^2 - \mu^2 \sum_{j=1}^{v(k)} \beta_j^2 \leq \|\mathbf{x}^{v(k)+1} - \mathbf{x}\|^2,$$

which means that the sequence  $\{\Gamma_{v(k)+1}\}_{k=1}^\infty$  is bounded above. Finally, by using the inequality (3.5), we obtain that  $\{\Gamma_k\}_{k=1}^\infty$  is bounded and, subsequently,  $\{\mathbf{x}^k\}_{k=1}^\infty$  is also bounded.  $\square$

A simple consequence of Lemma 3.3 is the boundedness of related sequences.

**Lemma 3.4.** *The sequences  $\{\nabla f(\mathbf{x}^k)\}_{k=1}^\infty$ ,  $\{\mathbf{d}^k\}_{k=1}^\infty$ , and  $\{\mathbf{y}^k\}_{k=1}^\infty$  are bounded.*

*Proof.* Let  $k \in \mathbb{N}$  and  $\mathbf{x} \in \bigcap_{i=1}^m \text{Fix } T_i$  be given. We first note from the  $L$ -smoothness of  $f$  that

$$\begin{aligned} \|\nabla f(\mathbf{x}^k)\| &\leq \|\nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x})\| + \|\nabla f(\mathbf{x})\| \\ &\leq L\|\mathbf{x}^k - \mathbf{x}\| + \|\nabla f(\mathbf{x})\| \leq L(M_1 + \|\mathbf{x}\|) + \|\nabla f(\mathbf{x})\| =: M, \end{aligned}$$

where  $M_1 := \sup_{k \in \mathbb{N}} \|\mathbf{x}^k\|$ . This means that the boundedness of the sequence  $\{\nabla f(\mathbf{x}^k)\}_{k=1}^\infty$  is now obtained.

Next, by the construction of  $\mathbf{d}^k$ , we note for all  $k \geq 1$  that

$$\|\mathbf{d}^{k+1}\| = \left\| -\nabla f(\mathbf{x}^{k+1}) + \varphi_{k+1} \frac{\mathbf{d}^k}{\max\{1, \|\mathbf{d}^k\|\}} \right\| \leq \|\nabla f(\mathbf{x}^{k+1})\| + \varphi_{k+1} \leq M + 1.$$

Thus, we have  $\|\mathbf{d}^k\| \leq \max\{M + 1, \|\mathbf{d}^1\|\}$  for all  $k \in \mathbb{N}$  and the boundedness of  $\{\mathbf{d}^k\}_{k=1}^\infty$  is obtained. Finally, these two obtained results immediately yield the boundedness of the sequence  $\{\mathbf{y}^k\}_{k=1}^\infty$ .  $\square$

**Lemma 3.5.** *For all  $k \geq 2$  and  $\mathbf{x} \in \bigcap_{i=1}^m \text{Fix } T_i$ , it holds that*

$$\|\mathbf{x}^{k+1} - \mathbf{x}\|^2 \leq (1 - \alpha_k)\|\mathbf{x}^k - \mathbf{x}\|^2 + \alpha_k \delta_k.$$

*Proof.* Let  $k \geq 2$  and  $\mathbf{x} \in \bigcap_{i=1}^m \text{Fix } T_i$  be given. We note from the inequality (3.1) that

$$\begin{aligned} \|\mathbf{x}^{k+1} - \mathbf{x}\|^2 &\leq \|\mathbf{y}^k - \mathbf{x}\|^2 \\ &= \left\| \mathbf{x}^k + \mu\beta_k \frac{\mathbf{d}^k}{\max\{1, \|\mathbf{d}^k\|\}} - \mathbf{x} \right\|^2 \\ &= \left\| \mathbf{x}^k + \frac{\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \left( -\nabla f(\mathbf{x}^k) + \varphi_k \frac{\mathbf{d}^{k-1}}{\max\{1, \|\mathbf{d}^{k-1}\|\}} \right) - \mathbf{x} \right\|^2 \\ &= \left\| \left( \mathbf{x}^k - \frac{\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \nabla f(\mathbf{x}^k) \right) - \left( \mathbf{x} - \frac{\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \nabla f(\mathbf{x}) \right) \right. \\ &\quad \left. + \frac{\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \left( \frac{\varphi_k \mathbf{d}^{k-1}}{\max\{1, \|\mathbf{d}^{k-1}\|\}} - \nabla f(\mathbf{x}) \right) \right\|^2 \\ &\leq \left\| \left( \mathbf{x}^k - \frac{\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \nabla f(\mathbf{x}^k) \right) - \left( \mathbf{x} - \frac{\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \nabla f(\mathbf{x}) \right) \right\|^2 \\ &\quad + 2 \left\langle \frac{\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \left( \frac{\varphi_k \mathbf{d}^{k-1}}{\max\{1, \|\mathbf{d}^{k-1}\|\}} - \nabla f(\mathbf{x}) \right), \mathbf{y}^k - \mathbf{x} \right\rangle \\ &\leq \left( 1 - \frac{\beta_k \tau}{\max\{1, \|\mathbf{d}^k\|\}} \right) \|\mathbf{x}^k - \mathbf{x}\|^2 \end{aligned}$$

$$\begin{aligned}
& + 2 \frac{\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \left\langle \frac{\varphi_k \mathbf{d}^{k-1}}{\max\{1, \|\mathbf{d}^{k-1}\|\}} - \nabla f(\mathbf{x}), \mathbf{y}^k - \mathbf{x} \right\rangle \\
= & \left( 1 - \frac{\beta_k \tau}{\max\{1, \|\mathbf{d}^k\|\}} \right) \|\mathbf{x}^k - \mathbf{x}\|^2 \\
& + \frac{\beta_k \tau}{\max\{1, \|\mathbf{d}^k\|\}} \left( \frac{2\mu}{\tau} \left( \frac{\varphi_k}{\max\{1, \|\mathbf{d}^{k-1}\|\}} \langle \mathbf{d}^{k-1}, \mathbf{y}^k - \mathbf{x} \rangle - \langle \nabla f(\mathbf{x}), \mathbf{y}^k - \mathbf{x} \rangle \right) \right) \\
= & (1 - \alpha_k) \|\mathbf{x}^k - \mathbf{x}\|^2 + \alpha_k \delta_k,
\end{aligned}$$

where the second inequality holds by the fact that  $\|\mathbf{x} + \mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + 2\langle \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle$  for all  $\mathbf{x}, \mathbf{y} \in \mathcal{H}$ , and the third inequality holds by Fact 1.3.  $\square$

The following proposition will be a key tool for obtaining the convergence result in Theorem 2.4. The idea and its proof can be consulted in [19, Lemma 2.6] and [20, Lemma 2.4].

**Proposition 3.6.** *Let  $\{a_k\}_{k=1}^\infty$  be a nonnegative real sequence,  $\{\delta_k\}_{k=1}^\infty$  be a real sequence, and  $\{\alpha_k\}_{k=1}^\infty$  be a real sequence in  $[0, 1]$  such that  $\sum_{k=1}^\infty \alpha_k = \infty$ . Suppose that*

$$a_{k+1} \leq (1 - \alpha_k)a_k + \alpha_k \delta_k \quad \text{for all } k \in \mathbb{N}.$$

*If  $\limsup_{j \rightarrow \infty} \delta_{k_j} \leq 0$  for every subsequence  $\{a_{k_j}\}_{j=1}^\infty$  of  $\{a_k\}_{k=1}^\infty$  satisfying*

$$\liminf_{j \rightarrow \infty} (a_{k_{j+1}} - a_{k_j}) \geq 0,$$

*then  $\lim_{k \rightarrow \infty} a_k = 0$ .*

Now, we are in a position to prove Theorem 2.4.

*Proof.* Let  $\mathbf{x}^*$  be the unique solution to the problem 1.1. For simplicity, we denote  $a_k := \|\mathbf{x}^k - \mathbf{x}^*\|^2$  for all  $k \in \mathbb{N}$ . Now, by using the facts obtained in Lemmata 3.3 and 3.4 and the fact that  $\lim_{k \rightarrow \infty} \beta_k = 0$ , we obtain

$$0 \leq \lim_{k \rightarrow \infty} |\varepsilon_k| = \lim_{k \rightarrow \infty} \left| \mu^2 \beta_k^2 + \frac{2\mu\beta_k}{\max\{1, \|\mathbf{d}^k\|\}} \langle \mathbf{x}^k - \mathbf{x}^*, \mathbf{d}^k \rangle \right| \leq 0,$$

which implies that  $\lim_{k \rightarrow \infty} \varepsilon_k = 0$ .

Next, we let a subsequence  $\{a_{k_j}\}_{j=1}^\infty$  of the sequence  $\{a_k\}_{k=1}^\infty$  satisfying

$$\liminf_{j \rightarrow \infty} (a_{k_{j+1}} - a_{k_j}) \geq 0$$

or, equivalently,

$$\limsup_{j \rightarrow \infty} (a_{k_j} - a_{k_{j+1}}) \leq 0.$$

By utilizing the inequality obtained in Lemma 3.1 and the fact  $\lim_{k \rightarrow \infty} \varepsilon_k = 0$ , we obtain

$$\begin{aligned}
0 \leq \limsup_{j \rightarrow \infty} \frac{\lambda_{k_j}(2 - \lambda_{k_j})}{4m} \sum_{i=1}^m \|S_i(\mathbf{y}^{k_j}) - S_{i-1}(\mathbf{y}^{k_j})\|^2 &\leq \limsup_{j \rightarrow \infty} (a_{k_j} - a_{k_j+1} + \varepsilon_{k_j}) \\
&= \limsup_{j \rightarrow \infty} (a_{k_j} - a_{k_j+1}) + \lim_{j \rightarrow \infty} \varepsilon_{k_j} \leq 0.
\end{aligned}$$

This implies that

$$\lim_{j \rightarrow \infty} \frac{\lambda_{k_j}(2 - \lambda_{k_j})}{4m} \sum_{i=1}^m \|S_i(\mathbf{y}^{k_j}) - S_{i-1}(\mathbf{y}^{k_j})\|^2 = 0.$$

Since  $\varepsilon^2 \leq \lambda_k(2 - \lambda_k)$ , we obtain that

$$\lim_{j \rightarrow \infty} \|S_i(\mathbf{y}^{k_j}) - S_{i-1}(\mathbf{y}^{k_j})\| = 0, \text{ for all } i = 1, 2, \dots, m. \quad (3.7)$$

On the other hand, since  $\{\mathbf{y}^k\}_{k=1}^\infty$  is a bounded sequence, so is the sequence  $\{\langle \mathbf{y}^{k_j} - \mathbf{x}^*, \nabla f(\mathbf{x}^*) \rangle\}_{j=1}^\infty$ . Now, let  $\{\mathbf{y}^{k_{j_\ell}}\}_{\ell=1}^\infty$  be a sequence of  $\{\mathbf{y}^{k_j}\}_{j=1}^\infty$  such that

$$\liminf_{j \rightarrow \infty} \langle \mathbf{y}^{k_j} - \mathbf{x}^*, \nabla f(\mathbf{x}^*) \rangle = \lim_{\ell \rightarrow \infty} \langle \mathbf{y}^{k_{j_\ell}} - \mathbf{x}^*, \nabla f(\mathbf{x}^*) \rangle.$$

Due to the boundedness of the sequence  $\{\mathbf{y}^{k_j}\}_{j=1}^\infty$ , there exists a weakly cluster point  $\mathbf{z} \in \mathcal{H}$  and a subsequence  $\{\mathbf{y}^{k_{j_\ell}}\}_{\ell=1}^\infty$  of  $\{\mathbf{y}^{k_j}\}_{j=1}^\infty$  such that  $\mathbf{y}^{k_{j_\ell}} \rightharpoonup \mathbf{z} \in \mathcal{H}$ . According to the obtained fact in (3.7), we note that

$$\lim_{\ell \rightarrow \infty} \|T_1(\mathbf{y}^{k_{j_\ell}}) - \mathbf{y}^{k_{j_\ell}}\| = \lim_{\ell \rightarrow \infty} \|S_1(\mathbf{y}^{k_{j_\ell}}) - S_0(\mathbf{y}^{k_{j_\ell}})\| = 0.$$

Since  $T_1$  satisfies the demi-closedness principle, we obtain that  $\mathbf{z} \in \text{Fix } T_1$ . Furthermore, we note that the facts  $\mathbf{y}^{k_{j_\ell}} \rightharpoonup \mathbf{z}$  and

$$\lim_{\ell \rightarrow \infty} \|(T_1(\mathbf{y}^{k_{j_\ell}}) - T_1(\mathbf{z})) - (\mathbf{y}^{k_{j_\ell}} - \mathbf{z})\| = 0$$

yield that  $T_1(\mathbf{y}^{k_{j_\ell}}) \rightharpoonup T_1(\mathbf{z}) = \mathbf{z}$ . Furthermore, we observe that

$$\lim_{\ell \rightarrow \infty} \|(T_2(T_1(\mathbf{y}^{k_{j_\ell}})) - T_1(\mathbf{y}^{k_{j_\ell}}))\| = \lim_{\ell \rightarrow \infty} \|S_2(\mathbf{y}^{k_{j_\ell}}) - S_1(\mathbf{y}^{k_{j_\ell}})\| = 0.$$

By invoking the assumption that  $T_2$  satisfies the demi-closedness principle, we also obtain that  $\mathbf{z} \in \text{Fix } T_2$ . By continuing the same argument used in the above proving lines, we obtain that  $\mathbf{z} \in \text{Fix } T_i$  for all  $i = 1, 2, \dots, m$ , then  $\mathbf{z} \in \bigcap_{i=1}^m \text{Fix } T_i$ . Since  $\mathbf{x}^*$  is the unique solution to the problem (1.1), we note from the optimality condition in Fact 1.2 that

$$\liminf_{j \rightarrow \infty} \langle \mathbf{y}^{k_j} - \mathbf{x}^*, \nabla f(\mathbf{x}^*) \rangle = \lim_{\ell \rightarrow \infty} \langle \mathbf{y}^{k_{j_\ell}} - \mathbf{x}^*, \nabla f(\mathbf{x}^*) \rangle = \langle \mathbf{z} - \mathbf{x}^*, \nabla f(\mathbf{x}^*) \rangle \geq 0. \quad (3.8)$$

Now, the assumption that  $\lim_{k \rightarrow \infty} \varphi_k = 0$ , the boundedness of the sequences  $\{\mathbf{y}^k\}_{k=1}^\infty$  and  $\{\mathbf{d}^k\}_{k=1}^\infty$ , and the relation (3.8) yield that

$$\limsup_{j \rightarrow \infty} \delta_{k_j} = \limsup_{j \rightarrow \infty} \frac{2\mu}{\tau} \left( \frac{\varphi_{k_j}}{\max\{1, \|\mathbf{d}^{k_j}\|\}} \langle \mathbf{d}^{k_j-1}, \mathbf{y}^{k_j} - \mathbf{x}^* \rangle - \langle \nabla f(\mathbf{x}^*), \mathbf{y}^{k_j} - \mathbf{x}^* \rangle \right) \leq 0.$$

Hence, by applying Proposition 3.6, we conclude that  $\lim_{k \rightarrow \infty} a_k = 0$ . The proof is completed.  $\square$

#### 4. Some illustrated consequences

In this section, we will provide some simple consequences of the main theorem. We start with the minimization problem over the system of linear inequalities. This constraint is nothing else but the intersection of linear half-spaces. We subsequently show that the proposed algorithm and convergence result are also applicable in the well-known support vector machine learning.

##### 4.1. Minimization problem over system of linear inequalities

In this subsection, we assume that the function  $f : \mathcal{H} \rightarrow \mathbb{R}$  is  $\eta$ -strongly convex and  $L$ -smooth as given in the assumption (A1). Now, let  $\mathbf{a}_i \in \mathcal{H}$  with  $\mathbf{a}_i \neq \mathbf{0}$  and  $b_i \in \mathbb{R}$ ,  $i = 1, \dots, m$ . We consider the minimization of a strongly convex smooth function over the intersection of nonempty closed and convex sets, which is in the following form:

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) \\ & \text{subject to} && \langle \mathbf{a}_i, \mathbf{x} \rangle \leq b_i, i = 1, \dots, m. \end{aligned} \quad (4.1)$$

We may assume the consistency of the system of linear inequalities and we also denote the solution point to the problem (4.1) by  $\mathbf{x}^*$ . Now, for each  $i = 1, \dots, m$ , we let  $H_i := \{\mathbf{x} \in \mathcal{H} : \langle \mathbf{a}_i, \mathbf{x} \rangle \leq b_i\}$  be the half-space corresponding to  $\mathbf{a}_i$  and  $b_i$ . Furthermore, we set  $T_i := P_{H_i}$ , the metric projection onto the half-space  $H_i$ , for all  $i = 1, \dots, m$ . It is worth noting that the metric projection onto a half-space has a closed-form expression, that is,

$$T_i := P_{H_i}(\mathbf{x}) = \mathbf{x} - \frac{\max\{\langle \mathbf{a}_i, \mathbf{x} \rangle - b_i, 0\}}{\|\mathbf{a}_i\|^2} \mathbf{a}_i,$$

see [17, Subsection 4.1.3] for further details. Note that  $H_i$  is a closed and convex set and the fixed-point set. Fix  $T_i = H_i$  for all  $i = 1, \dots, m$ . To derive an iterative method for solving the problem (4.1), we recall the notations  $T := T_m T_{m-1} \cdots T_1$ ,  $S_0 := Id$ , and  $S_i := T_i T_{i-1} \cdots T_1$ , for all  $i = 1, 2, \dots, m$ . Now, for every  $\mathbf{x} \in \mathcal{H}$ , by setting  $\mathbf{u}^i := S_i(\mathbf{x})$ , we have  $\mathbf{u}^0 = \mathbf{x}$  and  $\mathbf{u}^i = T_i(\mathbf{u}^{i-1})$ , for all  $i = 1, 2, \dots, m$ . In this case, the stepsize function  $\sigma : \mathcal{H} \rightarrow (0, \infty)$  can be written as the following:

$$\sigma(\mathbf{x}) := \begin{cases} \frac{\sum_{i=1}^m (\langle \mathbf{a}_i, \mathbf{x} \rangle - b_i) \left( \frac{\max\{\langle \mathbf{a}_i, \mathbf{u}^i \rangle - b_i, 0\}}{\|\mathbf{a}_i\|^2} \right)}{\|\mathbf{u}^m - \mathbf{x}\|^2}, & \text{if } \mathbf{x} \notin \bigcap_{i=1}^m H_i, \\ 1, & \text{otherwise,} \end{cases}$$

see [13, Section 4.2] for further details.

In order to solve the problem (4.1), we consider a particular situation of MESCOm-CGD as the following algorithm.

---

**Algorithm 2:** MESCoM-CGD for minimizing over the system of linear inequalities
 

---

**Initialization:** Put nonnegative sequences  $\{\beta_k\}_{k=1}^\infty$ ,  $\{\varphi_k\}_{k=1}^\infty$  and  $\{\lambda_k\}_{k=1}^\infty$ , and a parameter  $\mu \in (0, \frac{2\eta}{L^2})$ . Choose an initial point  $\mathbf{x}^1 \in \mathcal{H}$  arbitrarily and set  $\mathbf{d}^1 = -\nabla f(\mathbf{x}^1)$ .

**Iterative Step** ( $k \in \mathbb{N}$ ): For a current iterate  $\mathbf{x}^k \in \mathcal{H}$  and direction  $\mathbf{d}^k \in \mathcal{H}$ , repeat the following steps:

**Step 1.** Compute the iterate  $\mathbf{y}^k \in \mathcal{H}$  as

$$\mathbf{y}^k := \mathbf{x}^k + \mu\beta_k \frac{\mathbf{d}^k}{\max\{1, \|\mathbf{d}^k\|\}}.$$

**Step 2.** Put  $\mathbf{u}^{k,0} := \mathbf{y}^k$ . For each  $i = 1, 2, \dots, m$ , compute

$$\mathbf{u}^{k,i} := \mathbf{u}^{k,i-1} - \frac{\max\{\langle \mathbf{a}_i, \mathbf{u}^{k,i-1} \rangle - b_i, 0\}}{\|\mathbf{a}_i\|^2} \mathbf{a}_i.$$

Compute the stepsize  $\sigma(\mathbf{y}^k)$  as

$$\sigma(\mathbf{y}^k) := \begin{cases} \frac{\sum_{i=1}^m (\langle \mathbf{a}_i, \mathbf{y}^k \rangle - b_i) \left( \frac{\max\{\langle \mathbf{a}_i, \mathbf{u}^{k,i} \rangle - b_i, 0\}}{\|\mathbf{a}_i\|^2} \right)}{\|\mathbf{u}^{k,m} - \mathbf{y}^k\|^2}, & \text{if } \mathbf{y}^k \notin \bigcap_{i=1}^m H_i, \\ 1, & \text{otherwise.} \end{cases}$$

**Step 3.** Compute the recurrence  $\mathbf{x}^{k+1} \in \mathcal{H}$  and the search direction  $\mathbf{d}^{k+1} \in \mathcal{H}$  as

$$\mathbf{x}^{k+1} := \mathbf{y}^k + \lambda_k \sigma(\mathbf{y}^k) (\mathbf{u}^{k,m} - \mathbf{y}^k)$$

and

$$\mathbf{d}^{k+1} := -\nabla f(\mathbf{x}^{k+1}) + \varphi_{k+1} \frac{\mathbf{d}^k}{\max\{1, \|\mathbf{d}^k\|\}}.$$

**Step 4.** Update  $k := k + 1$  and return to **Step 1**.

---

We obtain an immediate consequence of Theorem 2.4 in the following corollary.

**Corollary 4.1.** *Let the sequence  $\{\mathbf{x}^k\}_{k=1}^\infty$  be generated by Algorithm 2 and assume that assumption (A1) and hypotheses (H1)–(H3) hold, then the sequence  $\{\mathbf{x}^k\}_{k=1}^\infty$  converges strongly to the unique solution  $\mathbf{x}^*$  of the problem (4.1).*

To examine the behavior of Algorithm 2 and the convergence given in Corollary 4.1, we consider the solving of the minimum-norm problem to the system of homogeneous linear inequalities. We assume that the whole space  $\mathcal{H}$  is finite dimensional so that  $\mathcal{H} = \mathbb{R}^n$ . Given a matrix  $\mathbf{A} = [\mathbf{a}_1 | \dots | \mathbf{a}_m]^T$  of predictors  $\mathbf{a}_i = (\mathbf{a}_{1i}, \dots, \mathbf{a}_{ni}) \in \mathbb{R}^n$ , for all  $i = 1, \dots, m$ ,  $\mathbf{b} = (b_1, \dots, b_m) \in \mathbb{R}^m$  is a vector. The



minimum-norm problem is to find the vector  $\mathbf{x} \in \mathbb{R}^n$  that solves the problem

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|\mathbf{x}\|^2 \\ & \text{subject to} && \mathbf{Ax} \leq \mathbf{b} \end{aligned} \quad (4.2)$$

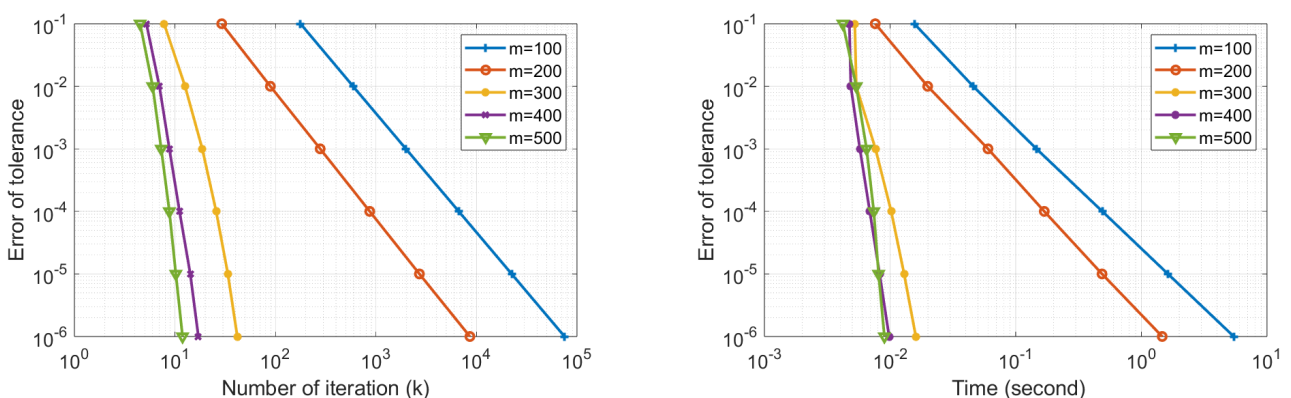
or, equivalently, in the explicit form

$$\begin{aligned} & \text{minimize} && \frac{1}{2} \|\mathbf{x}\|^2 \\ & \text{subject to} && \langle \mathbf{a}_i, \mathbf{x} \rangle \leq b_i, i = 1, \dots, m. \end{aligned}$$

Now, by putting the constrained sets  $H_i := \{\mathbf{x} \in \mathbb{R}^n : \langle \mathbf{a}_i, \mathbf{x} \rangle \leq b_i\}$ ,  $i = 1, \dots, m$  as half spaces and  $T_i := P_{H_i}$ ,  $i = 1, \dots, m$ , the metric projections onto  $H_i$  with  $\text{Fix } T_i = H_i$  satisfy the demi-closed principle. Furthermore, the function  $f(\mathbf{x}) := \frac{1}{2} \|\mathbf{x}\|^2$  is 1-strongly convex with 1-smooth. In this situation, the problem (4.2) is nothing else but a special case of the problem (4.1), which yields that Algorithm 2 and the convergence given in Corollary 4.1 are applicable.

To perform a numerical illustration in solving the problem (4.1), we generate the matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , where  $m = 200$  and  $n = 100$  by uniformly distributed random generating between  $(-5, 5)$ , and set the vector  $\mathbf{b} = \mathbf{0}$ . According to Remark 2.3 (i)-(iii), we examine the influence of parameters  $\beta_0 \in [0.8, 1]$ ,  $\varphi_0 \in [0.1, 1]$ , and  $\lambda_0 \in [0.1, 1.9]$ . We fix the corresponding parameter  $\mu = 1.9$ . We terminated the experiment with the error of norm, that is,  $\|\mathbf{x}^{k+1} - \mathbf{x}^k\| < \epsilon$ . We manually choose the choice of parameters  $\beta_0 = 1$ ,  $\varphi_0 = 1$ , and  $\lambda_0 = 1$  with the smallest number of iterations when the error of tolerance  $\epsilon = 10^{-4}$  is met.

Next, we show behavior of Algorithm 2 when solving the problem (4.2) for various error of tolerance  $\epsilon$ . We set number of variables to be  $n = 100$  and consider several number of inequalities, that is,  $m = 100, 200, 300, 400$ , and  $500$ . According to the above experiments, we set the parameters  $\beta_0 = 1$ ,  $\varphi_0 = 1$ , and  $\lambda_0 = 1$ . We plot the number of iterations and computational time in seconds in Figure 1.



**Figure 1.** Behavior of Algorithm 2 for various errors of tolerance.

It can be noticed from Figure 1 that a smaller number  $m$  needed a larger number of iterations for all errors of tolerance. Moreover, for the numbers  $m = 100, 200$ , and  $300$ , it can be seen that a smaller number  $m$  needed a larger computational time.

#### 4.2. Support vector machine learning problem

In this subsection, we consider the constructing of a classifier in binary classification problem starting by a given training datasets of two classes. More precisely, the  $i$ th training data  $\mathbf{a}_i \in \mathbb{R}^n$  and the  $i$ th label  $b_i \in \{-1, +1\}$  of the  $i$ th training data, for all  $i = 1, \dots, m$ . The support vector machine learning is to train a weight  $\mathbf{u} \in \mathbb{R}^n$  so that a (linear) classifier  $c(\mathbf{a}) := \langle \mathbf{a}, \mathbf{u} \rangle$  can give a corrected class of every new tested data  $\mathbf{a} \in \mathbb{R}^n$ . In this situation,  $\mathbf{a}$  will be identified to be in the class  $-1$  if  $c(\mathbf{a}) < 0$ , and to the class  $+1$ , otherwise. Mathematically, the support vector machine learning can be form as the following minimization problem:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} \|\mathbf{u}\|^2 + \frac{1}{2} \sum_{i=1}^m \xi_i^2 \\ \text{subject to} \quad & b_i \langle \mathbf{a}_i, \mathbf{u} \rangle \geq 1 - \xi_i, \forall i = 1, \dots, m, \\ & \xi_i \geq 0, \forall i = 1, \dots, m, \end{aligned} \quad (4.3)$$

where  $\xi_i \geq 0$  is nonnegative slack variable corresponding to the  $i$ th training data for all  $i = 1, \dots, m$ . By introducing a new variable  $\mathbf{x} := (\mathbf{u}, \xi_1, \dots, \xi_m) \in \mathbb{R}^{n+m}$  and using the idea given in, for instance [15, 21], the problem (4.3) can be written as

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} \|\mathbf{x}\|^2 \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \leq \mathbf{b}, \end{aligned} \quad (4.4)$$

where the matrix  $\mathbf{A} \in \mathbb{R}^{2m \times (n+m)}$  is given by

$$\mathbf{A} = \begin{bmatrix} -b_1 \mathbf{a}_1^\top & -1 & 0 & \cdots & 0 \\ -b_2 \mathbf{a}_2^\top & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -b_m \mathbf{a}_m^\top & 0 & 0 & \cdots & -1 \\ \mathbf{0}_{\mathbb{R}^n}^\top & -1 & 0 & \cdots & 0 \\ \mathbf{0}_{\mathbb{R}^n}^\top & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_{\mathbb{R}^n}^\top & 0 & 0 & \cdots & -1 \end{bmatrix} \in \mathbb{R}^{2m \times (n+m)}$$

and the vector  $\mathbf{b} \in \mathbb{R}^{2m}$  is given by

$$\mathbf{b} = \begin{pmatrix} -\mathbf{1}_{\mathbb{R}^m} \\ \mathbf{0}_{\mathbb{R}^m} \end{pmatrix}.$$

This problem (4.4) is nothing else but a particular case of the problem (4.2) and, subsequently, Algorithm 2 and the convergence given in Corollary 4.1 are also applicable.

In the first experiment, we aim to classify the  $28 \times 28$  images on gray scale pixels of the handwritten digits from the MNIST dataset, which was provided as <https://cs.nyu.edu/roweis/data.html>. We used the dataset of 5000 images for the handwritten digit 9 and the dataset of 5000 images for the handwritten digits 0 – 8. The images are labeled by the classes  $+1$  and  $-1$ , respectively. We perform the 10-fold cross-validation on the given datasets. Actually, in each class, we put a fold of 1000 images to be the testing data and the remaining 9 folds consisting of 9000 images to be the training data. We perform the cross-validation process repeatedly for 10 times so that each fold is set to be the testing data.

Recalling that the class of digit 9 and the class of digits 0 – 8 are labeled by +1 and –1, which are positive and negative, respectively, we denote the numbers of a tested data. It is classified as positive by true positive (TP); if classified as negative, then it is denoted by false negative (FN). If a tested data is labeled as negative and is classified as negative, then it is denoted as true negative (TN); if it is classified as positive, it is denoted as false positive (FP). These numbers are summarized as the following details:

Actual Class	Classified Class	
	Positive (+1)	Negative (-1)
Positive (+1)	TP := True Positive	FN := False Negative
Negative (-1)	FP := False Positive	TN := True Negative

To measure the performance of each obtained classifier performed by each cross-validation, we consider classification performance metrics, namely, accuracy, precision, recall, specificity, and F-measure. These performance metrics are computed by the following details:

$$\begin{aligned}
 \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\
 \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\
 \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\
 \text{Specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \\
 \text{F-measure} &= \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}.
 \end{aligned}$$

In the experiment, we terminate the experiment when the number of iterations is  $k = 50$  and, subsequently, average each performance metric after the cross-validation process repeatedly for 10 times. The classification performance metrics of each parameter are presented in Table 1.

**Table 1.** Classification performance metrics for varying parameters  $\beta_0$  and  $\varphi_0$ .

$\beta_0$	$\varphi_0$	Accuracy	Precision	Recall	Specificity	F-measure	Time
0.5	0.1	0.9325	0.9507	0.9173	0.9489	0.9337	340.23
	0.3	0.9325	0.9507	0.9173	0.9489	0.9337	339.46
	0.5	0.9325	0.9507	0.9173	0.9489	0.9337	347.93
	0.7	0.9325	0.9507	0.9173	0.9489	0.9337	341.14
	0.9	0.9325	0.9507	0.9173	0.9489	0.9337	352.12
0.6	0.1	0.9323	0.9507	0.9171	0.9489	0.9335	338.62
	0.3	0.9323	0.9507	0.9171	0.9489	0.9335	339.44
	0.5	0.9323	0.9507	0.9171	0.9489	0.9335	343.61
	0.7	0.9323	0.9507	0.9171	0.9489	0.9335	341.39
	0.9	0.9323	0.9507	0.9171	0.9489	0.9335	356.79
0.7	0.1	0.9321	0.9507	0.9167	0.9489	0.9333	339.02
	0.3	0.9321	0.9507	0.9167	0.9489	0.9333	340.37
	0.5	0.9321	0.9507	0.9167	0.9489	0.9333	343.67
	0.7	0.9321	0.9507	0.9167	0.9489	0.9333	341.12
	0.9	0.9321	0.9507	0.9167	0.9489	0.9333	357.62
0.8	0.1	0.9322	0.9509	0.9167	0.9491	0.9335	341.26
	0.3	0.9322	0.9509	0.9167	0.9491	0.9335	341.02
	0.5	0.9322	0.9509	0.9167	0.9491	0.9335	342.17
	0.7	0.9322	0.9509	0.9167	0.9491	0.9335	347.23
	0.9	0.9322	0.9509	0.9167	0.9491	0.9335	342.92
0.9	0.1	0.9322	0.9509	0.9167	0.9491	0.9335	339.05
	0.3	0.9322	0.9509	0.9167	0.9491	0.9335	340.88
	0.5	0.9322	0.9509	0.9167	0.9491	0.9335	340.90
	0.7	0.9322	0.9509	0.9167	0.9491	0.9335	346.44
	0.9	0.9322	0.9509	0.9167	0.9491	0.9335	342.06
1.0	0.1	0.9323	0.9511	0.9168	0.9493	0.9336	338.86
	0.3	0.9323	0.9511	0.9168	0.9493	0.9336	340.88
	0.5	0.9323	0.9511	0.9168	0.9493	0.9336	340.48
	0.7	0.9323	0.9511	0.9168	0.9493	0.9336	346.35
	0.9	0.9323	0.9511	0.9168	0.9493	0.9336	341.66

The results given in Table 1 shows that the parameters  $\beta_0$  and  $\varphi_0$  slightly effected the classification performances. The highest accuracy and F-measure values were obtained for the case  $\beta_0 = 0.5$ . Actually, we can observe that these five metrics as well as the computation running times were not much different. This may yield the stability of the proposed method in the sense that the corresponding parameters do not hugely effect the convergence of the proposed method.

## 5. Conclusions

We presented the modified version of the extrapolated sequential constraint method proposed in [11] for solving the minimization problem over the intersection of the fixed-point sets of cutter operators.

We not only presented a simple version of the strong convergence of the proposed method, but also omitted the boundedness assumption used in [11]. We examined the proposed method to solve the binary classification by using support vector machines.

### Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

### Acknowledgments

The authors are grateful to three anonymous reviewers for their comments and suggestions that improved the paper's quality and presentation. This work has received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B05F650018]. D. Rakjarungkiat gratefully acknowledges the research capability enhancement program through graduate student scholarship, Faculty of Science, Khon Kaen University.

### Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

### References

1. H. H. Bauschke, The approximation of fixed points of compositions of nonexpansive mappings in Hilbert space, *J. Math. Anal. Appl.*, **202** (1996), 150–159. <https://doi.org/10.1006/jmaa.1996.0308>
2. I. Yamada, The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings, In: *Inherently parallel algorithms in feasibility and optimization and their applications*, Elsevier, 2001, 473–504.
3. H. K. Xu, T. H. Kim, Convergence of hybrid steepest-descent methods for variational inequalities, *J. Optim. Theory Appl.*, **119** (2003), 185–201. <https://doi.org/10.1023/B:JOTA.0000005048.79379.b6>
4. A. Cegielski, Extrapolated simultaneous subgradient projection method for variational inequality over the intersection of convex subsets, *J. Nonlinear Convex Anal.*, **15** (2014), 211–218.
5. A. Cegielski, A. Gibali, S. Reich, R. Zalas, An algorithm for solving the variational inequality problem over the fixed point set of a quasi-nonexpansive operator in Euclidean space, *Numer. Funct. Anal. Optim.*, **34** (2013), 1067–1096. <https://doi.org/10.1080/01630563.2013.771656>
6. A. Cegielski, R. Zalas, Methods for variational inequality problem over the intersection of fixed point sets of quasi-nonexpansive operators, *Numer. Funct. Anal. Optim.*, **34** (2013), 255–283. <https://doi.org/10.1080/01630563.2012.716807>
7. S. Sabach, S. Shtern, A first order method for solving convex bilevel optimization problems, *SIAM J. Optim.*, **27** (2017), 640–660. <https://doi.org/10.1137/16M105592X>

8. B. Tan, S. Li, Strong convergence of inertial Mann algorithms for solving hierarchical fixed point problems, *J. Nonlinear Var. Anal.*, **4** (2020), 337–355. <http://dx.doi.org/10.23952/jnva.4.2020.3.02>
9. B. Tan, X. Qin, A. Gibali, Three approximation methods for solving constraint variational inequalities and related problems, *Pure Appl. Funct. Anal.*, **8** (2023), 965–986.
10. M. Prangprakhon, N. Nimana, N. Petrot, A sequential constraint method for solving variational inequality over the intersection of fixed point sets, *Thai J. Math.*, **18** (2020), 1105–1123.
11. M. Prangprakhon, N. Nimana, Extrapolated sequential constraint method for variational inequality over the intersection of fixed-point sets, *Numer. Algorithms*, **88** (2021), 1051–1075. <https://doi.org/10.1007/s11075-021-01067-z>
12. H. Iiduka, I. Yamada, A use of conjugate gradient direction for the convex optimization problem over the fixed point set of a nonexpansive mapping, *SIAM J. Optim.*, **19** (2009), 1881–1893. <https://doi.org/10.1137/070702497>
13. A. Cegielski, Y. Censor, Extrapolation and local acceleration of an iterative process for common fixed point problems, *J. Math. Anal. Appl.*, **394** (2012), 809–818. <https://doi.org/10.1016/j.jmaa.2012.04.072>
14. A. Cegielski, N. Nimana, Extrapolated cyclic subgradient projection methods for the convex feasibility problems and their numerical behaviour, *Optimization*, **68** (2019), 145–161. <https://doi.org/10.1080/02331934.2018.1509214>
15. N. Petrot, M. Prangprakhon, P. Promsinchai, N. Nimana, A dynamic distributed conjugate gradient method for variational inequality problem over the common fixed-point constraints. *Numer. Algorithms*, **93** (2023), 639–668. <https://doi.org/10.1007/s11075-022-01430-8>
16. A. Beck, *First-ordered methods in optimization*, Philadelphia: SIAM, 2017.
17. A. Cegielski, *Iterative methods for fixed point problems in Hilbert spaces*, Berlin, Heidelberg: Springer, 2012. <https://doi.org/10.1007/978-3-642-30901-4>
18. P. E. Mainge, Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization, *Set Valued Anal.*, **16** (2008), 899–912. <https://doi.org/10.1007/s11228-008-0102-z>
19. S. Saejung, P. Yotkaew, Approximation of zeros of inverse strongly monotone operators in Banach spaces, *Nonlinear Anal.*, **75** (2012), 742–750. <https://doi.org/10.1016/j.na.2011.09.005>
20. C. Jaipranop, S. Saejung, On Halpern-type sequences with applications in variational inequality problems, *Optimization*, **71** (2020), 675–710. <https://doi.org/10.1080/02331934.2020.1812065>
21. R. I. Boş, E. R. Csetnek, N. Nimana, Gradient-type penalty method with inertial effects for solving constrained convex optimization problems with smooth data, *Optim. Lett.*, **12** (2018), 17–33. <https://doi.org/10.1007/s11590-017-1158-1>

