




Article

# Proximal Gradient Method for Solving Bilevel Optimization Problems

Seifu Endris Yimer <sup>1,2</sup> , Poom Kumam <sup>1,3,\*</sup>  and Anteneh Getachew Gebrie <sup>2</sup> 

<sup>1</sup> KMUTT Fixed Point Research Laboratory, SCL 802 Fixed Point Laboratory & Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), 126 Pracha-Uthit Road, Bang Mod, Thung Khru, Bangkok 10140, Thailand; seifu.endris@mail.kmutt.ac.th

<sup>2</sup> Department of Mathematics, College of Computational and Natural Science, Debre Berhan University, P.O. Box 445 Debre Berhan, Ethiopia; antgetm@gmail.com

<sup>3</sup> Center of Excellence in Theoretical and Computational Science (TaCS-CoE), Science Laboratory Building, Faculty of Science, King Mongkut's University of Technology Thonburi (KMUTT), 126 Pracha-Uthit Road, Bang Mod, Thung Khru, Bangkok 10140, Thailand

\* Correspondence: poom.kum@mail.kmutt.ac.th

Received: 1 June 2020; Accepted: 17 August 2020; Published: 4 October 2020



**Abstract:** In this paper, we consider a bilevel optimization problem as a task of finding the optimum of the upper-level problem subject to the solution set of the split feasibility problem of fixed point problems and optimization problems. Based on proximal and gradient methods, we propose a strongly convergent iterative algorithm with an inertia effect solving the bilevel optimization problem under our consideration. Furthermore, we present a numerical example of our algorithm to illustrate its applicability.

**Keywords:** bilevel problem; fixed point problem; proximal method; gradient method

## 1. Introduction

Let  $H$  be a real Hilbert space and consider the *constrained minimization problem*:

$$\begin{aligned} \min h \\ \text{s.t. } x \in C \end{aligned} \quad (1)$$

where  $C$  is a nonempty closed convex subset of  $H$  and  $h : H \rightarrow \mathbb{R}$  is a convex and continuously differentiable function. The gradient–projection algorithm (GPA, for short) is usually applied to solve the minimization problem (1) and has been studied extensively by many authors; see, for instance, [1–3] and references therein. This algorithm generates a sequence  $\{x_n\}$  through the recursion:

$$x_{n+1} = P_C(x_n - \gamma \nabla h(x_n)), \quad (2)$$

where  $\nabla h$  is the gradient of  $h$ ,  $x_0$  is the initial guess chosen arbitrarily from  $C$ ,  $\gamma$  is a stepsize which may be chosen in different ways, and  $P_C$  is the metric projection from  $H$  onto  $C$ . By the optimality condition on problem (1), it follows that

$$\bar{x} \in C \text{ solves (1) if and only if } \langle \nabla h(\bar{x}), y - \bar{x} \rangle \geq 0, \quad \forall y \in C.$$

If  $\nabla h$  is Lipschitz continuous and strongly monotone, i.e., there exists  $L_h > 0$  and  $\sigma > 0$  such that for all  $x, y \in H$ ,

$$\|\nabla h(x) - \nabla h(y)\| \leq L_h \|x - y\| \text{ and } \langle \nabla h(x) - \nabla h(y), x - y \rangle \geq \sigma \|x - y\|^2,$$

then the operator  $T_\gamma = P_C(I - \gamma \nabla h)$  is a contraction provided that  $0 < \gamma < \frac{2\sigma}{L_h^2}$ . Therefore, for  $0 < \gamma < \frac{2\sigma}{L_h^2}$ , we can apply Banach’s contraction principle to get that the sequence  $\{x_n\}$  defined by (2) converges strongly to the unique fixed point of  $T_\gamma$  (or the unique solution of the minimization (1)). Moreover, if you set  $C = H$  in (1), then we have an unconstrained optimization problem, and hence the gradient algorithm

$$x_{n+1} = x_n - \gamma \nabla h(x_n)$$

generates a sequence  $\{x_n\}$  strongly convergent to the global minimizer point of  $h$ .

Consider the other most well-known problem called *unconstrained minimization problem*:

$$\begin{aligned} & \min g \\ & \text{s.t. } x \in H, \end{aligned} \tag{3}$$

where  $H$  is a real Hilbert space and  $g : H \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper, convex, lower semicontinuous function. An analogous method for solving (3) with better properties is based on the notion of proximal mapping introduced by Moreau [4], i.e., the proximal operator of the function  $g$  with scaling parameter  $\lambda > 0$  is a mapping  $\text{prox}_{\lambda g} : H \rightarrow H$  given by

$$\text{prox}_{\lambda g}(x) = \arg \min_{y \in H} \{g(y) + \frac{1}{2\lambda} \|x - y\|^2\}.$$

Proximal operators are firmly nonexpansive and the optimality condition of (3) is

$$\bar{x} \in H \text{ solves (3) if and only if } \text{prox}_{\lambda g}(\bar{x}) = \bar{x}.$$

Many properties of proximal operator can be found in [5] and the references therein. We know that the so called *proximal point algorithm*, i.e.,  $x_{n+1} = \text{prox}_{\lambda g}(x_n)$ , is the most popular method solving optimization problem (3) (introduced by Martinet [6,7] and later by Rockafellar [8]).

The split inverse problem (SIP) [9] is formulated by linking problems installed in two different places  $X$  and  $Y$  connected by a linear transformations, i.e., SIP is a problem of finding a point in space  $X$  solving a problem IP1 installed in  $X$  and its image under linear transformation solves a problem IP2 installed in another space  $Y$ . The presence of step size choice dependent on operator norm is not quite recommended in the iterative method of solving SIPs, as it is not always easy to estimate the norm of an operator; see, for example, the Theorem of Hendrickx and Olshevsky in [10]. For example, in the early study of the iterative method of solving the split feasibility problem [11–13], the determination of the step-size depends on the operator norm (or at least estimate value of the operator norm) and this is not as easy of a task. To overcome this difficulty, Lopez et al. [14] introduced a new way of selecting the step sizes that the information of operator norm is not necessary for solving a split feasibility problem (SFP):

$$\text{find } \bar{x} \in C \text{ such that } A\bar{x} \in Q$$

where  $C$  and  $Q$  are closed convex subsets of real Hilbert spaces  $H_1$  and  $H_2$ , respectively. To be precise, Lopez et al. [14] introduced an iterative algorithm that generates a sequence  $\{x_n\}$  by

$$x_{n+1} = P_C(I - \gamma_n A^*(I - P_Q)A)x_n. \tag{4}$$

The parameter  $\gamma_n$  appeared in (4) by  $\gamma_n = \frac{\rho_n l(x_n)}{\|\nabla l(x_n)\|^2}$ ,  $\forall n \geq 1$ , where  $\rho_n \in (0, 4)$ ,  $l(x_n) = \frac{1}{2} \|(I - P_Q)Ax_n\|^2$  and  $\nabla l(x_n) = A^*(I - P_Q)Ax_n$ .

A bilevel problem is a two-level hierarchical problem such that the solution of the lower level problem determines the feasible space of the upper level problem. In general, Yimer et al. [15] presented a bilevel problem as an archetypal model given by

$$\text{find } \bar{x} \in S \subset X \text{ that solves problem P1 installed in a space } X, \tag{5}$$

where  $S$  is the solution set of the problem

$$\text{find } x^* \in Y \subset X \text{ that solves problem P2 installed in a space } X. \tag{6}$$

According to [16], the bilevel problem (problem (5) and (6)) is a hierarchical game of two players as decision makers who make their decisions according to a hierarchical order. The problem is also called the leader’s and follower’s problem where the problem (5) is called the leader’s problem and (6) is called the follower’s problem, meaning, the first player (which is called the leader) makes his selection first and communicates it to the second player (the so-called follower). There are many studies for several type bilevel problems, see, for example, [15,17–24]. The bilevel optimization problem is a bilevel problem when the hierarchical structure involves the optimization problem. Bilevel optimization problems have become an increasingly important class of optimization problems during the last few years and decades due their to vast application of solving the real life problems. For example, in toll-setting problem [25], in chemical engineering [26], in electricity markets [27], and in supply chain problems [28].

Motivated by the above theoretical results and inspired by the applicability of the bilevel problem, we consider the following *bilevel optimization problem* given by

$$\begin{aligned} & \min h \\ & \text{s.t. } x \in \bigcap_{i=1}^N \text{Fix}U_i, \\ & A(x) \in \bigcap_{j=1}^M \arg \min g_j, \end{aligned} \tag{7}$$

where  $A : H_1 \rightarrow H_2$  is a linear transformation,  $h : H_1 \rightarrow \mathbb{R}$  is convex function,  $g_j : H_2 \rightarrow \mathbb{R} \cup \{+\infty\}$  is convex nonsmooth function, and  $\arg \min g_j = \{\bar{z} \in H_2 : g_j(\bar{z}) \leq g_j(z), \forall z \in H_2\}$  for  $j \in \{1, \dots, M\}$ ,  $U_i : H_1 \rightarrow H_1$  is demimetric mapping and  $\text{Fix}U_i = \{x \in H_1 : U_i(x) = x\}$  for  $i \in \{1, \dots, N\}$ , and  $H_1$  and  $H_2$  are two real Hilbert spaces.

For a real Hilbert space  $H$ , the mapping  $U : H \rightarrow H$  with  $\text{Fix}U \neq \emptyset$  is called  $\omega$ -demimetric if  $\omega \in (-\infty, 1)$  and

$$\langle x - \bar{x}, x - Ux \rangle \geq \frac{1 - \omega}{2} \|x - Ux\|^2, \quad \forall x \in H, \forall \bar{x} \in \text{Fix}U. \tag{8}$$

The demimetric mapping is introduced by Takahashi [29] in a smooth, strictly convex and reflexive Banach space. For a real Hilbert space  $H$ , (8) is equivalent to the following:

$$\|Ux - \bar{x}\|^2 \leq \|x - \bar{x}\|^2 + \omega \|x - Ux\|^2, \quad \forall x \in H, \forall \bar{x} \in \text{Fix}U,$$

and  $\text{Fix}U$  is a closed and convex subset of  $H$  [29]. The class of demimetric mappings contains the classes of strict pseudocontractions, firmly quasi-nonexpansive mappings, and quasi-nonexpansive mappings, see [29,30] and the references therein.

Assume that  $\Omega$  is the set of solutions of lower level problems of the bilevel optimization problem (7), that is,

$$\Omega = \left\{ x \in \bigcap_{i=1}^N \text{Fix}U_i : A(x) \in \bigcap_{j=1}^M \arg \min g_j \right\}. \tag{9}$$

Therefore, the *bilevel optimization problem* (7) is simply

$$\begin{aligned} & \min h \\ & \text{s.t. } x \in \Omega, \end{aligned}$$

where  $\Omega$  is given by (9). If  $H_1 = H_2 = H$ ,  $A = I$  (identity operator),  $g = g_j$  for all  $j \in \{1, \dots, M\}$ , the problem (7) is reduced to the bilevel optimization problem:

$$\begin{aligned} & \min h \\ & \text{s.t. } x \in \arg \min g. \end{aligned} \tag{10}$$

Bilevel problems like (10) have already been considered in the literature, for example, [23,31,32] for the case  $H = \mathbb{R}^p$ .

Note that, to the best of our knowledge, the bilevel optimization problem (7), with a finite intersection of fixed point sets of the broadest class of nonlinear mappings and finite intersection of minimize point sets of non-smooth functions as a lower level, has not been addressed before.

An inertial term is a two-step iterative method, and the next iterate is defined by making use of the previous two iterates. It is firstly introduced by Polyak [33] as an acceleration process in solving a smooth convex minimization problem. It is well known that combining algorithms with an inertial term speeds up or accelerates the rate of convergence of the sequence generated by the algorithm. In this paper, we introduce a proximal gradient inertial algorithm with a strong convergence result for approximating a bilevel optimization problem (7), where our algorithm is designed to address a way of selecting the step-sizes such that its implementation does not need any prior information about the operator norm.

## 2. Preliminary

Let  $C$  be a nonempty closed convex subset of a real Hilbert space  $H$ . The metric projection on  $C$  is a mapping  $P_C : H \rightarrow C$  defined by

$$P_C(x) = \arg \min\{\|y - x\| : y \in C\}, \quad x \in H.$$

For  $x \in H$  and  $z \in C$ , then  $z = P_C(x)$  if and only if  $\langle x - z, y - z \rangle \leq 0, \quad \forall y \in C$ .

Let  $T : H \rightarrow H$ . Then,

(a)  $T$  is  $L$ -Lipschitz if there exists  $L > 0$  such that

$$\|Tx - Ty\| \leq L\|x - y\|, \quad \forall x, y \in H.$$

If  $L \in (0, 1)$ , then we call  $T$  a contraction with constant  $L$ . If  $L = 1$ , then  $T$  is called a nonexpansive mapping.

(b)  $T$  is strongly monotone if there exists  $\sigma > 0$  such that

$$\langle Tx - Ty, x - y \rangle \geq \sigma\|x - y\|^2, \quad \forall x, y \in H.$$

In this case,  $T$  is called  $\sigma$ -strongly monotone.

(c)  $T$  is firmly nonexpansive if

$$\|Tx - Ty\|^2 \leq \|x - y\|^2 - \|(I - T)x - (I - T)y\|^2, \quad \forall x, y \in H,$$

which is equivalent to

$$\|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y \rangle, \quad \forall x, y \in H.$$

If  $T$  is firmly nonexpansive,  $I - T$  is also firmly nonexpansive.

Let  $H$  be a real Hilbert space. If  $G : H \rightarrow 2^H$  is maximal monotone set-valued mapping, then we define the resolvent operator  $J_\lambda^G$  associated with  $G$  and  $\lambda > 0$  as follows:

$$J_\lambda^G(x) = (I + \lambda G)^{-1}(x), \quad x \in H.$$

It is well known that  $J_\lambda^G$  is single-valued, nonexpansive, and 1-inverse strongly monotone (firmly nonexpansive). Moreover,  $0 \in G(\bar{x})$  if and only if  $\bar{x}$  is a fixed point of  $J_\lambda^G$  for all  $\lambda > 0$ ; see more about maximal monotone and its associated resolvent operator and examples of maximal monotone operators in [34].

The subdifferential of a convex function  $f : H \rightarrow \mathbb{R} \cup \{+\infty\}$  at  $x \in H$ , denoted by  $\partial f(x)$ , is defined by

$$\partial f(x) = \{\xi \in H : f(z) \geq f(x) + \langle \xi, z - x \rangle, \quad \forall z \in H\}.$$

If  $\partial f(x) \neq \emptyset$ ,  $f$  is said to be subdifferentiable at  $x$ . If the function  $f$  is continuously differentiable, then  $\partial f(x) = \{\nabla f(x)\}$ ; this is the gradient of  $f$ . If  $f$  is a proper, lower semicontinuous function, the subdifferential operator is a maximal monotone operator, and the proximal operator is the resolvent of the subdifferential operator (see, for example, in [5]), i.e.,

$$\text{prox}_{\lambda f} = J_\lambda^{\partial f} = (I + \lambda \partial f)^{-1}$$

Thus, this results in proximal operators being firmly nonexpansive, and a point  $\bar{x}$  minimizes  $f$  if and only if  $\text{prox}_{\lambda f}(\bar{x}) = \bar{x}$ .

**Definition 1.** Let  $H$  be a real Hilbert space. A mapping  $U : H \rightarrow H$  is called demiclosed if, for a sequence  $\{x_n\}$  in  $H$  such that  $\{x_n\}$  converges weakly to  $\bar{x}$  and  $\lim_{n \rightarrow \infty} \|x_n - Ux_n\| = 0$ ,  $U\bar{x} = \bar{x}$  holds.

**Lemma 1.** For a real Hilbert space  $H$ , we have

- (i)  $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y \rangle, \quad \forall x, y \in H;$
- (ii)  $\|x + y\|^2 \leq \|x\|^2 + 2\langle y, x + y \rangle, \quad \forall x, y \in H;$
- (iii)  $\langle x, y \rangle = \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \frac{1}{2}\|x - y\|^2, \quad \forall x, y \in H.$

**Lemma 2.** [35] Let  $\{c_n\}$  and  $\{\gamma_n\}$  be a sequences of nonnegative real numbers,  $\{\beta_n\}$  be a sequences of real numbers such that

$$c_{n+1} \leq (1 - \alpha_n)c_n + \beta_n + \gamma_n, \quad n \geq 1,$$

where  $0 < \alpha_n < 1$  and  $\sum \gamma_n < \infty$ .

- (i) If  $\beta_n \leq \alpha_n M$  for some  $M \geq 0$ , then  $\{c_n\}$  is a bounded sequence.
- (ii) If  $\sum \alpha_n = \infty$  and  $\limsup_{n \rightarrow \infty} \frac{\beta_n}{\alpha_n} \leq 0$ , then  $c_n \rightarrow 0$  as  $n \rightarrow \infty$ .

**Definition 2.** Let  $\{\Gamma_n\}$  be a real sequence. Then,  $\{\Gamma_n\}$  decreases at infinity if there exists  $n_0 \in \mathbb{N}$  such that  $\Gamma_{n+1} \leq \Gamma_n$  for  $n \geq n_0$ . In other words, the sequence  $\{\Gamma_n\}$  does not decrease at infinity, if there exists a subsequence  $\{\Gamma_{n_t}\}_{t \geq 1}$  of  $\{\Gamma_n\}$  such that  $\Gamma_{n_t} < \Gamma_{n_t+1}$  for all  $t \geq 1$ .

**Lemma 3.** [36] Let  $\{\Gamma_n\}$  be a sequence of real numbers that does not decrease at infinity. In addition, consider the sequence of integers  $\{\varphi(n)\}_{n \geq n_0}$  defined by

$$\varphi(n) = \max\{k \in \mathbb{N} : k \leq n, \Gamma_k \leq \Gamma_{k+1}\}.$$

Then,  $\{\varphi(n)\}_{n \geq n_0}$  is a nondecreasing sequence verifying  $\lim_{n \rightarrow \infty} \varphi(n) = \infty$ , and, for all  $n \geq n_0$ , the following two estimates hold:

$$\Gamma_{\varphi(n)} \leq \Gamma_{\varphi(n)+1} \text{ and } \Gamma_n \leq \Gamma_{\varphi(n)+1}.$$

Let  $D$  be a closed, convex subset of a real Hilbert space  $H$  and  $g : D \times D \rightarrow \mathbb{R}$  be a bifunction. Then, we say that  $g$  satisfies condition CO on  $D$  if the following four assumptions are satisfied:

- (a)  $g(u, u) = 0$ , for all  $u \in D$ ;
- (b)  $g$  is monotone on  $D$ , i.e.,  $g(u, v) + g(v, u) \leq 0$ , for all  $u, v \in D$ ;
- (c) for each  $u, v, w \in D$ ,  $\limsup_{\alpha \downarrow 0} g(\alpha w + (1 - \alpha)u, v) \leq g(u, v)$ ;
- (d)  $g(u, \cdot)$  is convex and lower semicontinuous on  $D$  for each  $u \in D$ .

**Lemma 4.** [37] (Lemma 2.12) Let  $g$  satisfy condition CO on  $D$ . Then, for each  $r > 0$  and  $u \in H_2$ , define a mapping (called resolvent of  $g$ ), given by

$$T_r^g(u) = \{w \in D : g(w, v) + \frac{1}{r} \langle v - w, w - u \rangle \geq 0, \forall v \in D\}. \tag{11}$$

Then, the following holds:

- (i)  $T_r^g$  is single-valued;
- (ii)  $T_r^g$  is a firmly nonexpansive, i.e., for all  $u, v \in H$ ,

$$\|T_r^g(u) - T_r^g(v)\|^2 \leq \langle T_r^g(u) - T_r^g(v), u - v \rangle;$$

- (iii)  $\text{Fix}(T_r^g) = \text{SEP}(g, D) = \{\bar{x} \in D : g(\bar{x}, y) \geq 0, \forall y \in D\}$ , where  $\text{Fix}(T_r^g)$  is the fixed point set of  $T_r^g$ ;
- (iv)  $\text{SEP}(g, D)$  is closed and convex.

### 3. Main Results

Our approach here is based on taking an existing algorithm on (1), (3), and the fixed point problem of nonlinear mapping, and determining how it can be used in the setting of bilevel optimization problem (7) considered in this paper. We present a self-adaptive proximal gradient algorithm with an inertial effect for generating a sequence that converges to the unique solution of the bilevel optimization problem (7) under the the following basic assumptions.

**Assumption 1.** Assume that  $A, h, g_j$  ( $j \in \{1, \dots, M\}$ ) and  $U_i$  ( $i \in \{1, \dots, N\}$ ) in a bilevel optimization problem (7) satisfies

- A1.** Each  $A$  is nonzero bounded linear operator;
- A2.**  $h$  is proper, convex, continuously differentiable, and the gradient  $\nabla h$  is a  $\sigma$ -strongly monotone operator and  $L_h$ -Lipschitz continuous;
- A3.** Each  $U_i$  is  $\omega_i$ -demimetric and demiclosed mapping for all  $i \in \{1, \dots, N\}$ ;
- A4.** Each  $g_j$  is a proper, convex, lower semicontinuous function for all  $j \in \{1, \dots, M\}$ .

**Assumption 2.** Let  $\theta \in [0, 1)$  and  $\gamma$  be a real number, and the real sequences  $\{\beta_n^{(i)}\}$  ( $i \in \{1, \dots, N\}$ ),  $\{\delta_n^{(j)}\}$  ( $j \in \{1, \dots, M\}$ ),  $\{\alpha_n\}$ ,  $\{\rho_n\}$ ,  $\{\varepsilon_n\}$  satisfy the following conditions:

- (C1)  $\gamma \in (0, \frac{2\sigma}{L_h^2})$ .
- (C2)  $0 < \liminf_{n \rightarrow \infty} \zeta_n^{(i)} \leq \limsup_{n \rightarrow \infty} \zeta_n^{(i)} < 1, \forall i \in \{1, \dots, N\}$ , and  $\sum_{i=1}^N \zeta_n^{(i)} = 1$ .
- (C3)  $0 < \liminf_{n \rightarrow \infty} \delta_n^{(j)} \leq \limsup_{n \rightarrow \infty} \delta_n^{(j)} < 1, \forall j \in \{1, \dots, M\}$ , and  $\sum_{j=1}^M \delta_n^{(j)} = 1$ .
- (C4)  $0 < \liminf_{n \rightarrow \infty} \beta_n \leq \limsup_{n \rightarrow \infty} \beta_n < \beta = \min\{1 - \omega_1, \dots, 1 - \omega_N\}$ .
- (C5)  $0 < \alpha_n < 1, \lim_{n \rightarrow \infty} \alpha_n = 0$  and  $\sum_{n=1}^{\infty} \alpha_n = \infty$ .
- (C6)  $0 < \rho_n < 4$  and  $\liminf_{n \rightarrow \infty} \rho_n(4 - \rho_n) > 0$ .
- (C7)  $\varepsilon_n > 0$  and  $\varepsilon_n = o(\alpha_n)$ .

Assuming that the Assumption 1 is satisfied, the solution set  $\Omega$  of the lower level problem of (7) is nonempty, and, for each  $j \in \{1, \dots, M\}$ , define  $l^{(j)}$  by

$$l^{(j)}(x) = \frac{1}{2} \|(I - \text{prox}_{\lambda g_j})A(x)\|^2.$$

Note that, from Aubin [38], if  $g_j$  is indicator function, then  $l^{(j)}$  is convex, w-lsc and differentiable for each  $j \in \{1, \dots, M\}$ , and  $\nabla l^{(j)}$  is given by

$$\nabla l^{(j)}(x) = A^*(I - \text{prox}_{\lambda g_j})A(x)$$

Next, we present and analyze the strong convergence of Algorithm 1 using  $l^{(j)}$  and  $\nabla l^{(j)}$  by assuming that  $l^{(j)}$  is differentiable.

---

**Algorithm 1:** Self-adaptive proximal gradient algorithm with inertial effect.

---

**Initialization:** Let the real number  $\gamma$  and the real sequences  $\{\beta_n^{(i)}\}$  ( $i \in \{1, \dots, N\}$ ),  $\{\delta_n^{(j)}\}$  ( $j \in \{1, \dots, M\}$ ),  $\{\alpha_n\}$ ,  $\{\beta_n\}$ ,  $\{\rho_n\}$  and  $\{\varepsilon_n\}$  satisfy the conditions in Assumption 2 (C1)–(C7).

Choose  $x_0, x_1 \in H$  arbitrarily and proceed with the following computations:

**Step 1.** Given the iterates  $x_{n-1}$  and  $x_n$  ( $n \geq 1$ ), choose  $\theta_n$  such that  $0 \leq \theta_n \leq \bar{\theta}_n$ , where

$$\bar{\theta}_n := \begin{cases} \min \left\{ \theta, \frac{\varepsilon_n}{\|x_{n-1} - x_n\|} \right\}, & \text{if } x_{n-1} \neq x_n \\ \theta, & \text{otherwise.} \end{cases}$$

**Step 2.** Evaluate  $y_n = x_n + \theta_n(x_n - x_{n-1})$ .

**Step 3.** Evaluate

$$s_n = \sum_{i=1}^N \zeta_n^{(i)} ((1 - \beta_n)I - \beta_n U_i) y_n.$$

**Step 4.** Find

$$z_n = s_n - \sum_{j=1}^M \delta_n^{(j)} \tau_n^{(j)} \nabla l^{(j)}(s_n),$$

where  $\tau_n^{(j)} = \rho_n \frac{l^{(j)}(s_n)}{(\eta_n^{(j)})^2}$  for  $\eta_n^{(j)} = \max\{1, \|\nabla l^{(j)}(s_n)\|\}$ .

**Step 5.** Find  $x_{n+1} = \alpha_n(y_n - \gamma \nabla h(y_n)) + (1 - \alpha_n)z_n$ .

**Step 6.** Set  $n := n + 1$  and go to **Step 1**.

---

**Remark 1.** From Condition (C7) and Step 1 of Algorithm 1, we have that

$$\frac{\theta_n}{\alpha_n} \|x_n - x_{n-1}\| \rightarrow 0, \quad n \rightarrow \infty.$$

Since  $\{\alpha_n\}$  is bounded, we also have  $\theta_n \|x_n - x_{n-1}\| \rightarrow 0, \quad n \rightarrow \infty$ . Note that Step 1 of Algorithm 1 is easily implemented in numerical computation since the value of  $\|x_n - x_{n-1}\|$  is a priori known before choosing  $\theta_n$ .

Note that: Let  $V_\gamma = I - \gamma \nabla h$ , where  $\gamma \in (0, \frac{2\sigma}{L_h^2})$ . Then, we have

$$\begin{aligned} \|V_\gamma(x) - V_\gamma(y)\|^2 &\leq \|x - y\|^2 + \gamma^2 \|\nabla h(x) - \nabla h(y)\|^2 \\ &\quad - 2\gamma \langle \nabla h(x) - \nabla h(y), x - y \rangle \\ &\leq \|x - y\|^2 + \gamma^2 L_h^2 \|x - y\|^2 - 2\gamma\sigma \|x - y\|^2 \\ &= \mu^2 \|x - y\|^2, \quad \forall x, y \in H_1, \end{aligned} \tag{12}$$

where for  $\mu = \sqrt{1 - \gamma(2\sigma - \gamma L_h^2)}$ . Therefore, for  $\gamma \in (0, \frac{2\sigma}{L_h^2})$ , the mapping  $V_\gamma$  is a contraction mapping with constant  $\mu$ . Consequently, the mapping  $P_\Omega V_\gamma$  is also a contraction mapping with constant  $\mu$ , i.e.,  $\|P_\Omega V_\gamma(x) - P_\Omega V_\gamma(y)\| \leq \|V_\gamma(x) - V_\gamma(y)\| \leq \mu\|x - y\|, \forall x, y \in H_1$ . Hence, by the Banach contraction principle, there exists a unique element  $\bar{x} \in H_1$  such that  $\bar{x} = P_\Omega V_\gamma(\bar{x})$ . Clearly,  $\bar{x} \in \Omega$  and we have

$$\bar{x} = P_\Omega V_\gamma(\bar{x}) \Leftrightarrow \langle \nabla h(\bar{x}), y - \bar{x} \rangle \geq 0, \forall y \in \Omega.$$

**Lemma 5.** For the sequences  $\{s_n\}, \{y_n\}$  and  $\{z_n\}$  generated by Algorithm 1 and for  $\bar{x} \in \Omega$ , we have

$$(i) \|z_n - \bar{x}\|^2 \leq \|y_n - \bar{x}\|^2 - \sum_{i=1}^N \zeta_n^{(i)} \beta_n (1 - \omega_i - \beta_n) \|(I - U_i)y_n\|^2 - \rho_n(4 - \rho_n) \sum_{j=1}^M \delta_n^{(j)} \frac{(l^{(j)}(s_n))^2}{(\eta_n^{(j)})^2}.$$

$$(ii) \|z_n - \bar{x}\| \leq \|s_n - \bar{x}\| \leq \|y_n - \bar{x}\|.$$

**Proof.** Let  $\bar{x} \in \Omega$ . Now, since  $I - \text{prox}_{\lambda g_j}$  are firmly nonexpansive, and since  $A(\bar{x})$  is the minimizer of each  $g_j$ , we have for all  $x \in H_1$

$$\begin{aligned} \langle \nabla l^{(j)}(x), x - \bar{x} \rangle &= \langle A^*(I - \text{prox}_{\lambda g_j})A(x), x - \bar{x} \rangle \\ &= \langle (I - \text{prox}_{\lambda g_j})A(x), A(x) - A_k(\bar{x}) \rangle \\ &\geq \|(I - \text{prox}_{\lambda g_j})A(x)\|^2 = 2l^{(j)}(x). \end{aligned} \tag{13}$$

By the definition of  $z_n$ , we get

$$\begin{aligned} \|s_n - \bar{x}\|^2 &= \left\| \sum_{i=1}^N \zeta_n^{(i)} ((1 - \beta_n)I - \beta_n U_i)y_n - \bar{x} \right\|^2 \\ &\leq \sum_{i=1}^N \zeta_n^{(i)} \|((1 - \beta_n)I - \beta_n U_i)y_n - \bar{x}\|^2 \\ &= \sum_{i=1}^N \zeta_n^{(i)} \left( \|y_n - \bar{x}\|^2 + \beta_n^2 \|(I - U_i)y_n\|^2 - 2\beta_n \langle y_n - \bar{x}, (I - U_i)y_n \rangle \right) \\ &= \sum_{i=1}^N \zeta_n^{(i)} \left( \|y_n - \bar{x}\|^2 + \beta_n^2 \|(I - U_i)y_n\|^2 - \beta_n(1 - \omega_i) \|(I - U_i)y_n\|^2 \right) \\ &= \|y_n - \bar{x}\|^2 - \sum_{i=1}^N \zeta_n^{(i)} \beta_n (1 - \omega_i - \beta_n) \|(I - U_i)y_n\|^2. \end{aligned} \tag{14}$$



Using the definition of  $y_n$ , Lemma 1 (ii), and (13), we have

$$\begin{aligned}
 \|z_n - \bar{x}\|^2 &= \left\| s_n - \sum_{j=1}^M \delta_n^{(j)} \tau_n^{(j)} \nabla l^{(j)}(s_n) - \bar{x} \right\|^2 \\
 &\leq \sum_{j=1}^M \delta_n^{(j)} \left\| s_n - \tau_n^{(j)} \nabla l^{(j)}(s_n) - \bar{x} \right\|^2 \\
 &\leq \sum_{j=1}^M \delta_n^{(j)} \left( \|s_n - \bar{x}\|^2 + \|\tau_n^{(j)} \nabla l^{(j)}(s_n)\|^2 - 2\tau_n^{(j)} \langle \nabla l^{(j)}(s_n), s_n - \bar{x} \rangle \right) \\
 &\leq \sum_{j=1}^M \delta_n^{(j)} \left( \|s_n - \bar{x}\|^2 + (\tau_n^{(j)} \|\nabla l^{(j)}(s_n)\|)^2 - 4\tau_n^{(j)} l^{(j)}(s_n) \right) \\
 &\leq \sum_{j=1}^M \delta_n^{(j)} \left( \|s_n - \bar{x}\|^2 + (\tau_n^{(j)} \eta_n^{(j)})^2 - 4\tau_n^{(j)} l^{(j)}(s_n) \right) \\
 &= \sum_{j=1}^M \delta_n^{(j)} \left\{ \|s_n - \bar{x}\|^2 + \left( \rho_n \frac{l^{(j)}(s_n)}{(\eta_n^{(j)})^2} \eta_n^{(j)} \right)^2 - 4\rho_n \frac{l^{(j)}(s_n)}{(\eta_n^{(j)})^2} l^{(j)}(s_n) \right\} \\
 &= \|s_n - \bar{x}\|^2 - \rho_n(4 - \rho_n) \sum_{j=1}^M \delta_n^{(j)} \frac{(l^{(j)}(s_n))^2}{(\eta_n^{(j)})^2}. \tag{15}
 \end{aligned}$$

The result (i) follows from (14) and (15), and, in view of (C2)–(C6), the result (ii) follows from (14) and (15). □

**Theorem 1.** *The sequence  $\{x_n\}$  generated by Algorithm 1 converges strongly to the solution of problem (7).*

**Proof. Claim 1:** *The sequences  $\{x_n\}$ ,  $\{y_n\}$  and  $\{z_n\}$  are bounded.*

Let  $\bar{x} \in \Omega$ . Now, from the definition of  $y_n$ , we get

$$\|y_n - \bar{x}\| = \|x_n + \theta_n(x_n - x_{n-1}) - \bar{x}\| \leq \|x_n - \bar{x}\| + \theta_n \|x_n - x_{n-1}\|. \tag{16}$$

Using (16) and the definition of  $x_{n+1}$ , we get

$$\begin{aligned}
 \|x_{n+1} - \bar{x}\| &= \|(1 - \alpha_n)(z_n - \bar{x}) + \alpha_n(V_\gamma(y_n) - V_\gamma(\bar{x})) + \alpha_n(V_\gamma(\bar{x}) - \bar{x})\| \\
 &= (1 - \alpha_n)\|z_n - \bar{x}\| + \alpha_n\|V_\gamma(y_n) - V_\gamma(\bar{x})\| + \alpha_n\|V_\gamma(\bar{x}) - \bar{x}\| \\
 &= (1 - \alpha_n)\|z_n - \bar{x}\| + \alpha_n\mu\|y_n - \bar{x}\| + \alpha_n\|V_\gamma(\bar{x}) - \bar{x}\| \\
 &\leq (1 - \alpha_n(1 - \mu))\|y_n - \bar{x}\| + \alpha_n\|V_\gamma(\bar{x}) - \bar{x}\| \\
 &\leq (1 - \alpha_n(1 - \mu))\|x_n - \bar{x}\| + (1 - \alpha_n(1 - \mu))\theta_n\|x_n - x_{n-1}\| \\
 &\quad + \alpha_n\|V_\gamma(\bar{x}) - \bar{x}\| \\
 &\leq (1 - \alpha_n(1 - \mu))\|x_n - \bar{x}\| \\
 &\quad + \alpha_n(1 - \mu) \left\{ \frac{(1 - \alpha_n(1 - \mu))\theta_n}{1 - \mu} \|x_n - x_{n-1}\| + \frac{\|V_\gamma(\bar{x}) - \bar{x}\|}{1 - \mu} \right\}. \tag{17}
 \end{aligned}$$

Observe that, by (C6) and Remark 1, we see that

$$\lim_{n \rightarrow \infty} \frac{(1 - \alpha_n(1 - \mu))\theta_n}{1 - \mu} \frac{\theta_n}{\alpha_n} \|x_n - x_{n-1}\| = 0.$$

Let  $\hat{L} = 2 \max \left\{ \frac{\|V_\gamma(\bar{x}) - \bar{x}\|}{1 - \mu}, \sup_{n \geq 1} \frac{(1 - \alpha_n(1 - \mu))\theta_n}{1 - \mu} \frac{\theta_n}{\alpha_n} \|x_n - x_{n-1}\| \right\}$ .

Then, (17) becomes

$$\|x_{n+1} - \bar{x}\| \leq (1 - \alpha_n(1 - \mu))\|x_n - \bar{x}\| + \alpha_n(1 - \mu)\hat{L}.$$

Thus, by Lemma 2, the sequence  $\{x_n\}$  is bounded. As a consequence,  $\{y_n\}$ ,  $\{z_n\}$ ,  $\{V_\gamma(y_n)\}$  and  $\{s_n\}$  are also bounded.

**Claim 2:** The sequence  $\{x_n\}$  converges strongly to  $\bar{x} \in \Omega$ , where  $\bar{x} = P_\Omega V_\gamma(\bar{x})$ .

Now,

$$\begin{aligned} \|y_n - \bar{x}\|^2 &= \|x_n + \theta_n(x_n - x_{n-1}) - \bar{x}\|^2 \\ &= \|x_n - \bar{x}\|^2 + \theta_n^2 \|x_n - x_{n-1}\|^2 + 2\theta_n \langle x_n - \bar{x}, x_n - x_{n-1} \rangle. \end{aligned} \tag{18}$$

From Lemma 1 (iii), we have

$$\langle x_n - \bar{x}, x_n - x_{n-1} \rangle = \frac{1}{2} \|x_n - \bar{x}\|^2 - \frac{1}{2} \|x_{n-1} - \bar{x}\|^2 + \frac{1}{2} \|x_n - x_{n-1}\|^2. \tag{19}$$

From (18) and (19) and since  $0 \leq \theta_n < 1$ , we get

$$\begin{aligned} \|y_n - \bar{x}\|^2 &= \|x_n - \bar{x}\|^2 + \theta_n^2 \|x_n - x_{n-1}\|^2 \\ &\quad + \theta_n (\|x_n - \bar{x}\|^2 - \|x_{n-1} - \bar{x}\|^2 + \|x_n - x_{n-1}\|^2) \\ &\leq \|x_n - \bar{x}\|^2 + 2\theta_n \|x_n - x_{n-1}\|^2 \\ &\quad + \theta_n (\|x_n - \bar{x}\|^2 - \|x_{n-1} - \bar{x}\|^2). \end{aligned} \tag{20}$$

Using the definition of  $x_{n+1}$  and Lemma 1 (ii), we have

$$\begin{aligned} \|x_{n+1} - \bar{x}\|^2 &= \|\alpha_n V_\gamma(y_n) + (1 - \alpha_n)z_n - \bar{x}\|^2 \\ &= \|\alpha_n (V_\gamma(y_n) - \bar{x}) + (1 - \alpha_n)(z_n - \bar{x})\|^2 \\ &\leq (1 - \alpha_n)^2 \|z_n - \bar{x}\|^2 + 2\alpha_n \langle V_\gamma(y_n) - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &\leq \|z_n - \bar{x}\|^2 + 2\alpha_n \langle V_\gamma(y_n) - \bar{x}, x_{n+1} - \bar{x} \rangle. \end{aligned} \tag{21}$$

Lemma 5 (i) together with (20) and (21) give

$$\begin{aligned} \|x_{n+1} - \bar{x}\|^2 &\leq \|z_n - \bar{x}\|^2 + 2\alpha_n \langle V_\gamma(y_n) - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &\leq \|y_n - \bar{x}\|^2 + 2\alpha_n \langle V_\gamma(y_n) - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &\quad - \sum_{i=1}^N \zeta_n^{(i)} \beta_n (1 - \omega_i - \beta_n) \|(I - U_i)y_n\|^2 \\ &\quad - \rho_n (4 - \rho_n) \sum_{j=1}^M \delta_n^{(j)} \frac{(I^{(j)}(s_n))^2}{(\eta_n^{(j)})^2} \\ &\leq \|x_n - \bar{x}\|^2 + 2\theta_n \|x_n - x_{n-1}\|^2 \\ &\quad + \theta_n (\|x_n - \bar{x}\|^2 - \|x_{n-1} - \bar{x}\|^2) \\ &\quad + 2\alpha_n \langle V_\gamma(y_n) - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &\quad - \sum_{i=1}^N \zeta_n^{(i)} \beta_n (1 - \omega_i - \beta_n) \|(I - U_i)y_n\|^2 \\ &\quad - \rho_n (4 - \rho_n) \sum_{j=1}^M \delta_n^{(j)} \frac{(I^{(j)}(s_n))^2}{(\eta_n^{(j)})^2}. \end{aligned} \tag{22}$$

Since the sequence  $\{x_n\}$  and  $\{V_\gamma(y_n)\}$  are bounded, there exists  $M_1$  such that  $2 \langle V_\gamma(y_n) - \bar{x}, x_{n+1} - \bar{x} \rangle \leq M_1$  for all  $n \geq 1$ . Thus, from (22), we obtain

$$\begin{aligned} \|x_{n+1} - \bar{x}\|^2 &\leq \|x_n - \bar{x}\|^2 + 2\theta_n \|x_n - x_{n-1}\|^2 \\ &\quad + \theta_n (\|x_n - \bar{x}\|^2 - \|x_{n-1} - \bar{x}\|^2) + \alpha_n M_1 \\ &\quad - \sum_{i=1}^N \zeta_n^{(i)} \beta_n (1 - \omega_i - \beta_n) \|(I - U_i)y_n\|^2 \\ &\quad - \rho_n (4 - \rho_n) \sum_{j=1}^M \delta_n^{(j)} \frac{(I^{(j)}(s_n))^2}{(\eta_n^{(j)})^2}. \end{aligned} \tag{23}$$

Let us distinguish the following two cases related to the behavior of the sequence  $\{\Gamma_n\}$  where  $\Gamma_n = \|x_n - \bar{x}\|^2$ .

*Case 1.* Suppose the sequence  $\{\Gamma_n\}$  decreases at infinity. Thus, there exists  $n_0 \in \mathbb{N}$  such that  $\Gamma_{n+1} \leq \Gamma_n$  for  $n \geq n_0$ . Then,  $\{\Gamma_n\}$  converges and  $\Gamma_n - \Gamma_{n+1} \rightarrow 0$  as  $n \rightarrow \infty$ .

From (23), we have

$$\sum_{i=1}^N \zeta_n^{(i)} \beta_n (1 - \omega_i - \beta_n) \|(I - U_i)y_n\|^2 \leq (\Gamma_n - \Gamma_{n+1}) + \alpha_n M_1 + \theta_n (\Gamma_n - \Gamma_{n-1}) + 2\theta_n \|x_n - x_{n-1}\|^2, \tag{24}$$

and

$$\rho_n (4 - \rho_n) \sum_{j=1}^M \delta_n^{(j)} \frac{(l^{(j)}(s_n))^2}{(\eta_n^{(j)})^2} \leq (\Gamma_n - \Gamma_{n+1}) + \alpha_n M_1 + \theta_n (\Gamma_n - \Gamma_{n-1}) + 2\theta_n \|x_n - x_{n-1}\|^2. \tag{25}$$

Since  $\Gamma_n - \Gamma_{n+1} \rightarrow 0$  ( $\Gamma_{n-1} - \Gamma_n \rightarrow 0$ ) and using (C5), (C6), and Remark 1 (noting  $\alpha_n \rightarrow 0$ ,  $\theta_n \|x_n - x_{n-1}\| \rightarrow 0$ ,  $\{x_n\}$  is bounded and  $\liminf_{n \rightarrow \infty} \rho_n (4 - \rho_n) > 0$ ); we have, from (24) and (25),

$$\sum_{j=1}^M \delta_n^{(j)} \frac{(l^{(j)}(s_n))^2}{(\eta_n^{(j)})^2} \rightarrow 0 \text{ and } \sum_{i=1}^N \zeta_n^{(i)} \beta_n (1 - \omega_i - \beta_n) \|(I - U_i)y_n\|^2 \rightarrow 0. \tag{26}$$

In view of (26) and conditions (C2)–(C6), we have

$$\frac{(l^{(j)}(s_n))^2}{(\eta_n^{(j)})^2} \rightarrow 0 \text{ and } \|(I - U_i)y_n\| \rightarrow 0, \quad n \rightarrow \infty, \tag{27}$$

for all  $j \in \{1, \dots, M\}$  and for all  $i \in \{1, \dots, N\}$ .

Using (26), we have

$$\|s_n - z_n\|^2 \leq 16 \sum_{j=1}^M \delta_n^{(j)} \frac{(l^{(j)}(s_n))^2}{(\eta_n^{(j)})^2} \rightarrow 0, \quad n \rightarrow \infty. \tag{28}$$

Similarly, from (26), we have

$$\|y_n - s_n\|^2 \leq \beta^2 \sum_{i=1}^N \zeta_n^{(i)} \|(I - U_i)y_n\|^2 \rightarrow 0, \quad n \rightarrow \infty. \tag{29}$$

Using the definition of  $y_n$  and Remark 1, we have

$$\|x_n - y_n\| = \|x_n - x_n - \theta_n(x_n - x_{n-1})\| = \theta_n \|x_n - x_{n-1}\| \rightarrow 0, \quad n \rightarrow \infty. \tag{30}$$

Moreover, using the definition of  $x_{n+1}$  and boundedness of  $\{V_\gamma(y_n)\}$  and  $\{z_n\}$  together with condition (C5), we have

$$\|x_{n+1} - z_n\| = \alpha_n \|V_\gamma(y_n) - z_n\| \rightarrow 0, \quad n \rightarrow \infty. \tag{31}$$

Therefore, from (28)–(31), we have

$$\|x_{n+1} - x_n\| \leq \|x_{n+1} - z_n\| + \|z_n - s_n\| + \|s_n - y_n\| + \|x_n - y_n\| \rightarrow 0, \quad n \rightarrow \infty. \tag{32}$$

For each  $j \in \{1, \dots, M\}$ ,  $\nabla l^{(j)}(\cdot)$  are Lipschitz continuous with constant  $\|A\|^2$ . Therefore, the sequence  $\{(\eta_n^{(j)})^2\}_{n=1}^\infty$  is bounded sequence for each  $j \in \{1, \dots, M\}$ , and hence, using (27), we have  $\lim_{n \rightarrow \infty} l^{(j)}(s_n) = 0$  for all  $j \in \{1, \dots, M\}$ .

Let  $p$  be a weak cluster point of  $\{x_n\}$ ; there exists a subsequence  $\{x_{n_l}\}$  of  $\{x_n\}$  such that  $x_{n_l} \rightharpoonup p$  as  $l \rightarrow \infty$ . Since  $\|x_n - y_n\| \rightarrow 0$  as  $n \rightarrow \infty$  (from (30)), we have  $y_{n_l} \rightharpoonup p$  as  $l \rightarrow \infty$ . Hence, using  $y_{n_l} \rightharpoonup p$ ,

(27) and demiclosedness of  $U_i$ , we have  $p \in \text{Fix}U_i$  for all  $i \in \{1, \dots, N\}$ .

Moreover, since  $\|x_n - s_n\| \rightarrow 0$  as  $n \rightarrow \infty$  (from (29) and (30)), we have  $s_{n_l} \rightharpoonup p$  as  $l \rightarrow \infty$ . Hence, the weak lower-semicontinuity of  $l^{(j)}(\cdot)$  implies that

$$0 \leq l^{(j)}(p) \leq \liminf_{l \rightarrow \infty} l^{(j)}(s_{n_l}) = \lim_{n \rightarrow \infty} l^{(j)}(s_n) = 0$$

for all  $j \in \{1, \dots, M\}$ . That is,  $l^{(j)}(p) = \frac{1}{2} \|(I - \text{prox}_{\lambda g_j})A(p)\|^2 = 0$  for all  $j \in \{1, \dots, M\}$ . Thus,  $A(p) \in \Omega$ .

We now show that  $\limsup_{n \rightarrow \infty} \langle (I - V_\gamma)\bar{x}, \bar{x} - x_n \rangle \leq 0$ . Indeed, since  $\bar{x} = P_\Omega V_\gamma(\bar{x})$  and, from above,  $p$  is a weak cluster point of  $\{x_n\}$ , i.e.,  $x_{n_l} \rightharpoonup p$ , and  $p \in \Omega$ , we obtain that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \langle (I - V_\gamma)\bar{x}, \bar{x} - x_n \rangle &= \lim_{l \rightarrow \infty} \langle (I - V_\gamma)\bar{x}, \bar{x} - x_{n_l} \rangle \\ &= \langle (I - V_\gamma)\bar{x}, \bar{x} - p \rangle \leq 0. \end{aligned} \tag{33}$$

Since  $\|x_{n+1} - x_n\| \rightarrow 0$  from (32), from (33), we obtain

$$\limsup_{n \rightarrow \infty} \langle (I - V_\gamma)\bar{x}, \bar{x} - x_{n+1} \rangle \leq 0. \tag{34}$$

Now, using Lemma 5 (ii), we get

$$\begin{aligned} \|x_{n+1} - \bar{x}\|^2 &= \langle \alpha_n V_\gamma(y_n) + (1 - \alpha_n)z_n - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &= \alpha_n \langle V_\gamma(y_n) - V_\gamma(\bar{x}), x_{n+1} - \bar{x} \rangle + (1 - \alpha_n) \langle z_n - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &\quad + \alpha_n \langle V_\gamma(\bar{x}) - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &\leq \mu \alpha_n \|y_n - \bar{x}\| \|x_{n+1} - \bar{x}\| + (1 - \alpha_n) \|z_n - \bar{x}\| \|x_{n+1} - \bar{x}\| \\ &\quad + \alpha_n \langle V_\gamma(\bar{x}) - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &\leq (1 - \alpha_n(1 - \mu)) \|y_n - \bar{x}\| \|x_{n+1} - \bar{x}\| + \alpha_n \langle V_\gamma(\bar{x}) - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &\leq (1 - \alpha_n(1 - \mu)) \left( \frac{\|y_n - \bar{x}\|^2}{2} + \frac{\|x_{n+1} - \bar{x}\|^2}{2} \right) \\ &\quad + \alpha_n \langle V_\gamma(\bar{x}) - \bar{x}, x_{n+1} - \bar{x} \rangle. \end{aligned} \tag{35}$$

Therefore, from (35), we have

$$\begin{aligned} \|x_{n+1} - \bar{x}\|^2 &\leq \frac{1 - \alpha_n(1 - \mu)}{1 + \alpha_n(1 - \mu)} \|y_n - \bar{x}\|^2 + \frac{2\alpha_n}{1 + \alpha_n(1 - \mu)} \langle V_\gamma(\bar{x}) - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &= \left( 1 - \frac{2\alpha_n(1 - \mu)}{1 + \alpha_n(1 - \mu)} \right) \|y_n - \bar{x}\|^2 + \frac{2\alpha_n}{1 + \alpha_n(1 - \mu)} \langle V_\gamma(\bar{x}) - \bar{x}, x_{n+1} - \bar{x} \rangle. \end{aligned} \tag{36}$$

Combining (36) and

$$\|y_n - \bar{x}\| = \|x_n + \theta_n(x_n - x_{n-1}) - \bar{x}\| \leq \|x_n - \bar{x}\| + \theta_n \|x_n - x_{n-1}\|,$$

it holds that

$$\begin{aligned} \|x_{n+1} - \bar{x}\|^2 &\leq \left(1 - \frac{2\alpha_n(1-\mu)}{1+\alpha_n(1-\mu)}\right) (\|x_n - \bar{x}\| + \theta_n \|x_n - x_{n-1}\|)^2 \\ &\quad + \frac{2\alpha_n}{1+\alpha_n(1-\mu)} \langle V_\gamma(\bar{x}) - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &= \left(1 - \frac{2\alpha_n(1-\mu)}{1+\alpha_n(1-\mu)}\right) (\|x_n - \bar{x}\|^2 + \theta_n^2 \|x_n - x_{n-1}\|^2 \\ &\quad + 2\theta_n \|x_n - \bar{x}\| \|x_n - x_{n-1}\|) + \frac{2\alpha_n}{1+\alpha_n(1-\mu)} \langle V_\gamma(\bar{x}) - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &\leq \left(1 - \frac{2\alpha_n(1-\mu)}{1+\alpha_n(1-\mu)}\right) \|x_n - \bar{x}\|^2 + \theta_n^2 \|x_n - x_{n-1}\|^2 \\ &\quad + 2\theta_n \|x_n - \bar{x}\| \|x_n - x_{n-1}\| + \frac{2\alpha_n}{1+\alpha_n(1-\mu)} \langle V_\gamma(\bar{x}) - \bar{x}, x_{n+1} - \bar{x} \rangle. \end{aligned} \tag{37}$$

Since  $\{x_n\}$  is bounded, there exists  $M_2 > 0$  such that  $\|x_n - \bar{x}\| \leq M_2$  for all  $n \geq 1$ . Thus, in view of (37), we have

$$\begin{aligned} \Gamma_{n+1} &\leq \left(1 - \frac{2\alpha_n(1-\mu)}{1+\alpha_n(1-\mu)}\right) \Gamma_n + \theta_n \|x_n - x_{n-1}\| (\theta_n \|x_n - x_{n-1}\| + 2M_2) \\ &\quad + \frac{2\alpha_n}{1+\alpha_n(1-\mu)} \langle V_\gamma(\bar{x}) - \bar{x}, x_{n+1} - \bar{x} \rangle \\ &= (1 - a_n) \Gamma_n + a_n \vartheta_n, \end{aligned} \tag{38}$$

where  $a_n = \frac{2\alpha_n(1-\mu)}{1+\alpha_n(1-\mu)}$  and

$$\begin{aligned} \vartheta_n &= \frac{1+\alpha_n(1-\mu)}{2(1-\mu)} \left(\frac{\theta_n}{\alpha_n} \|x_n - x_{n-1}\|\right) \{\theta_n \|x_n - x_{n-1}\| + 2M_2\} \\ &\quad + \frac{1}{1-\mu} \langle V_\gamma(\bar{x}) - \bar{x}, x_{n+1} - \bar{x} \rangle. \end{aligned}$$

From (C5), Remark 1 and (34), we have  $\sum_{n=1}^\infty a_n = \infty$  and  $\limsup_{n \rightarrow \infty} \vartheta_n \leq 0$ . Thus, using Lemma 2 and (38), we get  $\Gamma_n \rightarrow 0$  as  $n \rightarrow \infty$ . Hence,  $x_n \rightarrow \bar{x}$  as  $n \rightarrow \infty$ .

Case 2. Assume that  $\{\Gamma_n\}$  does not decrease at infinity. Let  $\varphi : \mathbb{N} \rightarrow \mathbb{N}$  be a mapping for all  $n \geq n_0$  (for some  $n_0$  large enough) defined by

$$\varphi(n) = \max\{k \in \mathbb{N} : k \leq n, \Gamma_k \leq \Gamma_{k+1}\}.$$

By Lemma 3,  $\{\varphi(n)\}_{n=n_0}^\infty$  is a nondecreasing sequence,  $\varphi(n) \rightarrow \infty$  as  $n \rightarrow \infty$  and

$$\Gamma_{\varphi(n)} \leq \Gamma_{\varphi(n)+1} \text{ and } \Gamma_n \leq \Gamma_{\varphi(n)+1}, \quad \forall n \geq n_0. \tag{39}$$

In view of  $\|x_{\varphi(n)} - \bar{x}\|^2 - \|x_{\varphi(n)+1} - \bar{x}\|^2 = \Gamma_{\varphi(n)} - \Gamma_{\varphi(n)+1} \leq 0$  for all  $n \geq n_0$  and (23), we have for all  $n \geq n_0$

$$\begin{aligned} & \rho_{\varphi(n)}(4 - \rho_{\varphi(n)}) \sum_{j=1}^M \delta_{\varphi(n)}^{(j)} \frac{(I^{(j)}(s_{\varphi(n)}))^2}{(\eta_{\varphi(n)}^{(j)})^2} \\ & \leq (\Gamma_{\varphi(n)} - \Gamma_{\varphi(n)+1}) + \alpha_{\varphi(n)}M_1 + \theta_{\varphi(n)}(\Gamma_{\varphi(n)} - \Gamma_{\varphi(n)-1}) \\ & \quad + 2\theta_{\varphi(n)}\|x_{\varphi(n)} - x_{\varphi(n)-1}\|^2 \\ & \leq \alpha_{\varphi(n)}M_1 + \theta_{\varphi(n)}(\Gamma_{\varphi(n)} - \Gamma_{\varphi(n)-1}) + 2\theta_{\varphi(n)}\|x_{\varphi(n)} - x_{\varphi(n)-1}\|^2 \\ & \leq \alpha_{\varphi(n)}M_1 + \theta_{\varphi(n)}\|x_{\varphi(n)} - x_{\varphi(n)-1}\| \left( \sqrt{\Gamma_{\varphi(n)}} + \sqrt{\Gamma_{\varphi(n)-1}} \right) \\ & \quad + 2\theta_{\varphi(n)}\|x_{\varphi(n)} - x_{\varphi(n)-1}\|^2. \end{aligned} \tag{40}$$

Similarly, from (23), we have for all  $n \geq n_0$

$$\begin{aligned} & \sum_{i=1}^N \zeta_{\varphi(n)}^{(i)} \beta_{\varphi(n)}(1 - \omega_i - \beta_{\varphi(n)}) \|(I - U_i)y_{\varphi(n)}\|^2 \\ & \leq \alpha_{\varphi(n)}M_1 + \theta_{\varphi(n)}\|x_{\varphi(n)} - x_{\varphi(n)-1}\| \left( \sqrt{\Gamma_{\varphi(n)}} + \sqrt{\Gamma_{\varphi(n)-1}} \right) \\ & \quad + 2\theta_{\varphi(n)}\|x_{\varphi(n)} - x_{\varphi(n)-1}\|^2. \end{aligned} \tag{41}$$

Thus, for (40) and (41) together with (C3)–(C6) and Remark 1, we have for each  $j \in \{1, \dots, M\}$  and  $i \in \{1, \dots, N\}$ ,

$$\frac{(I^{(j)}(s_{\varphi(n)}))^2}{(\eta_{\varphi(n)}^{(j)})^2} \rightarrow 0, \text{ and } \|(I - U_i)y_{\varphi(n)}\| \rightarrow 0, \quad n \rightarrow \infty. \tag{42}$$

Using a similar procedure as above in Case 1, we have

$$\lim_{n \rightarrow \infty} \|x_{\varphi(n)} - s_{\varphi(n)}\| = \lim_{n \rightarrow \infty} \|x_{\varphi(n)} - y_{\varphi(n)}\| = \lim_{n \rightarrow \infty} \|x_{\varphi(n)+1} - x_{\varphi(n)}\| = 0.$$

By the similar argument as above in Case 1, since  $\{x_{\varphi(n)}\}$  is bounded, there exists a subsequence of  $\{x_{\varphi(n)}\}$  which converges weakly to  $p \in \Omega$  and this gives  $\limsup_{n \rightarrow \infty} \langle (I - V_\gamma)\bar{x}, \bar{x} - x_{\varphi(n)+1} \rangle \leq 0$ . Thus, from (38), we have

$$\Gamma_{\varphi(n)+1} \leq (1 - a_{\varphi(n)})\Gamma_{\varphi(n)} + a_{\varphi(n)}\vartheta_{\varphi(n)}, \tag{43}$$

where  $a_{\varphi(n)} = \frac{2\alpha_{\varphi(n)}(1-\mu)}{1+\alpha_{\varphi(n)}(1-\mu)}$  and

$$\begin{aligned} \vartheta_{\varphi(n)} = & \frac{1+\alpha_{\varphi(n)}(1-\mu)}{2(1-\mu)} \left( \frac{\theta_{\varphi(n)}}{\alpha_{\varphi(n)}} \|x_{\varphi(n)} - x_{\varphi(n)-1}\| \right) \{ \theta_{\varphi(n)} \|x_{\varphi(n)} - x_{\varphi(n)-1}\| \\ & + 2M_2 \} + \frac{1}{1-\mu} \langle V(\bar{x}) - \bar{x}, x_{\varphi(n)+1} - \bar{x} \rangle. \end{aligned}$$

Using  $\Gamma_{\varphi(n)} - \Gamma_{\varphi(n)+1} \leq 0$  for all  $n \geq n_0$  and  $\vartheta_{\varphi(n)} > 0$ , the last inequality gives

$$0 \leq -a_{\varphi(n)}\Gamma_{\varphi(n)} + a_{\varphi(n)}\vartheta_{\varphi(n)}.$$

Since  $a_{\varphi(n)} > 0$ , we obtain  $\|x_{\varphi(n)} - \bar{x}\|^2 = \Gamma_{\varphi(n)} \leq \vartheta_{\varphi(n)}$ . Moreover, since  $\limsup_{n \rightarrow \infty} \vartheta_{\varphi(n)} \leq 0$ , we have  $\lim_{n \rightarrow \infty} \|x_{\varphi(n)} - \bar{x}\| = 0$ . Thus,  $\lim_{n \rightarrow \infty} \|x_{\varphi(n)} - \bar{x}\| = 0$  together with  $\lim_{n \rightarrow \infty} \|x_{\varphi(n)+1} - x_{\varphi(n)}\| = 0$ , gives  $\lim_{n \rightarrow \infty} \Gamma_{\varphi(n)+1} = 0$ . Therefore, from (39), we obtain  $\lim_{n \rightarrow \infty} \Gamma_n = 0$ , that is,  $x_n \rightarrow \bar{x}$  as  $n \rightarrow \infty$ .  $\square$

For  $l(x) = \frac{1}{2}\|(I - \text{prox}_{\lambda g_j})(x)\|^2$  and  $\nabla l(x) = (I - \text{prox}_{\lambda g_j})(x)$ , we have the following results solving the bilevel problem (10):

**Corollary 1.** If  $\gamma \in (0, \frac{2\sigma}{L_h^2})$ , the sequence  $\{x_n\}$  generated by

$$\begin{cases} y_n = x_n + \theta_n(x_n - x_{n-1}), \\ \tau_n = \rho_n \frac{l(y_n)}{(\eta_n)^2}, \quad \eta_n = \max\{1, \|\nabla l(y_n)\|\}, \\ z_n = y_n - \tau_n \nabla l(y_n), \\ x_{n+1} = \alpha_n(y_n - \gamma \nabla h(y_n)) + (1 - \alpha_n)z_n, \end{cases}$$

converges strongly to the solution the bilevel problem (10) if  $\{\alpha_n\}$ ,  $\{\rho_n\}$  and  $\{\theta_n\}$  are real sequences such that

- (C1)  $0 < \alpha_n < 1$ ,  $\lim_{n \rightarrow \infty} \alpha_n = 0$  and  $\sum_{n=1}^{\infty} \alpha_n = \infty$ .
- (C2)  $0 < \rho_n < 4$  and  $\liminf_{n \rightarrow \infty} \rho_n(4 - \rho_n) > 0$ .
- (C3)  $\lim_{n \rightarrow \infty} \frac{\theta_n}{\alpha_n} \|x_n - x_{n-1}\| = 0$  where  $\theta_n \in [0, \theta)$  for  $\theta \in [0, 1)$ .

### 4. Applications

#### 4.1. Application to the Bilevel Variational Inequality Problem

Let  $H_1$  and  $H_2$  be two real Hilbert spaces. Assume that  $F : H_1 \rightarrow H_1$  is  $L_h$ -Lipschitz continuous and  $\sigma$ -strongly monotone on  $H_1$ ,  $A : H_1 \rightarrow H_2$  is a bounded linear operator,  $g_j : H_2 \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper, convex, lower semicontinuous function for all  $j \in \{1, \dots, M\}$ , and  $U_i : H_1 \rightarrow H_1$  is  $\omega_i$ -demimetric and demiclosed mapping for all  $i \in \{1, \dots, N\}$ . Then, replacing  $\nabla h$  by  $F$  in Algorithm 1, we obtain strong convergence for an approximation of a solution of the bilevel variational inequality problem

$$\text{find } \bar{x} \in \Omega \text{ such that } \langle F(\bar{x}), x - \bar{x} \rangle \geq 0, \quad \forall x \in \Omega,$$

where  $\Omega$  is the solution set of

$$\text{find } x \in \bigcap_{i=1}^N \text{Fix}U_i \text{ such that } A(x) \in \bigcap_{j=1}^M \arg \min g_j.$$

#### 4.2. Application to a Bilevel Optimization Problem with a Feasibility Set Constraint, Inclusion Constraint, and Equilibrium Constraint

Let  $H_1$  and  $H_2$  be two real Hilbert spaces,  $A : H_1 \rightarrow H_2$  be a linear transformation and  $h : H_1 \rightarrow \mathbb{R}$  be proper, convex, continuously differentiable, and the gradient  $\nabla h$  is  $\sigma$ -strongly monotone operator and  $L_h$ -Lipschitz continuous.

Now, consider the bilevel optimization problem with a feasibility set constraint

$$\begin{aligned} & \min h \\ & \text{s.t. } A(x) \in \bigcap_{i=1}^M Q_j, \end{aligned} \tag{44}$$

where each  $Q_j$  is a closed convex subset of  $H_2$  for  $j \in \{1, \dots, M\}$ . Replacing  $U_i = I$  for all  $i \in \{1, \dots, N\}$  and  $\text{prox}_{\lambda g_j}$  by projection mapping  $P_{Q_j}$  in Algorithm 1, we obtain strong convergence for an approximation of the solution of the bilevel problem (44).

Consider the bilevel optimization problem with inclusion constraint

$$\begin{aligned} & \min h \\ & \text{s.t. } 0 \in \bigcap_{i=1}^M G_j(A(x)), \end{aligned} \tag{45}$$

where  $G_j : H_2 \rightarrow 2^{H_2}$  is maximal monotone mapping for  $j \in \{1, \dots, M\}$ . Setting  $U_i = I$  for all  $i \in \{1, \dots, N\}$  and, replacing the proximal mapping  $g_j$  in Algorithm 1 by the resolvent operators

$J_\lambda^{G_j} = (I + \lambda G_j)^{-1}$  (for  $\lambda > 0$ ), and following the method of proof in theorems, we obtain a strong convergence result for approximation of the solution of the bilevel problem (45).

Consider the bilevel optimization problem with equilibrium constraint

$$\begin{aligned} & \min h \\ & \text{s.t. } A(x) \in \bigcap_{i=1}^M \text{SEP}(g_i, H_2), \end{aligned} \tag{46}$$

where  $g_j : H_2 \times H_2 \rightarrow \mathbb{R}$  is a bifunction and each  $g_j$  satisfies condition CO on  $H_2$ . We have strong convergence results solving (46) by setting  $U_i = I$  for all  $i \in \{1, \dots, N\}$  and replacing the proximal mappings by the resolvent operators  $T_r^{g_j}$  in Algorithm 1 (see (11) and properties of it in Lemma 4 (i)–(iv)).

### 5. Numerical Example

Taking the bilevel optimization problem (7) for  $H_1 = \mathbb{R}^p, H_2 = \mathbb{R}^q$ , the linear transformations  $A : \mathbb{R}^p \rightarrow \mathbb{R}^q$  are given by  $A(x) = G_{q \times p} x$ , where  $G_{q \times p}$  is  $q \times p$  matrix, and for  $x \in H_1 = \mathbb{R}^p, z \in H_2 = \mathbb{R}^q$ , we have

$$\begin{aligned} h(x) &= \frac{1}{2} x^T D x + \frac{1}{2} \|x\|_p^2, \\ U_i(x) &= \epsilon_i x, \quad i \in \{1, \dots, N\}, \\ g_1(z) &= \frac{1}{2} z^T B z, \quad g_2(z) = \|z\|_q, \quad g_3(z) = \sum_{t=1}^q \Phi(z_t), \end{aligned}$$

where  $D$  and  $B$  are invertible symmetric positive semidefinite  $p \times p$  and  $q \times q$  matrix, respectively,  $\epsilon_i \leq 1 \forall i \in \{1, \dots, N\}, z = (z_1, \dots, z_q) \in \mathbb{R}^q, \|\cdot\|_p$  is the Euclidean norm in  $\mathbb{R}^p, \|\cdot\|_q$  is the Euclidean norm in  $\mathbb{R}^q$ , and  $\Phi(z_t) = \max\{|z_t| - 1, 0\}$  for  $t = 1, 2, \dots, q$ .

Here,  $h(x) = f(x) + \frac{1}{2} \|x\|_p^2$  where  $f(x) = \frac{1}{2} x^T D x$  and hence the gradient  $\nabla f$  is  $\|D\|$ -Lipschitz. Thus, the gradient  $\nabla h$  is 1-strongly monotone and  $(\|D\| + 1)$ -Lipschitz. We choose  $\gamma = \frac{1}{(\|D\| + 1)^2}$ .

Now, for  $\lambda = 1$ , the proximal  $g_1, g_2$  and  $g_3$  is given by

$$\begin{aligned} \text{prox}_{\lambda g_1}(z) &= (I + B)^{-1}(z), \quad i \in \Phi, \\ \text{prox}_{\lambda g_2}(z) &= \begin{cases} \left(1 - \frac{1}{\|z\|_q}\right)z, & \|z\|_q \geq 1 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

and  $\text{prox}_{\lambda g_3}(z) = (\text{prox}_{\lambda \Phi}(z_1), \dots, \text{prox}_{\lambda \Phi}(z_q))$ , where

$$\text{prox}_{\lambda \Phi}(z_t) = \begin{cases} z_t, & \text{if } |z_t| < 1 \\ \text{sign}(z_t), & \text{if } 1 \leq |z_t| \leq 2 \\ \text{sign}(z_t - 1), & \text{if } |z_t| > 2. \end{cases}$$

We consider for  $p = q, \epsilon_i = \frac{1}{i+1}$  for  $i \in \{1, \dots, N\}$  and  $G_{q \times p} = I_{p \times p}$ , where  $I_{p \times p}$  is identity  $p \times p$  matrix. The parameters are chosen are  $\beta_n^{(i)} = \frac{i}{1+\dots+N}$  for  $i \in \{1, \dots, N\}, \delta_n^{(j)} = \frac{j}{6}$  for  $j \in \{1, 2, 3\}, \alpha_n = \frac{1}{n+1}, \epsilon_n = \frac{1}{(n+1)^2}, \rho_n = 1$  and  $\theta_n = \bar{\theta}_n$ .

For the purpose of testing our algorithm, we took the following data:

- $D$  and  $B$  are randomly generated invertible symmetric positive semidefinite  $p \times p$  matrices, respectively.
- $x_0$  and  $x_1$  are randomly generated starting points.
- The stopping criteria  $\frac{\|x_{n+1} - x_n\|}{\|x_2 - x_1\|} < TOL$ .



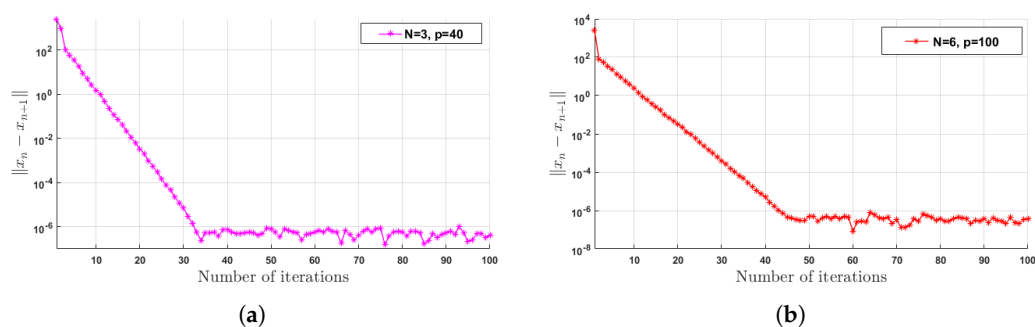
Tables 1 and 2 and Figure 1 illustrate the numerical results of our algorithms for this example under the parameters and data given above and for  $\theta = 0.5$ . The number of iterations ( $Iter(n)$ ), CPU time in seconds (CPU(s)), and the error  $err(n) = \|x_n - \bar{x}\|$ , where  $\bar{x}$  is the solution set of the bilevel optimization problem ( $\bar{x} = 0$  here in this example), are reported in Table 1.

**Table 1.** Performance of Algorithm 1 for different  $N$  and different dimensions  $p = q$  with  $TOL = 10^{-3}$ .

		Iter(n)	CPU(s)	err(n)
N = 3	p = 3	11	0.0600	0.3995
	p = 8	16	0.0854	0.4194
	p = 16	24	0.0801	0.3935
N = 6	p = 40	82	0.08920	0.5332
	p = 80	131	0.16109	0.8043
	p = 150	250	0.21723	0.9099
N = 10	p = 50	99	0.1295	0.6543
	p = 100	137	0.1463	0.7004
	p = 200	263	0.2969	0.7841

**Table 2.** Performance of Algorithm 1 for  $N = 8$  and different  $p = q$  with  $TOL = 10^{-4}$ .

Iter(n)	p = 5		p = 15	
	CPU(s)	err(n)	CPU(s)	err(n)
1		114.1709		372.5614
2		112.5227		367.2210
3		110.6359		361.0786
4		108.5059		354.1380
5		106.1671		346.5137
⋮		⋮		⋮
100		0.0428		3.2360
101		0.0416		3.0748
181		0.0385		2.8765
182	0.3702	0.0375		2.6385
⋮		⋮		⋮
207				0.0756
208			0.5311	0.0746



**Figure 1.** Algorithm 1 for different  $N$  and different dimensions  $p = q$ .

We now compare our algorithm for different  $\theta_n$ , i.e., for non-inertial accelerated case ( $\theta_n = 0$ ) and for inertial accelerated case ( $\theta_n \neq 0$ ). For the non-inertial accelerated case, we just simply take  $\theta = 0$ , and, for the inertial accelerated case, we take a very small  $\theta$  with  $\theta \in (0, 1)$  so that  $\theta_n = \bar{\theta}_n = \theta$ . Numerical comparisons of our proposed algorithm with inertial version ( $\theta_n \neq 0$ ) and its non-inertial version ( $\theta_n = 0$ ) are presented in Table 3.

**Table 3.** Performance of Algorithm 1 for for different dimensions  $\theta_n$ ,  $N = 3$  with  $TOL = 10^{-3}$ .

	$P = 4$			$P = 20$		
	Iter( $n$ )	CPU(s)	err( $_n$ )	Iter( $n$ )	CPU(s)	err( $_n$ )
$\theta = 0.1$	12	0.0811	0.3295	19	0.1005	0.4401
$\theta = 0.01$	11	0.0844	0.3112	17	0.0968	0.3224
$\theta = 0$	16	0.0960	0.5255	24	0.1203	0.4362

**Remark 2.** Tables 1 and 2 show that the CPU time and number of iterations of the algorithm increase linearly with the size or complexity of the problem (with the size of dimension  $p$  and  $q$ , number of mappings  $R$  and  $N$ , and number of functions  $M$ ). From Table 3, we can see that our algorithm has a better performance for the stepsize choice  $\theta_n \neq 0$ . This implies that the inertial version of our algorithm has a better convergence analysis.

## 6. Conclusions

In this paper, we have proposed the problem of minimizing a convex function over the solution set of the split feasibility problem of fixed point problems of demimetric mappings and constrained minimization problems of nonsmooth convex functions. We have showed that this problem can be solved by proximal and gradient methods where the gradient method is used for an upper level problem and the proximal method is used for a lower level problem. Most of the standard bilevel problems are particular cases of our framework.

**Author Contributions:** S.E.Y., P.K. and A.G.G. contributed equally in this research paper particularly on the conceptualization, methodology, validation, formal analysis, resource, and writing and preparing the original draft of the manuscript; however, P.K. fundamentally plays a great role in supervision and funding acquisition as well. Moreover, A.G.G. particularly wrote the code and run the algorithm in the MATLAB program. All authors have read and agreed to the published version of the manuscript.

**Funding:** Funding was received from the Petchra Pra Jom Klao Ph.D. Research Scholarship from King Mongkut's University of Technology Thonburi (KMUTT) and Theoretical and Computational Science (TaCS) Center. Moreover, Poom Kumam was supported by the Thailand Research Fund and the King Mongkut's University of Technology Thonburi under the TRF Research Scholar Grant No.RSA6080047. The Rajamangala University of Technology Thanyaburi (RMUTTT) (Grant No. NSF62D0604).

**Acknowledgments:** The authors acknowledge the financial support provided by the Center of Excellence in Theoretical and Computational Science (TaCS-CoE), KMUTT. Seifu Endris Yimer is supported by the Petchra Pra Jom Klao Ph.D. Research Scholarship from King Mongkut's University of Technology Thonburi (Grant No.9/2561).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Su, M.; Xu, H.K. Remarks on the gradient-projection algorithm. *J. Nonlinear Anal. Optim.* **2010**, *1*, 35–43.
2. Xu, H.K. Averaged mappings and the gradient-projection algorithm. *J. Optim. Theory Appl.* **2011**, *150*, 360–378. [[CrossRef](#)]
3. Ceng, L.C.; Ansari, Q.H.; Yao, J.C. Some iterative methods for finding fixed points and for solving constrained convex minimization problems. *Nonlinear Anal. Theor.* **2011**, *74*, 5286–5302. [[CrossRef](#)]
4. Rockafellar, R.T.; Wets, R.J. *Variational Analysis*; Springer Science & Business Media: Berlin, Germany, 2009.
5. Bauschke, H.H.; Combettes, P.L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*; Springer: New York, NY, USA, 2011.
6. Martinet, B. Brève communication. Régularisation d'inéquations variationnelles par approximations successives. *Revue Française d'Informatique et de Recherche Opérationnelle Série Rouge* **1970**, *4*, 154–158. [[CrossRef](#)]
7. Martinet, B. Détermination approchée d'un point fixe d'une application pseudo-contractante. *CR Acad. Sci. Paris* **1972**, *274*, 163–165.
8. Rockafellar, R.T. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.* **1976**, *14*, 877–898. [[CrossRef](#)]

9. Censor, Y.; Gibali, A.; Reich, S. Algorithms for the split variational inequality problem. *Numer. Algorithms* **2012**, *59*, 301–323. [[CrossRef](#)]
10. Hendrickx, J.M.; Olshevsky, A. Matrix p-Norms Are NP-Hard to Approximate If  $p \neq 1, 2, \infty$ . *SIAM J. Matrix Anal. A* **2010**, *31*, 2802–2812. [[CrossRef](#)]
11. Censor, Y.; Elfving, T. A multiprojection algorithm using Bregman projections in a product space. *Numer. Algorithms* **1994**, *8*, 221–239. [[CrossRef](#)]
12. Qu, B.; Xiu, N. A note on the CQ algorithm for the split feasibility problem. *Inverse Probl.* **2005**, *21*, 1655. [[CrossRef](#)]
13. Byrne, C. Iterative oblique projection onto convex sets and the split feasibility problem. *Inverse Probl.* **2002**, *18*, 441. [[CrossRef](#)]
14. López, G.; Martín-Márquez, V.; Wang, F.; Xu, H.K. Solving the split feasibility problem without prior knowledge of matrix norms. *Inverse Probl.* **2012**, *28*, 085004. [[CrossRef](#)]
15. Yimer, S.E.; Kumam, P.; Gebrie, A.G.; Wangkeeree, R. Inertial Method for Bilevel Variational Inequality Problems with Fixed Point and Minimizer Point Constraints. *Mathematics* **2019**, *7*, 841. [[CrossRef](#)]
16. Dempe, S.; Kalashnikov, V.; Pérez-Valdés, G.A.; Kalashnykova, N. Bilevel programming problems. In *Energy Systems*; Springer: Berlin, Germany, 2015.
17. Anh, T.V. Linesearch methods for bilevel split pseudomonotone variational inequality problems. *Numer. Algorithms* **2019**, *81*, 1067–1087. [[CrossRef](#)]
18. Anh, T.V.; Muu, L.D. A projection-fixed point method for a class of bilevel variational inequalities with split fixed point constraints. *Optimization* **2016**, *65*, 1229–1243. [[CrossRef](#)]
19. Yuying, T.; Van Dinh, B.; Plubtieng, S. Extragradient subgradient methods for solving bilevel equilibrium problems. *J. Inequal. Appl.* **2018**. [[CrossRef](#)]
20. Shehu, Y.; Vuong, P.T.; Zemkoho, A. An inertial extrapolation method for convex simple bilevel optimization. *Optim. Method Softw.* **2019**. [[CrossRef](#)]
21. Sabach, S.; Shtern, S. A first order method for solving convex bilevel optimization problems. *SIAM J. Optimiz.* **2017**, *27*, 640–660. [[CrossRef](#)]
22. Boţ, R.I.; Csetnek, E.R.; Nimana, N. An inertial proximal-gradient penalization scheme for constrained convex optimization problems. *Vietnam J. Math.* **2018**, *46*, 53–71. [[CrossRef](#)]
23. Solodov, M. An explicit descent method for bilevel convex optimization. *J. Convex Anal.* **2007**, *14*, 227.
24. Chang, S.S.; Quan, J.; Liu, J. Feasible iterative algorithms and strong convergence theorems for bi-level fixed point problems. *J. Nonlinear Sci. Appl.* **2016**, *9*, 1515–1528. [[CrossRef](#)]
25. Didi-Biha, M.; Marcotte, P.; Savard, G. Path-based formulations of a bilevel toll setting problem. In *Optimization with Multivalued Mappings*; Springer: Boston, MA, USA, 2006; pp. 29–50.
26. Mohideen, M.J.; Perkins, J.D.; Pistikopoulos, E.N. Optimal design of dynamic systems under uncertainty. *AIChE J.* **1996**, *42*, 2251–2272. [[CrossRef](#)]
27. Fampa, M.; Barroso, L.A.; Candal, D.; Simonetti, L. Bilevel optimization applied to strategic pricing in competitive electricity markets. *Comput. Optim. Appl.* **2008**, *39*, 121–142. [[CrossRef](#)]
28. Calvete, H.I.; Galé, C.; Oliveros, M.J. Bilevel model for production–distribution planning solved by using ant colony optimization. *Comput. Oper. Res.* **2011**, *38*, 320–327. [[CrossRef](#)]
29. Takahashi, S.; Takahashi, W. The split common null point problem and the shrinking projection method in Banach spaces. *Optimization* **2016**, *65*, 281–287. [[CrossRef](#)]
30. Song, Y. Iterative methods for fixed point problems and generalized split feasibility problems in Banach spaces. *J. Nonlinear Sci. Appl.* **2018**, *11*, 198–217. [[CrossRef](#)]
31. Cabot, A. Proximal point algorithm controlled by a slowly vanishing term: Applications to hierarchical minimization. *SIAM J. Optim.* **2005**, *15*, 555–572. [[CrossRef](#)]
32. Helou, E.S.; Simões, L.E.  $\epsilon$ -subgradient algorithms for bilevel convex optimization. *Inverse Probl.* **2017**, *33*, 5. [[CrossRef](#)]
33. Polyak, B.T. Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.* **1964**, *4*, 1–7. [[CrossRef](#)]
34. Lemaire, B. Which fixed point does the iteration method select? In *Recent Advances in Optimization*; Springer: Berlin/Heidelberg, Germany, 1997; pp. 154–167.
35. Xu, H.K. Iterative algorithms for nonlinear operators. *J. Lond. Math. Soc.* **2002**, *66*, 240–256. [[CrossRef](#)]

36. Maingé, P.E. Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization. *Set Valued Anal.* **2008**, *16*, 899–912. [[CrossRef](#)]
37. Combettes, P.L.; Hirstoaga, S.A. Equilibrium programming in Hilbert spaces. *J. Nonlinear Convex Anal.* **2005**, *6*, 117–136.
38. Aubin, J.P. *Optima and Equilibria: An Introduction to Nonlinear Analysis*; Springer Science & Business Media: Berlin, Germany, 2013.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).