Research Article

Anteneh Getachew Gebrie* and Rabian Wangkeeree

Strong convergence of an inertial extrapolation method for a split system of minimization problems

https://doi.org/10.1515/dema-2020-0025
received July 18, 2020; accepted October 26, 2020

Abstract: In this article, we propose an inertial extrapolation-type algorithm for solving split system of minimization problems: finding a common minimizer point of a finite family of proper, lower semicontinuous convex functions and whose image under a linear transformation is also common minimizer point of another finite family of proper, lower semicontinuous convex functions. The strong convergence theorem is given in such a way that the step sizes of our algorithm are selected without the need for any prior information about the operator norm. The results obtained in this article improve and extend many recent ones in the literature. Finally, we give one numerical example to demonstrate the efficiency and implementation of our proposed algorithm.

Keywords: minimization problem, Moreau-Yosida approximate, inertial term, strong convergence

MSC 2020: 65K10, 90C25, 49J52, 47H09

1 Introduction

Throughout this article, unless otherwise stated, we assume that H_1, H_2 and H are real Hilbert spaces, $A : H_1 \to H_2$ is nonzero bounded linear operator and I denotes the identity operator on a Hilbert space.

Assume C_i $(i = 1, \dots, N)$ and Q_i $(i = 1, \dots, M)$ are nonempty closed convex subsets of H_1 and H_2, respectively. The multiple-set split feasibility problem (MSSFP) which was introduced by Censor et al. [1] is formulated as finding a point

$$\bar{x} \in \bigcap_{i=1}^{N} C_i \text{ such that } A\bar{x} \in \bigcap_{j=1}^{M} Q_j. \tag{1}$$

In particular, if $N = M = 1$, then the MSSFP (1) is reduced to the problem known as the split feasibility problem (SFP) which was first introduced by Censor and Elfving [2] for modeling inverse problems in finite-dimensional Hilbert spaces. The SFP and MSSFP arise in many fields in the real world, and numerous methods have been proposed to solve the SFP, see for example [3–5] and references therein, and MSSFP, see for example [6–8] and references therein. Moreover, there are some studies of fixed point problems in the framework of the MSSFP, see for example [9–14].

* **Corresponding author: Anteneh Getachew Gebrie,** Department of Mathematics, Debre Berhan University, Debre Berhan, P.O. Box 445, Ethiopia, e-mail: antgetm@gmail.com
Rabian Wangkeeree: Department of Mathematics, Naresuan University, Phitsanulok 65000, Thailand, e-mail: rabianw@nu.ac.th

One of the most important problems in optimization theory and nonlinear analysis is the problem of approximating a solution of the unconstrained minimization problem. This can be stated as follows. Find $\bar{x} \in H$ such that

$$f(\bar{x}) = \min_{x \in H} f(x), \tag{2}$$

where $f : H \to \mathbb{R} \cup \{+\infty\}$ is proper, lower semicontinuous convex function. Our goal is to introduce a strong convergence iterative algorithm with inertial effect solving the MSSFP (1), where C_i and Q_j are solution sets of minimization problems of the form (2) for proper, lower semicontinuous convex functions f_i and g_j, respectively. We denote by arg min f the set of all minimizers of f on H, i.e.,

$$\arg\min f = \{\bar{x} \in H : f(\bar{x}) \le f(x), \forall x \in H\} = \{\bar{x} \in H : f(\bar{x}) = \min_{x \in H} f(x)\}.$$

If f is a smooth function (mostly if f is twice continuously differentiable), one of the numerical methods for finding approximate solutions of (2) is the Newton method, see [15,16]. Analogous method for solving (2) with better properties for the non-smooth case is based on the notion of proximal mapping introduced by Moreau [17], i.e., the proximal operator of the function f with scaling parameter $\lambda > 0$ is a mapping $\text{prox}_{\lambda f} : H \to H$ given by

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

The minimizers of f (points solving problem (2)) are precisely the fixed points of the proximal operator of f. Thus, solving the optimization problem (2) can be interpreted as finding fixed points of a proximal operator of f and proximal operators are firmly nonexpansive operators. This immediately suggests the most popular method

$$x_{n+1} = \text{prox}_{\lambda f}(x_n),$$

which is called the proximal minimization or the *proximal point algorithm* introduced by Martinet [18,19] and later by Rockafellar [20].

Let $f : H_1 \to \mathbb{R} \cup \{+\infty\}$, $g : H_2 \to \mathbb{R} \cup \{+\infty\}$ be two proper, convex, lower-semicontinuous functions, where g_λ is the Moreau-Yosida approximate [17] of the function g of parameter λ given by $g_\lambda(y) = \min_{u \in H_2}\left\{ g(u) + \frac{1}{2\lambda}\|y - u\|^2 \right\}$. In [21], Moudafi and Thakur introduced a weakly convergent algorithm solving the following minimization problem:

$$\min_{x \in H_1} \{ f(x) + g_\lambda(Ax) \}, \tag{3}$$

in case $\arg\min f \cap A^{-1}(\arg\min g) \neq \emptyset$. It should be noted that (3) is equivalent to the *split minimization problem* (SMP): finding a point $\bar{x} \in H_1$ with the property

$$\bar{x} \in \arg\min f \quad \text{such that} \quad A\bar{x} \in \arg\min g. \tag{4}$$

Operator norm is a global invariant and is often difficult to estimate, see for example the Theorem of Hendrickx and Olshevsky in [22]. However, in the several split inverse problem types in the literature, the implementation of the proposed iterative method requires the prior knowledge of operator norm to determine the step sizes. To overcome this difficulty, López et al. [4] introduced a new way of selecting the step sizes for solving the SFP such that the information of the operator norm is not necessary. Moudafi and Thakur [21] used the idea of López et al. [4] to introduce a new way of selecting the step sizes, given by

$$\theta_{\lambda\mu}(x) = \sqrt{\|A^*(I - \text{prox}_{\lambda g})Ax\|^2 + \|(I - \text{prox}_{\lambda\mu f})x\|^2}$$

with $h_\lambda(x) = \frac{1}{2}\|(I - \text{prox}_{\lambda g})Ax\|^2$ and $l_{\lambda\mu}(x) = \frac{1}{2}\|(I - \text{prox}_{\lambda\mu f})x\|^2$, such that the implementation of the iterative algorithm they proposed for solving (4) does not need any prior information about the operator norm. They proposed the following split proximal algorithm, which generates, from an initial point $x_1 \in H_1$ assume that x_n has been constructed and $\theta_\lambda(x_n) \neq 0$, then compute x_{n+1} via the rule

$$x_{n+1} = \operatorname{prox}_{\lambda\mu_n f}(x_n - \mu_n A^*(I - \operatorname{prox}_{\lambda g})Ax_n), \tag{5}$$

where step size $\mu_n = \rho_n \frac{h_\lambda(x_n) + l_{\lambda\mu_n}(x_n)}{\theta^2_{\lambda\mu_n}(x_n)}$ with $0 < \rho_n < 4$ and if $\theta_{\lambda\mu_n}(x_n) = 0$, then $x_{n+1} = x_n$ is a solution of SMP (4) and the iterative process stops; otherwise, we set $n := n + 1$ and go to (5). Based on Moudafi and Thakur [21] many iterative algorithms are proposed for solving SMP (4), see for example those by Abbas et al. in [23], Shehu et al. in [24], Shehu and Iyiola in [25–28] and Shehu and Ogbuisi in [29].

An inertial term is a two-step iterative method, and the next iterate is defined by making use of the previous two iterates. An *inertial extrapolation type algorithm*, i.e., an algorithm combining an inertial term, was first introduced by Polyak [30] as an acceleration process in solving a smooth convex minimization problem. It is well known that combining an algorithm with inertial term speeds up or accelerates the rate of convergence of the sequence generated by the algorithm. Consequently, a lot of research interest is now devoted to the inertial extrapolation-type algorithm, see [31–34] and references therein. Very recently, Shehu and Iyiola [25] proposed an inertial extrapolation-type algorithm for solving the SMP (4) using the setting

(a) $l(x) = \frac{1}{2}\|(I - \operatorname{prox}_{\lambda f})x\|^2$, $\nabla l(x) = (I - \operatorname{prox}_{\lambda f})x$,

(b) $h(x) = \frac{1}{2}\|(I - \operatorname{prox}_{\lambda g})Ax\|^2$, $\nabla h(x) = A^*(I - \operatorname{prox}_{\lambda g})Ax$ and $\theta(x_n) = \|\nabla l(x) + \nabla h(x)\|$.

They proposed the following weak convergence result.

Theorem 1.1. *Suppose the real parameters $\{\alpha_n\}$, $\{\beta_n\}$ and $\{\rho_n\}$ satisfy the following conditions:*
(c1) $\{\alpha_n\}$ is non-increasing sequence and $0 < \delta \le \alpha_n \le \frac{1}{2}$,
(c2) $\{\beta_n\}$ is non-increasing sequence and $0 \le \beta_n \le \frac{1-\kappa}{3} < \frac{1}{3}$ for some, $\kappa \in (0, 1)$,
(c3) $0 < \rho_n < 4$, $\liminf_{n\to\infty} \rho_n(4 - \rho_n) > 0$.
Then the sequence $\{x_n\}$ generated by the iterative algorithm

$$\begin{cases} x_0, x_1 \in H_1, \\ z_n = x_n + \beta_n(x_n - x_{n-1}), \\ y_n = z_n - \rho_n \dfrac{h(z_n) + l(z_n)}{\theta^2(z_n)}(\nabla l(z_n) + \nabla h_j(z_n)), \\ x_{n+1} = (1 - \alpha_n)z_n + \alpha_n y_n, \end{cases} \tag{6}$$

weakly converges to a point $\bar{x}$ solving the SMP (4).

Note that the proximal operator is a natural extension of the notion of a metric projection onto a closed convex set, i.e., $\operatorname{prox}_{\lambda f} = P_Q$, where $f = \delta_Q$ (f is the indicator function of a closed convex subset Q of H), and this perspective suggests various properties that we expect proximal operators to obey. However, there is a property that holds for the case of projection operators but not for the case of proximal operators in general. For example, consider a function h defined on H_2 given by $h(x) = \frac{1}{2}\|(I - \operatorname{prox}_{\lambda f})Ax\|^2$, where H_1 and H_2 are real Hilbert spaces, and $f : H_2 \to \mathbb{R} \cup \{+\infty\}$ is proper lower semicontinuous convex function. The function h is not differentiable at $x = \pm\lambda$ for the case $H_1 = H_2 = \mathbb{R}$, $A = I$ and $f(x) = |x|$, see [26]. However, if f is the indicator function of closed convex subset Q of H_2 ($f = \delta_Q$), then h is convex and weakly lower semicontinuous on H_1, and h is always differentiable and $\nabla h(x) = A^*(I - \operatorname{prox}_{\lambda f})Ax$, see [35].

Motivated by the above theoretical views, and inspired by results in [1,21,25], in this article we introduce the strong convergence theorem of an inertial extrapolation-type algorithm that incorporates a proximal operator, a viscosity method and an inertial term to solve the so-called *split system of minimization problem* (SSMP), given as a task finding a point $\bar{x} \in H_1$ with the property

$$\bar{x} \in \bigcap_{i \in \Phi} (\arg\min f_i) \text{ such that } A\bar{x} \in \bigcap_{j \in \Psi} (\arg\min g_j), \tag{7}$$

where $\Phi = \{1, \ldots, N\}$, $\Psi = \{1, \ldots, M\}$, $f_i : H_1 \to \mathbb{R} \cup \{+\infty\}$ and $g_j : H_2 \to \mathbb{R} \cup \{+\infty\}$ are proper, lower semicontinuous convex functions for $i \in \Phi$, $j \in \Psi$.

Let Γ be the solution set of SSMP (7), i.e.,

$$\Gamma = \{\bar{x} \in \bigcap_{i \in \Phi} (\arg\min f_i) : A\bar{x} \in \bigcap_{j \in \Psi} (\arg\min g_j)\}.$$

Note that if $f_i = f$ for all $i \in \Phi$ and $g_j = g$ for all $j \in \Psi$, then problem (7) reduces to the SMP (4) that is the problem considered in [21,23–29]. The aims of this study are twofold: to improve the weak convergence result of an inertial extrapolation-type algorithm proposed by Shehu and Iyiola [25] to a strong convergence result for an approximation of a solution of the SMP (4), and to accelerate and improve the results in [9,10] in solving the SSMP (7).

This article is organized in the following way. In Section 2, we collect some basic and useful definitions, lemmata, and theorems for further study. In Section 3, we propose an iterative method for the SSMP and analyze the strong convergence theorem of the proposed iterative method. In Section 4, we give a numerical example to discuss the performance of the proposed method. Finally, we give some conclusions.

2 Preliminary

In this section, in order to prove our result, we collect some facts and tools in a real Hilbert space H. The symbols "$\rightharpoonup$" and "$\to$" denote weak and strong convergence, respectively. Let C be a nonempty closed convex subset of H. The metric projection on C is a mapping $P_C : H \to C$ defined by

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

Lemma 2.1. *Let C be a closed convex subset of H. Given $x \in H$ and a point $z \in C$, then $z = P_C(x)$ if and only if $\langle x - z, y - z\rangle \le 0$, for all $y \in C$.*

Definition 2.1. Let $T : H \to H$. Then,

(a) T is L-Lipschitz if there exists $L > 0$ such that $\|Tx - Ty\| \le L\|x - y\|$, $\forall x, y \in H$. 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 firmly nonexpansive if

$$\|Tx - Ty\|^2 \le \|x - y\|^2 - \|(I - T)x - (I - T)y\|^2, \text{ for all } x, y \in H,$$

which is equivalent to $\|Tx - Ty\|^2 \le \langle Tx - Ty, x - y\rangle$, for all $x, y \in H$.

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

(c) T is strongly monotone if there exists a constant $\alpha > 0$ such that

$$\langle Tx - Ty, x - y\rangle \ge \alpha\|x - y\|^2, \text{ for all } x, y \in H.$$

(d) T is inverse strongly monotone if there exists a constant $\alpha > 0$ such that

$$\langle Tx - Ty, x - y\rangle \ge \alpha\|Tx - Ty\|^2, \text{ for all } x, y \in H.$$

Lemma 2.2. *For a real Hilbert space H, we have*

(i) $\|x + y\|^2 = \|x\|^2 + \|y\|^2 + 2\langle x, y\rangle$, *for all* $x, y \in H$;

(ii) $\|x + y\|^2 = \|x\|^2 + 2\langle y, x + y\rangle$, *for all* $x, y \in H$;

(iii) $\langle x, y\rangle \le \frac{1}{2}\|x\|^2 + \frac{1}{2}\|y\|^2 - \frac{1}{2}\|x - y\|^2$, *for all* $x, y \in H$.

A set-valued mapping $T : H \to 2^H$ is called monotone if, for all $x, y \in H$, $z \in Tx$ and $w \in Ty$ imply $\langle x - y, z - w\rangle \geq 0$. A monotone mapping $T : H \to 2^H$ is maximal if its graph $G(F) = \{(x, y) : y \in F(x),\ x \in H\}$ is not properly contained in the graph of any other monotone mapping. It is known that a monotone mapping T is maximal if, and only if, for all $(x, z) \in H \times H$, $\langle x - y, z - w\rangle \geq 0$ for all $(y, w) \in G(T)$, implies $z \in Tx$. If $T : H \to 2^H$ is a maximal monotone set-valued mapping, then we define the resolvent operator J_λ^T associated with T and $\lambda > 0$ as follows:

$$J_\lambda^T(x) = (I + \lambda T)^{-1}(x), \quad x \in H. \tag{8}$$

It is well known that J_λ^T is single-valued, nonexpansive (see, for example [37,36]) and 1-inverse strongly monotone (firmly nonexpansive). Moreover, $0 \in T(\bar{x})$ if and only if $\bar{x}$ is a fixed point of the resolvent operator J_λ^T for all $\lambda > 0$; see [38].

Let $f : H \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous convex function. The domain of f is denoted by dom f; that is, dom $f = \{x \in H : f(x) < \infty\}$. We denote the subdifferential of f at $x \in H$ by $\partial f(x)$, and is given by $\partial f(x) = \{y \in H : f(z) \geq f(x) + \langle y, z - x\rangle,\ \forall z \in H\}$. If $\partial f(x) \neq \emptyset$, f is said to be subdifferentiable at x. It is notable that a point $\bar{x} \in H$ minimizes f if and only if $0 \in \partial f(\bar{x})$. It is the classical result in operator theory that the subdifferential ∂f is a maximal monotone operator and $\text{prox}_{\lambda f} = (I + \lambda \partial f)^{-1}$, namely, for $x \in H$ we have the following equivalence between the subdifferential and proximal operator:

$$\text{prox}_{\lambda f}(x) = y \Leftrightarrow x - y \in \lambda \partial f(y).$$

Consequently, a point $\bar{x}$ minimizes f if and only if $\text{prox}_{\lambda f}(\bar{x}) = \bar{x}$. Hence, the convex minimization problem (2) can be formulated as finding fixed point of proximal operator.

Lemma 2.3. [39] *Let $\{c_n\}$ and $\{\gamma_n\}$ be sequences of nonnegative real numbers, $\{\beta_n\}$ be a sequence 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 $\limsup\limits_{n\to\infty} \frac{\beta_n}{\alpha_n} \leq 0$, then $c_n \to 0$ as $n \to \infty$.

Definition 2.2. Let $\{\Gamma_n\}$ be a real sequence. Then we say $\{\Gamma_n\}$ decrease 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 2.4. [40] *Let $\{\Gamma_n\}$ be a sequence of real numbers that does not decrease at infinity. Also 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 $\lim\limits_{n\to\infty} \varphi(n) = \infty$, 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 nonempty closed convex subset of H. Then we say that the bifunction $h : D \times D \to \mathbb{R}$ satisfies Condition CO on D if the following assumptions are satisfied:
(A1) $h(u, u) = 0$, for all $u \in D$;
(A2) h is monotone, i.e., $h(u, v) + h(v, u) \leq 0$, for all $u, v \in D$;
(A3) for each $u, v, w \in D$, $\limsup\limits_{\alpha\downarrow 0} h(\alpha w + (1 - \alpha)u, v) \leq h(u, v)$;
(A4) $h(u, .)$ is convex and lower semicontinuous on D for each $u \in D$.

Lemma 2.5. [41] *Let D be a nonempty closed convex subset of H and the bifunction $h : D \times D \to \mathbb{R}$ satisfies Condition CO on D. Then, for each $r > 0$ and $u \in H$, there exists $w \in D$ such that*

$$h(w, v) + \frac{1}{r}\langle v - w, w - u\rangle \geq 0, \text{ for all } v \in D.$$

The following lemma was given by Combettes and Hirstoaga in [42].

Lemma 2.6. [42] *If D is a nonempty closed convex subset of H and $h : D \times D \to \mathbb{R}$ is a bifunction satisfying Condition CO on D, then for each $r > 0$ and $u \in H$, the mapping $T_r^h : H \to \mathbb{R}$ (called the resolvent of h), given by*

$$T_r^h(u) = \left\{ w \in D : h(w, v) + \frac{1}{r}\langle v - w, w - u\rangle \geq 0,\ \forall v \in D \right\}$$

satisfies the following conditions:
(*a*) *T_r^h is single-valued and firmly nonexpansive;*
(*b*) *$Fix(T_r^h) = \{\bar{x} \in D : h(\bar{x}, y) \geq 0, \forall y \in D\}$, where $Fix(T_r^h)$ is the set of fixed points of T_r^h;*
(*c*) *$\{\bar{x} \in D : h(\bar{x}, y) \geq 0, \forall y \in D\}$ is closed and convex.*

3 Main result

First we extend the settings introduced by Moudafi and Thakur [21]. Let $\lambda > 0$. Then, for $x \in H_1$,
(I) for each $i \in \Phi$, define $l_i(x) = \frac{1}{2}\|(I - \mathrm{prox}_{\lambda f_i})x\|^2$ and $\nabla l_i(x) = (I - \mathrm{prox}_{\lambda f_i})x$,
(II) $l(x) = l_{i_x}(x)$ and $\nabla l(x) = \nabla l_{i_x}(x)$, where $i_x \in \arg\max\{l_i(x) : i \in \Phi\}$, i.e., $l(x) = \max\{l_i(x) : i \in \Phi\}$,
(III) for each $j \in \Psi$, define $h_j(x) = \frac{1}{2}\|(I - \mathrm{prox}_{\lambda g_j})Ax\|^2$ and $\nabla h_j(x) = A^*(I - \mathrm{prox}_{\lambda g_j})Ax$,
(IV) for each $j \in \Psi$, define $\theta_j(x) = \max\{\|\nabla h_j(x)\|, \|\nabla l(x)\|\}$.

Remark. From (I)–(IV) given above, $\|\nabla l_i(x)\| \leq \|\nabla l_{i_x}(x)\| = \|\nabla l(x)\|$, $l_i(x) = \frac{1}{2}\|\nabla l_i(x)\|^2$, $\|\nabla l(x)\| \leq \theta_j(x)$ and $\|\nabla h_j(x)\| \leq \theta_j(x)$ for all $i \in \Phi$ and for all $j \in \Psi$.

Consider the parameter sequences satisfying the following conditions.

Assumption 1. Suppose $\{\alpha_n\}$, $\{\varepsilon_n\}$, $\{\rho_n\}$, $\{\xi_n^{(j)}\}$ $(j \in \Psi)$ be real sequences satisfying the following conditions:
(C1) $0 < \alpha_n < 1$, $\lim\limits_{n\to\infty} \alpha_n = 0$ and $\sum_{n=1}^{\infty} \alpha_n = \infty$;
(C2) $\varepsilon_n > 0$ and $\varepsilon_n = o(\alpha_n)$;
(C3) $0 < \xi \leq \xi_n^{(j)} \leq 1$ and $\sum_{j\in\Psi} \xi_n^{(j)} = 1$ for each $n \geq 1$;
(C4) $0 < \rho_n < 2$ and $\liminf\limits_{n\to\infty} \rho_n(2 - \rho_n) > 0$.

We have plenty of choices for α_n and ε_n satisfying Conditions (C1) and (C2) of Assumption 1. For example, take $\alpha_n = \frac{1}{2n}$, $\varepsilon_n = \frac{1}{n^2}$. Thus, $0 < \alpha_n < 1$, $\lim\limits_{n\to\infty} \alpha_n = 0$ and $\lim\limits_{n\to\infty} \frac{\varepsilon_n}{\alpha_n} = 0$ (i.e., $\varepsilon_n = o(\alpha_n)$).

Using ∇l_i, l_i, l, ∇l, h_j, ∇h_j, θ_j given in (I)–(IV) and step sizes given in Assumption 1, we are now in a position to state our inertial extrapolation-type algorithm and prove its strong convergence to the solution of the SSMP (7) assuming that solution set Γ is nonempty.

Algorithm 1. Initialization: Let $V : H_1 \to H_1$ be a contraction mapping with constant γ. Choose $x_0, x_1 \in H_1$. Take arbitrary real numbers β and $\widehat{\Theta}$ such that $0 \leq \beta < 1$ and $\widehat{\Theta} > 0$. Let $\{\alpha_n\}$, $\{\varepsilon_n\}$, $\{\rho_n\}$, $\{\xi_n^{(j)}\}$ $(j \in \Psi)$ be real sequences satisfying Assumption 1.

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

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

Step 2. Evaluate $y_n = x_n + \beta_n(x_n - x_{n-1})$.
Step 3. For each $j \in \Psi$ find $l(y_n)$, $h_j(y_n)$, $\theta_j(y_n)$ and $\Psi_n = \{j \in \Psi\colon \theta_j(y_n) \neq 0\}$.
Step 4. For each $j \in \Psi$ evaluate $\mu_n^{(j)} = \rho_n \frac{h_j(y_n) + l(y_n)}{\Theta_j^2(y_n)}$, where

$$\Theta_j(y_n) = \begin{cases} \widehat{\Theta}, & \text{if } j \notin \Psi_n, \\ \theta_j(y_n), & \text{if } j \in \Psi_n. \end{cases}$$

Step 5. Evaluate

$$z_n = y_n - \frac{1}{2}\sum_{j\in\Psi}\{\xi_n^{(j)}\mu_n^{(j)}(\nabla h_j(y_n) + \nabla l(y_n))\}.$$

Step 6. Evaluate $x_{n+1} = \alpha_n V(y_n) + (1 - \alpha_n)z_n$.
Step 7. Set $n := n + 1$ and go to Step 1.

Remark. From Assumption 1 and Step 1 of Algorithm 1, we have that $\frac{\beta_n}{\alpha_n}\|x_n - x_{n-1}\| \to 0$, $n \to \infty$. Since $\{\alpha_n\}$ is bounded, we also have $\beta_n\|x_n - x_{n-1}\| \to 0$, $n \to \infty$.

Note that Step 1 of Algorithm 1 is easily implemented in numerical computation since the value of $\|x_n - x_{n-1}\|$ is a prior known before choosing β_n.

Remark. If $\Psi_n = \varnothing$, then

$$\begin{aligned} &\max\{\|\nabla h_j(y_n)\|, \|\nabla l(y_n)\|\} = 0 \Leftrightarrow \|\nabla h_j(y_n)\| = 0 = \|\nabla l(y_n)\|, \text{ for all } j \in \Psi, \\ &\quad\Leftrightarrow \|\nabla h_j(y_n)\| = 0 = \|\nabla l_i(y_n)\|, \text{ for all } i \in \Phi, j \in \Psi, \\ &\quad\Leftrightarrow A^*(I - \operatorname{prox}_{\lambda g_j})Ay_n = 0 = (I - \operatorname{prox}_{\lambda f_i})y_n, \text{ for all } i \in \Phi, j \in \Psi. \end{aligned}$$

Remark. The solution set Γ of problem (7) is closed convex set, because the set of minimizers of any proper, lower semicontinuous function is closed convex and A is bounded linear operator. Therefore, the metric projection P_Γ is well defined as we also assume that Γ is nonempty.

Lemma 3.1. *For the sequences $\{x_n\}$, $\{y_n\}$ and $\{z_n\}$ generated by Algorithm 1, we have*

$$\|z_n - \hat{x}\|^2 \le \|y_n - \hat{x}\|^2 + \rho_n(\rho_n - 2)\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n) + l(y_n))^2}{\Theta_j^2(y_n)}\right\}$$

for all $\hat{x} \in \Gamma$. Moreover, $\{x_n\}$, $\{y_n\}$ and $\{z_n\}$ are bounded sequences.

Proof. Let $\hat{x} \in \Gamma$. From the definition of y_n, we get

$$\|y_n - \hat{x}\| = \|x_n + \beta_n(x_n - x_{n-1}) - \hat{x}\| \le \|x_n - \hat{x}\| + \beta_n\|x_n - x_{n-1}\|. \tag{9}$$

Since $\operatorname{prox}_{\lambda f_i}$ and $\operatorname{prox}_{\lambda g_j}$ are firmly nonexpansive, $I - \operatorname{prox}_{\lambda f_i}$ and $I - \operatorname{prox}_{\lambda g_j}$ are also firmly nonexpansive, and since $\hat{x}$ verifies (7) (since minimizers of any function are exactly fixed points of its proximal mapping), we have for all $x \in H_1$

$$\langle \nabla l(x), x - \hat{x}\rangle = \langle \nabla l_{i_x}(x), x - \hat{x}\rangle = \langle (I - \operatorname{prox}_{\lambda f_{i_x}})x, x - \hat{x}\rangle \ge \|(I - \operatorname{prox}_{\lambda f_{i_x}})x\|^2 = 2l_{i_x}(x) = 2l(x) \tag{10}$$

and

$$\begin{aligned} \langle \nabla h_j(x), x - \hat{x}\rangle = \langle A^*(I - \operatorname{prox}_{\lambda g_j})Ax, x - \hat{x}\rangle = \langle (I - \operatorname{prox}_{\lambda g_j})Ax, Ax - A\hat{x}\rangle \ge \|(I - \operatorname{prox}_{\lambda g_j})Ax\|^2 = 2h_j(x), \\ \text{for all } j \in \Psi. \end{aligned} \tag{11}$$

Using the definition of z_n and Lemma 2.2 (i), we have

$$\begin{aligned}\|z_n - \hat{x}\|^2 &= \left\| y_n - \frac{1}{2}\sum_{j\in\Psi}\{\xi_n^{(j)}\mu_n^{(j)}(\nabla h_j(y_n) + \nabla l(y_n))\} - \hat{x} \right\|^2 \\ &\le \|y_n - \hat{x}\|^2 + \left\| \frac{1}{2}\sum_{j\in\Psi}\{\xi_n^{(j)}\mu_n^{(j)}(\nabla h_j(y_n) + \nabla l(y_n))\} \right\|^2 \\ &\quad - \left\langle \sum_{j\in\Psi}\{\xi_n^{(j)}\mu_n^{(j)}(\nabla h_j(y_n) + \nabla l(y_n))\}, y_n - \hat{x} \right\rangle.\end{aligned} \tag{12}$$

Noting $\|\nabla h_j(y_n)\| \le \Theta_j(y_n)$ and $\|\nabla l(y_n)\| \le \Theta_j(y_n)$ and using the convexity of $\|.\|^2$, we have

$$\begin{aligned}\left\| \frac{1}{2}\sum_{j\in\Psi}\{\xi_n^{(j)}\mu_n^{(j)}(\nabla h_j(y_n) + \nabla l(y_n))\} \right\|^2 &\le \frac{1}{2}\left\| \sum_{j\in\Psi}\xi_n^{(j)}\mu_n^{(j)}\nabla l(y_n) \right\|^2 + \frac{1}{2}\left\| \sum_{j\in\Psi}\xi_n^{(j)}\mu_n^{(j)}\nabla h_j(y_n) \right\|^2 \\ &\le \frac{1}{2}\sum_{j\in\Psi}\xi_n^{(j)}\|\mu_n^{(j)}\nabla l(y_n)\|^2 + \frac{1}{2}\sum_{j\in\Psi}\xi_n^{(j)}\|\mu_n^{(j)}\nabla h_j(y_n)\|^2 \\ &= \frac{1}{2}\sum_{j\in\Psi}\xi_n^{(j)}(\mu_n^{(j)})^2\|\nabla l(y_n)\|^2 + \frac{1}{2}\sum_{j\in\Psi}\xi_n^{(j)}(\mu_n^{(j)})^2\|\nabla h_j(y_n)\|^2 \\ &= \frac{1}{2}\sum_{j\in\Psi}\left\{\xi_n^{(j)}\left(\rho_n\frac{h_j(y_n) + l(y_n)}{\Theta_j^2(y_n)}\right)^2(\|\nabla l(y_n)\|^2 + \|\nabla h_j(y_n)\|^2)\right\} \\ &\le \sum_{j\in\Psi}\left\{\xi_n^{(j)}\left(\rho_n\frac{h_j(y_n) + l(y_n)}{\Theta_j^2(y_n)}\right)^2\Theta_j^2(y_n)\right\} \\ &= \rho_n^2\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n) + l(y_n))^2}{\Theta_j^2(y_n)}\right\}.\end{aligned} \tag{13}$$

From (10) and (11), we have

$$\begin{aligned}\left\langle \sum_{j\in\Psi}\{\xi_n^{(j)}\mu_n^{(j)}(\nabla h_j(y_n) + \nabla l(y_n))\}, y_n - \hat{x} \right\rangle &= \sum_{j\in\Psi}\{\xi_n^{(j)}\mu_n^{(j)}(\langle\nabla l(y_n), y_n - \hat{x}\rangle + \langle\nabla h_j(y_n), y_n - \hat{x}\rangle)\} \\ &\ge \sum_{j\in\Psi}\{\xi_n^{(j)}\mu_n^{(j)}(2l(y_n) + 2h_j(y_n))\} \\ &= \sum_{j\in\Psi}\left\{\xi_n^{(j)}\rho_n\frac{h_j(y_n) + l(y_n)}{\Theta_j^2(y_n)}(2l(y_n) + 2h_j(y_n))\right\} \\ &= 2\rho_n\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n) + l(y_n))^2}{\Theta_j^2(y_n)}\right\}.\end{aligned} \tag{14}$$

In view of (12), (13) and (14), we have

$$\begin{aligned}\|z_n - \hat{x}\|^2 &\le \|y_n - \hat{x}\|^2 + \rho_n^2\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n) + l(s_n))^2}{\Theta_j^2(y_n)}\right\} - 2\rho_n\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n) + l(y_n))^2}{\Theta_j^2(y_n)}\right\} \\ &= \|y_n - \hat{x}\|^2 + \rho_n(\rho_n - 2)\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n) + l(y_n))^2}{\Theta_j^2(y_n)}\right\}.\end{aligned} \tag{15}$$

Next show that the sequences $\{x_n\}$, $\{y_n\}$ and $\{z_n\}$ are bounded. From (15) and (C4) of Assumption 1, we have

$$\|z_n - \hat{x}\| \le \|y_n - \hat{x}\|. \tag{16}$$

Using (9), (16) and the definition of x_{n+1}, we get

$$\begin{aligned}\|x_{n+1}-\hat{x}\| &= \|\alpha_n V(y_n)+(1-\alpha_n)z_n-\hat{x}\| \\ &= \|(1-\alpha_n)(z_n-\hat{x})+\alpha_n(V(y_n)-V(\hat{x}))+\alpha_n(V(\hat{x})-\hat{x})\| \\ &\le (1-\alpha_n)\|z_n-\hat{x}\|+\alpha_n\|V(y_n)-V(\hat{x})\|+\alpha_n\|V(\hat{x})-\hat{x}\| \\ &= (1-\alpha_n)\|z_n-\hat{x}\|+\alpha_n\gamma\|y_n-\hat{x}\|+\alpha_n\|V(\hat{x})-\hat{x}\| \\ &\le (1-\alpha_n(1-\gamma))\|y_n-\hat{x}\|+\alpha_n\|V(\hat{x})-\hat{x}\| \\ &\le (1-\alpha_n(1-\gamma))\|x_n-\hat{x}\|+(1-\alpha_n(1-\gamma))\beta_n\|x_n-x_{n-1}\|+\alpha_n\|V(\hat{x})-\hat{x}\| \\ &\le (1-\alpha_n(1-\gamma))\|x_n-\hat{x}\|+\alpha_n(1-\gamma)\left\{\frac{(1-\alpha_n(1-\gamma))}{1-\gamma}\frac{\beta_n}{\alpha_n}\|x_n-x_{n-1}\|+\frac{\|V(\hat{x})-\hat{x}\|}{1-\gamma}\right\}.\end{aligned} \tag{17}$$

Observe that by (C1) of Assumption 1 and Remark 3, we see that

$$\lim_{n\to\infty}\frac{(1-\alpha_n(1-\gamma))}{1-\gamma}\frac{\beta_n}{\alpha_n}\|x_n-x_{n-1}\|=0.$$

Let

$$M=2\max\left\{\frac{\|V(\hat{x})-\hat{x}\|}{1-\gamma},\ \sup_{n\ge1}\frac{(1-\alpha_n(1-\gamma))}{1-\gamma}\frac{\beta_n}{\alpha_n}\|x_n-x_{n-1}\|\right\}.$$

Then (17) becomes

$$\|x_{n+1}-\hat{x}\|\le(1-\alpha_n(1-\gamma))\|x_n-\hat{x}\|+\alpha_n(1-\gamma)M.$$

Thus, by Lemma 2.3 the sequence $\{x_n\}$ is bounded. As a consequence, $\{y_n\}$, $\{V(y_n)\}$ and $\{z_n\}$ are also bounded. □

We now have the following strong convergence theorem for an approximation of solution of a Problem (7).

Theorem 3.2. *The sequence $\{x_n\}$ generated by Algorithm 1 converges strongly to $\bar{x}\in\Gamma$, where $\bar{x}=P_\Gamma V(\bar{x})$.*

Proof of Theorem 3.2.
Claim 1: *There exists a unique $\bar{x}\in H_1$ such that $\bar{x}=P_\Gamma V(\bar{x})$.*

As a result of

$$\|P_\Gamma V(x)-P_\Gamma V(y)\|\le\|V(x)-V(y)\|\le\gamma\|x-y\|,\ \text{for all } x,y\in H_1,$$

the mapping $P_\Gamma V$ is a contraction mapping of H_1 into itself. Hence, by the Banach contraction principle there exists a unique element $\bar{x}\in H_1$ such that $\bar{x}=P_\Gamma V(\bar{x})$. Clearly, $\bar{x}\in\Gamma$ and we have

$$\bar{x}=P_\Gamma V(\bar{x})\Leftrightarrow\langle\bar{x}-V(\bar{x}),y-\bar{x}\rangle\ge0,\ \text{for all } y\in\Gamma.$$

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

Let $\bar{x}\in\Gamma$, where $\bar{x}=P_\Gamma V(\bar{x})$. Now

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

From Lemma 2.2(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 $0\le\beta_n<1$, we get

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

Using the definition of x_{n+1} and Lemma 2.2(ii), we have

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

Lemma 3.1 together with (20) and (21) gives

$$\begin{aligned}\|x_{n+1}-\bar{x}\|^2 &= (1-\alpha_n)\|z_n-\bar{x}\|^2+2\alpha_n\langle V(y_n)-\bar{x},x_{n+1}-\bar{x}\rangle\\ &\le (1-\alpha_n)\|y_n-\bar{x}\|^2+2\alpha_n\langle V(y_n)-\bar{x},x_{n+1}-\bar{x}\rangle\\ &\quad+(1-\alpha_n)\rho_n(\rho_n-2)\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n)+l(y_n))^2}{\Theta_j^2(y_n)}\right\}\\ &\le (1-\alpha_n)\|x_n-\bar{x}\|^2+2(1-\alpha_n)\beta_n\|x_n-x_{n-1}\|^2\\ &\quad+(1-\alpha_n)\beta_n(\|x_n-\bar{x}\|^2-\|x_{n-1}-\bar{x}\|^2)\\ &\quad+2\alpha_n\langle V(y_n)-\bar{x},x_{n+1}-\bar{x}\rangle\\ &\quad+(1-\alpha_n)\rho_n(\rho_n-2)\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n)+l(y_n))^2}{\Theta_j^2(y_n)}\right\}.\end{aligned}\tag{22}$$

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

$$\begin{aligned}\|x_{n+1}-\bar{x}\|^2 &\le (1-\alpha_n)\|x_n-\bar{x}\|^2+2(1-\alpha_n)\beta_n\|x_n-x_{n-1}\|^2\\ &\quad+(1-\alpha_n)\beta_n(\|x_n-\bar{x}\|^2-\|x_{n-1}-\bar{x}\|^2)+\alpha_n M_1\\ &\quad+(1-\alpha_n)\rho_n(\rho_n-2)\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n)+l(y_n))^2}{\Theta_j^2(y_n)}\right\}.\end{aligned}\tag{23}$$

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 $\Gamma_{n+1}\le\Gamma_n$ for $n\ge n_0$. Then, $\{\Gamma_n\}$ converges and $\Gamma_n-\Gamma_{n+1}\to 0$ as $n\to 0$.

From (23) we have

$$\begin{aligned}(1-\alpha_n)\rho_n(2-\rho_n)\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n)+l(y_n))^2}{\Theta_j^2(y_n)}\right\} &\le (\Gamma_n-\Gamma_{n+1})+\alpha_n M_1+(1-\alpha_n)\beta_n(\Gamma_n-\Gamma_{n-1})\\ &\quad+2(1-\alpha_n)\beta_n\|x_n-x_{n-1}\|^2.\end{aligned}\tag{24}$$

Since $\Gamma_n-\Gamma_{n+1}\to 0$ alternatively $\Gamma_{n-1}-\Gamma_n\to 0$ and using (C1) of Assumption 1 and Remark 3 (noting $\alpha_n\to 0$, $0<\alpha_n<1$, $\beta_n\|x_n-x_{n-1}\|\to 0$ and $\{x_n\}$ is bounded), we have from (24)

$$(1-\alpha_n)\rho_n(2-\rho_n)\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n)+l(y_n))^2}{\Theta_j^2(y_n)}\right\}\to 0,\quad n\to\infty.\tag{25}$$

Conditions (C1) and (C4) of Assumption 1 (i.e., $0<\alpha_n<1$, $\alpha_n\to 0$ and $\liminf\limits_{n\to\infty}\rho_n(2-\rho_n)>0$) together with (25) yield

$$\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n)+l(y_n))^2}{\Theta_j^2(y_n)}\right\}\to 0,\quad n\to\infty.\tag{26}$$

In view of (26) and the restriction condition imposed on $\xi_n^{(j)}$ in (C3) of Assumption 1, we have

$$\frac{(h_j(y_n)+l(y_n))^2}{\Theta_j^2(y_n)}\to 0,\quad n\to\infty\tag{27}$$

for all $j\in\Psi$.

Now, using the definition of z_n and the convexity of $\|.\|^2$, we have

$$\begin{aligned}
\|y_n - z_n\|^2 &= \left\| \frac{1}{2}\sum_{j\in\Psi} \xi_n^{(j)}\mu_n^{(j)}(\nabla h_j(y_n) + \nabla l(y_n)) \right\|^2 \\
&\leq \frac{1}{2}\left\| \sum_{j\in\Psi} \xi_n^{(j)}\mu_n^{(j)}\nabla h_j(y_n) \right\|^2 + \frac{1}{2}\left\| \sum_{j\in\Psi} \xi_n^{(j)}\mu_n^{(j)}\nabla l(y_n) \right\|^2 \\
&\leq \frac{1}{2}\sum_{j\in\Psi} \xi_n^{(j)}(\mu_n^{(j)})^2\|\nabla h_j(y_n)\|^2 + \frac{1}{2}\sum_{j\in\Psi} \xi_n^{(j)}(\mu_n^{(j)})^2\|\nabla l(y_n)\|^2 \\
&= \frac{1}{2}\sum_{j\in\Psi}\{\xi_n^{(j)}(\mu_n^{(j)})^2(\|\nabla h_j(y_n)\|^2 + \|\nabla l(y_n)\|^2)\} \\
&\leq \sum_{j\in\Psi}\xi_n^{(j)}(\mu_n^{(j)})^2\Theta_j^2(y_n) = \sum_{j\in\Psi}\left\{\xi_n^{(j)}\left(\rho_n\frac{h_j(y_n)+l(y_n)}{\Theta_j^2(y_n)}\right)^2\Theta_j^2(y_n)\right\} \\
&\qquad = \rho_n^2\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n)+l(y_n))^2}{\Theta_j^2(y_n)}\right\} \\
&\qquad \leq 4\sum_{j\in\Psi}\left\{\xi_n^{(j)}\frac{(h_j(y_n)+l(y_n))^2}{\Theta_j^2(y_n)}\right\}.
\end{aligned} \tag{28}$$

Thus, (28) together with (26) gives

$$\|y_n - z_n\| \to 0, \quad n \to \infty. \tag{29}$$

Moreover, using the definition of y_n and Remark 3, we have

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

By (29) and (30), we get

$$\|x_n - z_n\| \leq \|x_n - y_n\| + \|y_n - z_n\| \to 0, \quad n \to \infty. \tag{31}$$

Using the definition of x_{n+1}, (C1) of Assumption 1 and noting that $\{V(y_n)\}$ and $\{z_n\}$ are bounded, we have The results from (31) and (32) give

$$\|x_{n+1} - x_n\| \leq \|x_{n+1} - z_n\| + \|z_n - x_n\| \to 0, \quad n \to \infty. \tag{33}$$

For each $i \in \Phi$ and for each $j \in \Psi$, $\nabla h_j(.)$ and $\nabla l_i(.)$ are Lipschitz continuous with constant $\|A\|^2$ and 1, respectively. Since the sequence $\{y_n\}$ is bounded and

$$\begin{aligned}
\|\nabla h_j(y_n)\| &= \|\nabla h_j(y_n) - \nabla h_j(\bar{x})\| \leq \|A\|^2\|y_n - \bar{x}\|, \text{ for all } j \in \Psi, \\
\|\nabla l_i(y_n)\| &= \|\nabla l_i(y_n) - \nabla l_i(\bar{x})\| \leq \|y_n - \bar{x}\|, \text{ for all } i \in \Phi,
\end{aligned}$$

we have the sequences $\{\|\nabla l_i(y_n)\|\}_{n=1}^{\infty}$ and $\{\|\nabla h_j(y_n)\|\}_{n=1}^{\infty}$ are bounded for each $i \in \Phi$ and $j \in \Psi$. Hence, the boundedness of $\{\|\nabla l_i(y_n)\|\}_{n=1}^{\infty}$ for all $i \in \Phi$ gives $\{\|\nabla l(y_n)\|\}_{n=1}^{\infty}$ is a bounded. Thus, we have $\{\theta_j^2(y_n)\}_{n=1}^{\infty}$ is bounded for each $j \in \Psi$. Therefore, the sequence $\{\Theta_j^2(y_n)\}_{n=1}^{\infty}$ is bounded sequence for each $j \in \Psi$. Consequently, using (27), we have

$$\lim_{n\to\infty}(h_j(y_n) + l(y_n)) = 0 \Leftrightarrow \lim_{n\to\infty} h_j(y_n) = \lim_{n\to\infty} l(y_n) = 0, \text{ for all } j \in \Psi.$$

By definition of $l(y_n)$, we have $l_i(y_n) \leq l(y_n)$, for all $i \in \Phi$. Therefore,

$$\lim_{n\to\infty} h_j(y_n) = \lim_{n\to\infty} l_i(y_n) = 0, \ \forall i \in \Phi, \text{ for all } j \in \Psi.$$

Let p be a weak cluster point of $\{x_n\}$, there exists a subsequence $\{x_{n_k}\}$ of $\{x_n\}$ such that $x_{n_k} \rightharpoonup p$ as $k \to \infty$. Then, in view of $x_{n_k} \rightharpoonup p$ and (30), we also see that $y_{n_k} \rightharpoonup p$. The weak lower-semicontinuity of $h_j(.)$ implies that

$$0 \leq h_j(p) \leq \liminf_{k\to\infty} h_j(y_{n_k}) = \lim_{n\to\infty} h_j(y_n) = 0, \text{ for all } j \in \Psi.$$

That is, $h_j(p) = \frac{1}{2}\|(I - \mathrm{prox}_{\lambda g_j})Ap\|^2 = 0$ for all $j \in \Psi$, i.e., Ap is a fixed point of the proximal mapping of each g_j, or equivalently, $0 \in \partial g_j(Ap)$ for all $j \in \Psi$. In other words, Ap is a minimizer of each g_j for all $j \in \Psi$.

Likewise, the weak lower-semicontinuity of $l_i(.)$ implies that

$$0 \le l_i(p) \le \liminf_{k\to\infty} l_i(y_{n_k}) = \lim_{n\to\infty} l_i(y_n) = 0, \text{ for all } i \in \Phi.$$

That is, $l_i(p) = \frac{1}{2}\|(I - \mathrm{prox}_{\lambda f_i})p\|^2 = 0$ for all $i \in \Phi$, i.e., p is a fixed point of the proximal mapping of each f_i or equivalently, $0 \in \partial f_i(p)$ for all $i \in \Phi$. In other words, p is a minimizer of each f_i for all $i \in \Phi$. Thus, $p \in \Gamma$.

Next, we show that $\limsup_{n\to\infty}\langle (I - V)\bar{x}, \bar{x} - x_n\rangle \le 0$. Indeed, since $\bar{x} = P_\Gamma V(\bar{x})$ and $p \in \Gamma$ we obtain

$$\limsup_{n\to\infty}\langle (I - V)\bar{x}, \bar{x} - x_n\rangle = \lim_{k\to\infty}\langle (I - V)\bar{x}, \bar{x} - x_{n_k}\rangle = \langle (I - V)\bar{x}, \bar{x} - p\rangle \le 0. \tag{34}$$

Since $\|x_{n+1} - x_n\| \to 0$ from (33), by (34) we have

$$\limsup_{n\to\infty}\langle (I - V)\bar{x}, \bar{x} - x_{n+1}\rangle \le 0.$$

Using Lemma 3.1 (the fact that $\|z_n - \bar{x}\| \le \|y_n - \bar{x}\|$, i.e., from (16)), we have

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

Therefore, from (35), we have

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

Combining (36) and

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

it follows that

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

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

$$\begin{aligned}
\Gamma_{n+1} &\le \left(1 - \frac{2\alpha_n(1 - \gamma)}{1 + \alpha_n(1 - \gamma)}\right)\Gamma_n + \beta_n\|x_n - x_{n-1}\|(\beta_n\|x_n - x_{n-1}\| + 2M_2) + \frac{2\alpha_n}{1 + \alpha_n(1 - \gamma)}\langle V(\bar{x}) - \bar{x}, x_{n+1} - \bar{x}\rangle \\
&= (1 - \delta_n)\Gamma_n + \delta_n\vartheta_n,
\end{aligned} \tag{38}$$

where $\delta_n = \frac{2\alpha_n(1-\gamma)}{1+\alpha_n(1-\gamma)}$ and

$$\vartheta_n = \frac{1+\alpha_n(1-\gamma)}{2(1-\gamma)}\left(\frac{\beta_n}{\alpha_n}\|x_n - x_{n-1}\|\right)\{\beta_n\|x_n - x_{n-1}\| + 2M_2\} + \frac{1}{1-\gamma}\langle V(\bar{x}) - \bar{x}, x_{n+1} - \bar{x}\rangle.$$

From (35), (C1) of Assumption 1 and Remark 3, we have $\sum_{n=1}^{\infty}\delta_n = \infty$ and $\limsup_{n\to\infty}\vartheta_n \le 0$. Thus, using Lemma 2.3 and (38), we get $\Gamma_n \to 0$ as $n \to \infty$. Hence, $x_n \to \bar{x}$ as $n \to \infty$.

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

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

By Lemma 2.4, $\{\varphi(n)\}_{n=n_0}^{\infty}$ is a nondecreasing sequence, $\varphi(n) \to \infty$ as $n \to \infty$ and

$$\Gamma_{\varphi(n)} \le \Gamma_{\varphi(n)+1} \text{ and } \Gamma_n \le \Gamma_{\varphi(n)+1}, \text{ for all } n \ge n_0. \tag{39}$$

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

$$\begin{aligned}
&(1-\alpha_{\varphi(n)})\rho_{\varphi(n)}(2-\rho_{\varphi(n)})\sum_{j\in\Psi}\left\{\xi_{\varphi(n)}^{(j)}\frac{(h_j(y_{\varphi(n)}) + l(y_{\varphi(n)}))^2}{\Theta_j^2(y_{\varphi(n)})}\right\}\\
&\le (\Gamma_{\varphi(n)} - \Gamma_{\varphi(n)+1}) + \alpha_{\varphi(n)}M_1 + (1-\alpha_{\varphi(n)})\beta_{\varphi(n)}(\Gamma_{\varphi(n)} - \Gamma_{\varphi(n)-1})\\
&\quad + 2(1-\alpha_{\varphi(n)})\beta_{\varphi(n)}\|x_{\varphi(n)} - x_{\varphi(n)-1}\|^2\\
&\le \alpha_{\varphi(n)}M_1 + (1-\alpha_{\varphi(n)})\beta_{\varphi(n)}(\Gamma_{\varphi(n)} - \Gamma_{\varphi(n)-1}) + 2(1-\alpha_{\varphi(n)})\beta_{\varphi(n)}\|x_{\varphi(n)} - x_{\varphi(n)-1}\|^2\\
&\le \alpha_{\varphi(n)}M_1 + (1-\alpha_{\varphi(n)})\beta_{\varphi(n)}\|x_{\varphi(n)} - x_{\varphi(n)-1}\|\left(\sqrt{\Gamma_{\varphi(n)}} + \sqrt{\Gamma_{\varphi(n)-1}}\right)\\
&\quad + 2(1-\alpha_{\varphi(n)})\beta_{\varphi(n)}\|x_{\varphi(n)} - x_{\varphi(n)-1}\|^2.
\end{aligned}\tag{40}$$

Thus, from (40) together with (C1) and (C2) from Assumption 1 and Remark 3, we have for each $j \in \Psi$,

$$\frac{(h_j(y_{\varphi(n)}) + l(y_{\varphi(n)}))^2}{\Theta_j^2(y_{\varphi(n)})} \to 0, \quad n \to \infty. \tag{41}$$

Using a similar procedure to that in Case 1, we have

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

Since $\{x_{\varphi(n)}\}$ is bounded, there exists a subsequence of $\{x_{\varphi(n)}\}$, still denoted by $\{x_{\varphi(n)}\}$ which converges weakly to p. By a similar argument to that in Case 1, we conclude immediately that $p \in \Gamma$. In addition, by the similar argument to that in Case 1, we have $\limsup_{n\to\infty}\langle (I-V)\bar{x}, \bar{x} - x_{\varphi(n)}\rangle \le 0$. Since $\lim_{n\to\infty}\|x_{\varphi(n)+1} - x_{\varphi(n)}\| = 0$, we get $\limsup_{n\to\infty}\langle (I-V)\bar{x}, \bar{x} - x_{\varphi(n)+1}\rangle \le 0$. From (38), we have

$$\Gamma_{\varphi(n)+1} \le (1-\delta_{\varphi(n)})\Gamma_{\varphi(n)} + \delta_{\varphi(n)}\vartheta_{\varphi(n)}, \tag{42}$$

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

$$\vartheta_{\varphi(n)} = \frac{1+\alpha_{\varphi(n)}(1-\gamma)}{2(1-\gamma)}\left(\frac{\beta_{\varphi(n)}}{\alpha_{\varphi(n)}}\|x_{\varphi(n)} - x_{\varphi(n)-1}\|\right)\{\beta_{\varphi(n)}\|x_{\varphi(n)} - x_{\varphi(n)-1}\| + 2M_2\} + \frac{1}{1-\gamma}\langle V(\bar{x}) - \bar{x}, x_{\varphi(n)+1} - \bar{x}\rangle.$$

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

$$0 \le -\delta_{\varphi(n)}\Gamma_{\varphi(n)} + \delta_{\varphi(n)}\vartheta_{\varphi(n)}.$$

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

This completes the proof. □

Remark.
(I). Our iterative scheme has relatively low computational complexity compared to iterative schemes in [9] and [10].
(II). The implementation of our algorithm does not need any prior information about the operator norm.
(III). We can also use $\theta_j(x) = \sqrt{\|\nabla h_j(x)\|^2 + \|\nabla l(x)\|^2}$ instead of $\theta_j(x) = \max\{\|\nabla h_j(x)\|, \|\nabla l(x)\|\}$ perhaps the proof for the strong convergence theorem is almost similar. It is clear to see that $\max\{\|\nabla h_j(x)\|, \|\nabla l(x)\|\} \leq \sqrt{\|\nabla h_j(x)\|^2 + \|\nabla l(x)\|^2}$, for $j \in \Psi$.

Remark. One of the main advantages of our algorithm is that the algorithm can be used to solve problems that can be converted to the fixed point problem of firmly nonexpansive mapping. The following are some examples.
(1) The *split system of inclusion problem*: Let $T_i : H_1 \to 2^{H_1}$, $U_j : H_2 \to 2^{H_2}$ be maximal monotone mappings for all $i \in \Phi$ and $j \in \Psi$. The split system of inclusion problem is to find $\bar{x} \in H_1$ such that

$$\begin{cases} 0 \in T_i(\bar{x}), & \text{for all } i \in \Phi, \\ 0 \in U_j(A\bar{x}), & \text{for all } j \in \Psi. \end{cases} \tag{43}$$

Replacing the proximal mappings of the convex functions f_i and g_j in Algorithm 1 by the resolvent operators $J_\lambda^{T_i}$ and $J_\lambda^{U_j}$ to the maximal monotone operators, and following the method of proof in Theorem 3.2, we obtain an inertial extrapolation-type algorithm with a strong convergence result for approximation of solution of a consistent split system of inclusion problem (43); see the resolvent operator defined in (8).
(2) The *MSSFP*: By taking $f_i = \delta_{C_i}$ and $g_j = \delta_{Q_j}$ (the indicator functions) for $i \in \Phi$, $j \in \Psi$, and replacing $\text{prox}_{\lambda f_i}$ by projection mapping P_{C_i}, and $\text{prox}_{\lambda g_j}$ by the projection mapping P_{Q_j} in Algorithm 1, we obtain an inertial extrapolation-type algorithm with strong convergence for an approximation of solution of the MSSFP (1).
(3) The *split system of equilibrium problem*: Let $f_i : H_1 \times H_1 \to \mathbb{R}$ and $g_j : H_2 \times H_2 \to \mathbb{R}$ be bifunctions, where $i \in \Phi$, $j \in \Psi$. Assume each bifunction f_i and g_j satisfy Condition CO on H_1 and H_2, respectively. The split system of equilibrium problem involves finding $\bar{x} \in H_1$ such that

$$\begin{cases} f_i(\bar{x}, x) \geq 0, & \text{for all } x \in H_1, \quad i \in \Phi, \\ g_j(A\bar{x}, u) \geq 0, & \text{for all } u \in H_2, \quad j \in \Psi. \end{cases} \tag{44}$$

Our result solves (44) by replacing the proximal mappings by the resolvent operators $T_\lambda^{f_i}$ and $T_\lambda^{g_j}$ in Algorithm 1 and then following the method of proof in Theorem 3.2; see the resolvent operator defined in Lemma 2.6.

It is worth mentioning that our approach also works for approximation of solution of SMP (4). Let Ω denote the solution set of (4), i.e., $\Omega = \{\bar{x} \in \arg\min f : A\bar{x} \in \arg\min g\}$, and assume that Ω is nonempty. Set $l(x) = \frac{1}{2}\|(I - \text{prox}_{\lambda f})x\|^2$, $\nabla l(x) = (I - \text{prox}_{\lambda f})x$, $h(x) = \frac{1}{2}\|(I - \text{prox}_{\lambda g})Ax\|^2$, $\nabla h(x) = A^*(I - \text{prox}_{\lambda g})Ax$ and $\theta(x) = \max\{\|\nabla h(x)\|, \|\nabla l(x)\|\}$. The following is an immediate consequence of our result.

Algorithm 2. Initialization: Let $V : H_1 \to H_1$ be a contraction with constant γ. Choose $x_0, x_1 \in H_1$ and take arbitrary real numbers β and $\widehat{\Theta}$ such that $0 \leq \beta < 1$ and $\widehat{\Theta} > 0$. Let $\{\alpha_n\}, \{\varepsilon_n\}, \{\rho_n\}$ be real sequences satisfying the following conditions:
(a) $0 < \alpha_n < 1$, $\lim\limits_{n\to\infty} \alpha_n = 0$ and $\sum_{n=1}^{\infty} \alpha_n = \infty$;
(b) $\varepsilon_n > 0$ and $\varepsilon_n = o(\alpha_n)$;
(c) $0 < \rho_n < 2$ and $\liminf\limits_{n\to\infty} \rho_n(2 - \rho_n) > 0$.

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

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

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

Step 3. Evaluate $\mu_n = \rho_n \frac{h(y_n) + l(y_n)}{\Theta^2(y_n)}$, where $\Theta(y_n) = \begin{cases} \widehat{\Theta}, & \text{if } \theta(y_n) = 0; \\ \theta(y_n), & \text{otherwise}. \end{cases}$

Step 4. Evaluate $z_n = y_n - \frac{1}{2}\mu_n(\nabla h(y_n) + \nabla l(y_n))$.

Step 5. Evaluate $x_{n+1} = \alpha_n V(y_n) + (1 - \alpha_n)z_n$.

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

Corollary 1. *The sequence $\{x_n\}$ generated by Algorithm 2 converges strongly to $\bar{x} \in \Omega$, where $\bar{x} = P_\Omega V(\bar{x})$.*

Proof. Setting $f_i = f$ for all $i \in \Phi$ and $g_j = g$ for all $j \in \Psi$ in Theorem 3.2, we obtain the desired result. □

Remark. The result in Corollary 1 is an improvement on inertial extrapolation-type algorithms in the sense that instead of weak convergence result proposed by Shehu and Iyiola in [25] we get strong convergence result.

4 Numerical results

In this section, we consider an example of the SSMP involving quadratic optimization problems. We study the behavior of our algorithm and compare with the proximal-type algorithms of [9] and [10]. The algorithm has been coded in MATLAB and is performed on a HP laptop with Intel(R) Core(TM) i5-7200U CPU @ 250GHz 2.70GHz and RAM 4.00GB.

Example. Consider the problem (7) for $H_1 = \mathbb{R}^p$, $H_2 = \mathbb{R}^q$, a linear transformation $A : \mathbb{R}^p \to \mathbb{R}^q$ and functions

$$f_i(x) = \frac{1}{2}x^T B_i x + x^T D_i \quad (i \in \Phi = \{1, \ldots, N\}), \quad g_1(u) = \|u\|_q \text{ and } g_2(u) = \sum_{k=1}^{q} h(u_k),$$

where A is $q \times p$ non-zero matrix, B_i is an invertible symmetric positive semidefinite $p \times p$ matrix and D_i is the zero vector in $\mathbb{R}^p$ for all $i \in \Phi$, $u = (u_1, \ldots, u_q) \in \mathbb{R}^q$, $\|.\|_q$ is the Euclidean norm in $\mathbb{R}^q$ and $h(u_k) = \max\{|u_k| - 1, 0\}$ for $k = 1, \ldots, q$.

In this example, it is clear to see that $\Gamma = \{0\}$. Now for $\lambda = 1$, the proximal operators are given by

$$\mathrm{prox}_{\lambda f_i}(x) = (I + B_i)^{-1}(x - D_i) = (I + B_i)^{-1}(x), \quad i \in \Phi,$$

$$\mathrm{prox}_{\lambda g_1}(u) = \begin{cases} \left(1 - \dfrac{1}{\|u\|_q}\right)u, & \|u\|_q \geq 1, \\ 0, & \text{otherwise} \end{cases}$$

and $\mathrm{prox}_{\lambda g_2}(u) = (\mathrm{prox}_{\lambda h}(u_1), \mathrm{prox}_{\lambda h}(u_2), \ldots, \mathrm{prox}_{\lambda h}(u_q))$, where

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

In this numerical experiment, we use $N = 3$, $p = q$, A is identity $p \times p$ matrix and B_1, B_2 and B_3 are randomly generated invertible symmetric positive semidefinite $p \times p$ matrices.

Experiment 1 (**Studying numerical behavior of Algorithm 1**)**:** Figures 1, 2 and 3 and Tables 1 and 2 describe the numerical results of our algorithm for this example, where $V : \mathbb{R}^p \to \mathbb{R}^p$ given by $Vx = \gamma x$ and $\gamma = 0.5$, $\alpha_n = \frac{1}{2n^t}$, $\varepsilon_n = \frac{1}{n^r}$, $\xi_n^{(j)} = \frac{j}{3}$, $\beta = 0.9$ and $\beta_n = \bar{\beta}_n$ for $0 < t \leq 1$, $r > t$, $j \in \Psi = \{1, 2\}$.

Tables 1 and 2 illustrate the execution time in second (CPU(s)) and the number of iterations (Iter(n)) of our algorithm when applied to this particular example. The stopping criterion in Tables 1 and 2 is defined as $\frac{\|x_n - x_{n-1}\|}{\|x_2 - x_1\|} \leq \text{TOL} = 10^{-4}$.

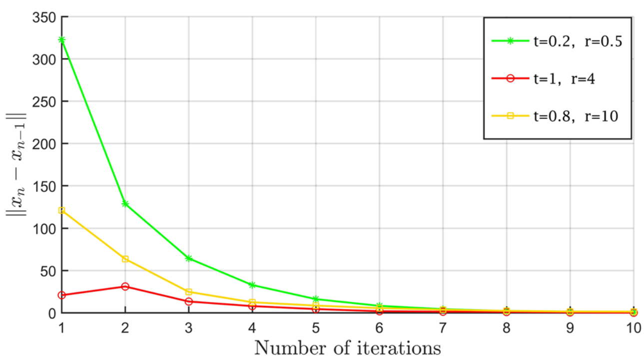

Figure 1: For $p = q = 6$ and for randomly generated starting points x_0 and x_1.

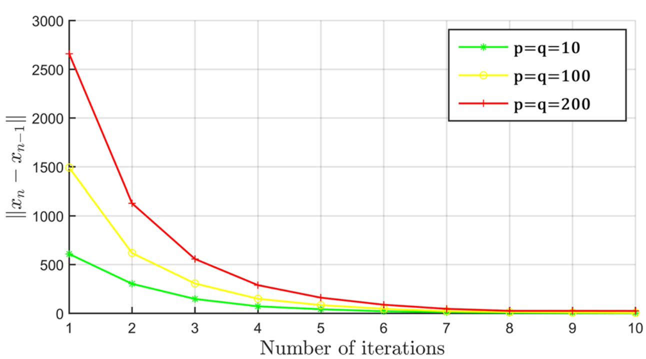

Figure 2: For $t = 0.1$, $r = 0.4$ and for randomly generated starting points x_0 and x_1.

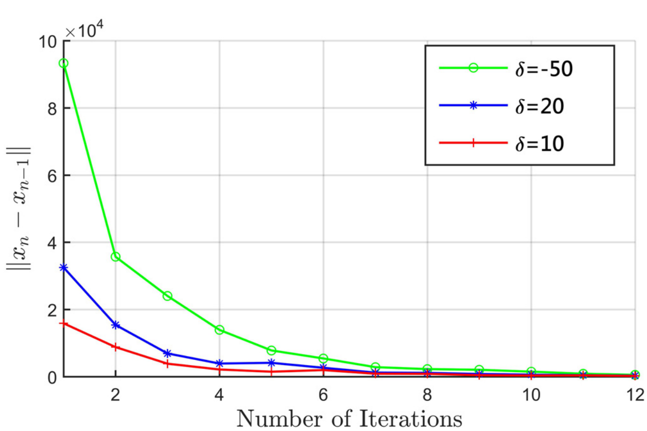

Figure 3: For $t = 0.2$, $r = 0.5$, $p = q = 500$, $x_1 = \delta x_0$, where $\delta \in \mathbb{R}$ and x_0 is randomly generated starting point.

Table 1: For randomly generated starting points x_0 and x_1

	$p = q = 3$		$p = q = 20$		$p = q = 80$	
	Iter(n)	**CPU(s)**	**Iter(n)**	**CPU(s)**	**Iter(n)**	**CPU(s)**
$t = 0.1$, $r = 0.2$	17	0.0127	14	0.0169	16	0.0269
$t = 0.95$, $r = 3$	21	0.0136	17	0.0185	21	0.0304
$t = 1$, $r = 10$	21	0.0129	16	0.0164	24	0.0299

Table 2: For $t = 0.55$, $r = 8$, $p = q = 4$

$x_0 = (1, 2, 3, 4)$, $x_1 = (-4, 7, -5, 10)$			$x_0 = (5, 6, 7, 8)$, $x_1 = (4, -7, 5, -10)$		
Iter(n)	**CPU(s)**	$\|x_n - x_{n-1}\|$	**Iter(n)**	**CPU(s)**	$\|x_n - x_{n-1}\|$
1		**12.2474**	**1**		**23.7486**
2		**5.6570**	**2**		**9.0177**
3		**3.5203**	**3**		**4.5572**
4		**1.9316**	**4**		**2.3349**
5		**1.1915**	**5**		**1.8997**
.					.
.					.
.					.
24		2.5606×10^{-4}	**24**		5.1148×10^{-4}
25	**0.0438**	1.3608×10^{-4}	**25**		3.6416×10^{-4}
			26		3.1116×10^{-4}
			27		2.4828×10^{-4}
			28		2.0318×10^{-4}
			29	**0.0613**	1.1041×10^{-4}

Experiment 2 (Comparison): We now compare our result with non-inertial extrapolated (non-accelerated) proximal-type algorithms [9] (ProxAL-A) and [10] (ProxAL-B). For this purpose, we use the following data:

Algorithm 1: $V : \mathbb{R}^p \to \mathbb{R}^p$ given by $Vx = \gamma x$ and $\gamma = 0.5$, $\alpha_n = \frac{1}{n}$, $\varepsilon_n = \frac{1}{n^2}$, $\xi_n^{(j)} = \frac{j}{3}$ $(j \in \Psi = \{1, 2\})$, $\beta = 0.8$, $\beta_n = \bar{\beta}_n$ and $\rho_n = \frac{1}{10}$.

ProxAL-A: $V = F = I$, $\mu = 1$, $\gamma = 0.5$, $\delta_n^{(i)} = \frac{i}{6}$ $(i \in \Phi = \{1, 2, 3\})$, $\xi_n^{(j)} = \frac{j}{3}$ $(j \in \Psi = \{1, 2\})$, $\alpha_n = \frac{1}{n+1}$ and $\rho_n = \frac{1}{10}$, see [9].

ProxAL-B: $\delta_n^{(i)} = \frac{i}{6}$ $(i \in \Phi = \{1, 2, 3\})$, $\xi_n^{(j)} = \frac{j}{3}$ $(j \in \Psi = \{1, 2\})$ and $\rho_n = \frac{1}{10}$, see [10].

Figures 4 and 5 along with Table 3 present the numerical results of our algorithm (Algorithm 1) in comparison with ProxAL-A and ProxAL-B. Figures 4 and 5 show the error $= \|x_n\|$ versus number of iterations, while Table 2 shows the CPU time exclusion (CPU(s)) and the number of iterations (Iter(n)) of Algorithm 1, ProxAL-A and ProxAL-B for the stopping criteria $\frac{\|x_n - x_{n-1}\|}{\| x_2 - x_1 \|} \le \mathrm{TOL} = 10^{-3}$.

From this preliminary numerical experiment, we observe that our algorithm crucially depends on step sizes, starting points and dimensions. Moreover, our proposed algorithm is efficient and easy to implement and outperforms the proposed algorithms in [9] and [10].

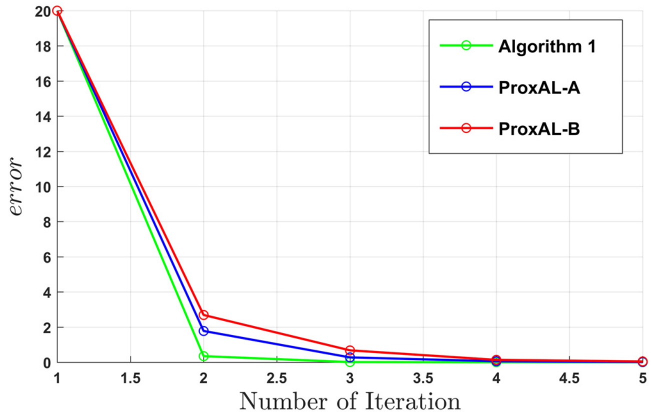

Figure 4: For $p = q = 4$ and $x_0 = (1, 1, 1, 1)$, $x_1 = 10x_0$.

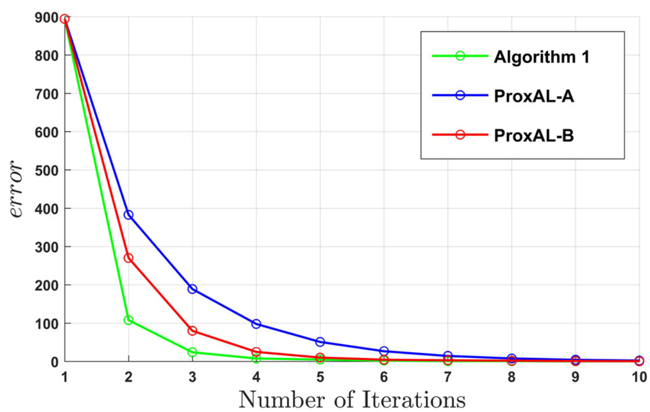

Figure 5: For $p = q = 80$ and $x_0 = (10, \ldots, 10)$, $x_1 = 10x_0$.

Table 3: For $x_0, x_1 \in \mathbb{R}^p$ with $x_0 = (-100, -100, \ldots, -100)$ and $x_1 = -2x_0$

	Algorithm 1		ProxAlg-A		ProxAlg-B	
Dimension	**Iter(n)**	**CPU(s)**	**Iter(n)**	**CPU(s)**	**Iter(n)**	**CPU(s)**
$p = q = 2$	7	0.0354	7	0.0401	9	0.0379
$p = q = 10$	12	0.0441	15	0.0616	21	0.1775
$p = q = 50$	27	0.1908	29	0.2168	29	0.1921

5 Conclusions

In this article, we introduce a strong convergence theorem for an inertial extrapolation-type algorithm for solving a *SSMP* (7). The problem we considered in this article is general for many of the problems considered in the literature concerning approximation of an unconstrained minimization problem, see for example [25–28,24,23]. Our result can also be applied to find a solution of the split system of inclusion problem, the MSSFP, and the split system of equilibrium problem. Furthermore, our result improves an inertial extrapolation-type algorithm proposed in [25] and also improves and accelerates algorithms in [9,10].

References

[1] Y. Censor, T. Elfving, N. Kopf, and T. Bortfeld, *The multiple-sets split feasibility problem and its applications for inverse problems*, Inverse Problems **21** (2005), no. 6, 2071–2084, DOI: https://doi.org/10.1088/0266-5611/21/6/017.

[2] Y. Censor and T. Elfving, *A multiprojection algorithm using Bregman projections in a product space*, Numer. Algorithms **8** (1994), no. 2, 221–239, DOI: https://doi.org/10.1007/BF02142692.

[3] Y. Dang, Y. Gao, and L. Li, *Inertial projection algorithms for convex feasibility problem*, J. Syst. Eng. Electron. **23** (2012), no. 5, 734–740, DOI: https://doi.org/10.1109/JSEE.2012.00090.

[4] G. López, V. Martín-Márquez, F. Wang, and H. K. Xu, *Solving the split feasibility problem without prior knowledge of matrix norms*, Inverse Problems **28** (2012), no. 8, 85–104, DOI: https://doi.org/10.1088/0266-5611/28/8/085004.

[5] B. Qu and N. Xiu, *A note on the CQ algorithm for the split feasibility problem*, Inverse Problems **21** (2005), no. 5, 1655–1665, DOI: https://doi.org/10.1088/0266-5611/21/5/009.

[6] A. Latif, J. Vahidi, and M. Eslamian, *Strong convergence for generalized multiple-set split feasibility problem*, Filomat **30** (2016), no. 2, 459–467, DOI: https://doi.org/10.2298/FIL1602459L.

[7] E. Masad and S. Reich, *A note on the multiple-set split convex feasibility problem in Hilbert space*, J. Nonlinear. Convex Anal. **8** (2007), no. 3, 367–371.

[8] H. K. Xu, *A variable Krasnosel'skii–Mann algorithm and the multiple-set split feasibility problem*, Inverse Problems **22** (2006), no. 6, 2021–2034, DOI: https://doi.org/10.1088/0266-5611/22/6/007.

[9] A. G. Gebrie and R. Wangkeeree, *Proximal method of solving split system of minimization problem*, J. Appl. Math. Comput. **63** (2020), 107–132, DOI: https://doi.org/10.1007/s12190-019-01310-w.

[10] A. G. Gebrie and R. Wangkeeree, *An iterative scheme for solving split system of minimization problems*, J. Comput. Anal. Appl. **28** (2020), no. 6, 968–980.

[11] A. G. Gebrie and R. Wangkeeree, *Parallel proximal method of solving split system of fixed point set constraint minimization problems*, Rev. R. Acad. Cienc. Exactas Fıs. Nat. Ser. A Mat. RACSAM **114** (2020), no. 1, DOI: https://doi.org/10.1007/s13398-019-00758-6.

[12] Y. C. Tang and L. W. Liu, *Several iterative algorithms for solving the split common fixed point problem of directed operators with applications*, Optimization **65** (2016), no. 1, 53–65, DOI: https://doi.org/10.1080/02331934.2014.984708.

[13] Y. Censor and A. Segal, *The split common fixed point problem for directed operators*, J. Convex Anal. **16** (2009), no. 2, 587–600.

[14] A. Abkar and E. Shahrosvand, *The split common fixed point problem of two infinite families of demicontractive mappings and the split common null point problem*, Filomat **31** (2017), no. 12, 3859–3874, DOI: https://doi.org/10.2298/FIL1712859A.

[15] M. M. Alves and B. F. Svaiter, *A proximal-Newton method for unconstrained convex optimization in Hilbert spaces*, Optimization **67** (2018), no. 1, 67–82, DOI: https://doi.org/10.1080/02331934.2017.1389942.

[16] P. E. Gill and W. Murray, *Newton-type methods for unconstrained and linearly constrained optimization*, Math. Program. **7** (1974), no. 1, 311–350, DOI: https://doi.org/10.1007/BF01585529.

[17] R. T. Rockafellar and R. J. Wets, *Variational Analysis*, Springer-Verlag, Berlin, Heidelberg, 2009.

[18] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, R.I.R.O. **4** (1970), no. R3, 154–158, DOI: https://doi.org/10.1051/m2an/197004R301541.

[19] B. Martinet, *Détermination approchée d'un point fixe d'une application pseudo-contractante*, CR Acad. Sci. Paris. **274** (1972), no. 2, 163–165.

[20] R. T. Rockafellar, *Monotone operators and the proximal point algorithm*, SIAM J. Control Optim. **14** (1976), no. 5, 877–898, DOI: https://doi.org/10.1137/0314056.

[21] A. Moudafi and B. S. Thakur, *Solving proximal split feasibility problems without prior knowledge of operator norms*, Optim. Lett. **8** (2014), no. 7, 2099–2110, DOI: https://doi.org/10.1007/s11590-013-0708-4.

[22] J. M. Hendrickx and A. Olshevsky, *Matrix p-norms are NP-hard to approximate if $p \neq 1,2,\infty$*, SIAM J. Matrix Anal. Appl. **31** (2010), no. 5, 2802–2812.

[23] M. Abbas, M. AlShahrani, Q. H. Ansari, O. S. Iyiola, and Y. Shehu, *Iterative methods for solving proximal split minimization problems*, Numer. Algorithms **78** (2018), no. 1, 193–215, DOI: https://doi.org/10.1007/s11075-017-0372-3.

[24] Y. Shehu, G. Cai, and O. S. Iyiola, *Iterative approximation of solutions for proximal split feasibility problems*. Fixed Point Theory Appl. **2015** (2015), art. 123, DOI: https://doi.org/10.1186/s13663-015-0375-5.

[25] Y. Shehu and O. S. Iyiola, *Convergence analysis for the proximal split feasibility problem using an inertial extrapolation term method*, J. Fixed Point Theory Appl. **19** (2017), no. 4, 2483–2510, DOI: https://doi.org/10.1007/s11784-017-0435-z.

[26] Y. Shehu and O. S. Iyiola, *Strong convergence result for proximal split feasibility problem in Hilbert spaces*, Optimization **66** (2017), no. 12, 2275–2290, DOI: https://doi.org/10.1080/02331934.2017.1370648.

[27] Y. Shehu and O. S. Iyiola, *Accelerated hybrid viscosity and steepest-descent method for proximal split feasibility problems*, Optimization **67** (2018), no. 4, 475–492, DOI: https://doi.org/10.1080/02331934.2017.1405955.

[28] Y. Shehu and O. S. Iyiola, *Nonlinear iteration method for proximal split feasibility problems*, Math. Meth. Appl. Sci. **41** (2018), no. 2, 781–802.

[29] Y. Shehu and F. U. Ogbuisi, *Convergence analysis for proximal split feasibility problems and fixed point problems*, J. Appl. Math. Comp. **48** (2015), no. 1–2, 221–239, DOI: https://doi.org/10.1007/s12190-014-0800-7.

[30] B. T. Polyak, *Some methods of speeding up the convergence of iteration methods*, USSR Comput. Math. Math. Phys. **4** (1964), no. 5, 1–17, DOI: https://doi.org/10.1016/0041-5553(64)90137-5.

[31] H. Attouch, J. Peypouquet, and P. Redont, *A dynamical approach to an inertial forward-backward algorithm for convex minimization*, SSIAM J. Optim. **24** (2014), no. 1, 232–256, DOI: https://doi.org/10.1137/130910294.

[32] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci. **2** (2009), no. 1, 183–202, DOI: https://doi.org/10.1137/080716542.

[33] C. Chen, R. H. Chan, S. Ma, and J. Yang, *Inertial proximal ADMM for linearly constrained separable convex optimization*, SIAM J. Imaging Sci. **8** (2015), no. 4, 2239–2267, DOI: https://doi.org/10.1137/15100463X.

[34] P. Ochs, T. Brox, and T. Pock, *iPiasco: Inertial proximal algorithm for strongly convex optimization*, J. Math. Imaging Vision. **53** (2015), no. 2, 171–181, DOI: https://doi.org/10.1007/s10851-015-0565-0.

[35] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer International, New York, 2011.

[36] H. Brezis, *Operateurs Maximaux Monotones: Et Semi-Groupes De Contractions Dans Les Espaces De Hilbert*, North-Holland Mathematics Studies, vol. 5, American Elsevier Publishing Company, Inc., New York, 1973.

[37] R. T. Rockafellar, *On the maximality of sums of nonlinear monotone operators*, Trans. Amer. Math. Soc. **149** (1970), no. 1, 75–88, DOI: https://doi.org/10.2307/1995660.

[38] B. Lemaire, *Which fixed point does the iteration method select?*, in: P. Gritzmann, R. Horst, E. Sachs, R. Tichatschke, (eds), Recent Advances in Optimization, Lecture Notes in Economics and Mathematical Systems, vol. 452, Springer, Berlin, Heidelberg, pp. 154–167, DOI: https://doi.org/10.1007/978-3-642-59073-3_11

[39] H. K. Xu, *Iterative algorithms for nonlinear operators*, Lond. Math. Soc. **66** (2002), no. 1, 240–256.

[40] P. E. Maingé, *Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization*, Set-Valued Anal. **16** (2008), no. 7–8, 899–912, DOI: https://doi.org/10.1007/s11228-008-0102-z.

[41] E. Blum and W. Oettli, *From optimization and variational inequalities to equilibrium problems*, Math. Student **63** (1994), 123–145.

[42] P. L. Combettes and S. A. Hirstoaga, *Equilibrium programming in Hilbert spaces*, J. Nonlinear Convex Anal. **6** (2005), no. 1, 117–136.