Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 9(8): 22366–22392.
DOI:10.3934/math.20241088
Received: 21 April 2024
Revised: 01 July 2024
Accepted: 09 July 2024
Published: 18 July 2024

Research article

A new accelerated algorithm with a linesearch technique for convex bilevel optimization problems with applications

Adisak Hanjing[1], **Panadda Thongpaen**[2] **and Suthep Suantai**[3,*]

[1] Department of Science and Mathematics, Rajamangala University of Technology Isan Surin Campus, Surin 32000, Thailand

[2] Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

[3] Research Center in Optimization and Computational Intelligence for Big Data Prediction, Department of Mathematics, Chiang Mai University, Chiang Mai 50200, Thailand

* **Correspondence:** Email: suthep.s@cmu.ac.th.

Abstract: We considered a convex bilevel optimization problem when the outer level problem was to find a minimizer of a strongly convex function over the set of solutions of the inner level problem which was in the form of minimization of the sum of a convex differentiable function and a nonsmooth convex function. In this work, we proposed a novel accelerated algorithm by employing both linesearch and inertial techniques for solving a convex bilevel optimization problem. Then, we proved the strong convergence of the sequence generated by our proposed algorithm to an optimal solution of the convex bilevel optimization problems without the continuity assumption on the gradient of the objective function. Moreover, we presented the convergence behavior of the proposed method by some numerical experiments addressing image restoration problems and data classification problems with least squares constraints. Finally, the performances of the restorative image and the data classification of the proposed method were compared with other existing algorithms in the literature. According to the experiment, our proposed algorithm had a better convergence behavior than the others in the literature.

Keywords: accelerated algorithm; convex bilevel optimization problem; inertial step; linesearch technique; data classification; image restoration
Mathematics Subject Classification: 65K05, 90C25, 90C30

1. Introduction

Convex bilevel optimization problems play an important role in many real-word applications such as image-signal processing, data science, data prediction, data classification, and artificial intelligence. For some interesting applications, we refer to the recent papers [1, 2]. More precisely, we recall the concept of the convex bilevel optimization problem as the following. Let ψ and ϕ be two proper convex and lower semi-continuous functions from a real Hilbert space H into $\mathbb{R} \cup \{+\infty\}$, and ϕ is a smooth function. In this work, we consider the following convex bilevel optimization problem:

$$\min_{z \in S^\star} h(z), \tag{1.1}$$

where h is a strongly convex differentiable function of the form H into $\mathbb{R}$ with parameter s, and $S^\star$ is the solution set of the problem:

$$\min_{z \in H} \{\phi(z) + \psi(z)\}. \tag{1.2}$$

Problems (1.1) and (1.2) are known as outer-level and inner-level problems, respectively. It is well-known that if $z^\star$ satisfies the variational inequality:

$$\langle \nabla h(z^\star), z - z^\star \rangle \geq 0, \quad \forall z \in S^\star,$$

then $z^\star$ is a solution of the outer-level problem (1.1); for more details, see [3]. Generally, the solution of problem (1.2) usually exists under the assumption that $\nabla\phi$ is Lipschitz continuous with parameter L_ϕ, that is, there exists $L_\phi > 0$ such that $\|\nabla\phi(w) - \nabla\phi(v)\| \leq L_\phi \|w - v\|$ for all $w, v \in H$.

The proximity operator, $\text{prox}_{\mu\psi}(z^\star) = J_\mu^{\partial\psi}(z^\star) = (I + \mu\partial\psi)^{-1}(z^\star)$, where I is an identity mapping and $\partial\psi$ is a subdifferential of ψ, is crucial in solving problem (1.2). It is known that a point $z^\star$ in $S^\star$ is a fixed point of proximity operator $\text{prox}_{\mu\psi}(I - \mu\nabla\phi)$. The following classical forward-backward splitting algorithm:

$$x_{k+1} = \text{prox}_{\mu_k\psi}(x_k - \mu_k \nabla\phi(x_k)) \tag{1.3}$$

was proposed for solving problem (1.2). After that, Sabach and Shtern [4] introduced the bilevel gradient sequential averaging method (BiG-SAM), as seen in Algorithm 2. They also proved that sequence $\{x_k\}$ generated by BiG-SAM converges strongly to the optimal point $z^\star$ in the convex bilevel optimization problem (1.1) and (1.2). Later, to speed up the rate of convergence of BiG-SAM, Shehu et al. [5] employed an inertial technique proposed by Polyak [6], as defined by Algorithm 3 (iBiGSAM). Moreover, they proved a strong convergence theorem of Algorithm 3 under some weaker assumptions on $\{\lambda_k\}$ given in [7], that is, $\lim_{k\to\infty} \lambda_k = 0$ and $\sum_{k=1}^{\infty} \lambda_k = +\infty$. Moreover, the convergence rate of the iBiG-SAM was consecutively improved by adapting the inertial technique, which is called the alternated inertial bilevel gradient sequential averaging method [8] (aiBiG-SAM), as seen in Algorithm 4. It was shown by some examples in [8] that the convergence behavior of aiBiG-SAM is better than BiG-SAM and iBiG-SAM. Recently, Jolaoso et al. [9] proposed a double inertial technique to accelerate the convergence rate of the strongly convergent 2-step inertial PPA algorithm solving for a zero of the sum of two maximal monotone operators. Yao et al. [10] also introduced a method for solving such a problem, called the weakly convergent FRB algorithm with momentum. This problem is just the inner-level problem in this work.

It is worth noting that all methods mentioned above desire a Lipschitz continuity assumption of $\nabla\phi$. However, finding a Lipschitz constant of $\nabla\phi$ is sometimes too difficult. To solve the inner-level problem without computing the Lipschitz constant of gradient $\nabla\phi$, Cruz and Nghia [11] presented a linesearch technique (Linesearch 1) for finding some suitable step size for a forward-backward splitting algorithm. This notion provides weaker assumptions on the gradient of ϕ, as seen in the following criteria:

A1. $\phi, \psi : H \to (-\infty, +\infty]$ are proper lower semicontinuous convex functions with $\operatorname{dom}\psi \subseteq \operatorname{dom}\phi$;

A2. ϕ is differentiable on an open set containing $\operatorname{dom}\psi$, and $\nabla\phi$ is uniformly continuous on any bounded subset of $\operatorname{dom}\psi$ and maps any bounded subset of $\operatorname{dom}\psi$ to a bounded set in H.

It is observed that assumption A2 is weaker than the Lipschitz continuity assumption on $\nabla\phi$. Under assumptions A1 and A2, they proved that the sequence $\{x_k\}$ generated by (1.3), where μ_k is derived from Linesearch 1 (see more detail in the appendix), converges weakly to the optimal solution of the inner level problem (1.2). Inspired by [11], several algorithms with the linesearch technique were proposed in order to solve problem (1.2); see [12–17], for examples. Recently, Hanjing et al. [17] introduced a new linesearch technique (Linesearch 2) and a new algorithm (Algorithm 6), called the forward-backward iterative method with the inertial technical term and linesearch technique, to solve the inner-level problem (1.2). For more details on Linesearch 2 and Algorithm 6, see the appendix. They proved that the sequence $\{x_k\}$ generated by Algorithm 7 converges weakly to a solution of problem (1.2) under some control conditions.

Note that Algorithm 7 was employed to find a solution of the inner-level problem (1.2) and it provided only weak convergence, but the strong convergence is more desired. Inspired by all of the mentioned works, we aim to develop Algorithm 7 for solving the convex bilevel problems (1.1) and (1.2) by employing Linesearch 2 together with the viscosity approximation methods. The strong convergence theorem of our developed algorithm is established under some suitable conditions and assumptions. Furthermore, we apply our proposed algorithm to solve image restoration and data classification problems including comparison of its performance with other algorithms.

2. Notations and instruments for convergent analysis

In this section, we provide some important definitions, propositions, and lemmas which will be used in the next section. Let H be a real Hilbert space and X be a nonempty closed convex subset of H. Then, for each $w \in H$, there exists a unique element $P_X w$ in X satisfying

$$\|w - P_X w\| \le \|w - z\|, \quad \forall z \in X.$$

The mapping P_X is known as the metric projection of H onto X. Moreover,

$$\langle w - P_X w, z - P_X w\rangle \le 0 \tag{2.1}$$

holds for all $w \in H$ and $z \in X$. A mapping $f : X \to H$ is called Lipschitz continuous if there exists $L_f > 0$ such that

$$\|f(v) - f(w)\| \le L_f\|v - w\|, \quad \forall v, w \in X.$$

If $L_f \in [0, 1)$, then f is called a contraction. Moreover, f is nonexpansive if $L_f = 1$. The domain of function $f : H \to [-\infty, +\infty]$ is denoted by $\operatorname{dom} f$, when $\operatorname{dom} f := \{v \in H : f(v) < \infty\}$. Let $\{x_k\}$ be a sequence in H, and we adopt the following notations:

1) $x_k \rightharpoonup w$ denotes that sequence $\{x_k\}$ converges weakly to $w \in H$;
2) $x_k \to w$ denotes that $\{x_k\}$ converges strongly to $w \in H$.

For each $v, w \in H$, the following conditions hold:

1) $\|v \pm w\|^2 = \|v\|^2 \pm 2\langle v, w\rangle + \|w\|^2$;
2) $\|v + w\|^2 \leq \|v\|^2 + 2\langle w, v + w\rangle$;
3) $\|tv + (1-t)w\|^2 = t\|v\|^2 + (1-t)\|w\|^2 - t(1-t)\|v - w\|^2, \quad \forall t \in \mathbb{R}$.

Let $\psi : H \to (-\infty, +\infty]$ be a proper function. The subdifferential $\partial\psi$ of ψ is defined by

$$\partial\psi(u) := \{v \in H : \langle v, w - u\rangle + \psi(u) \leq \psi(w),\ \forall w \in H\}, \quad \forall u \in H.$$

If $\partial\psi(u) \neq \emptyset$, then ψ is subdifferentiable at u, and the elements of $\partial\psi(u)$ are the subgradients of ψ at u. The proximal operator, $\text{prox}_\psi : H \to \text{dom}\,\psi$ with $\text{prox}_\psi(x) := (I + \partial\psi)^{-1}$, is single-valued with a full domain. Moreover, we have from [18] that for each $x \in H$ and $\mu > 0$,

$$\frac{x - \text{prox}_{\mu\psi}(x)}{\mu} \in \partial\psi(\text{prox}_{\mu\psi}(x)). \tag{2.2}$$

Let us next revisit some important properties for this work.

Lemma 1 ([19]). *Let $\partial\psi$ be a subdifferential of ψ. Then, the following hold:*

1) $\partial\psi$ is maximal monotone,
2) $Gph(\partial\psi) := \{(v, w) \in H \times H : w \in \partial\psi(v)\}$ is demiclosed, i.e., if $\{(v_k, w_k)\}$ is a sequence in $Gph(\partial\psi)$ such that $v_k \rightharpoonup v$ and $w_k \to w$, then $(v, w) \in Gph(\partial\psi)$.

Using the same idea of [4, Proposition 3], the following result can be proven.

Proposition 2. *Suppose $h : H \to \mathbb{R}$ is strongly convex with parameter $s > 0$ and continuously differentiable such that ∇h is Lipschitz continuous with constant L_h. Then, the mapping $I - t\nabla h$ is c-contraction for all $0 < t \leq \frac{2}{L_h + s}$, where $c = \sqrt{1 - \frac{2stL_h}{s + L_h}}$ and I is the identity operator.*

Proof: For any $x, y \in H$, we obtain

$$\begin{aligned}\left\|(x - t\nabla h(x)) - (y - t\nabla h(y))\right\|^2 = \|x - y\|^2 &- 2t\,\langle \nabla h(x) - \nabla h(y), x - y\rangle \\ &+ t^2\|\nabla h(x) - \nabla h(y)\|^2.\end{aligned} \tag{2.3}$$

Using the same proof as in the case of $H = \mathbb{R}^n$ on [20, Theorem 2.1.12], we get

$$\langle \nabla h(x) - \nabla h(y), x - y\rangle \geq \frac{sL_h}{s + L_h}\|x - y\|^2 + \frac{1}{s + L_h}\|\nabla h(x) - \nabla h(y)\|^2. \tag{2.4}$$

From (2.3) and (2.4), we get

$$\begin{aligned}\left\|(x - t\nabla h(x)) - (y - t\nabla h(y))\right\|^2 &\leq \left(1 - \frac{2stL_h}{s + L_h}\right)\|x - y\|^2 + t\left(t - \frac{2}{s + L_h}\right)\|\nabla h(x) - \nabla h(y)\|^2 \\ &\leq \sqrt{1 - \frac{2stL_h}{s + L_h}}\|x - y\|.\end{aligned}$$

Lemma 3 ([21]). *Let $\{a_k\}$ be a sequence of nonnegative real numbers satisfying*

$$a_{k+1} \leq (1 - b_k)a_k + b_k s_k, \quad \forall k \in \mathbb{N},$$

where $\{b_k\}$ is a sequence in $(0, 1)$ such that $\sum_{k=1}^{\infty} b_k = +\infty$ and $\{s_k\}$ is a sequence satisfying $\limsup_{k\to\infty} s_k \leq 0$. *Then,* $\lim\limits_{k\to\infty} a_k = 0$.

Lemma 4 ([22]). *Let $\{u_k\}$ be a sequence of real numbers such that there exists a subsequence $\{u_{m_j}\}$ of $\{u_k\}$ such that $u_{m_j} < u_{m_j+1}$ for all $j \in \mathbb{N}$. Then there exists a nondecreasing sequence $\{k_\ell\}$ of $\mathbb{N}$ such that* $\lim_{\ell\to\infty} k_\ell = \infty$ *and for all sufficiently large $\ell \in \mathbb{N}$, the following holds:*

$$u_{k_\ell} \leq u_{k_\ell+1} \quad \textit{and} \quad u_\ell \leq u_{k_\ell+1}.$$

3. Main convergence theorems

We begin this section by introducing a new accelerated algorithm (Algorithm 1) by using a linesearch technique together with some modifications of Algorithm 6 for solving bilevel convex minimization problems (1.1) and (1.2). Throughout this section, we let Ω be the set of all solutions of convex bilevel problems (1.1) and (1.2), and we assume that $h : H \to \mathbb{R}$ is a strongly convex differentiable function with parameter s such that ∇h is L_h-Lipschitz continuous and $t \in \left(0, \frac{2}{L_h+s}\right]$. Suppose $f : \operatorname{dom}\psi \to \operatorname{dom}\psi$ is a c-contraction for some $c \in (0, 1)$. Let $\{\gamma_k\}$ be a real positive sequence, $\{\xi_k\}$ a positive sequence, and $\{\lambda_k\}$ be a sequence in $(0, 1)$. We propose the following Algorithm 1:

Algorithm 1 Accelerated viscosity forward-backward algorithm with Linesearch 1.

```
We are given $x_1 = y_0 \in \operatorname{dom}\psi$, $\sigma > 0$, $\theta \in (0,1)$, $\rho \in (0, \frac{1}{2}]$, and $\delta \in \left(0, \frac{\rho}{4}\right)$.
For each $k \geq 1$, define $\mu_k :=$ Linesearch 2 $(u_k, \sigma, \theta, \delta)$ and evaluate
    $u_k = \lambda_k f(x_k) + (1 - \lambda_k)x_k,$
    $v_k = \operatorname{prox}_{\mu_k\psi}(u_k - \mu_k \nabla\phi(u_k)),$
    $y_k = \operatorname{prox}_{\mu_k\psi}(v_k - \mu_k \nabla\phi(v_k)).$
Select $\eta_k \in (0, \bar{\eta}_k]$ such that
    $\bar{\eta}_k = \begin{cases} \min\left\{\gamma_k, \frac{\xi_k}{\|y_k - y_{k-1}\|}\right\} & \text{if } y_k \neq y_{k-1}, \\ \gamma_k, & \text{otherwise.} \end{cases}$    (3.1)
Compute
    $x_{k+1} = P_{\operatorname{dom}\psi}(y_k + \eta_k(y_k - y_{k-1})).$
```

Remark 1. *Our proposed algorithm uses a linesearch technique for finding the step size of the proximal gradient methods in order to relax the continuity assumption on the gradeint of f. Note that this linesearch technique employes two proximal evaluations which is appropriated from the algorithms consisting of two proximal evaluations at each iteration, see [12–17]. It is observed that those algorithms have a better convergence behavior than the others.*

To prove the convergence results of Algorithm 1, we need to find the following results.

Lemma 5. *Let $\{x_k\}$ be a sequence generated by Algorithm 1, and $v \in \operatorname{dom}\psi$. Then, the following inequality holds:*

$$\|u_k - v\|^2 - \|y_k - v\|^2 \geq 2\mu_k\Big((\phi+\psi)(y_k) + (\phi+\psi)(v_k) - 2(\phi+\psi)(v)\Big) + \left(1 - \frac{4\delta}{\rho}\right)\Big(\|y_k - v_k\|^2 + \|v_k - u_k\|^2\Big).$$

Proof: Let $v \in \operatorname{dom}\psi$. It follows from (2.2) that

$$\frac{u_k - v_k}{\mu_k} - \nabla\phi(u_k) = \frac{u_k - \operatorname{prox}_{\mu_k\psi}(u_k - \mu_k\nabla\phi(u_k))}{\mu_k} - \nabla\phi(u_k) \in \partial\psi(v_k),$$

and

$$\frac{v_k - y_k}{\mu_k} - \nabla\phi(v_k) = \frac{v_k - \operatorname{prox}_{\mu_k\psi}(v_k - \mu_k\nabla\phi(v_k))}{\mu_k} - \nabla\phi(v_k) \in \partial\psi(y_k).$$

Then, by the definition of $\partial\psi$, we get

$$\psi(v) - \psi(v_k) \geq \left\langle \frac{u_k - v_k}{\mu_k} - \nabla\phi(u_k), v - v_k \right\rangle = \frac{1}{\mu_k}\langle u_k - v_k, v - v_k\rangle + \langle \nabla\phi(u_k), v_k - v\rangle, \tag{3.2}$$

and

$$\psi(v) - \psi(y_k) \geq \left\langle \frac{v_k - y_k}{\mu_k} - \nabla\phi(v_k), v - y_k \right\rangle = \frac{1}{\mu_k}\langle v_k - y_k, v - y_k\rangle + \langle \nabla\phi(v_k), y_k - v\rangle. \tag{3.3}$$

By the convexity of ϕ, we have for every $x \in \operatorname{dom}\phi$ and $y \in \operatorname{dom}\psi$,

$$\phi(x) - \phi(y) \geq \langle \nabla\phi(y), x - y\rangle, \tag{3.4}$$

which implies

$$\phi(v) - \phi(u_k) \geq \langle \nabla\phi(u_k), v - u_k\rangle, \tag{3.5}$$

and

$$\phi(v) - \phi(v_k) \geq \langle \nabla\phi(v_k), v - v_k\rangle. \tag{3.6}$$

From (3.2), (3.3), (3.5), and (3.6), we have

$$\begin{aligned}
&2(\phi+\psi)(v) - (\phi+\psi)(v_k) - \phi(u_k) - \psi(y_k) \\
&\quad\geq \frac{1}{\mu_k}\langle u_k - v_k, v - v_k\rangle + \langle \nabla\phi(u_k), v_k - v\rangle + \langle \nabla\phi(u_k), v - u_k\rangle \\
&\qquad + \langle \nabla\phi(v_k), v - v_k\rangle + \frac{1}{\mu_k}\langle v_k - y_k, v - y_k\rangle + \langle \nabla\phi(v_k), y_k - v\rangle \\
&\quad= \frac{1}{\mu_k}\Big(\langle u_k - v_k, v - v_k\rangle + \langle v_k - y_k, v - y_k\rangle\Big) + \langle \nabla\phi(u_k), v_k - u_k\rangle + \langle \nabla\phi(v_k), y_k - v_k\rangle
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\mu_k}\Big(\langle u_k - v_k, v - v_k\rangle + \langle v_k - y_k, v - y_k\rangle\Big)\\
&\quad + \langle \nabla\phi(u_k) - \nabla\phi(v_k), v_k - u_k\rangle + \langle \nabla\phi(v_k), v_k - u_k\rangle\\
&\quad + \langle \nabla\phi(v_k) - \nabla\phi(y_k), y_k - v_k\rangle + \langle \nabla\phi(y_k), y_k - v_k\rangle\\
&\geq \frac{1}{\mu_k}\Big(\langle u_k - v_k, v - v_k\rangle + \langle v_k - y_k, v - y_k\rangle\Big)\\
&\quad - \|\nabla\phi(u_k) - \nabla\phi(v_k)\|\|v_k - u_k\| + \langle \nabla\phi(v_k), v_k - u_k\rangle\\
&\quad - \|\nabla\phi(v_k) - \nabla\phi(y_k)\|\|y_k - v_k\| + \langle \nabla\phi(y_k), y_k - v_k\rangle.
\end{aligned}
$$

This together with (3.4) gives

$$
\begin{aligned}
&2(\phi+\psi)(v) - (\phi+\psi)(v_k) - \phi(u_k) - \psi(y_k)\\
&\quad\geq \frac{1}{\mu_k}\Big(\langle u_k - v_k, v - v_k\rangle + \langle v_k - y_k, v - y_k\rangle\Big) - \|\nabla\phi(u_k) - \nabla\phi(v_k)\|\|v_k - u_k\|\\
&\qquad - \|\nabla\phi(v_k) - \nabla\phi(y_k)\|\|y_k - v_k\| + \phi(v_k) - \phi(u_k) + \phi(y_k) - \phi(v_k)\\
&\quad= \frac{1}{\mu_k}\Big(\langle u_k - v_k, v - v_k\rangle + \langle v_k - y_k, v - y_k\rangle\Big) - \|\nabla\phi(u_k) - \nabla\phi(v_k)\|\|v_k - u_k\|\\
&\qquad - \|\nabla\phi(v_k) - \nabla\phi(y_k)\|\|y_k - v_k\| - \phi(u_k) + \phi(y_k)\\
&\quad\geq \frac{1}{\mu_k}\Big(\langle u_k - v_k, v - v_k\rangle + \langle v_k - y_k, v - y_k\rangle\Big) - \phi(u_k) + \phi(y_k)\\
&\qquad - \|\nabla\phi(u_k) - \nabla\phi(v_k)\|(\|y_k - v_k\| + \|v_k - u_k\|)\\
&\qquad - \|\nabla\phi(v_k) - \nabla\phi(y_k)\|(\|y_k - v_k\| + \|v_k - u_k\|)\\
&\quad= \frac{1}{\mu_k}\Big(\langle u_k - v_k, v - v_k\rangle + \langle v_k - y_k, v - y_k\rangle\Big) - \phi(u_k) + \phi(y_k)\\
&\qquad - (\|y_k - v_k\| + \|v_k - u_k\|)\Big(\|\nabla\phi(u_k) - \nabla\phi(v_k)\| + \|\nabla\phi(v_k) - \nabla\phi(y_k)\|\Big).
\end{aligned}
\tag{3.7}
$$

By the definition of Linesearch 1, we get

$$
\mu_k\Big((1-\rho)\|\nabla\phi(y_k) - \nabla\phi(v_k)\| + \rho\|\nabla\phi(v_k) - \nabla\phi(u_k)\|\Big) \leq \delta(\|y_k - v_k\| + \|v_k - u_k\|). \tag{3.8}
$$

From (3.7) and (3.8), we have

$$
\begin{aligned}
&\frac{1}{\mu_k}\Big(\langle u_k - v_k, v_k - v\rangle + \langle v_k - y_k, y_k - v\rangle\Big)\\
&\qquad\geq (\phi+\psi)(v_k) + \phi(u_k) + \psi(y_k) - 2(\phi+\psi)(v) - \phi(u_k) + \phi(y_k)\\
&\qquad\quad - (\|y_k - v_k\| + \|v_k - u_k\|)\Big(\|\nabla\phi(u_k) - \nabla\phi(v_k)\| + \|\nabla\phi(v_k) - \nabla\phi(y_k)\|\Big)\\
&\qquad\geq (\phi+\psi)(v_k) + (\phi+\psi)(y_k) - 2(\phi+\psi)(v)\\
&\qquad\quad - (\|y_k - v_k\| + \|v_k - u_k\|)\left(\Big(\frac{1}{\rho} - 1\Big)\|\nabla\phi(u_k) - \nabla\phi(v_k)\|\right) - (\|y_k - v_k\| + \|v_k - u_k\|)\|\nabla\phi(v_k) - \nabla\phi(y_k)\|\\
&\qquad= (\phi+\psi)(v_k) + (\phi+\psi)(y_k) - 2(\phi+\psi)(v)\\
&\qquad\quad - \frac{1}{\rho}(\|y_k - v_k\| + \|v_k - u_k\|)\Big((1-\rho)\|\nabla\phi(u_k) - \nabla\phi(v_k)\|\Big)
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{1}{\rho}(\|y_k - v_k\| + \|v_k - u_k\|)\Big(\rho\|\nabla\phi(v_k) - \nabla\phi(y_k)\|\Big)\\
&\geq (\phi+\psi)(v_k) + (\phi+\psi)(y_k) - 2(\phi+\psi)(v)\\
&\quad -\frac{1}{\rho}(\|y_k - v_k\| + \|v_k - u_k\|)\left(\frac{\delta}{\mu_k}(\|y_k - v_k\| + \|v_k - u_k\|)\right)\\
&= (\phi+\psi)(v_k) + (\phi+\psi)(y_k) - 2(\phi+\psi)(v) - \frac{\delta}{\rho\mu_k}(\|y_k - v_k\| + \|v_k - u_k\|)^2\\
&\geq (\phi+\psi)(v_k) + (\phi+\psi)(y_k) - 2(\phi+\psi)(v) - \frac{2\delta}{\rho\mu_k}\left(\|y_k - v_k\|^2 + \|v_k - u_k\|^2\right). \qquad (3.9)
\end{aligned}
$$

Moreover, we know that

$$
\langle u_k - v_k, v_k - v\rangle = \frac{1}{2}(\|u_k - v\|^2 - \|u_k - v_k\|^2 - \|v_k - v\|^2), \qquad (3.10)
$$

and

$$
\langle v_k - y_k, y_k - v\rangle = \frac{1}{2}(\|v_k - v\|^2 - \|v_k - y_k\|^2 - \|y_k - v\|^2). \qquad (3.11)
$$

By replacing (3.10) and (3.11) in (3.9), we obtain

$$
\begin{aligned}
\|u_k - v\|^2 - \|y_k - v\|^2 &\geq 2\mu_k\Big((\phi+\psi)(y_k) + (\phi+\psi)(v_k) - 2(\phi+\psi)(v)\Big)\\
&\quad - \frac{4\delta}{\rho}\left(\|y_k - v_k\|^2 + \|v_k - u_k\|^2\right) + \|u_k - v_k\|^2 + \|v_k - y_k\|^2\\
&= 2\mu_k\Big((\phi+\psi)(y_k) + (\phi+\psi)(v_k) - 2(\phi+\psi)(v)\Big)\\
&\quad + \left(1 - \frac{4\delta}{\rho}\right)\left(\|y_k - v_k\|^2 + \|v_k - u_k\|^2\right). \qquad (3.12)
\end{aligned}
$$

Lemma 6. *Let $\{x_k\}$ be a sequence generated by Algorithm 1 and $S^\star \neq \emptyset$. Suppose that $\lim_{k\to\infty}\frac{\xi_k}{\lambda_k} = 0$. Then $\{x_k\}$ is bounded. Furthermore, $\{f(x_k)\}$, $\{u_k\}$, $\{y_k\}$, and $\{v_k\}$ are also bounded.*

Proof: Let $v^\star \in S^\star$. By Lemma 5, we have

$$
\begin{aligned}
\|u_k - v^\star\|^2 - \|y_k - v^\star\|^2 &\geq 2\mu_k\Big((\phi+\psi)(y_k) + (\phi+\psi)(v_k) - 2(\phi+\psi)(v^\star)\Big)\\
&\quad + \left(1 - \frac{4\delta}{\rho}\right)\left(\|y_k - v_k\|^2 + \|v_k - u_k\|^2\right)\\
&\geq \left(1 - \frac{4\delta}{\rho}\right)\left(\|y_k - v_k\|^2 + \|v_k - u_k\|^2\right) \;\geq 0, \qquad (3.13)
\end{aligned}
$$

which implies

$$
\|u_k - v^\star\| \geq \|y_k - v^\star\|. \qquad (3.14)
$$

By the definition of u_k and since f is a contraction with constant c, we get

$$
\begin{aligned}
\|u_k - v^\star\| &= \left\|\lambda_k f(x_k) + (1-\lambda_k)x_k - v^\star\right\| \qquad (3.15)\\
&\leq \lambda_k\|f(x_k) - f(v^\star)\| + \lambda_k\|f(v^\star) - v^\star\| + (1-\lambda_k)\|x_k - v^\star\|
\end{aligned}
$$

$$\begin{aligned}
&\leq c\lambda_k\|x_k - v^\star\| + \lambda_k\|f(v^\star) - v^\star\| + (1-\lambda_k)\|x_k - v^\star\| \\
&= (1 - \lambda_k(1-c))\|x_k - v^\star\| + \lambda_k\|f(v^\star) - v^\star\|. \qquad (3.16)
\end{aligned}$$

This together with (3.14) gives

$$\begin{aligned}
\|x_{k+1} - v^\star\| &= \|\mathrm{P}_{\mathrm{dom}\,\psi}((y_k + \eta_k(y_k - y_{k-1})) - \mathrm{P}_{\mathrm{dom}\,\psi}(v^\star)\| \\
&\leq \|(y_k - v^\star) + \eta_k(y_k - y_{k-1})\| \qquad (3.17) \\
&\leq \|y_k - v^\star\| + \eta_k\|y_k - y_{k-1}\| \qquad (3.18) \\
&\leq \|u_k - v^\star\| + \eta_k\|y_k - y_{k-1}\| \qquad (3.19) \\
&\leq (1 - \lambda_k(1-c))\|x_k - v^\star\| + \lambda_k\|f(v^\star) - v^\star\| + \eta_k\|y_k - y_{k-1}\| \qquad (3.20) \\
&= (1 - \lambda_k(1-c))\|x_k - v^\star\| + \lambda_k(1-c)\left(\frac{\|f(v^\star) - v^\star\|}{1-c} + \frac{\eta_k}{\lambda_k(1-c)}\|y_k - y_{k-1}\|\right) \\
&\leq \max\left\{\|x_k - v^\star\|, \frac{\|f(v^\star) - v^\star\|}{1-c} + \frac{\eta_k}{\lambda_k(1-c)}\|y_k - y_{k-1}\|\right\}.
\end{aligned}$$

From (3.1), we have

$$\frac{\eta_k}{\lambda_k}\|y_k - y_{k-1}\| \leq \frac{\xi_k}{\|y_k - y_{k-1}\|} \cdot \frac{\|y_k - y_{k-1}\|}{\lambda_k} = \frac{\xi_k}{\lambda_k}.$$

Using $\lim\limits_{k\to\infty} \dfrac{\xi_k}{\lambda_k} = 0$, we obtain $\lim\limits_{k\to\infty} \dfrac{\eta_k}{\lambda_k}\|y_k - y_{k-1}\| = 0$. Therefore, there exists $N > 0$ such that $\frac{\eta_k}{\lambda_k}\|y_k - y_{k-1}\| \leq N$ for all $k \in \mathbb{N}$. The above inequality implies

$$\|x_{k+1} - v^\star\| \leq \max\left\{\|x_k - v^\star\|, \frac{\|f(v^\star) - v^\star\|}{1-c} + \frac{N}{1-c}\right\}.$$

By induction, we have $\|x_{k+1} - v^\star\| \leq \max\left\{\|x_1 - v^\star\|, \frac{\|f(v^\star)-v^\star\|}{1-c} + \frac{N}{1-c}\right\}$, and so $\{x_k\}$ is bounded. It follows that $\{f(x_k)\}$ is bounded. Combining this with the definition of u_k, we obtain that $\{u_k\}$ is bounded. It follows by (3.14) that $\{y_k\}$ and $\{v_k\}$ are also bounded.

Theorem 7. *Let $\{x_k\}$ be a sequence generated by Algorithm 1 and $S^\star \neq \emptyset$. Suppose ϕ and ψ satisfy A1 and A2 and the following conditions hold:*

1) $\{\lambda_k\}$ is a positive sequence in $(0, 1)$;
2) $\mu_k \geq \mu$ for some $\mu \in \mathbb{R}_+$;
3) $\lim_{k\to\infty} \lambda_k = 0$ and $\sum_{k=1}^\infty \lambda_k = +\infty$;
4) $\lim_{k\to\infty} \frac{\xi_k}{\lambda_k} = 0$.

Then, $x_k \to v^\star \in S^\star$ such that $v^\star = P_{S^\star} f(v^\star)$. Moreover, if $f := I - t\nabla h$, then $x_k \to v^\star \in \Omega$.
Proof: Let $v^\star \in S^\star$ be such that $v^\star = P_{S^\star} f(v^\star)$. By (3.17), Algorithm 1, and the fact that f is a contraction with constant c, we have

$$\begin{aligned}
\|x_{k+1} - v^\star\|^2 &\leq \|(y_k - v^\star) + \eta_k(y_k - y_{k-1})\|^2 \\
&\leq \|y_k - v^\star\|^2 + 2\eta_k\|y_k - v^\star\|\|y_k - y_{k-1}\| + \eta_k^2\|y_k - y_{k-1}\|^2 \\
&\leq \|u_k - v^\star\|^2 + 2\eta_k\|y_k - v^\star\|\|y_k - y_{k-1}\| + \eta_k^2\|y_k - y_{k-1}\|^2
\end{aligned}$$

$$
\begin{aligned}
&= \left\|\lambda_k f(x_k) + (1-\lambda_k)x_k - v^\star\right\|^2 + 2\eta_k\|y_k - v^\star\|\|y_k - y_{k-1}\| + \eta_k^2\|y_k - y_{k-1}\|^2 \\
&= \left\|\lambda_k(f(x_k) - f(v^\star)) + (1-\lambda_k)(x_k - v^\star) + \lambda_k(f(v^\star) - v^\star)\right\|^2 \\
&\quad + \eta_k\|y_k - y_{k-1}\|\big(2\|y_k - v^\star\| + \eta_k\|y_k - y_{k-1}\|\big) \\
&\le \left\|\lambda_k(f(x_k) - f(v^\star)) + (1-\lambda_k)(x_k - v^\star)\right\|^2 + 2\lambda_k\langle f(v^\star) - v^\star, u_k - v^\star\rangle \\
&\quad + \eta_k\|y_k - y_{k-1}\|\big(2\|y_k - v^\star\| + \eta_k\|y_k - y_{k-1}\|\big) \\
&= \lambda_k\|f(x_k) - f(v^\star)\|^2 + (1-\lambda_k)\|x_k - v^\star\|^2 + 2\lambda_k\langle f(v^\star) - v^\star, u_k - v^\star\rangle \\
&\quad - \lambda_k(1-\lambda_k)\|(f(x_k) - f(v^\star)) - (x_k - v^\star)\|^2 + \eta_k\|y_k - y_{k-1}\|\big(2\|y_k - v^\star\| + \eta_k\|y_k - y_{k-1}\|\big) \\
&\le \lambda_k\|f(x_k) - f(v^\star)\|^2 + (1-\lambda_k)\|x_k - v^\star\|^2 + 2\lambda_k\langle f(v^\star) - v^\star, u_k - v^\star\rangle \\
&\quad + \eta_k\|y_k - y_{k-1}\|\big(2\|y_k - v^\star\| + \eta_k\|y_k - y_{k-1}\|\big) \\
&\le c^2\lambda_k\|x_k - v^\star\|^2 + (1-\lambda_k)\|x_k - v^\star\|^2 + 2\lambda_k\langle f(v^\star) - v^\star, u_k - v^\star\rangle \\
&\quad + \eta_k\|y_k - y_{k-1}\|\big(2\|y_k - v^\star\| + \eta_k\|y_k - y_{k-1}\|\big) \\
&= (1 - \lambda_k(1-c^2))\|x_k - v^\star\|^2 + 2\lambda_k\langle f(v^\star) - v^\star, u_k - v^\star\rangle \\
&\quad + \eta_k\|y_k - y_{k-1}\|\big(2\|y_k - v^\star\| + \eta_k\|y_k - y_{k-1}\|\big) \\
&\le (1 - \lambda_k(1-c))\|x_k - v^\star\|^2 + 2\lambda_k\langle f(v^\star) - v^\star, u_k - v^\star\rangle \\
&\quad + \eta_k\|y_k - y_{k-1}\|\big(2\|y_k - v^\star\| + \eta_k\|y_k - y_{k-1}\|\big). \qquad (3.21)
\end{aligned}
$$

Since $\lim\limits_{k\to\infty}\eta_k\|y_k - y_{k-1}\| = \lim\limits_{k\to\infty}(\lambda_k)\dfrac{\xi_k}{\lambda_k} = 0$, there exists $N_1 > 0$ such that $\eta_k\|y_k - y_{k-1}\| \le N_1$ for all $k \in \mathbb{N}$. From Lemma 6, we have $\|y_k - v^\star\| \le N_2$ for some $N_2 > 0$. Choose $\bar{N} = \sup_{k\in\mathbb{N}}\{N_1, N_2\}$. By (3.21), we get

$$
\begin{aligned}
\|x_{k+1} - v^\star\|^2 &\le (1-\lambda_k(1-c))\|x_k - v^\star\|^2 + 2\lambda_k\langle f(v^\star) - v^\star, u_k - v^\star\rangle + 3\bar{N}\eta_k\|y_k - y_{k-1}\| \\
&= (1 - \lambda_k(1-c))\|x_k - v^\star\|^2 \\
&\quad + \lambda_k(1-c)\left(\frac{2}{1-c}\langle f(v^\star) - v^\star, u_k - v^\star\rangle + \frac{3\bar{N}\eta_k}{\lambda_k(1-c)}\|y_k - y_{k-1}\|\right). \qquad (3.22)
\end{aligned}
$$

In order to verify the convergence of $\{x_k\}$, we analyze the proof into the following two cases.

Case 1. Suppose there exists $M \in \mathbb{N}$ such that $\|x_{k+1} - v^\star\| \le \|x_k - v^\star\|$ for all $k \ge M$. This implies $\lim\limits_{k\to\infty}\|x_k - v^\star\|$ exists. From (3.22), we set $a_k = \|x_k - v^\star\|^2, b_k = \lambda_k(1-c)$, and $s_k = \frac{2}{1-c}\langle f(v^\star) - v^\star, u_k - v^\star\rangle + \frac{3\bar{N}\eta_k}{\lambda_k(1-c)}\|y_k - y_{k-1}\|$. It follows from $\sum\limits_{k=1}^{\infty}\lambda_k = +\infty$ that $\sum\limits_{k=1}^{\infty}b_k = (1-c)\sum\limits_{k=1}^{\infty}\lambda_k = +\infty$. In addition,

$$
\frac{3\bar{N}\eta_k}{\lambda_k(1-c)}\|y_k - y_{k-1}\| \le \frac{3\bar{N}}{1-c}\frac{\xi_k}{\|y_k - y_{k-1}\|}\cdot\frac{\|y_k - y_{k-1}\|}{\lambda_k} = \frac{3\bar{N}}{1-c}\left(\frac{\xi_k}{\lambda_k}\right).
$$

Then, by $\lim\limits_{k\to\infty}\dfrac{\xi_k}{\lambda_k} = 0$, we get $\lim\limits_{k\to\infty}\dfrac{3\bar{N}\eta_k}{\lambda_k(1-c)}\|y_k - y_{k-1}\| = 0$.

To employ Lemma 3, we need to guarantee that $\limsup_{k\to\infty} s_k \leq 0$. Since $\{u_k\}$ is bounded, there exists a subsequence $\{u_{k_j}\}$ of $\{u_k\}$ such that $u_{k_j} \rightharpoonup w$, for some $w \in H$, and

$$\limsup_{k\to\infty}\langle f(v^\star) - v^\star, u_k - v^\star\rangle = \lim_{j\to\infty}\langle f(v^\star) - v^\star, u_{k_j} - v^\star\rangle = \langle f(v^\star) - v^\star, w - v^\star\rangle.$$

Next, we show that $w \in S^\star$. We have from (3.19) and (3.20) that

$$\lim_{k\to\infty}\|x_k - v^\star\| = \lim_{k\to\infty}\|u_k - v^\star\|. \tag{3.23}$$

Combining this with (3.18) and (3.20), we obtain

$$\lim_{k\to\infty}\|x_k - v^\star\| = \lim_{k\to\infty}\|y_k - v^\star\|. \tag{3.24}$$

From (3.13), we have

$$\|u_k - v^\star\|^2 - \|y_k - v^\star\|^2 \geq \left(1 - \frac{4\delta}{\rho}\right)\left(\|y_k - v_k\|^2 + \|v_k - u_k\|^2\right) \geq \left(1 - \frac{4\delta}{\rho}\right)\|y_k - v_k\|^2 \quad \geq 0,$$

and

$$\|u_k - v^\star\|^2 - \|y_k - v^\star\|^2 \geq \left(1 - \frac{4\delta}{\rho}\right)\left(\|y_k - v_k\|^2 + \|v_k - u_k\|^2\right) \geq \left(1 - \frac{4\delta}{\rho}\right)\|v_k - u_k\|^2 \quad \geq 0.$$

From (3.23) and (3.24), we obtain

$$\lim_{k\to\infty}\|y_k - v_k\| = 0, \tag{3.25}$$

and

$$\lim_{k\to\infty}\|v_k - u_k\| = 0. \tag{3.26}$$

Moreover, we know that

$$\frac{u_{k_j} - v_{k_j}}{\mu_{k_j}} - \nabla\phi(u_{k_j}) + \nabla\phi(v_{k_j}) \in \partial\psi(v_{k_j}) + \nabla\phi(v_{k_j}) = \partial(\phi + \psi)(v_{k_j}).$$

The uniform convexity of $\nabla\phi$ and (3.26) yield

$$\lim_{k\to\infty}\|\nabla\phi(v_k) - \nabla\phi(u_k)\| = 0. \tag{3.27}$$

It implies, by assumption (2), that

$$\begin{aligned}\left\|\frac{u_{k_j} - v_{k_j}}{\mu_{k_j}} - \nabla\phi(u_{k_j}) + \nabla\phi(v_{k_j})\right\| &\leq \frac{1}{\mu_{k_j}}\|u_{k_j} - v_{k_j}\| + \|\nabla\phi(v_{k_j}) - \nabla\phi(u_{k_j})\| \\ &\leq \frac{1}{\mu}\|u_{k_j} - v_{k_j}\| + \|\nabla\phi(v_{k_j}) - \nabla\phi(u_{k_j})\|.\end{aligned}$$

This together with (3.26) and (3.27) yields

$$\left\|\frac{u_{k_j} - v_{k_j}}{\mu_{k_j}} - \nabla\phi(u_{k_j}) + \nabla\phi(v_{k_j})\right\| \to 0 \quad \text{as} \quad k \to \infty.$$

By the demiclosed nature of $\operatorname{Gph}(\partial(\phi + \psi))$, $0 \in \partial(\phi + \psi)(w)$, and so $w \in S^\star$. It derives from (2.1) that

$$\limsup_{k\to\infty}\langle f(v^\star) - v^\star, u_k - v^\star\rangle = \langle f(v^\star) - v^\star, w - v^\star\rangle = \langle f(v^\star) - P_{S^\star} f(v^\star), w - P_{S^\star} f(v^\star)\rangle \leq 0,$$

which implies that $\limsup_{k\to\infty} s_k \leq 0$. By Lemma 3, we obtain

$$\lim_{k\to\infty} \|x_k - v^\star\|^2 = 0.$$

Case 2. Suppose that there exists a subsequence $\{x_{m_j}\}$ of $\{x_k\}$ such that

$$\|x_{m_j} - v^\star\| < \|x_{m_j+1} - v^\star\|,$$

for all $j \in \mathbb{N}$. By Lemma 4, there is a nondecreasing sequence $\{k_\ell\}$ of $\mathbb{N}$ such that $\lim_{\ell\to\infty} k_\ell = \infty$, and for all sufficiently large $\ell \in \mathbb{N}$, the following formula holds:

$$\left\|x_{k_\ell} - v^\star\right\| \leq \left\|x_{k_\ell+1} - v^\star\right\| \quad \text{and} \quad \left\|x_\ell - v^\star\right\| \leq \left\|x_{k_\ell+1} - v^\star\right\|. \tag{3.28}$$

We have from (3.25) and (3.26) that

$$\lim_{\ell\to\infty} \left\|y_{k_\ell} - v_{k_\ell}\right\| = 0 \quad \text{and} \quad \lim_{\ell\to\infty} \left\|v_{k_\ell} - u_{k_\ell}\right\| = 0. \tag{3.29}$$

Since $\{u_{k_\ell}\}$ is bounded, there exists a weakly convergent subsequence $\{u_{k_{\ell_i}}\}$ of $\{u_{k_\ell}\}$ such that $u_{k_{\ell_i}} \rightharpoonup w^\star$ for some $w^\star \in H$, and

$$\limsup_{\ell\to\infty}\langle f(v^\star) - v^\star, u_{k_\ell} - v^\star\rangle = \lim_{i\to\infty}\langle f(v^\star) - v^\star, u_{k_{\ell_i}} - v^\star\rangle = \langle f(v^\star) - v^\star, w^\star - v^\star\rangle.$$

The uniform convexity of $\nabla\phi$ and (3.29) imply

$$\lim_{i\to\infty} \|\nabla\phi(v_{k_{\ell_i}}) - \nabla\phi(u_{k_{\ell_i}})\| = 0. \tag{3.30}$$

Moreover, we know that

$$\frac{u_{k_{\ell_i}} - v_{k_{\ell_i}}}{\mu_{k_{\ell_i}}} - \nabla\phi(u_{k_{\ell_i}}) + \nabla\phi(v_{k_{\ell_i}}) \in \partial\psi(v_{k_{\ell_i}}) + \nabla\phi(v_{k_{\ell_i}}) = \partial(\phi + \psi)(v_{k_{\ell_i}}).$$

It implies, by assumption (2), that

$$\begin{aligned}\left\|\frac{u_{k_{\ell_i}} - v_{k_{\ell_i}}}{\mu_{k_{\ell_i}}} - \nabla\phi(u_{k_{\ell_i}}) + \nabla\phi(v_{k_{\ell_i}})\right\| &\leq \frac{1}{\mu_{k_{\ell_i}}}\left\|u_{k_{\ell_i}} - v_{k_{\ell_i}}\right\| + \left\|\nabla\phi(v_{k_{\ell_i}}) - \nabla\phi(u_{k_{\ell_i}})\right\| \\ &\leq \frac{1}{\mu}\left\|u_{k_{\ell_i}} - v_{k_{\ell_i}}\right\| + \left\|\nabla\phi(v_{k_{\ell_i}}) - \nabla\phi(u_{k_{\ell_i}})\right\|.\end{aligned}$$

Using (3.29) and (3.30), we get

$$\left\|\frac{u_{k_{\ell_i}} - v_{k_{\ell_i}}}{\mu_{k_{\ell_i}}} - \nabla\phi(u_{k_{\ell_i}}) + \nabla\phi(v_{k_{\ell_i}})\right\| \to 0 \quad \text{as} \quad i \to \infty.$$

By demiclosedness of $\mathrm{Gph}(\partial(\phi+\psi))$, we obtain $0 \in \partial(\phi+\psi)(w^\star)$ and thus $w \in S^\star$. This implies that

$$\limsup_{\ell\to\infty}\langle f(v^\star)-v^\star, u_{k_\ell}-v^\star\rangle = \langle f(v^\star)-v^\star, w^\star-v^\star\rangle \le 0.$$

We derive from (3.22) and $\left\|x_{k_\ell}-v^\star\right\| \le \left\|x_{k_\ell+1}-v^\star\right\|$ that

$$\begin{aligned}\|x_{k_\ell+1}-v^\star\|^2 &\le (1-\lambda_{k_\ell}(1-c))\left\|x_{k_\ell}-v^\star\right\|^2\\ &\quad+\lambda_{k_\ell}(1-c)\left(\frac{2}{1-c}\langle f(v^\star)-v^\star, u_{k_\ell}-v^\star\rangle+\frac{3\bar{N}\eta_{k_\ell}}{\lambda_{k_\ell}(1-c)}\left\|y_{k_\ell}-y_{k_\ell-1}\right\|\right)\\ &\le (1-\lambda_{k_\ell}(1-c))\left\|x_{k_\ell+1}-v^\star\right\|^2\\ &\quad+\lambda_{k_\ell}(1-c)\left(\frac{2}{1-c}\langle f(v^\star)-v^\star, u_{k_\ell}-v^\star\rangle+\frac{3\bar{N}\eta_{k_\ell}}{\lambda_{k_\ell}(1-c)}\left\|y_{k_\ell}-y_{k_\ell-1}\right\|\right),\end{aligned}$$

which implies

$$\lambda_{k_\ell}(1-c)\|x_{k_\ell+1}-v^\star\|^2 \le \lambda_{k_\ell}(1-c)\left(\frac{2}{1-c}\langle f(v^\star)-v^\star, u_{k_\ell}-v^\star\rangle+\frac{3\bar{N}\eta_{k_\ell}}{\lambda_{k_\ell}(1-c)}\left\|y_{k_\ell}-y_{k_\ell-1}\right\|\right).$$

Consequently,

$$\left\|x_{k_\ell+1}-v^\star\right\|^2 \le \frac{2}{1-c}\langle f(v^\star)-v^\star, u_{k_\ell}-v^\star\rangle+\frac{3\bar{N}\eta_{k_\ell}}{\lambda_{k_\ell}(1-c)}\left\|y_{k_\ell}-y_{k_\ell-1}\right\|.$$

From the above inequality and $\left\|x_\ell-v^\star\right\| \le \left\|x_{k_\ell+1}-v^\star\right\|$, we obtain

$$0 \le \limsup_{\ell\to\infty}\left\|x_\ell-v^\star\right\|^2 \le \limsup_{\ell\to\infty}\left\|x_{k_\ell+1}-v^\star\right\|^2 \le 0.$$

Therefore, we can conclude that $x_k \to v^\star$. Finally, we show that $v^\star$ is the solution of problem (1.1). Since $f := I - t\nabla h$, it follows that $f(v^\star) := v^\star - t\nabla h(v^\star)$, which implies

$$\begin{aligned}0 &\le \langle P_{S^\star}f(v^\star)-f(v^\star), x-P_{S^\star}f(v^\star)\rangle\\ &= \langle v^\star-f(v^\star), x-v^\star\rangle\\ &= \langle v^\star-(v^\star-t\nabla h(v^\star)), x-v^\star\rangle\\ &= t\left\langle \nabla h(v^\star), x-v^\star\right\rangle,\end{aligned}$$

for all $x \in S^\star$. This together with $0 < t$ give us that $0 \le \left\langle \nabla h(v^\star), x-v^\star\right\rangle$ for all $x \in S^\star$. Hence $v^\star$ is the solution of the outer-level problem (1.1).

4. Applications

In this section, we present an experiment on image restoration and data classification problems by using our algorithm, and compare the performance of the proposed algorithm with BiG-SAM, iBiG-SAM, and aiBiG-SAM. We apply MATLAB 9.6 (R2019a) to perform all numerical experiments throughout this work. It runs on a MacBook Air 13.3-inch, 2020, with an Apple M1 chip processor and 8-core GPU, configured with 8 GB of RAM.

4.1. Image restoration problems

In this section, we apply the proposed algorithm to solve the true RGB image restoration problems, and compare its performance with BiG-SAM, iBiG-SAM, and aiBiG-SAM. Let A be a blurring operator, and x be an original image. If b represents an observed image, then a linear image restoration problem is defined by

$$Av = b + w, \tag{4.1}$$

where $x \in \mathbb{R}^{n\times 1}$ and w denotes an additive noise. In the traditional way, we apply the least absolute shrinkage and selection operator (LASSO) [23] method to approximate the original image x. It is given by

$$\min_x \left\{ \|Ax - b\|_2^2 + \alpha\|x\|_1 \right\}, \tag{4.2}$$

where α denotes a positive regularization parameter, $\|x\|_1 = \sum_{k=1}^n |x_k|$, and $\|x\|_2 = \sqrt{\sum_{k=1}^n |x_k|^2}$. We see that (4.2) is the inner-level problem (1.2) when $\phi(x) = \|Ax - b\|_2^2$ and $\psi(x) = \beta\|x\|_1$. When the true RGB image is transformed as the matrix on the LASSO model, we see that the size of matrix A and x as well as their members have an effect on the computation for the multiplication of Ax and $\|x\|$. To prevent this effect, we adopt the 2-D fast Fourier transform to convert the true RGB images into matrices instead. If $\mathcal{W}$ represents the 2-D fast Fourier transform, and $\mathcal{B}$ denotes the blurring matrix such that the blurring operator $\mathcal{A} = \mathcal{B}\mathcal{W}$, then problem (4.2) is transformed to the following problem:

$$\min_x \{ \|\mathcal{A}x - b\|_2^2 + \alpha\|\mathcal{W}x\|_1 \}, \tag{4.3}$$

where $b \in \mathbb{R}^{m\times n}$ is the observed image of size $m\times n$, and $\alpha > 0$ is a regularization parameter. Therefore, our proposed algorithm can be applied to solve an image restoration problem (4.1) by setting the inner-level problem as follows: $\phi(x) = \|\mathcal{A}x - b\|_2^2$, $\psi(x) = \alpha\|\mathcal{W}x\|_1$, and we choose the outer-level problem as $h(x) = \frac{1}{2}\|x\|^2$. Next, we select all of the parameters satisfying the convergence theorem of each algorithm as seen in Table 1.

Table 1. Chosen parameters of each algorithm.

Algorithm	Parameters									
	t	μ	α	λ_k	γ_k	ξ_k	δ	θ	σ	ρ
BiG-SAM	0.01	$\frac{k}{(k+1)L_\phi}$	-	$\frac{1}{k+2}$	-	-	-	-	-	-
iBiG-SAM	0.01	$\frac{k}{(k+1)L_\phi}$	3	$\frac{1}{50k}$	-	$\frac{10^{50}}{k^2}$	-	-	-	-
aiBiG-SAM	0.01	$\frac{k}{(k+1)L_\phi}$	3	$\frac{1}{k+2}$	-	$\frac{\lambda_k}{k^{0.01}}$	-	-	-	-
Algorithm 1	0.01	-	-	$\frac{1}{50k}$	$\frac{t_k-1}{t_{k+1}}$	$\frac{10^{50}}{k^2}$	0.124	0.1	0.9	0.5

Also, the Lipschitz constant L_ϕ of $\nabla\phi$ for BiG-SAM, iBiG-SAM, and aiBiG-SAM is calculated by the maximum eigenvalue of the matrix A^TA. The efficiency of a restorative image is measured by the peak signal-to-noise ratio (PSNR) in decibel (dB), which is given by

$$\text{PSNR}(x_k) = 10\log\left(\frac{255^2}{\text{mse}}\right),$$

where mse $= \frac{1}{K}\|x_k - v^\star\|_2^2$, K denotes the number of image samples, and $v^\star$ indicates the original image. We select the regularization parameter $\alpha = 0.00001$ and consider the original image (Wat Lok Moli) of size 256×256 pixels from [24]. We employ a Gaussian blur to construct blurred and noisy images of size 9×9 with the standard deviation $\sigma = 4$. The original and blurred images are shown in Figure 2. The results of deblurring the image of Wat Lok Moli at 500 iterations is demonstrated in Table 2.

Table 2. The values of PSNR at x_{10}, x_{50}, x_{100}, and x_{500}.

	The peak signal-to-noise ratio (PSNR)			
Iteration No.	BiG-SAM	iBiG-SAM	aiBiG-SAM	Algorithm 1
1	20.4661	20.5577	20.4661	20.6308
10	21.2325	21.7491	21.2327	22.9166
50	22.5011	25.0760	22.5015	26.4285
100	23.3503	26.5096	23.3508	27.7760
500	25.3727	30.8838	25.6802	31.4100

As seen in Table 2, our proposed algorithm (Algorithm 1) gives a higher value of PSNR than the others, which means that our algorithm has the best performance of the image restoration compared with others. The graph of PSNR for deblurring images at the 500th iteration are shown in Figure 1.

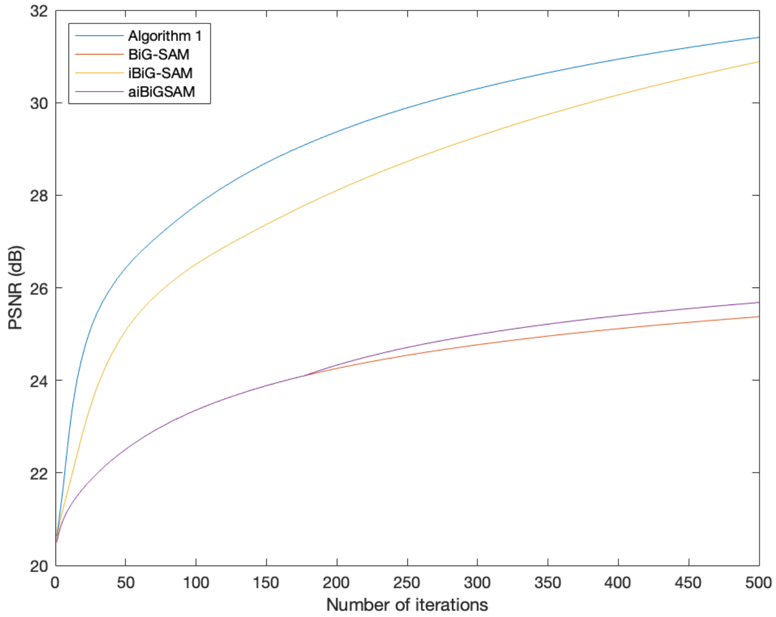
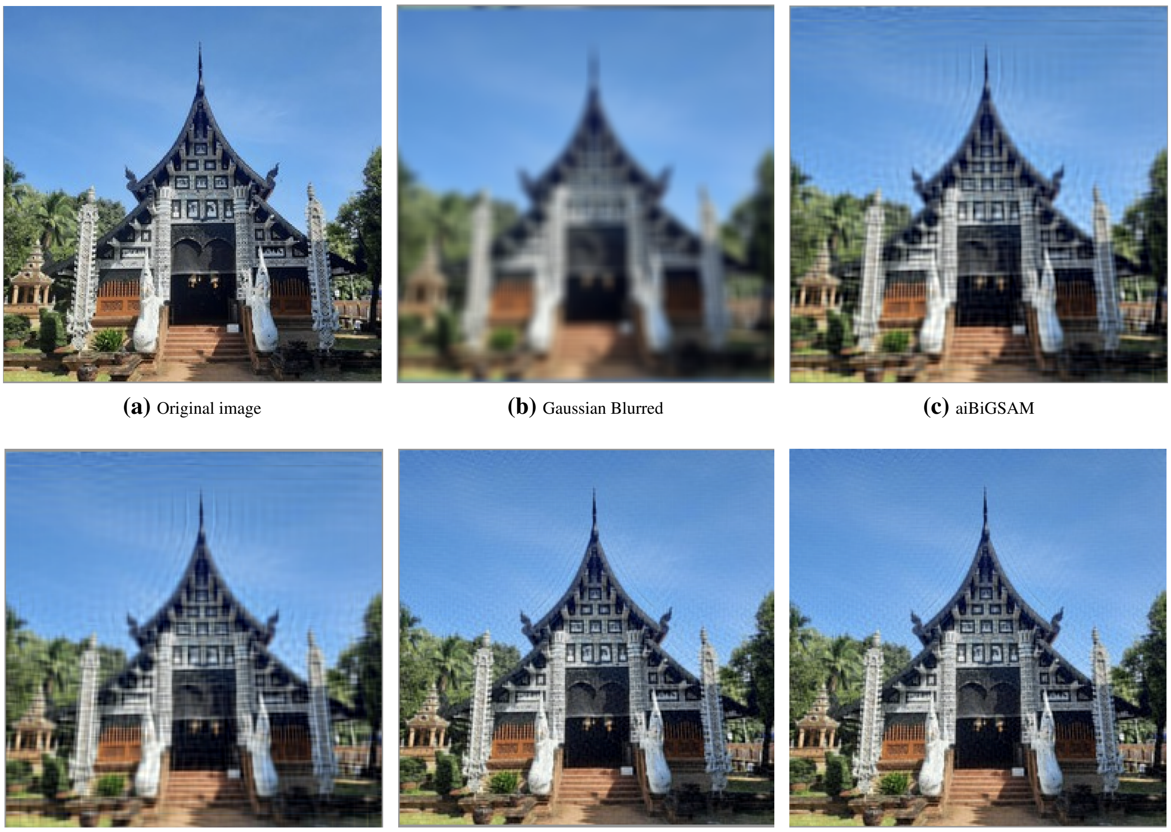


Figure 1. The graph of PSNR for Wat Lok Moli.

All restoration images of Wat Lok Moli of each algorithm at the 500th iteration are shown in Figure 2.



(a) Original image (b) Gaussian Blurred (c) aiBiGSAM

(d) BiGSAM (e) iBiGSAM (f) Algorithm 1

Figure 2. Results for image restoration at 500th iterations.

4.2. Data classification

Machine learning is crucial because it allows computers to learn from data and make decisions or predictions. There are three types of machine learning such as supervised learning, unsupervised learning, and reinforcement learning. Our work uses supervised learning which uses the extreme learning machine (ELM) [25] and a single-layer feedback neural network (SLFNs) model while the reinforcement learning is typically used for decision-making problems where an agent learns to perform actions in an environment to maximize cumulative rewards (see more information in [26, 27]). However, it is not commonly used directly for data classification, which is more traditionally tackled using supervised learning techniques.

In this work, we aim to use the studied algorithm to solve a binary data classification problem. We focus on classifying the patient datasets of heart disease [28] and breast cancer [29] into classes. The details of the studied datasets are given in Table 3.

Table 3. Details of datasets.

Datasets	Samples	Attributes	Classes
Heart disease	303	13	2
Breast cancer	699	11	2

Here, we accessed the above datasets on June 12, 2022 from `https://archive.ics.uci.edu`. We first start with a necessary notion for data classification problems. Now, we recall a concept of ELM. Suppose $p_k \in \mathbb{R}^n$ is an input data, and $q_k \in \mathbb{R}^m$ is the target. The training set of N samples is given by $S := \{(p_k, q_k) : p_k \in \mathbb{R}^n, q_k \in \mathbb{R}^m, k = 1, 2, \ldots, N\}$. The output of the i-th hidden node for any single hidden layer of ELM is

$$h_i(p) = \mathcal{G}(\langle w_i, p\rangle + r_i), \tag{4.4}$$

where $\mathcal{G}$ is an activate function, r_i is a bias, and w_i is the weight vector connecting the i-th hidden node and the input node. If M denotes the amount of the hidden nodes, then ELM for SLFNs gives the output function as:

$$o_j = \sum_{i=1}^{M} m_i h_i(p_j), \quad j = 1, 2, \ldots, N,$$

where m_i is the weight vector connecting the i-th hidden node and the output node. Thus, an output matrix of hidden layer $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} h_1(p_1) & h_2(p_1) & \cdots & h_M(p_1) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(p_N) & h_2(p_N) & \cdots & h_M(p_N) \end{bmatrix}.$$

A main purpose of ELM is to find a weight $m = [m_1^T, \ldots, m_M^T]^T$ such that

$$\mathbf{A}m = \mathbf{Q}, \tag{4.5}$$

where $\mathbf{Q} = [q_1^T, \ldots, q_N^T]^T$ is the training data. We observe from (4.5) that $m = \mathbf{A}^\dagger \mathbf{Q}$ whenever the Moore–Penrose generalized inverse $\mathbf{A}^\dagger$ of $\mathbf{A}$ exists. In some situations, if $\mathbf{A}^\dagger$ does not exist, it may be difficult to find wight m, which satisfies (4.5). In order to overcome this situation, we utilize the following convex minimization problem (4.6) to solve m:

$$\min_{m} \ \|\mathbf{A}m - \mathbf{Q}\|_2^2 + \beta\|m\|_1, \tag{4.6}$$

where β is the regularized parameter and $\|(m_1, m_2, \ldots, m_p)\|_1 = \sum_{i=1}^{p} |m_i|$. It derives from (4.6) that $\phi(m) := \|\mathbf{A}m - \mathbf{Q}\|_2^2$ and $\psi := \beta\|m\|_1$ are inner-level functions of problem (1.2). To employ the proposed algorithm, BiG-SAM, iBiG-SAM, and aiBiG-SAM for solving data classification, we choose the outer-level function $h(m) = \frac{1}{2}\|m\|^2$ for problem (1.1). With datasets from Table 3, we select an activation function $\mathcal{G}$ as sigmoid, and set the hidden node $M = 30$. Choose $t_0 = 1$ and $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$, for all $k \geq 0$. All parameters of each algorithm are chosen as in Table 4.

Table 4. Chosen parameters of each algorithm.

Algorithm	Parameters									
	t	μ	α	λ_k	γ_k	ξ_k	δ	θ	σ	ρ
BiG-SAM	0.01	$\frac{1}{L_\phi}$	-	$\frac{1}{k+2}$	-	-	-	-	-	-
iBiG-SAM	0.01	$\frac{1}{L_\phi}$	3	$\frac{1}{50k}$	-	$\frac{10^{50}}{k^2}$	-	-	-	-
aiBiG-SAM	0.01	$\frac{1}{L_\phi}$	3	$\frac{1}{k+2}$	-	$\frac{\lambda_k}{k^{0.01}}$	-	-	-	-
Algorithm 1	0.01	-	-	$\frac{1}{50k}$	$\frac{t_k-1}{t_{k+1}}$	$\frac{10^{50}}{k^2}$	0.124	0.1	0.9	0.5

Also, the Lipschitz gradient L_ϕ of $\nabla\phi$ for BiG-SAM, iBiG-SAM, and aiBiG-SAM can be calculated by $2\|\mathbf{A}\|^2$. In order to measure the performance of the accuracy for prediction, we use the following formula:

$$\text{Accuracy (Acc)} = \frac{TP + TN}{TP + TN + FP + FN} \times 100,$$

where TP is the number of cases correctly identified as patient, TN represent the number of cases correctly identified as healthy, FN means the number of cases incorrectly identified as healthy, and FP denotes the number of cases incorrectly identified as patient. In what follows, **Acc Train** refers to the accuracy of training on the dataset, while **Acc Test** indicates the accuracy of testing on the dataset. We present the iteration numbers and training time on the learning model for each algorithm in Table 5.

Table 5. The iteration number and training time of each algorithm with the highest accuracy on each dataset.

Dataset	Algorithm	Iteration no.	Training time	Acc train	Acc test
Heart Disease	BiG-SAM	1421	0.0207	85.24	79.57
	iBiG-SAM	410	0.0069	87.14	82.80
	aiBiG-SAM	1421	0.0321	85.24	79.57
	Algorithm 1	243	0.0871	87.14	82.80
Breast Cancer	BiG-SAM	587	0.0185	95.71	99.04
	iBiG-SAM	114	0.0041	96.12	99.04
	aiBiG-SAM	587	0.0191	95.71	99.04
	Algorithm 1	48	0.0428	96.12	99.04

As seen in Table 5, we observe that the training time of Algorithm 1 is not significantly different compared with the other algorithms. However, it needs to compute parameter μ_k occurring from the lineserch technique, while the other algorithms do not have this process. Note that under the linesearch technique, our algorithm has better convergence behavior than the others in terms of the number of iterations. This means that the proposed algorithm provides the best optimal weight compared with the others. To evaluate the performance of each algorithm, we construct a 10-fold cross validation. The 10-fold cross validation splits data into training sets and testing sets, as seen in Table 6.

Table 6. The number of samples in each fold for all datasets.

	Heart disease		Breast cancer	
	Train	Test	Train	Test
Fold 1	273	30	630	69
Fold 2	272	31	629	70
Fold 3	272	31	629	70
Fold 4	272	31	629	70
Fold 5	273	30	629	70
Fold 6	273	30	629	70
Fold 7	273	30	629	70
Fold 8	273	30	629	70
Fold 9	273	30	629	70
Fold 10	273	30	629	70

In addition, we use the following formula in order to measure the success probability of making a correct positive class classification, which is defined by

$$\text{Precision (Pre)} = \frac{TP}{TP + FP}.$$

Also, the sensitivity of the model toward identifying the positive class is estimated by

$$\text{Recall (Rec)} = \frac{TP}{TP + FN}.$$

The appraising tool is the average accuracy which is given by

$$\text{Average Acc} = \sum_{i=1}^{N} \frac{u_i}{v_i} \times 100\%/N,$$

where N is the number of sets considered during the cross validation ($N = 10$), u_i is the number of correctly predicted data at fold i, and v_i is the number of all data at fold i.

Let $\mathbf{Err}_M$ be the sum of errors in all 10 training sets, $\mathbf{Err}_K$ be the sum of errors in all 10 testing sets, M be the sum of all data in 10 training sets, and K be the sum of all data in 10 testing sets. Then,

$$\mathbf{Error}_{\%} = \frac{\mathbf{error}_{M\%} + \mathbf{error}_{K\%}}{2},$$

where $\mathbf{error}_{M\%} = \frac{\mathbf{Err}_M}{M} \times 100\%$ and $\mathbf{error}_{K\%} = \frac{\mathbf{Err}_K}{K} \times 100\%$.

We show the performance of each algorithm for patient prediction of heart disease and breast cancer with the 300th iteration in Tables 7 and 8.

Table 7. Experiment results in each fold for heart disease at the 300th iteration.

Heart disease		BiG-SAM		iBiG-SAM		aiBiG-SAM		Algorithm 1	
		Train	Test	Train	Test	Train	Test	Train	Test
Fold 1	Pre	0.79	0.88	0.82	0.94	0.79	0.88	0.83	0.87
	Rec	0.86	0.88	0.91	0.94	0.86	0.88	0.93	0.76
	Acc	79.85	86.67	84.25	93.33	79.85	86.67	85.71	80.00
Fold 2	Pre	0.79	0.78	0.82	0.82	0.79	0.78	0.84	0.83
	Rec	0.86	0.88	0.93	0.88	0.86	0.88	0.91	0.94
	Acc	80.15	80.65	84.56	83.87	80.15	80.65	86.03	87.10
Fold 3	Pre	0.79	0.78	0.82	0.78	0.79	0.78	0.84	0.81
	Rec	0.88	0.88	0.93	0.88	0.88	0.88	0.92	0.81
	Acc	80.88	80.65	85.29	80.65	80.88	80.65	86.03	80.65
Fold 4	Pre	0.81	0.74	0.82	0.79	0.81	0.74	0.84	0.74
	Rec	0.87	0.88	0.91	0.94	0.87	0.88	0.92	0.88
	Acc	81.62	77.42	84.56	83.87	81.62	77.42	86.40	77.42
Fold 5	Pre	0.79	0.77	0.82	0.81	0.79	0.77	0.84	0.81
	Rec	0.85	1.00	0.91	1.00	0.85	1.00	0.91	1.00
	Acc	79.85	83.33	84.62	86.67	79.85	83.33	85.35	86.67
Fold 6	Pre	0.79	0.82	0.82	0.80	0.79	0.82	0.86	0.74
	Rec	0.87	0.82	0.92	0.94	0.87	0.82	0.92	0.82
	Acc	80.59	80.00	84.98	83.33	80.59	80.00	87.18	73.33
Fold 7	Pre	0.78	0.84	0.82	0.76	0.78	0.84	0.83	0.94
	Rec	0.86	0.94	0.91	0.94	0.86	0.94	0.91	0.94
	Acc	79.49	86.67	84.62	80.00	79.49	86.67	84.62	93.33
Fold 8	Pre	0.81	0.71	0.83	0.76	0.81	0.71	0.83	0.79
	Rec	0.87	0.71	0.93	0.76	0.87	0.71	0.91	0.88
	Acc	82.05	66.67	85.71	73.33	82.05	66.67	85.35	80.00
Fold 9	Pre	0.81	0.70	0.83	0.75	0.81	0.70	0.83	0.83
	Rec	0.86	0.82	0.91	0.88	0.86	0.82	0.91	0.88
	Acc	81.32	70.00	84.98	76.67	81.32	70.00	85.35	83.33
Fold 10	Pre	0.80	0.83	0.82	0.83	0.80	0.83	0.82	0.84
	Rec	0.86	0.88	0.92	0.88	0.86	0.88	0.92	0.94
	Acc	80.59	83.33	84.98	83.33	80.59	83.33	84.98	86.67
Average Pre		0.80	0.79	0.82	0.81	0.80	0.79	0.84	0.82
Average Rec		0.87	0.87	0.92	0.90	0.87	0.87	0.91	0.89
Average Acc		80.64	79.54	84.86	82.51	80.64	79.54	85.70	82.85
$\text{Error}_{\%}$		19.91		16.32		19.91		15.73	

Table 8. Experiment results in each fold for breast cancer at the 300th iteration.

		BiG-SAM		iBiG-SAM		aiBiG-SAM		Algorithm 1	
Breast cancer		Train	Test	Train	Test	Train	Test	Train	Test
Fold 1	Pre	0.97	0.97	0.99	0.97	0.97	0.97	0.99	1.00
	Rec	0.98	0.89	0.98	0.89	0.98	0.89	0.98	0.89
	Acc	96.35	91.30	97.62	91.30	96.35	91.30	97.62	92.75
Fold 2	Pre	0.97	1.00	0.97	1.00	0.97	1.00	0.97	1.00
	Rec	0.97	0.98	0.98	0.98	0.97	0.98	0.98	0.98
	Acc	96.50	98.57	96.50	98.57	96.50	98.57	96.66	98.57
Fold 3	Pre	0.97	1.00	0.97	1.00	0.97	1.00	0.97	1.00
	Rec	0.97	0.98	0.97	0.98	0.97	0.98	0.97	0.98
	Acc	96.34	98.57	96.18	98.57	96.34	98.57	96.34	98.57
Fold 4	Pre	0.97	0.94	0.96	0.96	0.97	0.94	0.97	0.96
	Rec	0.97	1.00	0.97	1.00	0.97	1.00	0.97	1.00
	Acc	96.03	95.71	95.87	97.14	96.03	95.71	96.50	97.14
Fold 5	Pre	0.98	0.98	0.98	0.98	0.98	0.98	0.99	0.98
	Rec	0.98	1.00	0.97	1.00	0.98	1.00	0.97	1.00
	Acc	96.82	98.57	96.66	98.57	96.82	98.57	97.14	98.57
Fold 6	Pre	0.97	0.98	0.98	0.98	0.97	0.98	0.98	0.98
	Rec	0.97	0.98	0.98	0.98	0.97	0.98	0.97	0.98
	Acc	96.03	97.14	96.82	97.14	96.03	97.14	96.66	97.14
Fold 7	Pre	0.96	0.98	0.98	1.00	0.96	0.98	0.98	1.00
	Rec	0.98	0.98	0.97	0.98	0.98	0.98	0.97	0.98
	Acc	96.03	97.14	96.66	98.57	96.03	97.14	96.82	98.57
Fold 8	Pre	0.98	0.98	0.99	1.00	0.98	0.98	0.99	1.00
	Rec	0.97	0.96	0.97	0.96	0.97	0.96	0.97	0.96
	Acc	96.82	95.71	97.30	97.14	96.82	95.71	97.62	97.14
Fold 9	Pre	0.98	0.94	0.98	0.98	0.98	0.94	0.98	0.98
	Rec	0.97	1.00	0.97	1.00	0.97	1.00	0.97	1.00
	Acc	96.50	95.71	96.50	98.57	96.50	95.71	96.98	98.57
Fold 10	Pre	0.96	0.98	0.97	0.98	0.96	0.98	0.98	0.98
	Rec	0.97	1.00	0.98	0.98	0.97	1.00	0.97	0.98
	Acc	95.87	98.55	96.34	97.10	95.87	98.55	96.82	95.71
Average Pre		0.97	0.97	0.97	0.98	0.97	0.97	0.98	0.99
Average Rec		0.97	0.98	0.97	0.97	0.97	0.98	0.97	0.97
Average Acc		96.33	96.70	96.65	97.27	96.33	96.70	96.92	97.41
Error$_{\%}$		3.55		3.11		3.55		2.90	

According to Tables 7 and 8, Algorithm 1 gives the best average accuracy of training and testing datasets compared with BiG-SAM, iBiG-SAM, and aiBiG-SAM. We also see that our algorithm provides higher the recall and precision for diagnosis of heart disease and breast cancer. Furthermore, the proposed algorithm has the lowest percent error on prediction.

5. Conclusions

Recently, there are various algorithms for solving convex bilevel optimization problems (1.1) and (1.2). These methods require the Lipschitz continuity assumption of the gradient of the objective function on problem (1.2). To relax this criteria, the linesearch technique is applied. In this work, we proposed a novel accelerated algorithm employing both linesearch and inertial techniques for solving convex bilevel optimization problems (1.1) and (1.2). The convergence theorem of the proposed algorithm was analyzed under some suitable conditions. Furthermore, we applied our algorithm to solve image restoration and data classification problems. According to our experiment, we obtained that the proposed algorithm has more efficiency on image restoration and data classification than the others.

It is worth mentioning that in real-world application, if we appropriately choose the objective function of the outer-level problem (1.1), our algorithm can provide more benefit and accuracy for the specific objective of data classifications. Note that we use $\frac{1}{2}\|x\|^2$ as the outer-level objective function, so our solution is a minimum norm problem. In order to improve the accuracy for prediction, in future work, we need a new mathematical model and deep learning algorithm. Very recently, a deep extreme learning machine is an appropriate model for improving accuracy for prediction, see [30,31]. However, deep extreme learning algorithms are also challenging to study and discuss. Moreover, we would like to employ our method for prediction of noncommunicable diseases of patient data from the Sriphat Medical Center, Faculty of Medicine, Chiang Mai University.

Author contributions

Adisak Hanjing: formal analysis, investigation, resources, methodology, writing-review & editing, validation, data curation, and funding acquisition; Panadda Thongpaen: formal analysis, invesigation, writing-original draft, software, visualization, data curation; Suthep Suantai: conceptualization, supervision, project administration, methodology, validation, and formal funding acquisition. All authors have read and agreed to the published version of the manuscript.

Use of AI tools declaration

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

Acknowledgments

This work was partially supported by Chiang Mai University and the Fundamental Fund 2024 (FF030/2567), Chiang Mai University. The first author was supported by the Science Research and Innovation Fund, agreement no. FF67/P1-012. The authors would also like to thank the Rajamangala University of Technology Isan for partial financial support.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. P. Thongpaen, W. Inthakon, T. Leerapun, S. Suantai, A new accelerated algorithm for convex bilevel optimization problems and applications in data classification, *Symmetry*, **14** (2022), 2617. https://doi.org/10.3390/sym14122617
2. P. Thongsri, B. Panyanak, S. Suantai, A new accelerated algorithm based on fixed point method for convex bilevel optimization problems with applications, *Mathematics*, **11** (2023), 702. https://doi.org/10.3390/math11030702
3. A. Dhara, J. Dutta, *Optimality conditions in convex optimization: a finite-dimensional view*, Boca Raton: CRC Press, 2011. https://doi.org/10.1201/b11156
4. S. Sabach, S. Shtern, A first order method for solving convex bilevel optimization problems, *SIAM J. Optimiz.*, **27** (2017), 640–660. https://doi.org/10.1137/16M105592X
5. Y. Shehu, P. T. Vuong, A. Zemkoho, An inertial extrapolation method for convex simple bilevel optimization, *Optim. Method. Softw.*, **2021** (36), 1–19. https://doi.org/10.1080/10556788.2019.1619729
6. B. T. Polyak, Some methods of speeding up the convergence of iteration methods, *Comp. Math. Math. Phys.*, **4** (1964), 1–7. https://doi.org/10.1016/0041-5553(64)90137-5
7. H. K. Xu, Viscosity approximation methods for nonexpansive mappings, *J. Math. Anal. Appl.*, **298** (2004), 279–291. https://doi.org/10.1016/j.jmaa.2004.04.059
8. P. C. Duan, Y. Q. Zhang, Alternated and multi-step inertial approximation methods for solving convex bilevel optimization problems, *Optimization*, **72** (2023), 2517–2545. https://doi.org/10.1080/02331934.2022.2069022
9. L. O. Jolaoso, Y. Shehu, J. C. Yao, Strongly convergent inertial proximal point algorithm without on-line rule, *J. Optim. Theory Appl.*, **200** (2024), 555–584. https://doi.org/10.1007/s10957-023-02355-5
10. Y. H. Yao, A. Adamu, Y. Shehu, Forward–reflected–backward splitting algorithms with momentum: weak, linear and strong convergence results, *J. Optim. Theory Appl.*, **201** (2024), 1364–1397. https://doi.org/10.1007/s10957-024-02410-9
11. J. Y. B. Cruz, T. T. A. Nghia, On the convergence of the forward–backward splitting method with linesearches, *Optim. Method. Softw.*, **31** (2016), 1209–1238. https://doi.org/10.1080/10556788.2016.1214959
12. K. Kankam, N. Pholasa, P. Cholamjiak, On convergence and complexity of the modified forward-backward method involving new linesearches for convex minimization, *Math. Method. Appl. Sci.*, **42** (2019), 1352–1362. https://doi.org/10.1002/mma.5420
13. S. Suantai, K. Kankam, P. Cholamjiak, A novel forward-backward algorithm for solving convex minimization problem in Hilbert spaces, *Mathematics*, **8** (2020), 42. https://doi.org/10.3390/math8010042
14. W. Inthakon, S. Suantai, P. Sarnmeta, D. Chumpungam, A new machine learning algorithm based on optimization method for regression and classification problems, *Mathematics*, **8** (2020), 1007. https://doi.org/10.3390/math8061007

15. S. Suantai, M. A. Noor, K. Kankam, P. Cholamjiak, Novel forward–backward algorithms for optimization and applications to compressive sensing and image inpainting, *Adv. Differ. Equ.*, **2021** (2021), 265. https://doi.org/10.1186/s13662-021-03422-9

16. D. Chumpungam, P. Sarnmeta, S. Suantai, A new forward–backward algorithm with line searchand inertial techniques for convex minimization problems with applications, *Mathematics*, **9** (2021), 1562. https://doi.org/10.3390/math9131562

17. A. Hanjing, P. Jailoka, S. Suantai, An accelerated forward-backward algorithm with a new linesearch for convex minimization problems and its applications, *AIMS Mathematics*, **6** (2021), 6180–6200. https://doi.org/10.3934/math.2021363

18. H. H. Bauschke, P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, 2 Eds., Cham: Springer, 2017. https://doi.org/10.1007/978-3-319-48311-5

19. R. S. Burachik, A. N. Iusem, *Set-valued mappings and enlargements of monotone operators*, New York: Springer, 2008. https://doi.org/10.1007/978-0-387-69757-4

20. Y. Nesterov, *Introductory lectures on convex optimization: A basic course*, New York: Springer, 2004. https://doi.org/10.1007/978-1-4419-8853-9

21. H. K. Xu, Another control condition in an iterative method for nonexpansive mappings, *B. Aust. Math. Soc.*, **65** (2002), 109–113. https://doi.org/10.1017/S0004972700020116

22. P. E. Mainge, Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization, *Set-Valued Anal.*, **16** (2008), 899–912. https://doi.org/10.1007/s11228-008-0102-z

23. R. Tibshirani, Regression shrinkage and selection via the Lasso, *J. R. Stat. Soc. B*, **58** (1996), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

24. P. Thongpaen, W. Inthakon, A. Kaewkhao, S. Suantai, Convex minimization problems based on an accelerated fixed point algorithm with applications to image restoration problems, *J. Nonlinear Var. Anal.*, **7** (2023), 87–101. https://doi.org/10.23952/jnva.7.2023.1.06

25. G. B. Huang, Q. Y. Zhu, C. K. Siew, Extreme learning machine: theory and applications, *Neurocomputing*, **70** (2006), 489–501. https://doi.org/10.1016/j.neucom.2005.12.126

26. Z. K. Yao, X. L. Liang, G. P. Jiang, J. Y. Yao, Model-based reinforcement learning control of electrohydraulic position servo systems, *IEEE-ASME T. Mech.*, **28** (2023), 1446–1455. https://doi.org/10.1109/TMECH.2022.3219115

27. Z. K. Yao, F. Y. Xu, G. P. Jiang, J. Y. Yao, Data-driven control of hydraulic manipulators by reinforcement learning, *IEEE-ASME T. Mech.*, **99** (2023), 1–12. https://doi.org/10.1109/TMECH.2023.3336070

28. R. Detrano, A. Janosi, W. Steinbrunn, M. Pfisterer, J. J. Schmid, S. Sandhu, et al., International application of a new probability algorithm for the diagnosis of coronary artery disease, *Am. J. Cardiol.*, **64** (1989), 304–310. https://doi.org/10.1016/0002-9149(89)90524-9

29. W. N. Street, W. H. Wolberg, O. L. Mangasarian, Nuclear feature extraction for breast tumor diagnosis, *IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology*, San Jose, CA, United States, 1993, 861–870. https://doi.org/10.1117/12.148698

30. M. D. Tissera, M. D. McDonnell, Deep extreme learning machines: supervised autoencoding architecture for classification, *Neurocomputing*, **174** (2016), 42–49. https://doi.org/10.1016/j.neucom.2015.03.110

31. B. Jia, D. Li, Z. S. Pan, G. Y. Hu, Two-dimensional extreme learning machine, *Math. Probl. Eng.*, **2015** (2015), 491587. https://doi.org/10.1155/2015/491587

Appendix A

In this section, we discuss the specific details of algorithms related to our work. These algorithms were proposed for solving convex bilevel optimization problems as follows:

Algorithm 2 BiG-SAM: Bilevel gradient sequential averaging method [4].

```
Initialization Step: Select the sequence $\{\lambda_k\} \subset (0,1]$ corresponding to criteria assumed in [7], and take arbitrary $x_1 \in \mathbb{R}^n$. Consider the step sizes $\mu \in \left(0, \frac{1}{L_\phi}\right]$ and the parameter $t \in \left(0, \frac{2}{L_h+s}\right]$.
Iterative Step: For all $k \geq 1$, set $y_k = \operatorname{prox}_{\mu\psi}(I - \mu\nabla\phi)(x_k)$ and define
    $u_k = (I - t\nabla h)(x_k)$
    $x_{k+1} = \lambda_k u_k + (1-\lambda_k) y_k.$
```

Algorithm 3 iBiG-SAM: Inertial with bilevel gradient sequential averaging method.

```
Initialization Step: Select the sequence $\{\lambda_k\} \subset (0,1)$, and take arbitrary $x_1, x_0 \in \mathbb{R}^n$. Consider the step sizes $\mu \in \left(0, \frac{2}{L_\phi}\right)$, the parameter $t \in \left(0, \frac{2}{L_h+s}\right]$, and $\alpha \geq 3$.
Iterative Step: For all $k \geq 1$, set $z_k := x_k + \eta_k(x_k - x_{k-1})$ while $\eta_k \in [0, \bar{\eta}_k]$ corresponding to
    $\bar{\eta}_k = \begin{cases} \min\left\{\frac{k}{k+\alpha-1}, \frac{\xi_k}{\|x_k - x_{k-1}\|}\right\} & \text{if } x_k \neq x_{k-1}, \\ \frac{k}{k+\alpha-1} & \text{otherwise,} \end{cases} \quad (.1)$
    and define
    $y_k = \operatorname{prox}_{\mu\psi}(I - \mu\nabla\phi)(z_k),$
    $u_k = (I - t\nabla h)(z_k)$
    $x_{k+1} = \lambda_k u_k + (1-\lambda_k) y_k.$
```

Algorithm 4 aiBiG-SAM: The alternated inertial bilevel gradient sequential averaging method.

```
Initialization Step: Select the sequence $\{\lambda_k\} \subset (0,1)$ corresponding to criteria assumed in [5], and take arbitrary $x_1, x_0 \in H$. Consider the step sizes $\mu \in \left(0, \frac{2}{L_\phi}\right)$, the parameter $t \in \left(0, \frac{2}{L_h+s}\right]$, and $\alpha \geq 3$.
Initialization Step: For $k \geq 1$, if $k$ is odd, evaluate
    $$z_k := x_k + \eta_k(x_k - x_{k-1}),$$
    where $0 \leq |\eta_k| \leq \bar{\eta}_k$ while $\bar{\eta}_k$ corresponds to
    $$\bar{\eta}_k := \begin{cases} \min\left\{\frac{k}{k+\alpha-1}, \frac{\xi_k}{\|x_k - x_{k-1}\|}\right\} & \text{if } x_k \neq x_{k-1}, \\ \frac{k}{k+\alpha-1} & \text{if } x_k = x_{k-1}. \end{cases}$$
    When $k$ is even, set $z_k := x_k$. After that, define
    $$\begin{aligned} y_k &= \text{prox}_{\mu\psi}(I - \mu\nabla\phi)(z_k), \\ u_k &= (I - t\nabla h)(z_k) \\ x_{k+1} &= \lambda_k u_k + (1 - \lambda_k) y_k. \end{aligned}$$
```

Next, the details of the linesearch technique related to this work are provided as follows:

Algorithm 5 Linesearch 1 $(x, \sigma, \theta, \delta)$.

```
Initialization Step: Take arbitrary point $x \in \text{dom}\,\psi$, and set $L(x,\mu) = \text{prox}_{\mu\psi}(x - \mu\nabla\phi(x))$.
Choose $\theta \in (0,1)$ and $\delta \in \left(0, \frac{1}{2}\right)$.
Computation Step: Select $\sigma > 0$ and define the first value $\mu = \sigma$.
while
    $$\mu\|\nabla\phi(L(x,\mu)) - \nabla\phi(x)\| > \delta\|L(x,\mu) - x\|$$
 do
    $\mu = \theta\mu$,
    $L(x,\mu) = L(x,\theta\mu),\ S(x,\mu) = S(x,\theta\mu)$.
end while
Output $\mu$.
```

Algorithm 6 Linesearch 2 $(x, \sigma, \theta, \delta)$.

```
Initialization Step: Take arbitrary point $x \in \mathrm{dom}\psi$, and set $L(x,\mu) = \mathrm{prox}_{\mu\psi}(x - \mu\nabla\phi(x))$ and $S(x,\mu) = \mathrm{prox}_{\mu\psi}(L(x,\mu) - \mu\nabla\phi(L(x,\mu)))$.
Choose $\theta \in (0,1), \rho \in \left(0, \frac{1}{2}\right]$, and $\delta \in \left(0, \frac{\rho}{8}\right)$.
Computation Step: Select $\sigma > 0$ and define the first value $\mu = \sigma$.
while
    $$\mu\Big((1-\rho)\|\nabla\phi(S(x,\mu)) - \nabla\phi(L(x,\mu))\| + \rho\|\nabla\phi(L(x,\mu)) - \nabla\phi(x)\|\Big) > \delta\Big(\|S(x,\mu) - L(x,\mu)\| + \|L(x,\mu) - x\|\Big)$$
 do
    $\mu = \theta\mu$,
    $L(x,\mu) = L(x,\theta\mu),\ S(x,\mu) = S(x,\theta\mu)$.
end while
Output $\mu$.
```

Algorithm 7 FBIL: The forward-backward iterative method with the inertial technical term and linesearch technique.

```
Initialization Step: Take arbitrary points $x_1 = y_0 \in \mathrm{dom}\psi$.
For $k \geq 1$, calculate $\mu_k :=$ Linesearch 1 $(x_k, \sigma, \theta, \delta)$, and define
    $$z_k = \mathrm{prox}_{\mu_k\psi}(x_k - \mu_k\nabla\phi(x_k)),$$
    $$y_k = \mathrm{prox}_{\mu_k\psi}(z_k - \mu_k\nabla\phi(z_k)),$$
    $$x_{k+1} = \mathrm{P}_{\mathrm{dom}_\psi}\left(y_k + \eta_k(y_k - y_{k-1})\right),$$
where $\mathrm{P}_{\mathrm{dom}_\psi}$ is a metric projection mapping and $\eta_k \geq 0$.
```

AIMS Press