Thai Journal of Mathematics
Volume 21 Number 3 (2023)
Pages 625–645

http://thaijmath.in.cmu.ac.th

ISSN 1686-0209

A Novel Algorithm for Convex Bi-level Optimization in Hilbert Spaces with Applications to Data Classification Problems

Puntita Sae-jia[1] **and Suthep Suantai**[2,*]

[1] *Graduate Master's Degree Program in Mathematics, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand*
e-mail : puntita_sae@cmu.ac.th

[2] *Research Center in Mathematics and Applied Mathematics, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand*
e-mail : suthep.s@cmu.ac.th

Abstract In this paper, we present a new accelerated algorithm for solving convex bi-level optimization problems in Hilbert spaces and analyse convergence behavior of the proposed algorithms. We prove the strong convergence theorems of the proposed algorithms under some suitable conditions. As an application, we apply our main results to solve some data classification problems.

MSC: 49K35; 47H10; 20M12

Keywords: convex bi-level optimization problems; accelerated algorithm; common fixed point; nonexpansive mappings; classification problems

Submission date: 27.04.2023 / Acceptance date: 10.05.2023

1. Introduction

Fixed point theory is the one of the most powerful tools of mathematics. Fixed point techniques are applied extensively in various areas such as image processing, engineering, physics, computer science, economics, telecommunication, and other sciences. Fixed point theory focuses on two important problems which are existence and approximation problems.

Throughout this paper, let H be a real Hilbert space which inner product and the associated norm are denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. Let C be a nonempty closed convex subset of H and $\{x_n\}$ be a sequence in H, we denote the strong convergence and the weak convergence of $\{x_n\}$ to $x \in H$ by $x_n \to x$ and $x_n \rightharpoonup x$, respectively.

A mapping $T : C \to C$ is said to be a *contraction* if there exists a constant $k \in [0, 1)$ such that

$$\|T(x) - T(y)\| \leq k\|x - y\|, \ \forall x, y \in C.$$

*Corresponding author.

Published by The Mathematical Association of Thailand.

If $k = 1$ in above inequality, T is called *nonexpansive*.

There are various iteration methods for finding a fixed point of nonexpansives and other nonlinear mappings. A classical iteration process introduced in 1953 by Mann [1] was as Mann iteration process and it was defined by

$$x_{n+1} = \alpha_n x_n + (1-\alpha_n)Tx_n, \forall n \geq 1, \tag{1.1}$$

where $\{\alpha_n\}$ is a real sequence in $[0,1]$ and the initial value of $x_1 \in H$ is arbitrarily chosen. However, convergence of $\{x_n\}$ is in general not strong. Later, Halpern [2] introduced the method defined as follows:

$$x_{n+1} = \alpha_n x_0 + (1-\alpha_n)Tx_n, \forall n \geq 1, \tag{1.2}$$

where $\{\alpha_n\} \subset [0,1]$ and $x_0, x_1 \in C$. Under some conditions on $\{\alpha_n\}$, he obtained a strong convergence theorem of (1.2).

In 1974, Ishikawa [3] modified the Mann iteration, called the Ishikawa iteration process, given by

$$\begin{cases} y_n &= (1-\alpha_n)x_n + \alpha_n Tx_n, \\ x_{n+1} &= (1-\beta_n)x_n + \beta_n Ty_n, \forall n \geq 1, \end{cases} \tag{1.3}$$

where $\{\alpha_n\}, \{\beta_n\} \subset [0,1]$ and the initial value $x_1 \in H$.

In 2000, Moufafi [4] introduced a well-known viscosity approximation method for a nonexpansive mapping as follows: for the initial value of $x_1 \in H$,

$$x_{n+1} = \alpha_n f(x_n) + (1-\alpha_n)Tx_n, \forall n \geq 1, \tag{1.4}$$

where $\{\alpha_n\} \subset [0,1]$ and f is a contraction mapping. Under some suitable conditions, he proved that $\{x_n\}$ generated by (1.4) converges strongly to a fixed point of T, when T is a nonexpansive mapping.

In [5], Agarwal et al. extended Ishikawa iteration, called S-iteration process, by the following method:

$$\begin{cases} y_n &= (1-\alpha_n)x_n + \alpha_n Tx_n, \\ x_{n+1} &= (1-\beta_n)Tx_n + \beta_n Ty_n, \forall n \geq 1, \end{cases} \tag{1.5}$$

where $\{\alpha_n\}, \{\beta_n\} \subset [0,1]$ and the initial value of x_1. They showed that the convergence behavior of S-iteration is progressive than the iterations of Mann and Ishikawa.

Now, let $\{T_n\}$ be a family of nonexpansive mappings of C into itself. Over the past two decades, many authors have presented fixed point iteration process for finding a common fixed point of $\{T_n\}$.

To find a common fixed point of a countable family of nonexpansive mapping, Aoyama et al. [6] introduced a Halpern type iterative sequence, defined as follows: for the initial value of x_1,

$$x_{n+1} = \alpha_n x + (1-\alpha_n)T_n x_n, \forall n \geq 1, \tag{1.6}$$

where $\{\alpha_n\} \subset [0,1]$ and $x \in C$ is arbitrary.

After that, Takahashi [7] presented the following iteration process:

$$x_{n+1} = \alpha_n f(x_n) + (1-\alpha_n)T_n x_n, \forall n \geq 1, \tag{1.7}$$

where $\{\alpha_n\} \subset [0,1]$ and f is a contraction mapping. Under some condition on $\{\alpha_n\}$, he obtained a strong convergence theorem of (1.7).

In 2010, Klin-eam and Suantai [8] presented and studied the following algorithm: for $x_1 \in C$,

$$\begin{cases} y_n &= \alpha_n f(x_n) + (1-\beta_n)T_n x_n, \\ x_{n+1} &= (1-\beta_n)y_n + \beta_n T_n y_n, \forall n \geq 1, \end{cases} \tag{1.8}$$

where $\{\alpha_n\} \subset [0,1]$ and f is a contraction mapping. Moreover, under some suitable condition, they proved that $\{x_n\}$ generated by (1.8) converges strongly to a common fixed point of $\{T_n\}$.

To speed up the convergence behavior of the iteration processes, Polyak [9] presented an inertial technique to improve the convergence behavior of the method. Thence, inertial technique was used widely to accelerate the convergence behavior of the studied methods.

In 2009, Beck and Teboulle [10] introduced a fast iterative shrinkage-thresholding algorithm (FISTA), defined as follows: for the initial points $x_1 = y_0 \in \mathbb{R}^n$ and $t_1 = 1$,

$$\begin{cases} y_n &= Tx_n, \\ t_{n+1} &= \frac{1+\sqrt{1+4t_n^2}}{2}, \\ \theta_n &= \frac{t_n - 1}{t_{n+1}}, \\ x_{n+1} &= y_n + \theta_n(y_n - y_{n-1}), \forall n \geq 1, \end{cases} \tag{1.9}$$

where $T = \text{prox}_{\lambda g}(I - \lambda \nabla f)$ for $\lambda > 0$.

Let $\{T_n\}$ and $\mathcal{T}$ be families of nonexpansive mappings of H into itself such that $\emptyset \neq F(\mathcal{T}) \subset \bigcap_{n=1}^{\infty} F(T_n)$, where $F(\mathcal{T})$ is the set of all common fixed points of $\mathcal{T}$. We say that $\{T_n\}$ satisfies NST- condition(I) with $\mathcal{T}$ if for each bounded sequence $\{x_n\}$ such that $\lim_{n\to\infty} \|x_n - T_n x_n\| = 0$, it follows

$$\lim_{n\to\infty} \|x_n - Tx_n\| = 0 \text{ for all } T \in \mathcal{T}.$$

In particular, if $\mathcal{T}$ consists of one mapping T, i.e., $\mathcal{T} = \{T\}$, then $\{T_n\}$ is said to satisfy NST- condition(I) with T.

In 2020, Puangpee and Suantai [11] introduced a new accelerated viscosity algorithm (NAVA), defined as follows: for the initial value of $x_0, x_1 \in H$,

$$\begin{cases} y_n &= x_n + \theta_n(x_n - x_{n-1}), \\ z_n &= (1-\sigma_n)y_n + \sigma_n T_n y_n, \\ x_{n+1} &= \alpha_n f(x_n) + \beta_n T_n y_n + \gamma_n T_n z_n, \forall n \geq 1, \end{cases} \tag{1.10}$$

where $\{\sigma_n\}, \{\alpha_n\}, \{\beta_n\}$ and $\{\gamma_n\} \subset (0,1)$. Under some control conditions, they obtained a strong convergence theorem of (1.10).

Motivated by these works mentioned above we aim to construct a new accelerated algorithm for finding a common fixed point of a countable family of nonexpansive mappings and prove its convergence theorem. We also aim to apply our obtained results for solving some convex minimization problems.

In this paper, we are interested in the following bi-level optimization problem. The *outer* level given by the following constrained minimization problem,

$$\min_{x \in X^*} \omega(x), \qquad \text{(MNP)} \quad (1.11)$$

where ω is a strongly convex and differntiable function while X^* is the set of the minimizers of the *inner* level problem, which is the classical convex composite model, given by

$$\min_{x\in\mathbb{R}^n}\{\varphi(x) := f(x) + g(x)\} \quad \text{(P)} \quad (1.12)$$

where f is convex and continuously differentiable function and g is proper, lower semi-continuous and convex function from $\mathbb{R}^n$ to $(-\infty, \infty]$. There exists some direct methods for solving the (MNP) problem.

In 2005, Solodov [12] proposed an explicit and more tractable proximal point method for solving problem (MNP). Since then, various proximal point algorithm have been developed to solve the problem under different type of framework. Another direct approach to solve problem (MNP) is the *Hybrid Steepest Descent Method* (HSDM) present in [13] by Yamada et al.

In 2014, Beck and Sabach [14] proposed a new direct first order method for solving problem (MNP), called the Minimal Norm Gradient (MNG), for which the authors proved an $O(1/\sqrt{k})$ rate convergence result, in the term of the inner objective function value.

Motivated by the result in [14], Sabach and Shtern [15] proposed the new method, called *Sequential Averaging Method* (SAM), developed in [16]. It was employed in [15] for solving the problem in a more general setting. The proposed method was proved to have the rate of convergence of $O(1/k)$ in term of the function f. It is called the *Bi-Level Gradient Sequential Averaging Method* (BiG-SAM) defined as follows:

$$\begin{cases} y_n &= \operatorname{prox}_{\gamma g}(x_{n-1} - \gamma\nabla f(x_{n-1})), \\ z_n &= x_{n-1} - s\nabla\omega(x_{n-1}), \\ x_n &= \alpha_n z_n + (1-\alpha_n)y_n, \ n \geq 1, \end{cases} \quad (1.13)$$

with $x_0 \in \mathbb{R}^n, \gamma \in (0, 1/L_f], s \in (0, 2/(L_\omega + \sigma)]$ and $\{\alpha_n\}_{n\in\mathbb{N}} \in (0, 1]$, where L_f and L_ω are the Lipschitz gradient of f and ω, respectively.

Later, Shehu et al. [17] introduced an inertial extrapolation step to BiG-SAM, called *inertial Bi-Level Gradient Sequential Averaging Method* (iBiG-SAM) defined as follows:

$$\begin{cases} \bar{\theta_n} &= \begin{cases} \min\{\frac{n-1}{n+\alpha-1}, \frac{\epsilon_n}{\|x_n - x_{n-1}\|}\} & \text{if } x_n \neq x_{n-1}, \\ \frac{n-1}{n+\alpha-1} & \text{otherwise}, \end{cases} \\ y_n &= x_n + \theta_n(x_n - x_{n-1}), \\ s_n &= \operatorname{prox}_{\gamma g}(y_n - \gamma\nabla f(y_n)), \\ z_n &= y_n - s\nabla h(y_n), \\ x_{n+1} &= \alpha_n z_n + (1-\alpha_n)s_n, \ n \geq 1. \end{cases} \quad (1.14)$$

with $x_0, x_1 \in \mathbb{R}^n, \alpha \geq 3, \gamma \in (0, \frac{2}{L_f})$ and $s \in (0, \frac{2}{L_h+\sigma}]$ and $0 \leq \theta_n \leq \bar{\theta_n}$, where $\{\alpha_n\}$ and $\{\epsilon_n\}$ satisfy the conditions in Assumption A.

Assumption A. Suppose $\{\alpha_n\}_{n=1}^{\infty}$ is a sequence in $(0, 1)$ and $\{\epsilon_n\}_{n=1}^{\infty}$ is a positive sequence satisfying the following conditions:

(a) $\lim_{n\to\infty} \alpha_n = 0$ and $\sum_{n=1}^{\infty} \alpha_n = \infty$.

(b) $\epsilon_n = o(\alpha_n)$, i.e. $\lim_{n\to\infty}(\epsilon_n/\alpha_n) = 0$ (e.g. $\epsilon_n = 1/(n+1)^2, \alpha - n = 1/(n+1)$).

Motivated and inspired by the research going on in this direction, we are interest to propose a novel algorithm for convex bi-level optimization problems in Hilbert spaces. Then prove a strong convergence result of the proposed algorithms under some suitable control conditions. As an application, we apply our algorithms to solving data classification problems.

The organization of this paper is as follows: In Section 2, we provide some basic definitions and useful lemmas. The main results of this paper are given in Section 3, in this section, we introduced a new accelerated algorithm for solving convex bi-level optimization problem and then prove a strong convergence of the proposed algorithm. And also apply our main result to solving data classification problems in Section 4. Finally, in Section 5, is the summary of our work.

2. Preliminaries

In this section, we recall some important definitions, lemmas and propositions which are useful to prove our main results.

Definition 2.1. A mapping $T : C \to C$ is said to be

1. *Lipschitzian* if there exists $\tau \geq 0$ such that
$$\|Tx - Ty\| \leq \tau\|x - y\|, \ \forall x, y \in C,$$
2. *contraction* if T is Lipschitzian with the coefficient $\tau \in [0, 1)$,
3. *nonexpansive* if T is Lipschitzian with the coefficient $\tau = 1$.

It is well-known that if T is nonexpansive, then $F(T)$ is closed and convex.

Definition 2.2. Let C be a nonempty subset of H and $x \in H$. If there exists a point $x^* \in C$ such that
$$\|x^* - x\| \leq \|y - x\|, \quad \forall y \in C,$$
then x^* is called a *metric projection* of x on C, denoted by $P_C x$. If $P_C x$ exists and is unique for all x, then the function P_C of H onto C is called the *metric projection.*

Proposition 2.3 ([18])**.** *Let C be a nonempty convex subset of H and let $x \in H, x^* \in C$. Then,*
$$x^* = P_C x \Leftrightarrow \langle x - x^*, y - x^* \rangle \leq 0, \quad \forall y \in C.$$

Let $\{T_n\}$ and $\mathcal{T}$ be families of nonexpansive mappings of H into itself such that $\emptyset \neq F(\mathcal{T}) \subset \bigcap_{n=1}^{\infty} F(T_n)$, where $F(\mathcal{T})$ is the set of all common fixed points of $\mathcal{T}$. We say that $\{T_n\}$ satisfies NST- condition(I) with $\mathcal{T}$ if for each bounded sequence $\{x_n\}$ such that $\lim_{n\to\infty} \|x_n - T_n x_n\| = 0$, it follows
$$\lim_{n\to\infty} \|x_n - Tx_n\| = 0 \text{ for all } T \in \mathcal{T}.$$
In particular, if $\mathcal{T}$ concists of one mapping T, i.e., $\mathcal{T} = \{T\}$, then $\{T_n\}$ is said to satisfy NST- condition(I) with T.

Definition 2.4. Let $g : \mathbb{R}^n \to \mathbb{R}$ is proper convex and lower semi-continuous function $\lambda > 0$. The *proximitor* of λg at v, denoted by $\text{prox}_{\lambda g}(v)$, is defined by
$$\text{prox}_{\lambda g}(v) = \arg\min_x \left\{ g(x) + \frac{\|x-v\|^2}{2\lambda} \right\}, v \in \mathbb{R}^n.$$

Let $f, g : H \to (-\infty, +\infty]$ be proper convex and lower semi-continuous and $\lambda > 0$. Suppose f is differentiable. The operator $T := \text{prox}_{\lambda g}(I - \lambda\nabla f)$ is known as the forward-backward operator of f and g with respect to λ.

The following lemmas and propositions are essential tools for proving the main results.

Lemma 2.5 ([18],[20])**.** *Let H be a real Hilbert space. For $x, y \in H$ and any arbitrary real number $\lambda \in [0, 1]$, the following results hold:*

i. $\|\lambda x + (1 - \lambda)y\|^2 = \lambda\|x\|^2 + (1 - \lambda)\|y\|^2 - \lambda(1 - \lambda)\|x - y\|^2$;

ii. $\|x \pm y\|^2 = \|x\|^2 \pm 2\langle x, y\rangle + \|y\|^2, \forall x, y \in H$;

iii. $\|x+y\|^2 \leq \|x\|^2 + 2\langle y, x+y\rangle$.

The identity in Lemma 2.5 (i) implies that the following equality holds:

$$\|\alpha x+\beta y+\gamma z\|^2 = \alpha\|x\|^2+\beta\|y\|^2+\gamma\|z\|^2-\alpha\beta\|x-y\|^2-\beta\gamma\|y-z\|^2-\alpha\gamma\|x-z\|^2, \tag{2.1}$$

for all $x, y, z \in H$ and $\alpha, \beta, \gamma \in [0,1]$ with $\alpha+\beta+\gamma = 1$.

Lemma 2.6 ([19]). *Let $g : H \to \mathbb{R} \cup \{\infty\}$ be proper convex and lower semi-continuous function, and $f : H \to \mathbb{R}$ be convex differentiable with gradient ∇f being L-Lipschitz constant for some $L > 0$. If $\{T_n\}$ is the forward-backward operator of f and g with respect to $c_n \in (0, 2/L)$ such that c_n converges to c, then $\{T_n\}$ satisfies NST-condition(I) with T, where T is the forward-backward operator of f and g with respect to $c \in (0, 2/L)$.*

Let $A : H \to 2^H$ be a maximal monotone operator and $\lambda > 0$. The *resolvent* J^A_λ of A is defined by

$$J^A_\lambda := (I + \lambda A)^{-1}.$$

Proposition 2.7 ([19]). *Let H be a real Hilbert space. Let $A : H \to 2^H$ be a maximally monotone operator and $B : H \to H$ an L-Lipschitz operator, where $L > 0$. Let $T_n = J^A_{\lambda_n}(I - \lambda_n B)$, where $0 < \lambda_n < \frac{2}{L}$ for all $n \geq 1$ and let $T = J^A_\lambda(I - \lambda B)$, where $0 < \lambda < \frac{2}{L}$ with $\lambda_n \to \lambda$. Then $\{T_n\}$ satisfies the NST-condition(I) with T.*

Lemma 2.8 ([11]). *Let $\{T_n\}$ be a family of nonexpansive mappings of H into itself and $T : H \to H$ a nonexpansive mapping with $\emptyset \neq F(T) \subset \bigcap_{n=1}^\infty F(T_n)$. One always has, if $\{T_n\}$ satisfies NST-condition(I) with T, then $\{T_t\}$ also satisfies NST-condition(I) with T, for any subsequences $\{t\}$ of positive integers.*

Proposition 2.9 ([15]). *Let $\omega : \mathbb{R}^n \to \mathbb{R}$ is strongly convex with parameter $\sigma > 0$ and let ω is a continuously differentiable function such that $\nabla\omega$ is Lipschitz continuous with constant L_ω. Then, the mapping defined by $S_s = I - s\nabla\omega$, where I is the identity operator, is a contraction for all $s \leq 2/(L_\omega + \sigma)$, that is*

$$\|x - s\nabla\omega(x) - (y - s\nabla\omega(y))\| \leq \sqrt{1 - \tfrac{2s\sigma L_\omega}{\sigma + L_\omega}}\|x - y\|, \forall x, y \in \mathbb{R}^n.$$

Lemma 2.10 ([21]). *Let H be a real Hilbert space and $T : H \to H$ a nonexpansive mapping with $F(T) \neq \emptyset$. Then the mapping $I - T$ is demiclosed at zero, i.e, for any sequences $\{x_n\}$ in H such that $x_n \rightharpoonup x \in H$ and $\|x_n - Tx_n\| \to 0$ imply $x \in F(T)$.*

Lemma 2.11 ([22],[23]). *Let $\{s_n\}, \{\xi_n\}$ be sequences of nonnegative real numbers, $\{\delta_n\}$ a sequence in $[0,1]$ and $\{t_n\}$ a sequence of real numbers such that*

$$s_{n+1} \leq (1 - \delta_n)s_n + \delta_n t_n + \xi_n,$$

for all $n \in \mathbb{N}$. If the following conditions hold:

1. $\sum_{n=1}^\infty \delta_n = \infty$;
2. $\sum_{n=1}^\infty \xi_n < \infty$;
3. $\limsup_{n\to\infty} t_n \leq 0$.

Then $\lim_{n\to\infty} s_n = 0$.

Lemma 2.12 ([24]). *Let $\{\Theta_n\}$ be a sequence of real numbers that dose not decrease at infinity in the sense that there exists a subsequence $\{\Theta_{n_i}\}$ of $\{\Theta_n\}$ which satisfies $\Theta_{n_i} < \Theta_{n_i+1}$ for all $i \in \mathbb{N}$. Define the sequence $\{\tau(n)\}_{n\geq n_0}$ of integers as follows:*

$$\tau(n) := \max\{k \le n : \Theta_k < \Theta_{k+1}\},$$

where $n_0 \in \mathbb{N}$ *such that* $\{k \le n_0 : \Theta_k < \Theta_{k+1}\} \neq \emptyset$. *Then the following hold:*

1. $\tau(n_0) \le \tau(n_0+1) \le \cdots$ *and* $\tau(n) \to \infty$;
2. $\Theta_{\tau(n)} \le \Theta_{\tau(n)+1}$ *and* $\Theta_n \le \Theta_{\tau(n)+1}$ *for all* $n \ge n_0$.

3. Main Results

In this section, we provide a novel accelerated common fixed point algorithm using the inertial technique together with the viscosity approxiamtion method for finding a common fixed point of a family of nonexpansive mappings in Hilbert space. Secondly, under some conditions, we prove its strong convergence theorem.

As follows, we provide a novel accelerated algorithm, Algorithm 1, for approximating a solution of a common fixed point problem.

Throughout this section, let $\{T_n\}$ be a family of nonexpansive mappings on H into itself. Let f be a k-contraction mapping on H with $k \in (0,1)$ and let $\{\eta_n\} \subset (0,\infty)$ and $\{\alpha_n\}, \{\beta_n\}, \{\gamma_n\} \subset (0,1)$.

Algorithm 1

Initialize : Take $x_0, x_1 \in H$. Let $\{\mu_n\} \subset (0,\infty)$.
For $n \ge 1$ **:**
Set

$$\theta_n = \begin{cases} \min\{\mu_n, \frac{\eta_n \alpha_n}{\|x_n - x_{n-1}\|}\} & \text{if } x_n \neq x_{n-1}; \\ \mu_n & \text{otherwise.} \end{cases}$$

Compute

$$\begin{cases} y_n &= x_n + \theta_n(x_n - x_{n-1}), \\ z_n &= \gamma_n f(y_n) + (1-\gamma_n)T_n y_n, \\ x_{n+1} &= (1-\alpha_n-\beta_n)y_n + \alpha_n T_n z_n + \beta_n T_n y_n. \end{cases}$$

Next, we prove the strong convergence theorem of the sequence generated by Algorithm 1.

Theorem 3.1. *Let* $T : H \to H$ *be a nonexpansive mapping with* $F(T) \neq \emptyset$ *and* $f : H \to H$ *be a contraction mapping with the constant* $k \in (0,1)$. *Assume that* $\emptyset \neq F(T) \subset \bigcap_{n=1}^{\infty} F(T_n)$ *and* $\{T_n\}$ *satisfies NST-condition(I) with* T. *Let* $\{x_n\}$ *be a sequence generated by Algorithm 1 such that the following additional conditions hold:*

1. $\lim_{n\to\infty} \eta_n = 0$,
2. $\lim_{n\to\infty} \gamma_n = 0$ *and* $\sum_{n=1}^{\infty} \gamma_n = \infty$,
3. $0 < a < \alpha_n$ *for some* $a \in \mathbb{R}$,
4. $0 < b < \beta_n < \alpha_n + \beta_n < c < 1$ *for some* $b, c \in \mathbb{R}$,

then the sequence $\{x_n\}$ *converges strongly to* $u \in F(T)$ *where* $u = P_{F(T)} f(u)$.

Proof. Let $u \in F(T)$ such that $u = P_{F(T)} f(u)$. First of all, we show that $\{x_n\}$ is bounded. By the definition of y_n and z_n, we have

$$\begin{aligned} \|y_n - u\| &= \|x_n + \theta_n(x_n - x_{n-1}) - u\| \\ &\le \|x_n - u\| + \theta_n \|x_n - x_{n-1}\|, \forall n \ge 1 \end{aligned} \tag{3.1}$$

and

$$\begin{aligned}
\|z_n - u\| &= \|\gamma_n f(y_n) + (1-\gamma_n)T_n y_n - u\| \\
&= \|\gamma_n (f(y_n) - u) + (1-\gamma_n)(T_n y_n - u)\| \\
&\leq \gamma_n \|f(y_n) - u\| + (1-\gamma_n)\|T_n y_n - u\| \\
&\leq \gamma_n \|f(y_n) - u\| + \gamma_n \|f(u) - u\| + (1-\gamma_n)\|y_n - u\| \\
&\leq \gamma_n k \|y_n - u\| + \gamma_n \|f(u) - u\| + (1-\gamma_n)\|y_n - u\| \\
&= (1-(1-k)\gamma_n)\|y_n - u\| + \gamma_n \|f(u) - u\| \\
&\leq \|y_n - u\| + \gamma_n \|f(u) - u\|, \forall n \geq 1.
\end{aligned} \tag{3.2}$$

From (3.1) and (3.2), we also have

$$\begin{aligned}
\|x_{n+1} - u\| &= \|\alpha_n T_n z_n + \beta_n T_n y_n + (1-\alpha_n-\beta_n)y_n - u\| \\
&= \|\alpha_n (T_n z_n - u) + \beta_n (T_n y_n - u) + (1-\alpha_n-\beta_n)(y_n - u)\| \\
&\leq \alpha_n \|T_n z_n - u\| + \beta_n \|T_n y_n - u\| + (1-\alpha_n-\beta_n)\|y_n - u\| \\
&\leq \alpha_n \|z_n - u\| + \beta_n \|y_n - u\| + (1-\alpha_n-\beta_n)\|y_n - u\| \\
&= \alpha_n \|z_n - u\| + (1-\alpha_n)\|y_n - u\| \\
&\leq \alpha_n ((1-(1-k)\gamma_n)\|y_n - u\| + \gamma_n \|f(u) - u\|) + (1-\alpha_n)\|y_n - u\| \\
&= (1-(1-k)\alpha_n\gamma_n)\|y_n - u\| + \alpha_n\gamma_n \|f(u) - u\| \\
&\leq (1-(1-k)\alpha_n\gamma_n)(\|x_n - u\| + \theta_n \|x_n - x_{n-1}\|) + \alpha_n\gamma_n \|f(u) - u\| \\
&= (1-(1-k)\alpha_n\gamma_n)\|x_n - u\| \\
&\quad + (1-(1-k)\alpha_n\gamma_n)\theta_n \|x_n - x_{n-1}\| + \alpha_n\gamma_n \|f(u) - u\| \\
&= (1-(1-k)\alpha_n\gamma_n)\|x_n - u\| \\
&\quad + (1-k)\alpha_n\gamma_n [\tfrac{(1-(1-k)\alpha_n\gamma_n)}{(1-k)\gamma_n} \cdot \tfrac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| + \tfrac{\|f(u)-u\|}{1-k}].
\end{aligned} \tag{3.3}$$

According to the definition of θ_n and the assumption (1), we have

$$\frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| \to 0 \text{ as } n \to \infty.$$

Then there exists a positive constant $M_1 > 0$ such that

$$\frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| \leq M_1, \qquad \forall n \geq 1.$$

From (3.3), we obtain

$$\begin{aligned}
\|x_{n+1} - u\| &\leq (1-(1-k)\alpha_n\gamma_n)\|x_n - u\| \\
&\quad + (1-k)\alpha_n\gamma_n [\tfrac{\tau}{(1-k)} \cdot \tfrac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| + \tfrac{\|f(u)-u\|}{1-k}] \\
&\leq (1-(1-k)\alpha_n\gamma_n)\|x_n - u\| + (1-k)\alpha_n\gamma_n [\tfrac{\tau M_1 + \|f(u)-u\|}{(1-k)}] \\
&\leq \max\left\{\|x_n - u\|, \tfrac{\tau M_1 + \|f(u)-u\|}{(1-k)}\right\} \\
&\quad \vdots \\
&\leq \max\left\{\|x_1 - u\|, \tfrac{\tau M_1 + \|f(u)-u\|}{(1-k)}\right\}, \forall n \geq 1
\end{aligned}$$

where $\tau = \sup\left\{\frac{(1-(1-k)\alpha_n\gamma_n)}{\gamma_n}\right\}$. This implies $\{x_n\}$ is bounded and so are $\{y_n\}, \{z_n\}$, $\{f(x_n)\}$ and $\{T_n y_n\}$.

On the other hand, we have

$$\begin{aligned}\|y_n - u\|^2 &= \|x_n + \theta_n(x_n - x_{n-1}) - u\|^2 \\ &= \|(x_n - u) + \theta_n(x_n - x_{n-1})\|^2 \\ &= \|x_n - u\|^2 + 2\theta_n\langle x_n - u, x_n - x_{n-1}\rangle + \theta_n^2\|x_n - x_{n-1}\|^2 \\ &\leq \|x_n - u\|^2 + 2\theta_n\|x_n - u\| \cdot \|x_n - x_{n-1}\| + \theta_n^2\|x_n - x_{n-1}\|^2.\end{aligned} \tag{3.4}$$

By Lemma 2.5 (3) and the inequality (3.4), we have

$$\begin{aligned}\|x_{n+1} - u\|^2 &= \|\alpha_n T_n z_n + \beta_n T_n y_n + (1 - \alpha_n - \beta_n)y_n - u\|^2 \\ &= \|\alpha_n(T_n z_n - u) + \beta_n(T_n y_n - u) + (1 - \alpha_n - \beta_n)(y_n - u)\|^2 \\ &\leq \alpha_n\|T_n z_n - u\|^2 + \beta_n\|T_n y_n - u\|^2 + (1 - \alpha_n - \beta_n)\|y_n - u\|^2 \\ &\leq \alpha_n\|z_n - u\|^2 + \beta_n\|y_n - u\|^2 + (1 - \alpha_n - \beta_n)\|y_n - u\|^2 \\ &= \alpha_n\|z_n - u\|^2 + (1 - \alpha_n)\|y_n - u\| \\ &= \alpha_n\|\gamma_n f(y_n) + (1 - \gamma_n)T_n y_n - u\|^2 + (1 - \alpha_n)\|y_n - u\|^2 \\ &= \alpha_n\|\gamma_n(f(y_n) - f(u)) + (1 - \gamma_n)(T_n y_n - u) + \gamma_n(f(u) - u)\|^2 \\ &\quad +(1 - \alpha_n)\|y_n - u\|^2 \\ &\leq \alpha_n\|\gamma_n(f(y_n) - f(u)) + (1 - \gamma_n)(T_n y_n - u)\|^2 \\ &\quad +2\alpha_n\gamma_n\langle f(u) - u, z_n - u\rangle + (1 - \alpha_n)\|y_n - u\|^2 \\ &\leq \alpha_n\gamma_n\|f(y_n) - f(u)\|^2 + \alpha_n(1 - \gamma_n)\|T_n y_n - u\|^2 \\ &\quad +2\alpha_n\gamma_n\langle f(u) - u, z_n - u\rangle + (1 - \alpha_n)\|y_n - u\|^2 \\ &\leq \alpha_n\gamma_n k\|y_n - u\|^2 + \alpha_n(1 - \gamma_n)\|y_n - u\|^2 + (1 - \alpha_n)\|y_n - u\|^2 \\ &\quad +2\alpha_n\gamma_n\langle f(u) - u, z_n - u\rangle \\ &= (1 - (1 - k)\alpha_n\gamma_n)\|y_n - u\|^2 + 2\alpha_n\gamma_n\langle f(u) - u, z_n - u\rangle \\ &\leq (1 - (1 - k)\alpha_n\gamma_n)\|x_n - u\|^2 \\ &\quad +(1 - (1 - k)\alpha_n\gamma_n)[2\theta_n\|x_n - u\| \cdot \|x_n - x_{n-1}\| + \theta_n^2\|x_n - x_{n-1}\|^2] \\ &\quad +2\alpha_n\gamma_n\langle f(u) - u, z_n - u\rangle \\ &= (1 - (1 - k)\alpha_n\gamma_n)\|x_n - u\|^2 \\ &\quad +(1 - (1 - k)\alpha_n\gamma_n)[\theta_n\|x_n - x_{n-1}\|(2\|x_n - u\| + \theta_n\|x_n - x_{n-1}\|)] \\ &\quad +2\alpha_n\gamma_n\langle f(u) - u, z_n - u\rangle\end{aligned} \tag{3.5}$$

Since

$$\theta_n\|x_n - x_{n-1}\| = \alpha_n \cdot \frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| \to 0 \text{ as } n \to \infty,$$

there exists a positive constant $M_2 > 0$ such that

$$\theta_n\|x_n - x_{n-1}\| \leq M_2, \qquad \forall n \geq 1.$$

From the inequality (3.5), we derive that for all $n \geq 1$,

$$\begin{aligned}\|x_{n+1} - u\|^2 &\leq (1 - (1 - k)\alpha_n\gamma_n)\|x_n - u\|^2 \\ &\quad +3M_3(1-(1-k)\alpha_n\gamma_n)\theta_n\|x_n - x_{n-1}\| + 2\alpha_n\gamma_n\langle f(u) - u, z_n - u\rangle \\ &= (1 - (1 - k)\alpha_n\gamma_n)\|x_n - u\|^2 \\ &\quad +(1 - k)\alpha_n\gamma_n\Big[\frac{3M_3(1-(1-k)\alpha_n\gamma_n)}{(1-k)\gamma_n} \cdot \frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| \\ &\quad +\frac{2}{1-k}\langle f(u) - u, z_n - u\rangle\Big] \\ &\leq (1 - (1 - k)\alpha_n\gamma_n)\|x_n - u\|^2 \\ &\quad +(1-k)\alpha_n\gamma_n[\frac{3M_3\tau}{(1-k)} \cdot \frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| + \frac{2}{1-k}\langle f(u) - u, z_n - u\rangle],\end{aligned} \tag{3.6}$$

where $M_3 = \max\{\sup_n \|x_n - u\|, M_2\}$. From above inequality, we set

$$s_n := \|x_n - u\|^2, \quad \delta_n := \alpha_n \gamma_n (1-k)$$

and

$$t_n := \tfrac{3M_3\tau}{(1-k)} \cdot \tfrac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| + \tfrac{2}{1-k}\langle f(u) - u, z_n - u\rangle, \qquad \forall n \geq 1.$$

So, we obtain

$$s_{n+1} \leq (1-\delta_n)s_n + \delta_n t_n, \tag{3.7}$$

for all $n \geq 1$.

Now, we consider two cases for the proof as follows:

Case 1. Suppose that there exists a natural number n_0 such that the sequence $\{\|x_n - u\|\}_{n \geq n_0}$ is nonincreasing. Hence $\{\|x_n - u\|\}$ converges due to it is bounded from below by 0. Using the assumption (2) and (3), we get that $\sum_{n=1}^{\infty} \delta_n = \infty$. We next claim that

$$\limsup_{n\to\infty} \langle f(u) - u, z_n - u\rangle \leq 0.$$

By the inequality (3.2), we have

$$\begin{aligned}\|z_n - u\|^2 - \|y_n - u\|^2 &= (\|y_n - u\| + \gamma_n\|f(u) - u\|)^2 - \|y_n - u\|^2 \\ &= \|y_n - u\|^2 + 2\gamma_n\|y_n - u\| \cdot \|f(u) - u\| \\ &\quad + \gamma_n^2\|f(u) - u\| - \|y_n - u\|^2 \\ &= 2\gamma_n\|y_n - u\| \cdot \|f(u) - u\| + \gamma_n^2\|f(u) - u\|^2.\end{aligned} \tag{3.8}$$

Coming back to the definition of x_{n+1}, by Lemma 2.5 (1), (3.4) and (3.8), one has that

$$\begin{aligned}\|x_{n+1} - u\|^2 &= \|\alpha_n T_n z_n + \beta_n T_n y_n + (1-\alpha_n-\beta_n)y_n - u\|^2 \\ &= \|\alpha_n(T_n z_n - u) + \beta_n(T_n y_n - u) + (1-\alpha_n-\beta_n)(y_n - u)\|^2 \\ &\leq \alpha_n\|T_n z_n - u\|^2 + \beta_n\|T_n y_n - u\|^2 + (1-\alpha_n-\beta_n)\|y_n - u\|^2 \\ &\quad -\beta_n(1-\alpha_n-\beta_n)\|y_n - T_n y_n\|^2 \\ &\leq \alpha_n\|z_n - u\|^2 + \beta_n\|y_n - u\|^2 + (1-\alpha_n-\beta_n)\|y_n - u\|^2 \\ &\quad -\beta_n(1-\alpha_n-\beta_n)\|y_n - T_n y_n\|^2 \\ &= \alpha_n\|z_n - u\|^2 + (1-\alpha_n)\|y_n - u\|^2 - \beta_n(1-\alpha_n-\beta_n)\|y_n - T_n y_n\|^2 \\ &\leq \alpha_n[\|z_n - u\|^2 - \|y_n - u\|^2] \\ &\quad + \|x_n - u\|^2 + 2\theta_n\|x_n - u\| \cdot \|x_n - x_{n-1}\| + \theta_n^2\|x_n - x_{n-1}\|^2 \\ &\quad -\beta_n(1-\alpha_n-\beta_n)\|y_n - T_n y_n\|^2 \\ &\leq 2\alpha_n\gamma_n\|y_n - u\| \cdot \|f(u) - u\| + \alpha_n\gamma_n^2\|f(u) - u\|^2 \\ &\quad + \|x_n - u\|^2 + 2\theta_n\|x_n - u\| \cdot \|x_n - x_{n-1}\| + \theta_n^2\|x_n - x_{n-1}\|^2 \\ &\quad -\beta_n(1-\alpha_n-\beta_n)\|y_n - T_n y_n\|^2.\end{aligned} \tag{3.9}$$

It implies that for all $n \geq 1$,

$$\begin{aligned}\beta_n(1-\alpha_n-\beta_n)\|y_n - T_n y_n\|^2 \;\leq\; & 2\alpha_n\gamma_n\|y_n - u\| \cdot \|f(u) - u\| + \alpha_n\gamma_n^2\|f(u) - u\|^2 \\ & + \|x_n - u\|^2 - \|x_{n+1} - u\|^2 \\ & + \theta_n\|x_n - x_{n-1}\|(2\|x_n - u\| + \theta_n\|x_n - x_{n-1}\|).\end{aligned} \tag{3.10}$$

It follows from assumption (2), (4) and the converges of the sequence $\{\|x_n - u\|\}$ and $\theta_n\|x_n - x_{n-1}\| \to 0$ that

$$\|y_n - T_n y_n\| \to 0 \text{ as } n \to \infty. \tag{3.11}$$

According to $\{T_n\}$ satisfies NST-condition(I) with T, we obtain

$$\|y_n - T y_n\| \to 0 \text{ as } n \to \infty. \tag{3.12}$$

By definition of y_n and z_n, we have

$$\begin{aligned} \|y_n - z_n\| &= \|y_n - \gamma_n f(y_n) - (1-\gamma_n)T_n y_n\| \\ &= \|\gamma_n(f(y_n) - y_n) + (1-\gamma_n)(T_n y_n - y_n)\| \\ &\le \gamma_n\|f(y_n) - y_n\| + (1-\gamma_n)\|T_n y_n - y_n\|. \end{aligned} \tag{3.13}$$

This implies by (3.11) and assumption (2) that

$$\|y_n - z_n\| \to 0 \text{ as } n \to \infty. \tag{3.14}$$

By definition of x_{n+1}, we have

$$\begin{aligned} \|x_{n+1} - y_n\| &\le \|x_{n+1} - T_n y_n\| + \|T_n y_n - y_n\| \\ &= \|\alpha_n T_n z_n + \beta_n T_n y_n + (1-\alpha_n-\beta_n)y_n - T_n y_n\| + \|T_n y_n - y_n\| \\ &= \|\alpha_n(T_n z_n - T_n y_n) + \beta_n(T_n y_n - T_n y_n) + (1-\alpha_n-\beta_n)(T_n y_n - y_n)\| \\ &\quad + \|T_n y_n - y_n\| \\ &\le \alpha_n\|T_n z_n - T_n y_n\| + (1-\alpha_n-\beta_n)\|T_n y_n - y_n\| + \|T_n y_n - y_n\| \\ &\le \alpha_n\|z_n - y_n\| + (2-\alpha_n-\beta_n)\|T_n y_n - y_n\|, \end{aligned} \tag{3.15}$$

which implies

$$\|x_{n+1} - y_n\| \to 0 \text{ as } n \to \infty. \tag{3.16}$$

By definition of y_n, we obtain

$$\|y_n - x_n\| = \theta_n\|x_n - x_{n-1}\| \to 0 \text{ as } n \to \infty. \tag{3.17}$$

Hence

$$\|x_{n+1} - x_n\| \le \|x_{n+1} - y_n\| + \|y_n - x_n\| \to 0 \text{ as } n \to \infty. \tag{3.18}$$

Let

$$v = \limsup_{n\to\infty}\langle f(u) - u, z_n - u\rangle. \tag{3.19}$$

So, there exists a subsequence $\{z_{n_k}\}$ of $\{z_n\}$ such that

$$v = \lim_{k\to\infty}\langle f(u) - u, z_{n_k} - u\rangle.$$

Since $\{z_{n_k}\}$ is bounded, there exists a subsequnce $\{z_{n'_k}\}$ of $\{z_{n_k}\}$ such that $z_{n'_k} \rightharpoonup w \in H$. Without loss of generality, we may assume that $z_{n_k} \rightharpoonup w \in H$ and

$$v = \lim_{k\to\infty}\langle f(u) - u, z_{n_k} - u\rangle.$$

Since $\|y_n - z_n\| \to 0$, we get $y_{n_k} \rightharpoonup w$. It implies by Lemma 2.10 that $w \in F(T)$. Moreover, using $u = P_{F(T)}f(u)$ and Proposition 2.3, we obtain

$$v = \lim_{k\to\infty} \langle f(u) - u, z_{n_k} - u\rangle = \langle f(u) - u, w - u\rangle \leq 0. \tag{3.20}$$

Then

$$v = \limsup_{n\to\infty} \langle f(u) - u, z_n - u\rangle \ \ \leq 0. \tag{3.21}$$

It implies from (3.21) with the fact of $\frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| \to 0$ that $\limsup_{n\to\infty} t_n \leq 0$.

So, from (3.7) and using Lemma 2.11, we obtain that $x_n \to u$.

Case 2. Suppose that sequence $\{\|x_n - u\|\}_{n\geq n_0}$ is not monotonically nonincreasing sequence for all n_0. We set

$$\Phi_n := \|x_n - u\|^2.$$

So, there exists a subsequence $\{\Phi_{n_j}\}$ of $\{\Phi_n\}$ such that $\Phi_{n_j} < \Phi_{n_j+1}$ for all $j \in \mathbb{N}$. In this case, we define $\tau : \{n : n \geq n_0\}$, by

$$\tau(n) := \max\{k \in \mathbb{N} : k \leq n, \Phi_k < \Phi_{k+1}\}.$$

By Lemma 2.12, we have that $\Phi_{\tau(n)} \leq \Phi_{\tau(n)+1}$ for all $n \geq n_0$. That is

$$\|x_{\tau(n)} - u\| \leq \|x_{\tau(n)+1} - u\|, \ \forall n \geq n_0.$$

As in **Case 1**, we can conclude that for all $n \geq n_0$,

$$\begin{aligned}
&\beta_{\tau(n)}(1-\alpha_{\tau(n)}-\beta_{\tau(n)})\|y_{\tau(n)} - T_{\tau(n)}y_{\tau(n)}\|^2 \\
&\leq 2\alpha_{\tau(n)}\gamma_{\tau(n)}\|y_{\tau(n)} - u\| \cdot \|f(u) - u\| \\
&\quad + \alpha_{\tau(n)}\gamma^2_{\tau(n)}\|f(u) - u\|^2 \\
&\quad + \|x_{\tau(n)} - u\|^2 - \|x_{\tau(n)+1} - u\|^2 \\
&\quad + \theta_n\|x_{\tau(n)} - x_{\tau(n)-1}\|(2\theta_{\tau(n)}\|x_{\tau(n)} - u\| + \theta_{\tau(n)}\|x_{\tau(n)} - x_{\tau(n)-1}\|) \\
&\leq 2\alpha_{\tau(n)}\gamma_{\tau(n)}\|y_{\tau(n)} - u\| \cdot \|f(u) - u\| \\
&\quad + \alpha_{\tau(n)}\gamma^2_{\tau(n)}\|f(u) - u\|^2 \\
&\quad + \theta_n\|x_{\tau(n)} - x_{\tau(n)-1}\|(2\theta_{\tau(n)}\|x_{\tau(n)} - u\| + \theta_{\tau(n)}\|x_{\tau(n)} - x_{\tau(n)-1}\|),
\end{aligned}$$

which implies

$$\|y_{\tau(n)} - T_{\tau(n)}y_{\tau(n)}\| \to 0 \text{ as } n \to \infty. \tag{3.22}$$

Similarly to the proof in **Case 1**, we have

$$\|y_{\tau(n)} - z_{\tau(n)}\| \to 0, \tag{3.23}$$

$$\|x_{\tau(n)+1} - y_{\tau(n)}\| \to 0, \tag{3.24}$$

and

$$\|y_{\tau(n)} - x_{\tau(n)}\| \to 0, \tag{3.25}$$

as $n \to \infty$ and hence

$$\|x_{\tau(n)+1} - x_{\tau(n)}\| \to 0 \text{ as } n \to \infty. \tag{3.26}$$

We next show that $\limsup\limits_{n\to\infty} \langle f(u) - u, z_{\tau(n)} - u\rangle \leq 0$. Put

$$v = \limsup_{n\to\infty} \langle f(u) - u, z_{\tau(n)} - u\rangle.$$

Without loss of generality, there exists a subsequence $\{z_{\tau(t)}\}$ of $\{z_{\tau(n)}\}$ such that $\{z_{\tau(t)}\} \rightharpoonup w \in H$ and

$$v = \lim_{t\to\infty} \langle f(u) - u, z_{\tau(t)} - u\rangle.$$

By Lemma 2.8, one has $\{T_{\tau(t)}\}$ satisfies NST-condition(I) with T, so according to the inequality (3.22), $\|y_{\tau(t)} - T_{\tau(t)}y_{\tau(t)}\| \to 0$ as $t \to \infty$, we obtain that

$$\|y_{\tau(n)} - Ty_{\tau(n)}\| \to 0 \text{ as } n \to \infty. \tag{3.27}$$

Since $\|y_{\tau(t)} - z_{\tau(t)}\| \to 0$, we get $y_{\tau(t)} \rightharpoonup w$ which implies, by Lemma 2.10, that $w \in F(T)$. Further using $u = P_{F(T)}f(u)$ and Proposition 2.3, we get

$$v = \lim_{t\to\infty} \langle f(u) - u, z_{\tau(t)} - u\rangle = \langle f(u) - u, w - u\rangle \le 0 \tag{3.28}$$

Then

$$v = \limsup_{n\to\infty} \langle f(u) - u, z_{\tau(n)} - u\rangle \le 0 \tag{3.29}$$

Since $\Phi_{\tau(n)} \le \Phi_{\tau(n)+1}$ and $(1-k)\alpha_{\tau(n)}\gamma_{\tau(n)} > 0$, as the proof in **Case 1**, we have for all $n \ge n_0$

$$\|x_{\tau(n)} - u\|^2 \le \frac{3M_3\tau}{(1-k)} \cdot \frac{\theta_{\tau(n)}}{\alpha_{\tau(n)}}\|x_{\tau(n)} - x_{\tau(n)-1}\| + \frac{2}{1-k}\langle f(u) - u, z_{\tau(n)} - u\rangle \tag{3.30}$$

It follows by the fact that $\frac{\theta_{\tau}(n)}{\alpha_{\tau}(n)}\|x_{\tau}(n) - x_{\tau(n)-1}\| \to 0$ and (3.29) that

$$\limsup_{n\to\infty} \|x_{\tau(n)} - u\|^2 \le 0,$$

and hence $\|x_{\tau(n)} - u\| \to 0$ as $n \to \infty$.

This implies by (3.26) that $\|x_{\tau(n)+1} - u\| \to 0$ as $n \to \infty$. By Lemma 2.12, we get

$$\|x_n - u\| \le \|x_{\tau(n)+1} - u\| \to 0 \text{ as } n \to \infty$$

Hence $x_n \to u$. The proof is complete. ■

We consider the following bi-level convex minimization problem:

$$\min_{x\in X^*} \omega(x), \tag{3.31}$$

where X^* is the optimal solution set of problem (3.32). For the objective function ω of problem (3.31) we make the following assumption. Let Ω be the set of all solutions of (3.31)

Assumption 1.

C1. $\omega : \mathbb{R}^m \to \mathbb{R}$ is strongly convex with parameter $\sigma > 0$,

C2. ω is a continuously differentiable function such that and $\nabla\omega$ is Lipschitz continuous with constant L_ω.

For the problem

$$X^* = \underset{x\in\mathbb{R}^m}{\arg\min}(f(x)+g(x)) \tag{3.32}$$

we assume:

Assumption 2.

A1. $f:\mathbb{R}^m\to\mathbb{R}$ is convex and continuously differentiable,
A2. ∇f is Lipschitz continuous with constant L_f,
A3. $g:\mathbb{R}^m\to(-\infty,\infty]$ is proper, lower semicontinuous, and convex.

Next, we introduce an algorithm for solving problem (3.31)
We obtain the following result as a consequence of Theorem 3.1

Algorithm 2 :

Input : $c_n\in(0,2/L_f), s\in(0,2/(L_\omega+\sigma))$.
Initialize : Take $x_0,x_1\in\mathbb{R}^m$. Let $\{\mu_n\}\subset(0,\infty)$.
For $n\geq 1$ **:**
Compute

$$\begin{cases} y_n &= x_n+\theta_n(x_n-x_{n-1}),\\ z_n &= \gamma_n(I-s\nabla\omega)(y_n)+(1-\gamma_n)\operatorname{prox}_{c_ng}(I-c_n\nabla f)y_n,\\ x_{n+1} &= (1-\alpha_n-\beta_n)y_n+\alpha_n\operatorname{prox}_{c_ng}(I-c_n\nabla f)z_n+\beta_n\operatorname{prox}_{c_ng}(I-c_n\nabla f)y_n.\end{cases}$$

Theorem 3.2. *Let $\omega:\mathbb{R}^m\to\mathbb{R}$ be a function satisfying the Assumption 1. Let $f:\mathbb{R}^m\to\mathbb{R}$ and $g:\mathbb{R}^m\to(-\infty,\infty]$ be a function satisfying the Assumption 2. Let $\{c_n\}$ be a sequence of positive real numbers in $(0,2/L_f)$ and let $c\in(0,2/L_f)$ such that $c_n\to c$ as $n\to\infty$. A sequence $\{x_n\}$ generated by Algorithm 2 with the same conditions as in Theorem 3.1 converges strongly to $u\in\Omega$.*

Proof. Put $T_n=\operatorname{prox}_{c_ng}(I-c_n\nabla f), n\in\mathbb{N}$, and $T=\operatorname{prox}_{cg}(I-c\nabla f)$. By Proposition 2.6 we know that $\{T_n\}$ satisfies the NST condition(I) with T. We also know that $\{T_n\}$ and T are nonexpansive mappings. It follows directly from Theorem 3.1 that $\{x_n\}$ converges to $u\in F(T)=\underset{x\in\mathbb{R}^m}{\arg\min}(f(x)+g(x))$. We also have that $f=I-s\nabla\omega$ is contraction with parameter $k=\sqrt{1-\frac{2s\sigma L_\omega}{\sigma+L_\omega}}$, whenever $s\in(0,2/(L_\omega+\sigma))$. It remains to show that $u\in\Omega$. By using $u=P_{F(T)}f(u)$ and Proposition 2.3, we obtain

$$\begin{aligned} u=P_{F(T)}f(u) \;&\Leftrightarrow\; \langle f(u)-u,z-u\rangle\leq 0,\;\forall z\in F(T)\\ &\Leftrightarrow\; \langle u-s\nabla\omega(u)-u,z-u\rangle\leq 0,\;\forall z\in F(T)\\ &\Leftrightarrow\; \langle -s\nabla\omega(u),z-u\rangle\leq 0,\;\forall z\in F(T)\\ &\Leftrightarrow\; \langle s\nabla\omega(u),z-u\rangle\geq 0,\;\forall z\in F(T)\\ &\Leftrightarrow\; \langle \nabla\omega(u),z-u\rangle\geq 0,\;\forall z\in F(T)=X^*=\underset{x\in\mathbb{R}^m}{\arg\min}(f(x)+g(x))\end{aligned}$$

Hence, u is the optimal solution for the problem (3.31) ∎

4. Application

In this section, we apply our algorithms, BiG-SAM and iBiG-SAM to solve some classification problems based on the method proposed by Huang et al [25], which is called extreme learning machine (ELM). It is formulated as follows:

Let $\{(x_j, t_j) : x_j \in \mathbb{R}^n, t_j \in \mathbb{R}^m), j = 1, 2, \ldots, N\}$ be a training set with N distinct samples where x_j is an input and t_j is a target. A simple mathematical model for the output of ELM for a single-layer feedforward neuron network (SLFNs) with M hidden nodes and activation function G are mathematically modeled as

$$o_j = \sum_{i=1}^{M} \eta_i G(\langle w_i, x_j\rangle + b_i),\ j = 1, \ldots, N,$$

where w_i is the weight connecting the i-th hidden node and the input node, η_i is the weight connecting the i-th hidden node and the output node, and b_i is the bias.

Let $\mathbf{H}$ be a matrix given by the following:

$$\mathbf{H} = \begin{bmatrix} G(\langle w_1, x_1\rangle + b_1) & \cdots & G(\langle w_M, x_1\rangle + b_M) \\ \vdots & \ddots & \vdots \\ G(\langle w_1, x_N\rangle + b_1) & \cdots & G(\langle w_M, x_N\rangle + b_M) \end{bmatrix}_{N\times M}.$$

This matrix $\mathbf{H}$ is known as the hidden-layer output matrix.

The target of standard SLFNs is to approximate these N sample with zero means that $\sum_{j=1}^{N} |o_j - t_j| = 0$, i.e., there exists η_i, w_i and b_i such that

$$t_j = \sum_{i=1}^{M} \eta_i G(\langle w_i, x_j\rangle + b_i),\ j = 1, \ldots, N.$$

From N equations above, we can formulate a simple equation as

$$\mathbf{H}x = \mathbf{T},$$

where $x = [\eta_1^T, \cdots, \eta_M^T]^T, \mathbf{T} = [t_1^T, \cdots, t_N^T]^T$.

The main objective of a standard SLFNs is estimate η_i, w_i and b_i for solving (1.12) while ELM aim to calculate only $x = [\eta_1^T, \cdots, \eta_M^T]^T$ with randomly w_i and b_i. If there is a pseudo-inverse $\mathbf{H}^+$ of $\mathbf{H}$, the solution of $\mathbf{H}x = \mathbf{T}$ is $x = \mathbf{H}^+\mathbf{T}$. If the solution is not exact, we can find the solution that is closest in the least squares sense, i.e.,

$$\textbf{Minimize: } \|\mathbf{H}x - \mathbf{T}\|_2^2. \tag{4.1}$$

In machine learning, fitness of model is very important for accuracy on training sets. Overfitting model cannot be used to predict unknown data. In order to avoid overfitting, we use most popular technique which is called the least absolute shrinkage and selection operator (LASSO). It can be formulated as follows:

$$\textbf{Minimize: } \|\mathbf{H}x - \mathbf{T}\|_2^2 + \lambda\|x\|_1, \tag{4.2}$$

where $\|\cdot\|_1$ is l_1-norm defined by $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ and $\lambda > 0$ is a regularization parameter. If we set $f(x) := \|\mathbf{H}x - \mathbf{T}\|_2^2$ and $g(x) := \lambda\|x\|_1$, then the problem (4.2) is reduced to the problem (3.32) as follows:

$$\min_x\{\|\mathbf{H}x - \mathbf{T}\|_2^2 + \lambda\|x\|_1\} := \min_x\{f(x) + g(x)\}.$$

We know that $\nabla f(x) = 2\mathbf{H}^T(\mathbf{H}x - \mathbf{T})$ and Lipschitz constant of ∇f is $L_f = \lambda_{\max}(\mathbf{H}^*\mathbf{H})$, where H^* is the conjugate transpose of $\mathbf{H}$.

Hence, we can use our algorithm as a learning algorithm to find the optimal weight x and solve classification problems.

We consider two datasets:

1. **Iris dataset :** This dataset contains 3 classes of 50 instances where each class refer to a type of iris plant. The purpose is to divide each type of iris plant (iris setosa, iris versicolour and iris virginica) from sepal and petal length.
2. **Heart disease dataset :** This dataset contains 303 samples, each of which has 13 attributes which refers to the presence of heart disease in the patient. The predicted attribute is purpose to classify the data into 2 classes.
3. **Breast cancer dataset :** This dataset contains 699 samples, each of which has 10 attributes which refers to the presence of breast cancer in the patient. The predicted attribute is purpose to classify the data into 2 classes.
4. **Wine dataset :** This dataset is the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. This dataset contains 178 sample, each of which has 13 attributes. In this dataset, we classify 3 classes of data.

   Data preparation technique : $k-$fold Cross-validation ($k = 10$)

Bi-level minimization problem

- **Outer :** $\min \omega(x)$
- **Inner :** $\min \|\mathbf{H}x - \mathbf{T}\|_2^2 + \lambda\|x\|_1$
  $\Rightarrow \nabla f(x) = 2\mathbf{H}^T(\mathbf{H}x - \mathbf{T})$
  $\Rightarrow L_f = \lambda_{\max}(\mathbf{H}^*\mathbf{H})$, the maximum eigenvalue of $\mathbf{H}^*\mathbf{H}$.
- **Quadratic function :** $\omega(x) = \frac{1}{2}x^T\mathbf{A}x$,
  where $[x]_{n\times 1}$ is a column vector, $\mathbf{A}$ is a symmetric positive definite $n \times n$ matrix such that $\mathbf{A} = U^TU$ where U is an upper triangular matrix with positive elements on the main diagonal.
  In this case, we have $\nabla\omega(x) = \frac{1}{2}(\mathbf{A} + \mathbf{A}^T)x = \mathbf{A}x$.
- **Lipschitz constant :** $L_\omega = \|\mathbf{A}\|$,

$$\|\nabla\omega(x) - \nabla\omega(x)\| = \|\mathbf{A}x - \mathbf{A}y\| \leq \|\mathbf{A}\|\cdot\|x - y\|.$$

- **Strongly convex with constant** $\rho = \lambda_{\min}(\mathbf{A})$, the minimum eigenvalue of $\mathbf{A}$.

In case of $\mathbf{A} = \mathbf{I}_{n\times n}$, we can reduce the outer level to $\omega(x) = \frac{1}{2}\|x\|_2^2$ with $L_\omega = 1, \rho = 1$.

Algorithms :

1. **Our Algorithm** (Algorithm 3)
2. **BiG-SAM** (Algorithm 4)
3. **iBiG-SAM** (Algorithm 5)

Algorithm 3

Input. $x_0, x_1 \in \mathbb{R}^m, \mu_n, \eta_n \in (0,\infty), c_n \in (0, \frac{2}{L}), s \in (0, \frac{2}{L_h+\sigma}]$ and $\alpha_n, \beta_n, \gamma_n \in (0,1)$ for $n \in \mathbb{N}$,

$$\theta_n = \begin{cases} \min\{\mu_n, \frac{\eta_n\alpha_n}{\|x_n - x_{n-1}\|}\} & \text{if } x_n \neq x_{n-1}; \\ \mu_n & \text{otherwise.} \end{cases}$$

Compute

$$\begin{cases} y_n &= x_n + \theta_n(x_n - x_{n-1}), \\ z_n &= \gamma_n(I - s\nabla\omega)(y_n) + (1-\gamma_n)\operatorname{prox}_{c_n g}(I - c_n\nabla f)y_n, \\ x_{n+1} &= (1-\alpha_n-\beta_n)y_n + \alpha_n \operatorname{prox}_{c_n g}(I - c_n\nabla f)z_n + \beta_n \operatorname{prox}_{c_n g}(I - c_n\nabla f)y_n. \end{cases}$$

Algorithm 4 BiG-SAM

Input. $x_0 \in \mathbb{R}^m, \alpha_n \in (0,1), \gamma \in (0, \frac{1}{L_f}]$ and $s \in (0, \frac{2}{L_\omega+\sigma})$, for $n \in \mathbb{N}$.
Compute

$$\begin{cases} y_n &= \operatorname{prox}_{\gamma g}(x_n - \gamma\nabla f(x_n)), \\ x_{n+1} &= \alpha_n(x_n - s\nabla\omega(x_k)) + (1-\alpha_n)y_n. \end{cases}$$

Algorithm 5 iBiG-SAM

Input. $x_0, x_1 \in \mathbb{R}^m, \alpha_n \in (0,1), \gamma \in (0, \frac{2}{L_f})$ and $s \in (0, \frac{2}{L_h+\sigma}]$, for $n \in \mathbb{N}$.
Choose $\theta_n \in [0, \bar{\theta_n}]$ with $\bar{\theta_n}$ defined by

$$\bar{\theta_n} := \begin{cases} \min\{\frac{n-1}{n+\alpha-1}, \frac{\epsilon_n}{\|x_n - x_{n-1}\|}\} & \text{if } x_n \neq x_{n-1}, \\ \frac{n-1}{n+\alpha-1} & \text{otherwise.} \end{cases}$$

Compute

$$\begin{cases} y_n = x_n + \theta_n(x_n - x_{n-1}), \\ s_n = \operatorname{prox}_{\gamma g}(y_n - \gamma\nabla f(y_n)), \\ z_n = y_n - s\nabla h(y_n), \\ x_{n+1} = \alpha_n z_n + (1-\alpha_n)s_n. \end{cases}$$

Parameters	**Case (a)**	**Case (b)**	**Case (c)**	**Case (d)**	**Case (e)**	**Case (f)**
α_n	$0.1 + \frac{1}{33n}$	$0.5 + \frac{1}{33n}$	$0.5 + \frac{1}{33n}$	$0.5 + \frac{1}{33n}$	$0.5 + \frac{1}{33n}$	$0.5 + \frac{1}{33n}$
β_n	$0.2 + \frac{1}{33n}$	$0.2 + \frac{1}{33n}$	$0.9 - \alpha_n$	$0.9 - \alpha_n$	$0.9 - \alpha_n$	$0.9 - \alpha_n$
γ_n	$\frac{1}{33n}$	$\frac{1}{33n}$	$\frac{1}{33n}$	$\frac{1}{33n}$	$\frac{1}{33n}$	$\frac{1}{33n}$
c_n	$\frac{2}{3L_f}$	$\frac{2}{3L_f}$	$\frac{2}{3L_f}$	$\frac{1}{L_f}$	$\frac{1}{L_f}$	$\frac{1}{L_f}$
η_n	$\frac{33\cdot 10^{20}}{n}$	$\frac{33\cdot 10^{20}}{n}$	$\frac{33\cdot 10^{20}}{n}$	$\frac{33\cdot 10^{20}}{n}$	$\frac{33\cdot 10^{20}}{n}$	$\frac{33\cdot 10^{20}}{n}$
μ_n	0.5	0.5	0.5	0.5	0.09	0.9

Begin with the first experiment, we study convergence behavior of our algorithm, Algorithm 3 by considering the above six different cases.

It is clear that these control parameters satisfy all condition of Theorem 3.1. In this experiment, we take $s = 0.01$, that is $(I - s\nabla\omega)x = 0.99x$. The performance of each case is shown by the terms of average accuracy as seen in table 1.

TABLE 1. The performance of each case with 10-fold cv. on Breast cancer dataset at $x_1, x_5, x_{10}, x_{25}, x_{50}, x_{100}, x_{250}, x_{500}$.

No.	**Case (a)**		**Case (b)**		**Case (c)**		**Case (d)**		**Case (e)**		**Case (f)**	
(n)	acc.tr	acc.te	acc.tr	acc.te	acc.tr	acc.te	acc.tr	acc.te	acc.tr	acc.te	acc.tr	acc.te
1	65.52	65.52	65.52	65.52	65.52	65.52	65.52	65.52	65.52	65.52	65.52	65.52
5	74.07	72.40	76.81	75.83	80.46	78.69	88.67	89.70	80.00	78.26	93.61	92.98
10	71.42	69.68	91.80	91.41	92.96	92.14	94.91	94.27	91.99	91.70	95.47	96.41
25	90.45	90.84	95.57	95.12	95.72	95.55	95.99	96.27	95.53	94.84	96.53	96.84
50	94.87	94.27	96.01	96.56	96.09	96.56	96.31	96.56	96.03	96.56	96.58	96.84
100	95.82	95.84	96.44	96.56	96.49	96.56	96.61	96.56	96.41	96.56	96.55	96.84
250	96.31	96.70	96.61	96.56	96.63	96.70	96.61	96.84	96.61	96.56	96.61	96.99
500	96.58	96.56	96.65	96.84	96.65	96.84	96.60	96.84	96.65	96.84	96.71	96.99

The result of table 1 indicates that when μ_n is close to 1, the performance of the algorithm is better than those of smaller μ_n.

From table 1, we select the most advantageous option of each parameter for each algorithm to achieve the highest level of performance as follows:

- Regularization parameter : $\lambda = 0.00001$
- Hidden nodes : HidN $= 30$
- $n = 1000$, $s = 0.01$ and $\alpha = 3$

Algorithm 3 :

$$\theta_n = \begin{cases} \min\{0.9, \frac{\eta_n \alpha_n}{\|x_n - x_{n-1}\|}\} & \text{if } x_n \neq x_{n-1}; \\ 0.9 & \text{otherwise.} \end{cases}$$

$$\alpha_n = 0.5 + \frac{1}{33n},\ \beta_n = 0.9 - \alpha_n,\ \gamma_n = \frac{1}{33n},\ c_n = \frac{1}{L_f} \text{ and } \eta_n = \frac{33\cdot 10^{20}}{n}.$$

Algorithm 4 and 5 (BiG-SAM and iBiG-SAM) :

$$\gamma = \frac{1}{L_f},\ \alpha_n = \frac{1}{n}.$$

We compare the performance of each algorithm at the 1000th iteration on difference 4 datasets and obtain the following results, as seen in table 2, table 3, table 4 and table 5.

Our experiments show that Algorithm 3 provides a higher accuracy than the others.

TABLE 2. The performance of each algorithm at 1000th iteration with 10-fold cv. on Iris dataset.

	Algorithm 3		**Algorithm 4**		**Algorithm 5**	
	acc. train	acc.test	acc. train	acc.test	acc. train	acc.test
Fold 1	89.63	100.00	83.70	93.33	83.70	100.00
Fold 2	89.63	100.00	83.70	80.00	88.15	80.00
Fold 3	94.81	93.33	83.70	73.33	87.41	80.00
Fold 4	88.89	86.67	85.93	80.00	85.93	80.00
Fold 5	91.85	86.67	83.70	86.67	87.41	80.00
Fold 6	94.81	80.00	85.19	86.67	87.41	66.67
Fold 7	92.59	86.67	83.70	80.00	86.67	86.67
Fold 8	94.81	93.33	85.19	86.67	88.15	86.67
Fold 9	94.07	93.33	82.96	93.33	86.67	93.33
Fold 10	88.89	93.33	82.96	86.67	85.19	93.33
Average acc.	92.00	91.33	84.07	84.67	86.67	84.67

TABLE 3. The performance of each algorithm at 1000th iteration with 10-fold cv. on Heart disease dataset.

	Algorithm 3		**Algorithm 4**		**Algorithm 5**	
	acc. train	acc.test	acc. train	acc.test	acc. train	acc.test
Fold 1	84.25	93.33	82.42	90.00	84.25	93.33
Fold 2	84.93	80.65	81.62	83.87	84.56	80.65
Fold 3	85.29	77.42	83.09	80.65	85.29	77.42
Fold 4	84.93	87.10	83.82	87.10	84.56	87.10
Fold 5	84.62	86.67	81.68	90.00	84.62	86.67
Fold 6	84.98	83.33	82.78	80.00	85.35	83.33
Fold 7	86.45	83.33	82.05	86.67	84.98	83.33
Fold 8	85.71	73.33	84.62	70.00	85.71	73.33
Fold 9	86.08	80.00	82.78	70.00	85.35	80.00
Fold 10	86.08	83.33	83.15	83.33	86.08	83.33
Average acc.	85.33	82.85	82.80	82.16	85.07	82.85

TABLE 4. The performance of each algorithm at 1000th iteration with 10-fold cv. on Breast cancer dataset.

	Algorithm 3		**Algorithm 4**		**Algorithm 5**	
	acc. train	acc.test	acc. train	acc.test	acc. train	acc.test
Fold 1	97.62	92.75	97.46	91.30	97.62	92.75
Fold 2	96.66	98.57	96.82	98.57	96.66	98.57
Fold 3	96.50	98.57	96.18	98.57	96.50	98.57
Fold 4	96.34	97.14	96.03	95.71	96.34	97.14
Fold 5	97.14	98.57	96.66	97.14	97.14	98.57
Fold 6	96.66	97.14	96.34	97.14	96.66	97.14
Fold 7	96.82	98.57	96.34	98.57	96.82	98.57
Fold 8	97.46	97.14	97.14	97.14	97.30	97.14
Fold 9	96.66	98.57	96.82	97.14	96.66	98.57
Fold 10	96.82	95.71	96.66	97.14	96.66	95.71
Average acc.	96.87	97.28	96.65	96.84	96.84	97.28

Table 5. The performance of each algorithm at 1000th iteration with 10-fold cv. on Wine dataset.

	Algorithm 3		**Algorithm 4**		**Algorithm 5**	
	acc. train	acc.test	acc. train	acc.test	acc. train	acc.test
Fold 1	100.00	100.00	97.52	100.00	98.76	100.00
Fold 2	100.00	100.00	97.50	100.00	99.38	100.00
Fold 3	100.00	100.00	97.50	100.00	98.75	94.44
Fold 4	99.38	100.00	97.50	100.00	98.75	100.00
Fold 5	100.00	100.00	97.50	100.00	98.75	100.00
Fold 6	100.00	100.00	97.50	100.00	99.38	100.00
Fold 7	100.00	94.44	98.12	94.44	99.38	94.44
Fold 8	100.00	100.00	97.50	100.00	98.75	100.00
Fold 9	100.00	94.44	98.75	88.89	99.38	94.44
Fold 10	100.00	100.00	98.14	94.12	99.38	94.12
Average acc.	99.94	98.89	97.75	97.75	99.06	97.75

5. Conclusion

In this paper, we introduced a new accelerated algorithm for convex bi-level optimization problems in Hilbert space. First, we prove a strong convergence in Algorithm 1 under some suitable control conditions. Next, we prove strong convergence theorems in Algorithm 2. Finally, we apply an algorithm to solve the data classification problems. We find that our algorithm provide a higher accuracy than BiG-SAM and iBig-SAM.

References

[1] W.R. Mann, Mean value methods in iteration, Proc. Amer. Math. Soc. 4 (3) (1953) 506–510.

[2] B. Halpern, Fixed points of nonexpanding maps, Bull. Amer. Math. Soc. 73 (6) (1967) 957–961.

[3] S. Ishikawa, Fixed points by a new iteration method, Proc. Amer. Math. Soc. 44 (1) (1974) 147–150.

[4] A. Moudafi, Viscosity approximation methods for fixed points problems, J. Math. Anal. and Appl. 241 (1) (2000) 46–55.

[5] R.P. Agarwal, D. O'Regan, D.R. Sahu, Iterative construction of fixed points of nearly asymptotically nonexpansive mappings, J. Nonlinear Convex Anal. 8 (1) (2007) 61.

[6] K. Aoyama, Y. Kimura, W. Takahashi, M. Toyoda, Approximation of common fixed points of a countable family of nonexpansive mappings in a Banach space, Nonlinear Anal. 67 (8) (2007) 2350–2360.

[7] W. Takahashi, Viscosity approximation methods for countable family of nonexpansive mapping in Banach spaces, Nonlinear Anal. 70 (2009) 719–734.

[8] C. Klin-eam, S. Suantai, Strong convergence of composite iterative schemes for a countable family of nonexpansive mappings in Banach spaces, Nonlinear Anal. 73 (2010) 431–439.

[9] B.T. Polyak, Some methods of speeding up the convergence of iteration methods, USSR Comput. Math. & Math. Phys. 4 (5) (1964) 1–17.

[10] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sci. 2 (1) (2009) 183–202.

[11] J. Puangpee, S. Suantai, A new accelerated viscosity iterative method for an infinite family of nonexpansive mappings with applications to image restoration problems, Mathematics 8 (4) (2020) 615.

[12] M. Solodov, An explicit descent method for bilevel convex optimization, J. Convex Anal. 4 (2) (2007) 227–237.

[13] I. Yamada, M. Yukawa, M. Yamagishi, Minimizing the Moreau envelope of non-smooth convex functions over the fixed point set of certain quasi-nonexpansive mappings, In Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer, New York (2011), 345–390.

[14] A. Beck, S. Sabach, A first order method for finding minimal norm-like solutions of convex optimization problems, Math. Program. 147 (1–2, Ser. A) (2014) 25–46.

[15] S. Sabach, S. Shtern, A first order method for solving convex bi-level optimization problems, SIAM J. Optim. 27 (2) (2017) 640–660.

[16] H.K. Xu, Viscosity approximation methods for nonexpansive mappings, J. Math. Anal. Appl. 298 (1) (2004) 279–291.

[17] Y. Shehu, P.T. Vuong, A. Zemkoho, An inertial extrapolation method for convex simple bilevel optimization, Optim. Methods Softw. 36 (1) (2021) 1–19.

[18] W. Takahashi, Introduction to Nonlinear and Convex Analysis, Yokohama Publishers, 2009.

[19] L. Bussaban, S. Suantai, A. Kaewkhao, A paralle inertial S-iteration forward-backward algorithm for regression and classification problems, Carpathian J. Math. 36 (2020) 35–44.

[20] W. Takahashi, Nonlinear Functional Analysis, Yokohama Publishers, 2000.

[21] K. Goebel, W.A. Kirk, Topic in Metric Fixed Point Theory, Cambridge University Press, 1990.

[22] K. Aoyama, Y. Yasunori, W. Takahashi, M. Toyoda, On a strongly nonexpansive sequence in a Hilbert space, J. Nonlinear Convex Anal. 8 (3) (2007) 471–490.

[23] H.K. Xu, Another control condition in an iterative method for nonexpansive mappings, Bull. Aust. Math. Soc. 65 (1) (2002) 109–113.

[24] P.E. Mainge, Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization, Var. Anal. 16 (2008) 899–912.

[25] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: Theory and applications, Neurocomputing 70 (1–3) (2006) 489–501.

[26] A. Hanjing, L. Bussaban, S. Suantai, The modified viscosity approximation method with inertial technique and forward-backward algorithm for convex optimization method, Mathematics 10 (7) (2022) 1036.

[27] P. Jailoka, S. Suantai, A. Hanjing, A fast viscosity forward-backward algorithm for convex minimization problems with an application in image recovery, Carpathian J. Math. 37 (3) (2021) 449–461.