S. Suantai & S. Rozyyev, J. Nonlinear Anal. Optim. Vol. 14(2) (2023), 43-61

Journal of Nonlinear Analysis and Optimization
Volume 14(2) (2023)
http://www.math.sci.nu.ac.th
ISSN : 1906-9685

J. Nonlinear Anal. Optim.

ACCELERATED FIXED POINT ALGORITHM FOR CONVEX BI-LEVEL OPTIMIZATION PROBLEMS IN HILBERT SPACES WITH APPLICATIONS

SUTHEP SUANTAI*[1,2] AND SERDAR ROZYYEV[3]

[1] Data Science Research Center,Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

[2] Research Group in Mathematics and Applied Mathematics, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand

[3] Master Degree Program in Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand; serdar.rozik@gmail.com

ABSTRACT. In this thesis, we propose and analyze a new accelerated algorithm for solving bi-level convex optimization problems in Hilbert spaces in the form of the minimization of smooth and strongly convex function over the optimal solutions set which is the set of all minimizers of the sum of smooth and nonsmooth functions. In addition, we apply our algorithms to solve regression and classification problems by using machine learning models. Our experiments show that our proposed machine learning algorithm has a better convergence behaviour than the others.

KEYWORDS:Bi-level convex problem, nonexpansive mapping, fixed point.
AMS Subject Classification: :65F10; 65H10.

1. Introduction

Let H be a real Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle$. For a nonempty closed convex subset C of H and let $T : C \to C$. A point $x \in C$ is a fixed point of T if $Tx = x$. We let $F(T) := \{x \in C : Tx = x\}$, the fixed point set of T. A mapping $T : C \to C$ is said to be *nonexpansive* if $\|Tx - Ty\| \le \|x - y\|$, $\forall x, y \in C$. A mapping $S : C \to C$ is said to be a *contraction* if there exists $\delta \in [0, 1)$ such that $\|S(x) - S(y)\| \le \delta \|x - y\|$, $\forall x, y \in C$.

There are many iteration methods for finding a fixed point of nonexpansive mappings. In 1953, Mann[1] showed an iteration process, which is defined as

$$x_{n+1} = \alpha_n x_n + (1 - \alpha_n) T x_n, \ n \in \mathbb{N} \tag{1.1}$$

* *Corresponding author.*
Email address : suthep.s@cmu.ac.th.
Article history : Received 15 September 2023; Accepted 12 December 2023.

where the initial value of x_1 is taken in C and $\{\alpha_n\} \subset [0,1]$. He proved a weak convergence theorem of (1.1) in certain conditions on $\{\alpha_n\}$. Later, Halpern[2] presented the algorithm defined as follows

$$\begin{aligned} &x_0, x_1 \in C \\ &x_{n+1} = \alpha_n x_0 + (1-\alpha_n)Tx_n,\ n \in \mathbb{N} \end{aligned} \tag{1.2}$$

where $\{\alpha_n\} \subset [0,1]$. He obtained a strong convergence theorem of (1.2) under some conditions on $\{\alpha_n\}$. In 2000, Moudafi[3] introduced a viscosity approximation method for a nonexpansive mapping as follows:

$$\begin{aligned} &x_0 = x \in C, \\ &x_{n+1} = \alpha_n f(x_n) + (1-\alpha_n)Tx_n,\ n \geq 0 \end{aligned} \tag{1.3}$$

where $f : C \to C$ is a contraction. He proved a strong convergence theorem of (1.3) under some conditions on $\{\alpha_n\}$.

In 1974, Ishikawa [4] introduced two-steps algorithm given by

$$\begin{aligned} &x_1 \in C \\ &y_n = \beta_n x_n + (1-\beta_n)Tx_n, \\ &x_{n+1} = \alpha_n x_n + (1-\alpha_n)Ty_n,\ n \in \mathbb{N}. \end{aligned} \tag{1.4}$$

In [5], Agarwal et al, introduced a new algorithm, called S-iteration, given by

$$\begin{aligned} &x_1 \in C, \\ &y_n = (1-\beta_n)x_n + \beta_n Tx_n, \\ &x_{n+1} = (1-\alpha_n)Tx_n + \alpha_n Ty_n,\ n \in \mathbb{N} \end{aligned} \tag{1.5}$$

where $\{\alpha_n\}, \{\beta_n\} \subset [0,1]$. They showed that the convergence behavior of the (1.5) is better than the iterations of Mann and Ishikawa.

Now, let $\{T_n\}$ be a sequence of nonexpansive mappings $T_n : C \to C$ and let $F := \bigcap_{n=1}^{\infty} F(T_n)$ be the set of all common fixed points of $T_n, n \in \mathbb{N}$. Over the last two decades, many mathematicians have turned their attention to finding a common fixed points of $\{T_n\}$.

Aoyama et al.[6] introduced a Halpern-type iterative scheme for finding a common fixed point of a countable family of nonexpansive mappings $\{T_n\}$ as follows :

$$\begin{aligned} &x_0 \in C, \\ &x_{n+1} = \alpha_n x + (1-\alpha_n)T_n x_n,\ n \geq 0 \end{aligned} \tag{1.6}$$

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

Later, Takahashi [7] studied the following iteration scheme:

$$\begin{aligned} &x_0 \in C, \\ &x_{n+1} = \alpha_n S(x_n) + (1-\alpha_n)T_n x_n,\ n \geq 0, \end{aligned} \tag{1.7}$$

where S is a contraction on C. He obtained strong convergence theorem of (1.7) under some conditions.

In 2010, Klin-eam and Suantai [8] introduced and studied the following algorithm:

$$\begin{aligned} &x_0 \in C, \\ &y_n = \alpha_n S(x_n) + (1-\alpha_n)T_n x_n, \end{aligned}$$

$$x_{n+1} = (1-\beta_n)y_n + \beta_n T_n y_n,\ n \geq 0, \tag{1.8}$$

where $S : C \to C$ is a contraction and $\{\alpha_n\}, \{\beta_n\} \subset [0,1]$. They proved strong convergence theorem of the sequence $\{x_n\}$ generated by (1.8) to a common fixed point of $\{T_n\}$ under some suitable conditions.

To speed up the convergence behavior of the iteration methods, Polyak [9] introduced an inertial technique to improve the convergence behavior of the method. Since then, this technique was used widely to accelerate the convergence behavior of the studied methods.

Beck and Teboulle[10] introduced a fast iterative shrinkase-thresholding algorithm (FISTA) as follows:

$$\begin{aligned}
&x_1 = y_0 \in C,\ t_1 = 1,\\
&y_n = Tx_n,\\
&t_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2},\ \theta_n = \frac{t_n - 1}{t_{n+1}},\\
&x_{n+1} = y_n + \theta_n(y_n - y_{n-1}).
\end{aligned} \tag{1.9}$$

They proved that rate of convergence of FISTA is better than the others.

In 2017, Verma and Shukla [11] introduced a new accelerated proximal gradient algorithm (NAGA) as follows:

$$\begin{aligned}
&x_0, x_1 \in C\\
&y_n = x_n + \theta_n(x_n - x_{n-1}),\\
&x_{n+1} = T_n[(1-\alpha_n)y_n + \alpha_n T_n y_n],
\end{aligned} \tag{1.10}$$

where $\{\theta_n\}, \{\alpha_n\}$ are sequences in $(0,1)$ and $\frac{\|x_n - x_{n-1}\|_2}{\theta_n} \to 0$. They proved a convergence theorem of NAGA and applied this method to solve a non-smooth convex minimization problem with sparsity-inducing regularizers for the multitask learning framework.

Now, consider the the following bi-level optimization problem. Here is an *inner* level problem,

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

where f is convex and continuously differentiable function and g is an extended valued (possibly nonsmooth) function. The following *outer* level problem

$$(MNP)\quad \min_{x\in X^*} \omega(x)$$

where ω is a strongly convex and differentiable function while X^* is the, assumed nonempty, set of minimizers of the *inner* level problem.

In 2017, Sabach and Shtern in [12] introduced a new method based on existing algorithm, that they call it Sequential Averaging Method (SAM), which was developed in [13] for solving a certain class of fixed-point problems. For solving the bi-level optimization problems (P) and (MNP), the Bi-Level Gradient Sequential Averaging Method (BiG-SAM) was defined in [12].

Motivated by the above works, we are interested to introducing new accelerated fixed point algorithm to solve bi-level convex optimization problems in Hilbert space and analyze the convergence behavior of the proposed algorithm for data regression, and data classification.

This paper is divided into three sections. The motivation of this work and some literature reviews is given in the introduction section. In Section 2, we give some preliminary materials containing useful definitions and lemmas which will be used for proving our main results. Finally, the main results and conclusion are in Section 3.

2. Preliminaries

In this section, we introduce lemmas, definitions and theorems on a real Hilbert space that will be used in this work. $\mathbb{R}$ stands for the set of real numbers and $\mathbb{N}$ denotes the set of natural numbers. Throughout this section, we assume that H is a real Hilbert space equipped with an inner product $\langle \cdot, \cdot \rangle$. We denote that $x_n \to x^*$ and $x_n \rightharpoonup x^*$, the strong convergence and weak convergence of a sequence $\{x_n\}$ to a point x^*, respectively.

Definition 2.1. Let C be a nonempty sebset 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 onto C and is denoted by $P_C x$. The operator P_C is called the *metric projection*.

Theorem 2.1 ([17]). *Let C be a nonempty closed convex subset of H. Then, for any $x \in H$ there exists a metric projection $P_C x$ of x onto C and it is unique.*

Proposition 2.2 ([14]). *Let C be a nonempty convex subset of H and let $x \in H, x^* \in C$. Then,*

$$x^* = P_C x \Leftrightarrow \langle x - x^*, y - x^* \rangle \leq 0, \quad \forall y \in C.$$

Definition 2.3. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *strongly convex* with constant $\sigma > 0$ if for any $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$,

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

In the case that f is differentiable, then f is strongly convex if and only if for any $x, y \in \mathbb{R}^n$,

$$f(y) - f(x) \geq \langle \nabla f(x), y - x \rangle + \frac{\sigma}{2} \|y - x\|^2 .$$

Definition 2.4. A mapping $T : C \to C$ is said to be

(i) *Lipschitzian* if there exists $\tau \geq 0$ such that

$$\|Tx - Ty\| \leq \tau \|x - y\|, \quad \forall x, y \in C,$$

(ii) a *contraction* if T is Lipschitzian with the coefficient $\tau \in [0, 1)$,

(iii) *nonexpansive* if T is Lipschitzian with the coefficient $\tau = 1$.

Let $T : C \to C$ be a mapping. We say that an element $x \in C$ is a fixed point of T if $x = Tx$. The set of all fixed points of T is denoted by $Fix(T) = F(T) := \{x \in C : Tx = x\}$ and is called the *fixed point set* of T. Let $\{T_n\}$ and Ω be families of nonexpansive operators of C into C such that $\varnothing \neq F(\Omega) \subset \Gamma := \bigcap_{n=1}^{\infty} F(T_n)$, where $F(\Omega)$ is the set of all common fixed points of Ω, and let $\omega_\omega(x_n)$ denote the set of all weak-cluster point of a bounded sequence $\{x_n\}$ in C. A sequence $\{T_n\}$ is said to satisfy the *NST-condition(I)* with Ω [15, 7], if for every bounded sequence $\{x_n\}$ in C,

$$\lim_{n\to\infty} \|x_n - T_n x_n\| = 0 \ \text{ implies } \ \lim_{n\to\infty} \|x_n - T x_n\| \ \text{ for all } \ T \in \Omega.$$

If Ω is singleton, i.e., $\Omega = \{T\}$, then $\{T_n\}$ is siad to satisfy the *NST-condition(I)* with T. After that, Nakajo et al. [8] presented the NST*-condition which is more general than that of NST-condition. A sequence $\{T_n\}$ is siad to satisfy the *NST*-condition* if for every bounded sequence $\{x_n\}$ in C,

$$\lim_{n\to\infty} \|x_n - T_n x_n\| = \lim_{n\to\infty} \|x_n - x_{n+1}\| = 0 \text{ implies } \omega_\omega(x_n) \subset \Gamma.$$

Theorem 2.2 (**Banach Fixed Point Theorem**). *Let C be a nonempty closed subset of H. If $T : C \to C$ is a contraction, then T has a unique fixed point $u \in C$.*

Proposition 2.5 ([25]). *Let C be a nonempty closed convex subset of H and P_C the metric projection from H onto C. Then the following hold:*

(a) $P_C \in C$ for all $x \in H$,
(b) P_C is firmly nonexpansive: $\langle x - y, P_C x - P_C y\rangle \geq \|P_C x - P_C y\|^2$ for all $x, y \in H$,
(c) P_C is nonexpansive: $\|P_C x - P_C y\| \leqslant \|x - y\|$ for all $x, y \in H$,
(d) P_C is monotone: $\langle P_C x - P_C y, x - y\rangle \geq 0$ for all $x, y \in H$,
(e) P_C is demiclosed: $x_n \rightharpoonup x_0$ and $P_C x_n \to y \Rightarrow P_C x_0 = y_0$.

Lemma 2.6 ([18]). *For a real Hilbert space H, let $g : H \to \mathbb{R} \cup \{\infty\}$ be a 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 h 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 h with respect to $c \in (0, 2/L)$.*

Proposition 2.7 ([12]). *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*

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

Lemma 2.8 ([14],[17]). *Let H be a real Hilbert space. Then the following results hold:*

(i) for all $t \in [0, 1]$ and $x, y \in H$,

$$\|tx + (1-t)y\|^2 = t\|x\|^2 + (1-t)\|y\|^2 - t(1-t)\|x - y\|^2\,;$$

(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 2.8 (i) implies that the following equality holds:

$$\|\alpha x + \beta y + \gamma z\|^2 = \alpha\|x\|^2 + \beta\|y\|^2 + \gamma\|z\|^2 - \alpha\beta\|x - y\|^2 - \beta\gamma\|y - z\|^2 - \alpha\gamma\|x - z\|^2\,,$$

for all $x, y, z \in H$ and $\alpha, \beta, \gamma \in [0, 1]$ with $\alpha + \beta + \gamma = 1$. (2.1)

Lemma 2.9 ([21]). *Let $\{a_n\}$, $\{b_n\}$ and $\{\gamma_n\}$ be sequences of nonnegative real numbers such that*

$$a_{n+1} \leq (1 + \gamma_n)a_n + b_n,\ n \in \mathbb{N}.$$

If $\sum_{n=1}^{\infty} \gamma_n < \infty$ and $\sum_{n=1}^{\infty} b_n < \infty$, then $\lim_{n\longrightarrow\infty} a_n$ exists.

Lemma 2.10 ([16]). *Let H be a Hilbert space and $\{x_n\}$ be sequences in H such that there exists a nonempty set $\Gamma \subset H$ satisfying*

(i) for every $p \in \Gamma$, $\lim_{n\longrightarrow\infty} \|x_n - p\|$ exists;

(ii) each weak-cluster point of the sequence $\{x_n\}$ is in Γ.

Then there exists $x^* \in \Gamma$ such that $\{x_n\}$ weakly converges to x^*.

Proposition 2.11 ([18]). *Let H be a 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 2.12 ([19]). *Let H be a Hilbert space and $T : H \to H$ a nonexpansive mapping with $F(T) \neq \varnothing$. 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 2.13 ([20],[22]). *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:

(i) $\sum_{n=1}^{\infty} \delta_n = \infty$;
(ii) $\sum_{n=1}^{\infty} \xi_n < \infty$;
(iii) $\limsup_{n\longrightarrow\infty} t_n \leq 0$.

Then $\lim s_n = 0$.

Lemma 2.14 ([23]). *Let $\{\Phi_n\}$ be a sequence of real numbers that does not decrease at infinity in the sense that there exists a subsequence $\{\Phi_{n_j}\}$ of $\{\Phi_n\}$ which satisfies $\Phi_{n_j} < \Phi_{n_j+1}$ for all $i \in \mathbb{N}$. Define the sequence $\{\psi(n)\}_{n\geq n_0}$ of integers as follows:*

$$\psi(n) := \max \{k \leq n : \Phi_k < \Phi_{k+1}\},$$

where $n_0 \in \mathbb{N}$ such that $\{k \leq n_0 : \Phi_k < \Phi_{k+1}\} \neq \varnothing$. Then the following hold:

(i) $\psi(n_0) \leq \psi(n_0 + 1) \leq \cdots$ *and* $\psi(n) \longrightarrow \infty$;
(ii) $\Phi_{\psi(n)} \leq \Phi_{\psi(n)+1}$ *and* $\Phi_n \leq \Phi_{\psi(n)+1}$ *for all* $n \geq n_0$.

3. Main Results

In this section, we first introduces a new algorithm for finding a common fixed point of a family of nonexpansive mappings in a real Hilbert space and then prove its strong convergence under some suitable conditions.

Here we propose a new accelerated algorithm for approximating the solution of a common fixed point problem:

Let H be a real Hilbert space. 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 $\{\alpha_n\}, \{\beta_n\}, \{\gamma_n\} \subset (0, 1)$.

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

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

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

Algorithm 1 :

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

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

Compute

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

then the sequence $\{x_n\}$ *converges strongly to* $u \in F(T) \subset \bigcap_{n=1}^{\infty} F(T_n)$ *where* $u = P_{F(T)} f(u)$

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

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

and

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

From (3.1) and (3.2), we have

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

for all $n \geq 1$

As specified by 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

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

From (3.3), we obtain

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

where $\rho = sup\left\{\frac{(1-(1-k)\alpha_n\gamma_n)}{\gamma_n} : n \in \mathbb{N}\right\}$. This implies that the sequence $\{x_n\}$ is bounded, so are $\{y_n\}$, $\{z_n\}$, $\{f(y_n)\}$ and $\{T_n y_n\}$.
On the other hand, we have

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

By Lemma 2.8 (i), (3.1) and (3.2) we have

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

$$
\begin{aligned}
&\quad + 2\alpha_n\gamma_n \langle f(u) - u, x_{n+1} - u\rangle \\
&= (1-(1-k)\alpha_n\gamma_n) \|y_n - u\|^2 + 2\alpha_n\gamma_n \langle f(u) - u, x_{n+1} - u\rangle \\
&\leq (1-(1-k)\alpha_n\gamma_n) \left(\|x_n - u\|^2 + 2\theta_n \|x_n - u\| \|x_n - x_{n-1}\| + \theta_n^2 \|x_n - x_{n-1}\|^2\right) \\
&\quad + 2\alpha_n\gamma_n \langle f(u) - u, x_{n+1} - u\rangle \\
&\leq (1-(1-k)\alpha_n\gamma_n) \|x_n - u\|^2 + (1-(1-k)\alpha_n\gamma_n) \left[\theta_n \|x_n - x_{n-1}\| (2\|x_n - u\|\right. \\
&\quad \left. + \theta_n \|x_n - x_{n-1}\|\right] + 2\alpha_n\gamma_n \langle f(u) - u, x_{n+1} - u\rangle .
\end{aligned}
$$

Since

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

there exists positive constant $M_2 > 0$ such that

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

From the inequality (3.5), we get that for $n \in \mathbb{N}$,

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

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

$$
t_n := \frac{3M_3\rho}{(1-k)} \cdot \frac{\theta_n}{\alpha_n} \|x_n - x_{n-1}\| + \frac{2}{1-k} \langle f(u) - u, x_{n+1} - u\rangle, \quad \forall n \geq 1
$$

then, we obtain

$$
s_{n+1} \leq (1-\delta_n)s_n + \delta_n t_n, \quad \forall n \geq 1. \qquad (3.7)
$$

Now, we consider the following two cases.

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

$$
\limsup_{n\longrightarrow\infty} \langle f(u) - u, x_{n+1} - u\rangle \leqslant 0.
$$

Coming back to the definition of x_{n+1}, by Lemma 2.8 (i) and (3.4), one has that

$$
\begin{aligned}
\|x_{n+1} - u\|^2 &= \|\alpha_n z_n + \beta_n T_n y_n + (1-\alpha_n-\beta_n)y_n - u\|^2 \\
&\leq \alpha_n \|z_n - u\|^2 + \beta_n \|T_n y_n - u\|^2 + (1-\alpha_n-\beta_n) \|y_n - u\|^2
\end{aligned}
$$

$$\begin{aligned}
&- \beta_n(1-\alpha_n-\beta_n)\|y_n - T_n y_n\|^2 \\
&\le \alpha_n \|z_n - u\|^2 + \beta_n \|y_n - u\|^2 + (1-\alpha_n-\beta_n)\|y_n - u\|^2 \\
&\quad - \beta_n(1-\alpha_n-\beta_n)\|y_n - T_n y_n\|^2 \\
&\le \alpha_n \|z_n - u\|^2 + (1-\alpha_n)\|y_n - u\|^2 \\
&\quad - \beta_n(1-\alpha_n-\beta_n)\|y_n - T_n y_n\|^2 \\
&\le \alpha_n \left[\|z_n - u\|^2 - \|y_n - u\|^2\right] + \|x_n - u\|^2 \\
&\quad + 2\theta_n \|x_n - u\| \|x_n - x_{n-1}\| \\
&\quad + \theta_n^2 \|x_n - x_{n-1}\|^2 - \beta_n(1-\alpha_n-\beta_n)\|y_n - T_n y_n\|^2
\end{aligned} \tag{3.8}$$

It implies that for all, $n \in \mathbb{N}$,

$$\begin{aligned}
\beta_n(1-\alpha_n-\beta_n)\|y_n - T_n y_n\|^2 \le \alpha_n &\left[\|z_n - u\|^2 - \|y_n - u\|^2\right] \\
&+ \|x_n - u\|^2 - \|x_{n+1} - u\|^2 \\
&+ \theta_n \|x_n - x_{n-1}\| \left(2\|x_n - u\| + \theta_n \|x_n - x_{n-1}\|\right).
\end{aligned} \tag{3.9}$$

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

$$\|y_n - T_n y_n\| \longrightarrow 0 \text{ as } n \longrightarrow \infty. \tag{3.10}$$

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

$$\|y_n - T y_n\| \longrightarrow 0 \text{ as } n \longrightarrow \infty. \tag{3.11}$$

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

$$\begin{aligned}
\|x_{n+1} - y_n\| &= \|\alpha_n(z_n - y_n) + \beta_n(T_n y_n - y_n)\| \\
&\le \alpha_n \|z_n - y_n\| + \beta_n \|T_n y_n - y_n\| \\
&= \alpha_n \|\gamma_n(f(y_n) - y_n) + (1-\gamma_n)(T_n y_n - y_n)\| + \beta_n \|T_n y_n - y_n\| \\
&\le \alpha_n\gamma_n \|f(y_n) - y_n\| + (1-\gamma_n)\alpha_n \|T_n y_n - y_n)\| + \beta_n \|T_n y_n - y_n\|.
\end{aligned} \tag{3.12}$$

This implies by (3.11) and $\gamma_n \to 0$ that

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

By the definition of y_n, we obtain

$$\|y_n - x_n\| = \theta_n \|x_n - x_{n-1}\| \to 0 \text{ as } n \longrightarrow \infty. \tag{3.13}$$

Hence

$$\|x_{n+1} - x_n\| \le \|x_{n+1} - y_n\| + \|y_n - x_n\| \longrightarrow 0 \text{ as } n \longrightarrow \infty. \tag{3.14}$$

Let

$$v = \limsup_{n \longrightarrow \infty} \langle f(u) - u, x_{n+1} - u \rangle. \tag{3.15}$$

So, there exists a subsequence $\{x_t\}$ of $\{x_n\}$ such that

$$v = \limsup_{t \longrightarrow \infty} \langle f(u) - u, x_{t+1} - u \rangle. \tag{3.16}$$

Since $\{x_t\}$ is bounded, there exists a subsequence $\{x_t'\}$ of $\{x_t\}$ such that $x_t' \rightharpoonup w \in H$. Without loss of generality, we may assume that $x_t \rightharpoonup w$.

$$v = \limsup_{t \longrightarrow \infty} \langle f(u) - u, x_{t+1} - u \rangle.$$

From (3.11) and (3.13), we derive

$$\begin{aligned} \|x_n - Tx_n\| &\leq \|x_n - y_n\| + \|y_n - Ty_n\| + \|Ty_n - Tx_n\| \\ &\leq 2\|x_n - y_n\| + \|y_n - Ty_n\| \to 0, \text{as } \ n \to \infty. \end{aligned} \tag{3.17}$$

It implies by Lemma (2.12) that $w \in F(T)$. Since $\|x_{n+1} - x_n\| \longrightarrow 0$, we get $x_{t+1} \rightharpoonup w$. Moreover, using $u = P_{F(T)}f(u)$ and Proposition 2.2 , we obtain

$$v = \limsup_{t \longrightarrow \infty} \langle f(u) - u, x_{t+1} - u \rangle = \langle f(u) - u, w - u \rangle \leq 0. \tag{3.18}$$

Hence

$$v = \limsup_{n \longrightarrow \infty} \langle f(u) - u, x_{n+1} - u \rangle \leq 0. \tag{3.19}$$

This implies by (3.19) and using 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 (3.6) and using Lemma 2.13, we obtain that $x_n \longrightarrow u$.

Case 2. Suppose the sequence $\{\|x_n - u\|\}_{n \geq n_0}$ is not a monotonically decreasing sequence for a a sufficiently large n_0. We set

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

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

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

By Lemma 2.14, we have that $\Phi_{\psi(n)} \leq \Phi_{\psi(n)+1}$ for all $n \geq n_0$. That is

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

As we know from Case 1, we obtain that for all $n \geq n_0$,

$$\begin{aligned} \beta_{\psi(n)}(1 - \alpha_{\psi(n)} - \beta_{\psi(n)}) \left\|y_{\psi(n)} - T_{\psi(n)}y_{\psi(n)}\right\|^2 &\leq \alpha_{\psi(n)}\left[\left\|z_{\psi(n)} - u\right\|^2 - \left\|y_{\psi(n)} - u\right\|^2\right] \\ &\quad + \left\|x_{\psi(n)} - u\right\|^2 - \left\|x_{\psi(n)+1} - u\right\|^2 \\ &\quad + \theta_{\psi(n)}\left\|x_{\psi(n)} - x_{\psi(n)-1}\right\| \\ &\quad \times \left(2\left\|x_{\psi(n)} - u\right\|\right. \\ &\quad \left. + \theta_{\psi(n)}\left\|x_{\psi(n)} - x_{\psi(n)-1}\right\|\right) \\ &\leq \alpha_{\psi(n)}\left[\left\|z_{\psi(n)} - u\right\|^2 - \left\|y_{\psi(n)} - u\right\|^2\right] \\ &\quad + \theta_{\psi(n)}\left\|x_{\psi(n)} - x_{\psi(n)-1}\right\| \\ &\quad \times \left(2\left\|x_{\psi(n)} - u\right\|\right. \\ &\quad \left. + \theta_{\psi(n)}\left\|x_{\psi(n)} - x_{\psi(n)-1}\right\|\right) \end{aligned}$$

which implies

$$\left\|y_{\psi(n)} - T_{\psi(n)}y_{\psi(n)}\right\| \longrightarrow 0 \ \text{ as } \ n \longrightarrow \infty. \tag{3.20}$$

Similar to the proof in Case 1, we have

$$\left\|x_{\psi(n)+1} - y_{\psi(n)}\right\| \longrightarrow 0 \ \text{ as } \ n \longrightarrow \infty \tag{3.21}$$

and

$$\left\|y_{\psi(n)} - x_{\psi(n)}\right\| \longrightarrow 0 \ \text{ as } \ n \longrightarrow \infty. \tag{3.22}$$

Hence

$$\left\|x_{\psi(n)+1} - x_{\psi(n)}\right\| \longrightarrow 0 \ \text{ as } \ n \longrightarrow \infty. \tag{3.23}$$

Next, we show $\limsup_{n\longrightarrow\infty}\left\langle f(u)-u, x_{\psi(n)+1}-u\right\rangle \le 0$. Put

$$\upsilon = \limsup_{n\longrightarrow\infty}\left\langle f(u)-u, x_{\psi(n)+1}-u\right\rangle$$

Without loss of generality, there exists a subsequence $\left\{x_{\psi(t)}\right\}$ of $\left\{x_{\psi(n)}\right\}$ such that $\left\{x_{\psi(t)}\right\}$ converges weakly to some point $w \in H$ and

$$\upsilon = \lim_{t\longrightarrow\infty}\left\langle f(u)-u, x_{\psi(t)+1}-u\right\rangle .$$

By Lemma 2.12, one has $\left\{T_{\psi(t)}\right\}$ satisfies NST-condition(I) with T, so according to the equality (3.20), $\left\|y_{\psi(t)}-T_{\psi(t)}y_{\psi(t)}\right\| \longrightarrow 0$ as $n \longrightarrow \infty$, we obtain that

$$\left\|y_{\psi(t)}-Ty_{\psi(t)}\right\| \longrightarrow 0 \text{ as } n \longrightarrow \infty \tag{3.24}$$

which implies, by (3.22) and Lemma 2.12 again, that $w \in F(T)$. Since $\left\|x_{\psi(t)+1}-x_{\psi(t)}\right\| \to 0$, we get $x_{\psi(t)+1} \rightharpoonup w$. Further, using $u = P_{F(T)}f(u)$ and Proposition 2.2, we get

$$\upsilon = \lim_{t\longrightarrow\infty}\left\langle f(u)-u, x_{\psi(t)+1}-u\right\rangle = \langle f(u)-u, w-u\rangle \le 0 \tag{3.25}$$

Then

$$\upsilon = \limsup_{t\longrightarrow\infty}\left\langle f(u)-u, x_{\psi(t)+1}-u\right\rangle \le 0 \tag{3.26}$$

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

$$\left\|x_{\psi(t)}-u_0\right\|^2 \le \frac{3M_3\sigma}{(1-k)}\cdot\frac{\theta_n}{\alpha_n}\left\|x_{\psi(t)}-x_{\psi(t)-1}\right\| + \frac{2}{1-k}\left\langle f(u)-u, x_{\psi(t)+1}-u\right\rangle \tag{3.27}$$

In fact we have that $\frac{\theta_n}{\alpha_n}\|x_n - x_n - 1\| \longrightarrow 0$ and (3.26)

$$\limsup_{n\longrightarrow\infty}\left\|x_{\psi(n)}-u\right\|^2 \le 0$$

so we get $\left\|x_{\psi(n)}-u\right\| \longrightarrow 0$ as $n \longrightarrow \infty$.
This implies by (3.23) that $\left\|x_{\psi(n)+1}-u\right\| \longrightarrow 0$ as $n \longrightarrow \infty$.
By Lemma 2.14, we get

$$\|x_n - u\| \le \left\|x_{\psi(n)+1}-u\right\| \longrightarrow 0 \text{ as } n \longrightarrow \infty.$$

Hence $x_n \longrightarrow u$. The proof is completed. □

We now consider the following bi-level convex minimization problem:

$$\min_{x\in X^*}\omega(x), \tag{3.28}$$

where X^* is the optimal solution set of problem (3.29). We let Ω be the set of all solutions of (3.28). For the objective function ω of problem (3.28) we make the following assumption.

Assumption 1.

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

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

For the problem

$$X^* = arc\min_{x\in\mathbb{R}^n}(f(x)+g(x)) \tag{3.29}$$

we assume the following assumption:

Assumption 2.

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

Now, we are ready to introduce an algorithm for solving problem (3.28)

Algorithm 2 :

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

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

Compute

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

We obtain the following result as a consequence of Theorem 3.1

Theorem 3.2. *Let $\omega : \mathbb{R}^n \to \mathbb{R}$ be a function satisfying the assumption 1. Let $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to (-\infty, \infty]$ be 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 $c_n \longrightarrow c$ as $n \longrightarrow \infty$. Then the sequence $\{x_n\}$ generated by Algorithm 2 with the same conditions as in Theorem 3.1 converges strongly to $u \in \Omega$*

Proof. Put $T_n = prox_{c_n g}(I - c_n \nabla f), n \in \mathbb{N}$, and $T = prox_{cg}(I - c\nabla f)$. By Proposition 2.11 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 $u \in F(T) = arc\min\limits_{x \in \mathbb{R}^n} (f(x) + g(x))$ such that $u = P_{F(T)} f(u)$. By Proposition 2.7 $f := I - s\nabla\omega(x)$ is a contraction with parameter $k = \sqrt{1 - \frac{2s\sigma L_\omega}{\sigma + L_\omega}}$. It remains to show that $n \in \Omega$. By using $u = P_{F(T)} f(u)$ and Proposition 2.2, we obtain

$$\begin{aligned} u = P_{F(T)} f(u) &\Leftrightarrow \langle f(u) - u, z - u \rangle \leq 0, \quad \forall z \in F(T) \\ &\Leftrightarrow \langle u - s\nabla\omega(u) - u, z - u \rangle \leq 0, \quad \forall z \in F(T) \\ &\Leftrightarrow \langle -s\nabla\omega(u), z - u \rangle \leq 0, \quad \forall z \in F(T) \\ &\Leftrightarrow \langle s\nabla\omega(u), z - u \rangle \geq 0, \quad \forall z \in F(T) \\ &\Leftrightarrow \langle \nabla\omega(u), z - u \rangle \geq 0, \quad \forall z \in F(T) = X^* \end{aligned}$$

Hence, u is the optimal solution for the problem (3.28). Therefore, $x_n \longrightarrow u \in \Omega$. □

3.1. **Numerical Results.** In this section, we apply our algorithms, Forward-backward Algorithm, FISTA and NAGA to solve some classification problems based on the method proposed by Huang et al [24], which is called extreme learning machine (ELM). It is formulated as follows:

Let $\{(x_k, t_k) : x_k \in \mathbb{R}^n, t_k \in \mathbb{R}^m, k = 1, 2, ..., N\}$ be as set of N samples where x_k is an *input* and t_k is a *target*. A simple mathematical model for the output of ELM for SLFNs with M hidden nodes and activation function G is defined by

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

where w_i is the weight that connects 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. The hidden layer output matrix **H** is defined by

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

The main objective of ELM is to calculate an optimal weight $\eta = [\eta_1^T, \cdots, \eta_M^T]^T$ such that $\mathbf{H}\eta = \mathbf{T}$, where $\mathbf{T} = [t_1^T, \cdots, t_N^T]^T$ is the training target.

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:

$$\textbf{Minimize:} \|\mathbf{H}\eta - \mathbf{T}\|_2^2 + \lambda\|\eta\|_1, \tag{3.30}$$

where λ is a regularization parameter.

If we set $f(x) := \|\mathbf{H}\eta - \mathbf{T}\|_2^2$ and $g(x) := \lambda\|\eta\|_1$, then we know that $\nabla f(x) = 2\mathbf{H}^T(\mathbf{H}x - \mathbf{T})$ and Lipschitz constant of $\nabla f$is $L = 2\|\mathbf{H}\|^2$.

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

Following we consider two data sets:

(i) **Iris data set :** Each sample in this data set has 4 attributes, and the set contains 3 classes with 50 samples for each type.

(ii) **Heart disease data set :** This data set contains 303 samples each of which has 13 attributes and 2 classes of data.

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

Algorithms :

(i) **Our Algorithm** (Algorithm 3)

(ii) **Forward-backward Algorithm** (Algorithm 4)

(iii) **FISTA** (Algorithm 5)

(iv) **NAGA** (Algorithm 6)

Algorithm 3 :

1: **Input** $x_0, x_1 \in \mathbb{R}^n, \mu_n, \eta_n, \gamma_n \in (0, \infty), \rho_n \in (0, \frac{2}{L})$ and $\alpha_n, \beta_n, \gamma_n \in (0, 1)$, for $n \in \mathbb{N}$,

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

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

$$z_n = \gamma_n h(y_n) + (1 - \gamma_n) prox_{\rho_n g}(y_n - \rho_n \nabla f(y_n)),$$

$$x_{n+1} = (1 - \alpha_n - \beta_n) y_n + \alpha_n z_n + \beta_n prox_{c_n g}(I - c_n \nabla f) y_n,$$

Algorithm 4 Forward-backward Algorithm

1: **Input** $x_0 \in \mathbb{R}^n, \rho_n \in (0, \frac{2}{L})$, for $n \in \mathbb{N}$,

$$x_{n+1} = prox_{\rho_n g}(x_n - \rho_n \nabla f(x_n)),$$

Algorithm 5 FISTA

1: **Input** $y_1 = x_0 \in \mathbb{R}^n$, and $t_1 = 1$, for $n \in \mathbb{N}$,

$$y_n = prox_{\frac{1}{L} g}(x_n - \frac{1}{L} \nabla f(x_n)),$$

$$t_{n+1} = \frac{1 + \sqrt{1 + 4t_n^2}}{2}, \quad \theta_n = \frac{t_n - 1}{t_{n+1}},$$

$$x_{n+1} = y_n + \theta_n (y_n - y_{n-1}),$$

Algorithm 6 NAGA

1: **Input** $x_0, x_1 \in \mathbb{R}^n, \theta_n \geq 0, \alpha_n \in (0,1), \rho_n \in (0, \frac{2}{L})$, for $n \in \mathbb{N}$,

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

$$z_n = (1 - \alpha_n) y_n + \alpha_n prox_{\rho_n g}(y_n - \rho_n \nabla f(y_n)),$$

$$x_{n+1} = prox_{\rho_n g}(z_n - \rho_n \nabla f(z_n)),$$

- **Chosen parameters of each algorithm :**
  - Contraction mapping : $h(x) = 0.9x$
  - Regularization parameter : $\lambda = 0.0333$
  - Hidden nodes : m $= 30$
  - $n = 5000$ and $\rho_n = \frac{1}{L}$

$$\textbf{Algorithm 3}: \quad \theta_n = \begin{cases} \min\{0.9, \dfrac{10^8}{n^3\|x_n - x_{n-1}}\|\}, & x_n \neq x_{n-1}, \\ 0.9, & \text{otherwise.} \end{cases}$$

$$\alpha_n = \frac{1}{3n}, \ \beta_n = \frac{n}{3n+1}, \ \gamma_n = 0.5 + \frac{1}{n}.$$

$$\textbf{NAGA}: \quad \theta_n = 0.9, \ \alpha_n = \frac{9n}{10(n+1)}.$$

TABLE 1. The performance of each algorithm at 5000th iteration with 10-fold cv. on Iris data set

	Algorithm 3		Algorithm 4		Algorithm 5		Algorithm 6	
	acc.train	acc.test	acc.train	acc.test	acc.train	acc.test	acc.train	acc.test
Fold 1	99.26	100	96.30	100	97.78	100	98.52	100
Fold 2	98.52	100	97.04	93.33	97.78	100	97.78	100
Fold 3	99.26	100	96.30	100	97.78	93.33	97.78	93.33
Fold 4	99.26	100	95.56	100	97.78	100	97.78	100
Fold 5	99.26	100	96.30	100	98.52	93.33	99.26	93.33
Fold 6	100	93.33	97.78	86.67	99.26	93.33	100	93.33
Fold 7	99.26	100	97.04	100	97.78	100	97.78	100
Fold 8	97.78	93.33	95.56	86.67	97.78	100	97.78	100
Fold 9	99.26	100	96.30	86.67	98.52	93.33	98.52	93.33
Fold 10	99.26	100	96.30	100	97.78	100	98.52	100
Average acc.	99.11	98.67	96.44	95.33	98.07	97.33	98.37	97.33
Time	0.1750		0.0986		0.1055		0.1702	

We observe from Table 1 that Algorithm 3 has the highest accuracy. It performs better than the other three algorithms.

Next we focus on bi-level minimization problem

$$\min_{\omega \in X^*} \omega(\beta)$$

where $X^* = arc \min_{x \in \mathbb{R}^n} (f(x) + g(x))$ with $f(\beta) = \min \ \|\mathbf{H}\beta - \mathbf{T}\|_2^2$ and $g(\beta) = \lambda \|\beta\|_1$.

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

In our problem, we are interested to use $\omega(\beta) = \frac{1}{2}\|\beta\|_2^2$ and we aim to compare performance of our algorithm (Algorithm 2) and BiG-SAM (Algorithm 7)

Algorithm 7 BiG-SAM

1: **Input** $x_0 \in \mathbb{R}^n, \gamma_n \in (0,1), \alpha \in (0, \frac{1}{L_f}]$ and $s \in (0, \frac{2}{L_\omega + \kappa})$, for $n \in \mathbb{N}$,

$$y_n = prox_{\alpha g}(x_n - \alpha \nabla f(x_n)),$$
$$x_{n+1} = \gamma_n (x_k - s\nabla\omega(x_k)) + (1-\gamma_n) y_k.$$

- **Chosen parameters of each algorithm :**

  - Regularization parameter : $\lambda = 0.0333$
  - Hidden nodes : m = 30
  - $n = 300$, $\rho_n = \frac{1}{L_f}$ and $s = \frac{1}{5}$

$$\textbf{Algorithm 2}: \quad \theta_n = \begin{cases} \min\{0.9, \ \dfrac{10^8}{n^3\|x_n - x_{n-1}}\|\}, & x_n \neq x_{n-1}, \\ 0.9, & \text{otherwise.} \end{cases}$$

$$\alpha_n = \frac{1}{3n}, \ \beta_n = \frac{n}{3n+1}, \ \gamma_n = \frac{1}{n} + 0.5.$$

$$\textbf{BiG-SAM}: \quad \alpha = \frac{1}{L_f}, \ \gamma_n = \frac{1}{n}.$$

TABLE 2. The performance of each algorithm at 300th iteration with 10-fold cv. on Iris data set

	Algorithm 2		Algorithm 7	
	acc.train	acc.test	acc.train	acc.test
Fold 1	88.89	86.67	80.74	80
Fold 2	89.63	93.33	80.74	80
Fold 3	88.15	100	81.48	86.67
Fold 4	88.15	100	80.74	86.67
Fold 5	87.41	86.67	80.74	80
Fold 6	88.89	73.33	80.74	73.33
Fold 7	90.37	86.67	75.56	86.67
Fold 8	90.37	86.67	77.78	80
Fold 9	89.63	80.00	80.74	73.33
Fold 10	88.89	80.00	81.48	73.33
Average acc.	89.04	88.67	80.07	80
Time	0.0079		0.0032	

TABLE 3. The performance of each algorithm at 300th iteration with 10-fold cv. on Heart disease data set

	Algorithm 2		Algorithm 7	
	acc.train	acc.test	acc.train	acc.test
Fold 1	81.68	90.00	79.49	86.67
Fold 2	81.62	83.87	80.15	77.42
Fold 3	81.99	80.65	80.51	77.42
Fold 4	83.09	83.87	81.25	80.65
Fold 5	81.32	90.00	79.85	83.33
Fold 6	82.05	83.33	79.85	76.67
Fold 7	81.68	86.67	79.49	86.67
Fold 8	83.15	66.67	80.95	66.67
Fold 9	82.78	70.00	81.32	70.00
Fold 10	82.42	83.33	80.22	83.33
Average acc.	82.18	81.84	80.31	78.88
Time	0.0070		0.0033	

In Table 2 and Table 3, we compare accuracy of Algorithm 2 with Algorithm 7 for different data sets. For Iris data set we achieve a testing accuracy of 89.04 and for Heart disease data set we achieve a testing accuracy of 82.18. In both cases, our proposed algorithm (Algorithm 2) has a better accuracy than Algorithm 7.

4. CONCLUSION

We introduced a new accelerated fixed point algorithm to find a common fixed point of a family of nonexpansive mappings in a real Hilbert space. First, we prove a strong convergence in Algorithm 1. Next, we prove strong convergence theorems in Algorithm 2. We applied our algorithm to solve the regression and classification problems. From our study, we obtained highest performance than the other methods shown in Section 4.

5. Acknowledgements

References

1. W. R. Mann, Mean value methods in iteration, Proc. Amer. Math. Soc. 4(1953),506{510.
2. B. Halpern, Fixed points of nonexpanding maps. Bull. Amer. Math. Soc. 73(1967), 957{961.
3. A. Moudafi, Viscosity approximation methods for fixed points problems, J. Math. Anal. Appl. 241(2000),46{55.
4. S. Ishikawa, Fixed points by a new iteration method, Proc. Amer. Math. Sot. 44 (1974), 147-150.
5. R.P. Agarwal, D. O'Regan, and D. R. Sahu. Iterative construction of fixed point of nearly asymptotically nonexpansive mappings. J. Nonlinear Convex Anal., 8(1) (2007), 61{79.
6. K. Aoyama, Y. Kimura, W. Takahashi and M. Toyada, Approximation of common fixed points of a countable family of nonexpansive mappings in a Banach space, Nonlinear Anal. 67 (2007), 2350{2360.
7. W. Takahashi, Viscosity approximation methods for countable family of nonexpansive mappings in Banach spaces, Nonlinear Anal. 70 (2009), 719{734.
8. C. Klin-eam and S. Suantai, Strong convergence of composite iterative schemes for a countable family of nonexpansive mappings in Banach spaces, Nonlinear Analysis, 73 (2010), 431-439.
9. B. Polyak, Some methods of speeding up the convergence of iteration methods. USSR Computational Mathematics and Mathematical Physics. 4(5) (1964), 1 { 17.
10. A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences. 2(1) (2009), 183{202.
11. M. Vermaand and K. Shukla, A new accelerated proximal gradient technique for regularized multitask learning framework. Pattern Recognition Letters. 95 (2017), 98 { 103.
12. S. Sabach and S. Shtern, A first order method for solving convex bi-level optimization problems. Accepted in SIAM Journal on Optimization (2017).
13. H.-K. Xu. Viscosity approximation methods for nonexpansive mappings. J. Math. Anal. Appl., 298(1):279{291, 2004.
14. W. Takahashi, Introduction to nonlinear and convex analysis, Yokohama Publishers, Yokohama (2009).
15. K. Nakajo, K. Shimoji and W. Takahashi, Strong convergence to common fixed points of families of nonexpansive mappings in Banach spaces, J. nonlinear Convex Anal. 8(2007), pp. 11-34.
16. A. Moudafi and E. Al-Shemas, Simultaneous iterative methods for split equality problem, Trans. Math. Program. Appl. 1 (2013), pp. 1-11.
17. W. Takahashi, Nonlinear Functional Analysis; Yokohama Publishers: Yokohama, Japan, 2000.
18. L. Bussaban, S. Suantai and A. Kaewkhao, A parallel inertial S-iteration forward-backward algorithm for regression and classification problems. Carpathian. J. 2020, 36, 35{44.
19. K. Goebel; W.A. Kirk, Topics in Metric Fixed Point Theory; Cambridge Studies in Advanced Mathematics; Cambridge University Press: Cambridge, UK, 1990; Volume 28.
20. K. Aoyama ; Y. Yasunori ; W. Takahashi ; M. Toyoda, On a strongly nonexpansive sequence in a Hilbert space. J. Nonlinear Convex Anal. 2007, 8, 471{489.
21. K. Tan and H.K. Xu, Approximating fixed points of nonexpansive mappings by the Ishikawa iteration process, J. Math. Anal. Appl., 178(2) (1993), pp. 301-308.
22. Xu, H.K. Another control condition in an iterative method for nonexpansive mappings. Bull. Aust. Math. Soc.
23. P.E. Mainge, Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization. Set-Valued Anal. 2008, 16, 899{912.2002, 65, 109{113.
24. G.B. Huang ; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. Neurocomputing 2006, 70, 489{501.
25. R. P. Agarwal, D. O'Regan, D. R. Sahu, Fixed point theory for Lipschitzian-type mappings with applications, Topological Fixed Point Theory and Its Applications, Springer, New York, (2009). 2, 2.4