CARPATHIAN J. MATH.
Volume **40** (2024), No. 2,
Pages 477 - 491

Online version at https://www.carpathian.cunbm.utcluj.ro/
Print Edition: ISSN 1584 - 2851; Online Edition: ISSN 1843 - 4401
DOI: https://doi.org/10.37193/CJM.2024.02.16

A New Viscosity Approximation Method with Inertial Technique for Convex Bilevel Optimization Problems and Applications

PITI THONGSRI and SUTHEP SUANTAI

ABSTRACT. This paper presents and analyzes a new viscosity approximation method with the inertial technique for finding a common fixed point of a countable family of nonexpansive mappings and then its strong convergence theorem is established under some suitable conditions. As a consequence, we employ our proposed algorithm for solving some convex bilevel optimization problems and then apply it for solving regression of a graph of cosine function and classification of some noncommunicable diseases by using the extreme learning machine model. We perform a comparative analysis with other algorithms to demonstrate the performance of our approach. Our numerical experiments confirm that our proposed algorithm outperforms other methods in the literature.

1. INTRODUCTION

Bilevel optimization is a type of mathematical optimization problem where one optimization problem is contained within another optimization problem. The solution to the outer problem is dependent on the solution to the inner problem. The difficulty lies in finding the optimal solution to both the leader and the follower problem simultaneously. Bilevel optimization plays an important role in a variety of real-world applications, including resource allocation in supply chain management, machine learning models for regression and classification of some noncommunicable diseases, pricing strategies in economics, optimization of power systems and optimization in traffic management.

In recent years, bilevel optimization has gained much attention and is one of the active research areas which can be applied in various fields. Hence, bilevel optimization is considered as an important and relevant tool in many practical applications.

Let H be a Hilbert space over the real numbers and let f, g be functions that map from H to $\mathbb{R}$. A convex bilevel minimization problem is a special type of optimization problem for which one problem is embedded within another problem. The outer level problem is the constraint minimization problem of the following form:

$$\min_{x \in X^*} h(x), \tag{1.1}$$

where $h : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function such that ∇h is Lipschitz continuous with constant L_h and strongly convex with parameter $\sigma > 0$ while X^* is the set of all minimizers of the inner level optimization problem of the following form:

$$\min_{x \in \mathbb{R}^n} \{f(x) + g(x)\}. \tag{1.2}$$

For solving problem (1.2), we normally assume the following assumption:

Received: 06.09.2023. In revised form: 04.03.2024. Accepted: 12.03.2024
2010 *Mathematics Subject Classification.* 47H09, 90C25, 65K10.
Key words and phrases. *convex bi-level optimization problems, Hilbert spaces, nonexpansive mappings, forward-backward algorithm, fixed point algorithm, regression and classification problems, noncommunicable Diseases.*
Corresponding author: Suthep Suantai; Suthep.s@cmu.ac.th

(a) $f : \mathbb{R}^n \to \mathbb{R}$ is convex and differentiable for which ∇f is L_f-Lipschitz continuous, that is,

$$\|\nabla f(x) - \nabla f(y)\| \leq L_f \|x - y\| \text{ for all } x, y \in \mathbb{R}^n;$$

(b) $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is proper convex and lower semi-continuous.

The solution to equation (1.2) can be described according to Theorem 16.3 of Bauschke and Combettes [5] as follows:

$$\overline{p} \in X^* \text{ if and only if } 0 \in \partial g(\overline{p}) + \nabla f(\overline{p}),$$

where ∇f is the gradient of f and ∂g is the subdifferential of g. On the other hand, problem (1.2) is equivalent with following fixed point problem:

$$\overline{p} \in X^* \text{ if and only if } \overline{p} = prox_{cg}(I - c\nabla f)(\overline{p}),$$

where $c > 0$ and $prox_{cg}(x) = argmin_{y \in H}(g(y) + \frac{1}{2c}\|x-y\|^2)$. The operator $prox_{cg}(I-c\nabla f)$ is called the forward-backward operator of f and g with respect to c. We also know that $prox_{cg}(I-c\nabla f)$ is a nonexpansive operator when $c \in (0, 2/L)$ and L is a Lipschitz constant of ∇f. From basic principle of optimization, we know that $\overline{p} \in X^*$ is a minimizer of problem (1.1) if and only if

$$\langle \nabla h(\overline{p}), x - \overline{p} \rangle \geq 0 \text{ for all } x \in X^*. \tag{1.3}$$

Over the last ten years, numerous researchers have been interested to find the optimal solutions for problem (1.2). A technique known as Forward-Backward Splitting (FBS) was presented by Lions and Mercier [15] as a straightforward algorithm to solve problem (1.2). Their algorithm was given by

$$x_{n+1} = prox_{c_n g}(I - c_n \nabla f)(x_n), \tag{1.4}$$

where the step-size $c_n \in (0, 2/L)$.

The concept of the inertial technique was first introduced by Polyak [19] to speed up the convergence rate of algorithms. Since then, this technique has become widely utilized for this purpose.

For example, Beck and Teboulle [6] introduced a fast iterative shrinkage-thresholding algorithm (FISTA) by using this technique for solving problem (1.2) as described by the following:

$$\begin{aligned}
&x_1 = u_0 \in C, t_1 = 1,\\
&u_n = prox_{\alpha g}(I - \alpha \nabla f)(x_n),\ \alpha > 0,\\
&t_{n+1} = \frac{\sqrt{1 + 4t_n^2} + 1}{2},\\
&\theta_n = \frac{t_n - 1}{t_{n+1}},\\
&x_{n+1} = u_n + \theta_n(u_n - u_{n-1}).
\end{aligned}$$

Furthermore, they demonstrated that FISTA exhibits superior convergence behavior compared to other methods.

Recently, several researchers, such as Jailoka et al. [14], Puangpee and Suantai [20], Thongsri et al. [27], Bussaban et al. [9], J. A. Abuchu et al. [1], and F. Akutsah et al. [2], have incorporated the inertial technique into their work. They introduced common fixed point algorithms for a countable families of nonexpansive operators and proved convergence results under the NST-condition (I), NST*-condition, and the condition (Z).

Furthermore, they successfully applied their algorithms to solve convex minimization problems.

In 2017, Sabach and Shtern [21] proposed a novel technique named Sequential Averaging Method (SAM) to solve convex bilevel optimization problems. They adapted a technique from [30] that was used to solve a specific type of fixed point problem. Later, they proposed the Bilevel Gradient Sequential Averaging Method (BiG-SAM) to solve the convex bilevel optimization problems (1.1) and (1.2). BiG-SAM was defined by Algorithm 1.

Algorithm 1 Bilevel Gradient Sequential Averaging Method (BiG-SAM)

(1)**Input** : $c \in (0, 1/L_f)$, $s \in (0, 2/(\sigma + L_h))$ and $\{\alpha_k\}_{k\in\mathbb{N}} \subset (0, 1]$.
(2)**Initialization**:choose $x_1 \in \mathbb{R}^n$.
(3) **General step**: $(k = 1, 2, ...)$:

$$\begin{aligned} v_k &= prox_{cg}(x_k - c\nabla f(x_k)), \\ u_k &= x_k - s\nabla h(x_k), \\ x_{k+1} &= \alpha_k u_k + (1-\alpha_k)v_k, \end{aligned}$$

where ∇h is the gradient of h.

Then they proved that the sequence $\{x_k\}$ generated by the BiG-SAM algorithm converges to a solution of problem 1.1 and 1.2 under some control conditions.

In 2019, Shehu et al. [23] employed an inertial technique to enhance the convergence behavior of the BiG-SAM algorithm. They introduced a new algorithm called the inertial Bilevel Gradient Sequential Averaging Method (iBiG-SAM) which was defined as Algorithm 2:

Algorithm 2 Inertial Bilevel Gradient Sequential Averaging Method (iBiG-SAM)

(1) **Input**: $\alpha \geq 3$, $c \in (0, 1/L_f)$, and $s \in (0, 2/(\sigma + L_h))$.
(2) **Initialization**: choose $x_0, x_1 \in \mathbb{R}^n$.
(3) **Step 1** For $k = 1, 2, ...,$

$$\mu_k = \begin{cases} \min\{\frac{k}{k+\alpha-1}, \frac{\gamma_k}{\|x_k - x_{k-1}\|}\}, & \text{if } x_k \neq x_{k-1}, \\ \frac{k}{k+\alpha-1}, & \text{otherwise.} \end{cases}$$

(4) **Step 2** Compute:

$$\begin{aligned} v_k &= x_k + \mu_k(x_k - x_{k-1}), \\ y_k &= prox_{cg}(v_k - c\nabla f(v_k)), \\ z_k &= v_k - s\nabla h(v_k), \\ x_{k+1} &= \alpha_k z_k + (1-\alpha_k)y_k, \end{aligned}$$

where ∇h is the gradient of h.

Very recently, a novel algorithm for solving convex bilevel optimization problems was introduced by Duan and Zhang [11]. This algorithm, called the alternated inertial Bilevel Gradient Sequential Averaging Method (aiBiG-SAM), is based on the proximal gradient algorithm and it was defined by the following iterative algorithm.

Algorithm 3 The alternated inertial Bilevel Gradient Sequential Averaging Method (aiBiG-SAM)

(1) **Input**: $\alpha \geq 3$, $c \in (0, 1/L_f)$ and $s \in (0, 2/(\sigma + L_h))$. Set $\lambda > 0$.
(2) **Initialization**: choose $x_0, x_1 \in \mathbb{R}^n$.
(3) **Step 1** For $(k = 1, 2, ...)$:

$$u_k = \begin{cases} x_k + \mu_k(x_k - x_{k-1})), & \text{if } k \text{ is odd,} \\ x_k, & \text{if } k \text{ is even.} \end{cases}$$

When k is odd, choose μ_k such that $0 \leq |\mu_k| \leq \theta_k$ where θ_k is defined by

$$\theta_k = \begin{cases} \min\{\frac{k}{k+\alpha-1}, \frac{\gamma_k}{\|x_k - x_{k-1}\|}\}, & \text{if } x_k \neq x_{k-1}, \\ \frac{k}{k+\alpha-1}, & \text{otherwise.} \end{cases}$$

(4) **Step 2** Compute:

$$y_k = prox_{cg}(u_k - c\nabla f(u_k)),$$
$$z_k = u_k - s\nabla h(u_k),$$
$$x_{k+1} = \alpha_k z_k + (1 - \alpha_k)y_k.$$

(5) **Step 3** If $\|x_k - x_{k-1}\| < \lambda$, then stop.

They also discussed strong convergence behavior of the proposed method under some conditions.

Motivated by these previous works, our objective is to introduce a more efficient algorithm for solving convex bilevel problems (1.1) and (1.2). We aim to establish a strong convergence theorem for the proposed algorithm under some suitable conditions. Furthermore, we apply this algorithm to solve classification and data prediction problems. The paper is structured as follows. In Section 2, we provide a description of the notations and useful lemmas that will be employed in subsequent sections. In Section 3, we thoroughly discuss and analyze the convergence properties of our proposed algorithm. Moving forward, in Section 4, we present various applications of the fixed point results obtained in Section 3, specifically for solving regression and classification problems. Additionally, we include numerical experiments on regression and classification problems within Section 4. Finally, we present the concluding remarks of our paper in Section 5.

2. Preliminaries

Consider a real Hilbert space H equipped with a norm $\|\cdot\|$ and an inner product $\langle \cdot, \cdot \rangle$. Let C be a nonempty closed convex subset of H. The metric projection onto C, denote by P_C, is defined for each $x \in H$, $P_C x$ is the unique element in C such that $\|x - P_C x\| = \inf_{y \in C} \|x - y\|$. It is known that

$$\overline{x} = P_C x \;\Leftrightarrow\; \langle x - \overline{x}, y - \overline{x} \rangle \leq 0, \tag{2.5}$$

for all $y \in C$; see [25]. A mapping $T : C \to C$ is called an *L-Lipschitz operator* if there exists $L > 0$ such that $\|Tx - Ty\| \leq L\|x - y\|$ for all $x, y \in C$. If $L = 1$, the operator T is referred to as a *nonexpansive operator*. A point $x \in C$ is said to be a fixed point of T if $Tx = x$. The fixed point set of T is denoted by $F(T)$, where $F(T) := \{x \in C : Tx = x\}$. Let $\{T_n\}$ and Ω be families of nonexpansive mappings from C into itself, such that $\emptyset \neq F(\Omega) \subset \Gamma := \bigcap_{n=1}^{\infty} F(T_n)$. Here, $F(\Omega)$ represents the set of all common fixed points of the operators in Ω, and Γ denotes the intersection of the fixed point sets $F(T_n)$ for all $n \in \mathbb{N}$. The sequence

$\{T_n\}$ satisfies the *NST-condition(I)* with respect to Ω [18] if, for every bounded sequence $\{v_n\}$ in C,

$$\lim_{n\to\infty} \|v_n - T_n v_n\| = 0 \text{ implies } \lim_{n\to\infty} \|v_n - T v_n\| = 0$$

for all $T \in \Omega$. If Ω is singleton, denoted as $\Omega = T$, then $\{T_n\}$ satisfies the *NST-condition(I)* with respect to T. It is well known that if $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a proper lower semi-continuous convex function, then for all $x \in \mathbb{R}$ the $prox_g(x)$ exists and is unique [4]. The solution to problem (1.1) can be characterized by Theorem 16.3 of Bauschke and Combettes [5] as follows:

$$\overline{p} \text{ is a minimizer of } (f+g) \text{ if and only if } 0 \in \partial g(\overline{p}) + \nabla f(\overline{p})$$

where ∇f represents the gradient of f and ∂g denotes the subdifferential of g. The subdifferential of g at $\overline{p}$, denoted by $\partial g(\overline{p})$, is defined as

$$\partial g(\overline{p}) := \{u : g(x) \geq \langle u, x - \overline{p}\rangle + g(\overline{p}) \text{ for all } x\}.$$

Observe that the subdifferential operator ∂g is maximal monotone, as discussed in [8]. Furthermore, the solution to problem (1.1) can be expressed as the solution to the following fixed point problem:

$$\overline{p} \text{ is a minimizer of } (f+g) \text{ if and only if } \overline{p} = prox_{cg}(I - c\nabla f)(\overline{p}).$$

where $c > 0$ and $prox_{cg} = (I + \partial g)^{-1}$. It is also known that $prox_g(x)$ exists and is unique for each $x \in \mathbb{R}^n$, as discussed in [7], and $prox_{cg}(I - c\nabla f)$ is a nonexpansive mapping when $c \in (0, 2/L)$. The operator $prox_{cg}(I - c\nabla f)$ is referred to as the forward-backward operator of f and g with respect to c. The following lemma is needed to prove our main result.

Lemma 2.1. [9] *Let f be a convex differentiable function from $\mathbb{R}^n$ into $\mathbb{R}$ such that ∇f is Lipschitz continuous with constant $L > 0$, and g is a proper convex and lower semi-continuous function from $\mathbb{R}^n$ into $\mathbb{R} \cup \{\infty\}$. Let $T := prox_{cg}(I - c\nabla f)$ and $T_n := prox_{c_n g}(I - c_n\nabla f)$, where $c, c_n \in (0, 2/L)$ with $c_n \to c$ as $n \to \infty$. Then $\{T_n\}$ satisfies the NST-condition (I) with T.*

Definition 2.1. [3, 4] A sequence $\{T_n : H \to H\}$ with a nonempty common fixed point set is said to satisfy the condition (Z) if $\{x_n\}$ is a bounded sequence in H such that

$$\lim_{n\to\infty} \|x_n - T_n x_n\| = 0,$$

it follows that every weak cluster point of $\{x_n\}$ belongs to $\bigcap_{n=1}^{\infty} F(T_n)$.

It is well known that $I - T$ is demiclosed when $T : H \to H$ is a nonexpansive operator. The following remark is obtained directly by above fact.

Remark 2.1. If $\{T_n\}$ is a sequence of nonexpansive operators satisfies the *NST-condition (I)* with respect to T where T is the nonexpansive operator, then $\{T_n\}$ satisfies the condition (Z).

The following useful facts are crucial for proving our main result.

Lemma 2.2. [25] *Let H be a real Hilbert space. Then the following results hold:*

(i) $\|x \pm y\|^2 = \|x\|^2 \pm 2\langle x, y\rangle + \|y\|^2 \; \forall x, y \in H$;

(ii) $\|x + y\|^2 \leq \|x\|^2 + 2\langle y, x + y\rangle$;

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

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

Lemma 2.3. [22] *Let $\{a_n\}$ be a sequence of nonnegative real numbers and $\{b_n\}$ a sequence of real numbers. Let $\{t_n\}$ be a sequence of real numbers in $(0,1)$ such that $\sum_{n=1}^{\infty} t_n = \infty$. Assume that*

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

If $\limsup_{i\to\infty} b_{n_i} \leq 0$ for every subsequence $\{a_{n_i}\}$ of $\{a_n\}$ satisfying

$$\liminf_{i\to\infty}(a_{n_{i+1}} - a_n) \geq 0,$$

then $\lim_{n\to\infty} a_n = 0$.

Proposition 2.1. [21] *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is strongly convex with convexity parameter $\sigma > 0$ and continuously differentiable function such that ∇f is Lipschitz continuous with constant L_f. Then, the mapping $I - \sigma\nabla f$ is a contraction for all $\sigma \leq \frac{2}{L_f+\rho}$, where I is the identity operator. That is $\|x - \sigma\nabla f(x) - (y - \sigma\nabla f(y))\| \leq \sqrt{1 - \frac{2\sigma\rho L_f}{\rho+L_f}}\|x-y\|$, for all $x, y \in \mathbb{R}^n$.*

3. Main results

Throughout this section, we let $\{T_n : H \to H\}$ and Ω be families of nonexpansive operators on a real Hilbert space H such that $F(\Omega) \subset \Gamma := \bigcap_{n=1}^{\infty} F(T_n)$ and let $S : H \to H$ be a $k-$contraction, where $k \in (0,1)$.

In order to find a common fixed point for a countable family of nonexpansive operators, we propose a novel accelerated algorithm using inertial technique of FISTA and viscosity approximation method. Subsequently, we establish a strong convergence theorem under certain conditions. Let us now present our accelerated algorithm as follows:

Algorithm 4

(1) **Initial.** Take $x_0, x_1 \in H$ arbitrarily and $t_1 = 0$.
(2) **For** $n \geq 1$, set

$$\theta_n = \begin{cases} \min\{\frac{t_n-1}{t_{n+1}}, \frac{\gamma_n\alpha_n}{\|x_n - x_{n-1}\|}\}, & \text{if } x_n \neq x_{n-1}, \\ \frac{t_n-1}{t_{n+1}}, & \text{otherwise}, \end{cases} \tag{3.6}$$

where $t_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2}$ and $\{\gamma_n\} \subset [0,1)$.
(3) **Step1.** Calculate w_n, u_n and x_{n+1} using:

$$\begin{aligned} w_n &= x_n + \theta_n(x_n - x_{n-1}) \\ u_n &= \alpha_n S(w_n) + (1-\alpha_n)T_n w_n \\ x_{n+1} &= (1-\beta_n)u_n + \beta_n T_n u_n, \end{aligned} \tag{3.7}$$

where $\{\alpha_n\}$, $\{\beta_n\} \subset [0,1]$.
Then, update $n := n+1$ and return to Step 1.

Theorem 3.1. *Let H be a real Hilbert space, $\{T_n\}$ be a family of nonexpansive mappings such that $\emptyset \neq \Gamma := \bigcap_{n=1}^{\infty} F(T_n)$. Suppose that $\{T_n\}$ satisfies condition (Z). Let $\{\alpha_n\}$ and $\{\beta_n\}$ be sequences in $[0,1]$ which satisfy the following conditions:*

(i) $\lim_{n\to\infty} \alpha_n = 0$ *and* $\sum_{n=1}^{\infty} \alpha_n = \infty$,
(ii) $\beta_n \in (a,b)$ *for some* $a, b \in (0,1)$ *with* $a < b$,
(iii) $\lim_{n\to\infty} \gamma_n = 0$.

Let S be a contraction on H and $x_0, x_1 \in H$ be arbitrarily. Let $\{x_n\}$ be a sequence generated by Algorithm 4. Then $\{x_n\}$ converges strongly to an element $\overline{p} \in \Gamma$, where $\overline{p} = P_\Gamma S(\overline{p})$.

Proof. Let $\overline{p} \in \Gamma$ be such that $\overline{p} = P_\Gamma S(\overline{p})$. First, we show that $\{x_n\}$ is bounded. By the definition of w_n and u_n, we have

$$\begin{aligned} \|w_n - \overline{p}\| &= \|x_n + \theta_n(x_n - x_{n-1}) - \overline{p}\| \\ &\le \|x_n - \overline{p}\| + \theta_n\|x_n - x_{n-1}\| \end{aligned} \tag{3.8}$$

and

$$\begin{aligned} \|u_n - \overline{p}\| &= \|\alpha_n S(w_n) + (1-\alpha_n)T_n w_n - \overline{p}\| \\ &= \|\alpha_n(S(w_n) - S(\overline{p})) + \alpha_n(S(\overline{p}) - \overline{p}) + (1-\alpha_n)(T_n w_n - \overline{p})\| \\ &\le \alpha_n\|S(w_n) - S(\overline{p})\| + \alpha_n\|S(\overline{p}) - \overline{p}\| + (1-\alpha_n)\|T_n w_n - \overline{p}\| \\ &\le k\alpha_n\|w_n - \overline{p}\| + \alpha_n\|S(\overline{p}) - \overline{p}\| + (1-\alpha_n)\|w_n - \overline{p}\| \\ &= (1-(1-k)\alpha_n)\|w_n - \overline{p}\| + \alpha_n\|S(\overline{p}) - \overline{p}\| \\ &\le (1-(1-k)\alpha_n)(\|x_n - \overline{p}\| + \theta_n\|x_n - x_{n-1}\|) + \alpha_n\|S(\overline{p}) - \overline{p}\| \\ &\le (1-(1-k)\alpha_n)\|x_n - \overline{p}\| + \alpha_n\left(\frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| + \|S(\overline{p}) - \overline{p}\|\right). \end{aligned} \tag{3.9}$$

From (3.8) and (3.9), we obtain

$$\begin{aligned} \|x_{n+1} - \overline{p}\| &= \|(1-\beta_n)u_n + \beta_n T_n u_n - \overline{p}\| \\ &\le (1-\beta_n)\|u_n - \overline{p}\| + \beta_n\|T_n u_n - \overline{p}\| \\ &\le (1-\beta_n)\|u_n - \overline{p}\| + \beta_n\|u_n - \| \\ &= \|u_n - \overline{p}\|. \end{aligned} \tag{3.10}$$

From $\lim_{n\to\infty} \frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| = 0$, there exists a constant $M > 0$ such that $\frac{\theta_n}{\alpha_n}\|x_n - x_{n-1}\| \le M$, for all $n \ge 1$. Thus

$$\|x_{n+1} - \overline{p}\| \le (1-(1-k)\alpha_n)\|x_n - \overline{p}\| + \alpha_n\left(M + \|S(\overline{p}) - \overline{p}\|\right).$$

By mathematical induction, we get

$$\|x_{n+1} - \overline{p}\| \le \max\left\{\|x_0 - \overline{p}\|, \frac{M + \|S(\overline{p}) - \overline{p}\|}{1-k}\right\} \ \forall n \ge 1.$$

This implies that $\{x_n\}$ is bounded and $\{w_n\}, \{u_n\}, \{T_n w_n\}, \{T_n u_n\}, \{S(w_n)\}$ are also bounded.

By definition of w_n, for $n \in \mathbb{N}$, we have

$$\begin{aligned} \|w_n - \overline{p}\|^2 &= \|x_n + \theta_n(x_n - x_{n-1}) - \overline{p}\|^2 \\ &= \|x_n - \overline{p}\|^2 + 2\theta_n\langle x_n - \overline{p}, x_n - x_{n-1}\rangle + \theta_n^2\|x_n - x_{n-1}\|^2 \end{aligned} \tag{3.11}$$

By Lemma 2.2(ii), we get

$$\begin{aligned}\|u_n-\overline{p}\|^2 &= \|\alpha_n S(w_n)+(1-\alpha_n)T_n w_n-\overline{p}\|^2\\ &= \|\alpha_n(S(w_n)-S(\overline{p}))+\alpha_n(S(\overline{p})-\overline{p})+(1-\alpha_n)(T_n w_n-\overline{p})\|^2\\ &\le \|\alpha_n(S(w_n)-S(\overline{p}))+(1-\alpha_n)(T_n w_n-\overline{p})\|^2\\ &+2\left\langle \alpha_n(S(\overline{p})-\overline{p}),u_n-\overline{p}\right\rangle\\ &\le \alpha_n\|S(w_n)-S(\overline{p})\|^2+(1-\alpha_n)\|T_n w_n-\overline{p}\|^2\\ &+2\alpha_n\left\langle S(\overline{p})-\overline{p},u_n-\overline{p}\right\rangle\\ &\le \alpha_n k\|w_n-\overline{p}\|^2+(1-\alpha_n)\|w_n-\overline{p}\|^2+2\alpha_n\left\langle S(\overline{p})-\overline{p},u_n-\overline{p}\right\rangle\\ &\le (1-\alpha_n+\alpha_n k)(\|x_n-\overline{p}\|^2+2\left\langle x_n-\overline{p},\theta_n(x_n-x_{n-1})\right\rangle)\\ &+(1-\alpha_n+\alpha_n k)(\theta_n^2\|x_n-x_{n-1}\|^2)+2\alpha_n\left\langle S(\overline{p})-\overline{p},u_n-\overline{p}\right\rangle. \end{aligned} \tag{3.12}$$

By Lemma 2.2(iii) and (3.12), we have

$$\begin{aligned}\|x_{n+1}-\overline{p}\|^2 &= \|(1-\beta_n)u_n+\beta_n T_n u_n-\overline{p}\|^2\\ &= \|(1-\beta_n)(u_n-\overline{p})+\beta_n(T_n u_n-\overline{p})\|^2\\ &= (1-\beta_n)\|u_n-\overline{p}\|^2+\beta_n\|T_n u_n-\overline{p}\|^2\\ &-\beta_n(1-\beta_n)\|T_n u_n-u_n\|^2\\ &\le \|u_n-\overline{p}\|^2-\beta_n(1-\beta_n)\|T_n u_n-u_n\|^2\\ &= (1-\alpha_n+\alpha_n k)(\|x_n-\overline{p}\|^2+2\left\langle x_n-\overline{p},\theta_n(x_n-x_{n-1})\right\rangle)\\ &+(1-\alpha_n+\alpha_n k)(\theta_n^2\|x_n-x_{n-1}\|^2)+2\alpha_n\left\langle S(\overline{p})-\overline{p},u_n-\overline{p}\right\rangle\\ &-\beta_n(1-\beta_n)\|T_n u_n-u_n\|^2\\ &\le (1-\alpha_n+\alpha_n k)\|x_n-\overline{p}\|^2-\beta_n(1-\beta_n)\|T_n u_n-u_n\|^2\\ &+\alpha_n(1-k)b_n, \end{aligned} \tag{3.13}$$

where

$$\begin{aligned} b_n &= (\frac{1}{1-k})\{2\left\langle S(\overline{p})-\overline{p},u_n-\overline{p}\right\rangle+2\|x_n-\overline{p}\|\left(\frac{\theta_n}{\alpha_n}\|x_n-x_{n-1}\|\right)\\ &+\left(\frac{\theta_n}{\alpha_n}\|x_n-x_{n-1}\|\right)\theta_n\|x_n-x_{n-1}\|\}. \end{aligned}$$

It follows that

$$\beta_n(1-\beta_n)\|T_n u_n-u_n\|^2\le\|x_n-\overline{p}\|^2-\|x_{n+1}-\overline{p}\|^2+\alpha_n(1-k)M', \tag{3.14}$$

where $M'=\sup\{b_n : n\in\mathbb{N}\}$.

Finally, we show that $x_n\to\overline{p}$. To do this, we will apply Lemma 2.3 by setting $a_n := \|x_n-\overline{p}\|^2$ and $t_n:=\alpha_n(1-k)$. From (3.13), we have the following inequality:

$$a_{n+1}\le(1-t_n)a_n+t_n b_n.$$

Suppose $\{a_{n_i}\}$ is a subsequence of $\{a_n\}$ such that

$$\liminf_{i\to\infty}(a_{n_{i+1}}-a_{n_i})\ge 0. \tag{3.15}$$

From (3.14), (3.15) and condition (i), we get

$$\begin{aligned}\limsup_{i\to\infty}\beta_{n_i}(1-\beta_{n_i})\|T_{n_i}u_{n_i}-u_{n_i}\|^2 &\le \limsup_{i\to\infty}(a_{n_i}-a_{n_{i+1}}+\alpha_{n_i}(1-k)M') \\ &\le \limsup_{i\to\infty}(a_{n_i}-a_{n_{i+1}}) \\ &\quad + (1-k)M'\lim_{i\to\infty}\alpha_{n_i} \\ &= -\liminf_{i\to\infty}(a_{n_{i+1}}-a_{n_i}) \\ &\le 0.\end{aligned}$$

The condition (ii) and above inequality lead to

$$\lim_{i\to\infty}\|T_{n_i}u_{n_i}-u_{n_i}\|=0. \tag{3.16}$$

By the choice of θ_n in (3.6) together with the condition (iii), we note that $\frac{\theta_n}{\alpha_n}\|x_{n+1}-x_n\|\to 0$ as $n\to\infty$. We next show that $\limsup_{i\to\infty} b_{n_i}\le 0$. Obviously, it suffices to show that

$$\limsup_{n\to\infty}\langle S(\overline{p})-\overline{p},u_{n_i}-\overline{p}\rangle\le 0.$$

Let $\{u_{n_{i_j}}\}$ be subsequence of $\{u_{n_i}\}$ such that

$$\lim_{j\to\infty}\left\langle S(\overline{p})-\overline{p},u_{n_{i_j}}-\overline{p}\right\rangle=\limsup_{k\to\infty}\langle S(\overline{p})-\overline{p},u_{n_i}-\overline{p}\rangle$$

and $u_{n_{i_j}}\rightharpoonup w$ for some $w\in H$. By (3.16), it follows from the condition (Z) of $\{T_n\}$ that $w\in\Gamma$. Here, the equation $\overline{p}=P_\Gamma S(\overline{p})$ yields

$$\begin{aligned}\limsup_{n\to\infty}\langle S(\overline{p})-\overline{p},u_{n_i}-\overline{p}\rangle &= \lim_{k\to\infty}\left\langle S(\overline{p})-\overline{p},u_{n_{i_j}}-\overline{p}\right\rangle \\ &= \langle S(\overline{p})-\overline{p},w-\overline{p}\rangle \\ &\le 0.\end{aligned}$$

By Lemma 2.3, we can conclude that $x_n\to\overline{p}$ as $n\to\infty$. The proof is complete. □

The following result is a consequence of Theorem 3.1 which asserts that Algorithm 5 strongly converges to a solution of convex bilevel problems (1.1) and (1.2). From now on, we denote $\mathcal{A}$ the set of all solutions of problem (1.1).

Algorithm 5

(1) **Initial.** Take $x_0,x_1\in\mathbb{R}^n$ arbitrarily and $t_1=0$.
(2) **For** $n\ge 1$, set

$$\theta_n=\begin{cases}\min\{\frac{t_n-1}{t_{n+1}},\frac{\gamma_n\alpha_n}{\|x_n-x_{n-1}\|}\}, & \text{if } x_n\ne x_{n-1},\\ \frac{t_n-1}{t_{n+1}}, & \text{otherwise,}\end{cases} \tag{3.17}$$

where $t_{n+1}=\frac{1+\sqrt{1+4t_n^2}}{2}$ and $\{\gamma_n\}\subset[0,1)$.
(3) **Step1.** Calculate u_n, z_n and x_{n+1} using:

$$\begin{aligned}w_n &= x_n+\theta_n(x_n-x_{n-1})\\ u_n &= \alpha_n(I-s\nabla h)(w_n)+(1-\alpha_n)prox_{c_ng}(I-c_n\nabla f)w_n\\ x_{n+1} &= (1-\beta_n)u_n+\beta_n prox_{c_ng}(I-c_n\nabla f)u_n,\end{aligned} \tag{3.18}$$

where $\{\alpha_n\}$, $\{\beta_n\}\subset[0,1]$ and $s\in(0,2/(\sigma+L_h))$.
Then, update $n:=n+1$ and return to Step 1.

Theorem 3.2. *Let $h : \mathbb{R}^n \to \mathbb{R}$ be a strongly convex function with a parameter $\sigma > 0$. Assume that h is continuously differentiable and that its gradient ∇h is Lipschitz continuous with a constant L_h. Suppose that f and g satisfy the assumptions of problem (1.2). Let $\{c_n\}$ be a sequence in $(0, 2/L_f)$ such that $c_n \to c$ as $n \to \infty$ where $c \in (0, 2/L_f)$ and let $\{x_n\}$ be a sequence generated by Algorithm 5. Then $\{x_n\}$ converges strongly to $\overline{p} \in \mathcal{A}$. Suppose the following conditions hold:*

(i) $\lim_{n\to\infty} \alpha_n = 0$ *and* $\sum_{n=1}^{\infty} \alpha_n = \infty$,
(ii) $\beta_n \in (a, b)$ *for some* $a, b \in (0, 1)$ *with* $a < b$,
(iii) $\lim_{n\to\infty} \gamma_n = 0$.

Proof. Let $T = prox_{cg}(I - c\nabla f)$ and $T_n = prox_{c_n g}(I - c_n \nabla f),\ n \in \mathbb{N}$. By Lemma 2.1 and Remark 2.1, we know that $\{T_n\}$ satisfies the condition (Z). By Theorem 3.1, we get that $\{x_n\}$ converges to $\overline{p} \in \Gamma = X^* = arg\min_{x\in\mathbb{R}^n}(f(x) + g(x))$. By Proposition 2.1 , $S = I - s\nabla h(x)$ is a k-contraction with parameter $k = \sqrt{1 - \frac{2s\sigma L_h}{\sigma + L_h}}$, whenever $s \in (0, 2/(\sigma + L_h))$. It remains to show that $\overline{p} = arcmin_{x\in X^*} h(x)$. By using $\overline{p} = P_{X^*}S(\overline{p})$ and (2.5), we have, for $z \in X^*$,

$$
\begin{aligned}
\overline{p} = P_\Gamma S(\overline{p}) &\Leftrightarrow \langle S(\overline{p}) - \overline{p}, z - \overline{p}\rangle \le\ 0 \\
&\Leftrightarrow \langle \overline{p} - s\nabla h(\overline{p}) - \overline{p}, z - \overline{p}\rangle \le\ 0 \\
&\Leftrightarrow \langle s\nabla h(\overline{p}), z - \overline{p}\rangle \ge\ 0 \\
&\Leftrightarrow s\langle \nabla h(\overline{p}), z - \overline{p}\rangle \ge\ 0 \\
&\Leftrightarrow \langle \nabla h(\overline{p}), z - \overline{p}\rangle \ge 0.
\end{aligned}
$$

Thus, $\overline{p}$ is an optimal solution for the problem (1.1). That is, $x_n \to \overline{p} \in \mathcal{A}$. □

4. Application

In this section, we utilize Algorithm 5 as a machine learning algorithm for regression of a graph of the Cosine function. Additionally, we apply this algorithm for data classification using Extreme Learning Machine models and Single Hidden Layer Feedforward Neural Networks. All computations and results are conducted using the MATLAB computing environment on a system equipped with a laptop computer (Intel Core-i5 gen 8/8.00 GB RAM/Windows 11/64-bit).

Moreover we compare performance of our proposed algorithm with Big-SAM, iBig-SAM, and aiBig-SAM.

Let us recall a concept of Extreme Learning Machine. Extreme Learning Machine (ELM) [13] can be defined as follows: Let $D = \{(x_k, q_k) : x_k \in \mathbb{R}^n, q_k \in \mathbb{R}^m,\ k = 1, 2, \dots, N\}$ be a training set of N distinct samples, x_k is an input data and q_k is a target. A mathematical model for standard Single-hidden Layer Feedforward Networks (SLFNs) with activation function $\varphi(x)$ and M hidden nodes is given by

$$
\sum_{j=1}^{M} \xi_j \varphi(\langle a_j, x_k\rangle + c_j) = o_k,\ k = 1, \dots, N,
$$

where ξ_j represents the weight vector connecting the j-th hidden node to the output node, while a_j denotes the weight vector connecting the j-th hidden node to the input node, c_j is the bias and o_k is the output from the model. The goal of SLFNs is to predict these N outputs such that $\sum_{d=1}^{N} \|o_k - q_k\| = 0$. That is,

$$
\sum_{j=1}^{M} \xi_j \varphi(\langle a_j, x_k\rangle + c_j) = q_k,\ k = 1, \dots, N. \tag{4.19}
$$

We can rewrite the above system of linear equation by the following matrix equation:

$$D\xi = Q, \tag{4.20}$$

where

$$D = \begin{bmatrix} \varphi(\langle a_1, x_1 \rangle + c_1) & \cdots & \varphi(\langle a_M, x_1 \rangle + c_M) \\ \vdots & \ddots & \vdots \\ \varphi(\langle a_1, x_N \rangle + c_1) & \cdots & \varphi(\langle a_M, x_N \rangle + c_M) \end{bmatrix}_{N \times M},$$

$$\xi = [\xi_1^T, \ldots, \xi_M^T]^T_{m \times M}, \; Q = [q_1^T, \ldots, q_N^T]^T_{m \times N}.$$

The objective of an SLFNs is estimating ξ_j, a_j and c_j for solving (4.19) while ELM aims to find only ξ_j with randomly a_j and c_j.

To solve the problem (4.20), we consider the following convex minimization problem:

$$\min_{\xi} \|D\xi - Q\|_2^2 + \lambda \|\xi\|_1 , \tag{4.21}$$

where $\lambda > 0$ is called the regularization parameter. Let X^* be a set of all solutions of (4.21). We are interested to used the function $h(\xi) = \frac{1}{2}\|\xi\|_2^2$ to select a solution in X^* which satisfies the outer level convex minimization problem: $\min_{\xi \in X^*} h(\xi)$. By setting, $g(\xi) = \lambda \|\xi\|_1$ and $f(\xi) = \|D\xi - Q\|_2^2$, we employ Algorithm 5 to solve the convex bilevel optimization problems (4.21).

4.1. **Regression of Cosine Function.** In our regression experiment involving a graph of the Cosine function, we formed a training set by randomly choosing 10 distinct points. To perform the regression, we employed the sigmoid function as our chosen activation function. Additionally, we fixed the number of hidden nodes at $M = 100$ and set the regularization parameter to $\lambda = 1 \times 10^{-5}$. We use mean squared error (MSE) to measure performance for regression of each studied algorithm. MSE is given by the following:

$$Mean\ squared\ error(MSE) = \frac{1}{N} \sum_{k=1}^{N} \|o_k - q_k\|^2.$$

We set all control conditions for each algorithm as in Table 1.

TABLE 1. Algorithms and their setting control conditions.

Methods	Setting
Algorithm 5	$s = 0.01$, $c_n = \frac{1}{L_f}$, $t_1 = 0$, $\alpha_n = \frac{1}{88n}$, $\gamma_n = \frac{88 \cdot 10^{20}}{n}$, $\beta_n = \frac{0.9(n+1)}{n}$
BiG-SAM	$s = 0.01$, $c_n = \frac{1}{L_f}$, $\alpha_n = \frac{2(0.1)}{1 - \frac{2 + c_n L_f}{4}}$, $\gamma_n = \frac{\alpha_n}{n^{0.01}}$
iBiG-SAM	$\alpha = 3$, $s = 0.01$, $c_n = \frac{1}{L_f}$, $\alpha_n = \frac{2(0.1)}{1 - \frac{2 + c_n L_f}{4}}$, $\gamma_n = \frac{\alpha_n}{n^{0.01}}$
aiBiG-SAM	$\alpha = 3$, $s = 0.01$, $c_n = \frac{1}{L_f}$, $\alpha_n = \frac{1}{k+2}$, $\gamma_n = \frac{\alpha_n}{n^{0.01}}$

We then obtain the numerical experiments which show MSE and computational time of each studied algorithm as in Table 2.

From Table 2 and Figure 1, we observe that Algorithm 5 gives a better performance to predict a Cosine function than others while there is no significant difference in computational time.

TABLE 2. Comparison of studied algotrithm for regression a cosine function with 300 iterations.

Methods	Computational time	MSE
Algorithm 5	0.0139	0.0483196
BIG-SAM	0.0121	0.5554438
iBIG-SAM	0.0125	0.5554769
aiBIG-SAM	0.0125	0.5583825

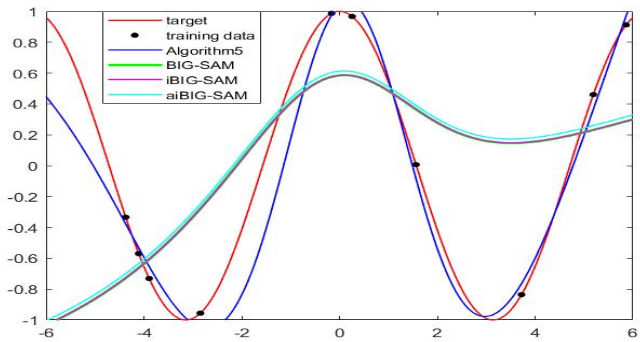

FIGURE 1. A regression of the cosine function at 300th iteration

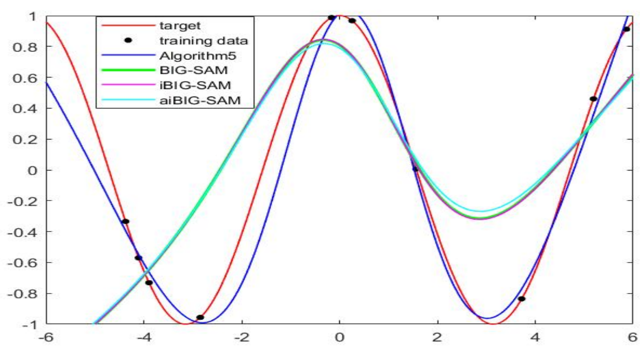

FIGURE 2. A regression of the cosine function at 2000th iteration

4.2. **Data Classification.** In this section, we employ our proposed Algorithm 5 for data classification of noncommunicable diseases and compare its performance with the others. For our experiment, we use five datasets of noncommunicable diseases from "`https://www.kaggle.com/`, accessed on 23 July 2021" and "`https://archive.ics.uci.edu/`, accessed on 23 July 2021" as follows:

Breast Cancer dataset [28]: The dataset contains 11 attributes. This dataset involves the classification of data into 2 distinct classes.

Diabetes dataset [24]: The dataset contains 9 attributes. This dataset is comprised of two distinct classes for classification purposes.

Heart Disease UCI dataset [10]: The dataset contains 14 attributes. This dataset also involves the classification of data into 2 classes.

Parkinsons dataset [16]: The data set comprises 23 attributes. Within this particular data set, we categorize information into two distinct classes.

Indian Liver Patient Dataset (ILPD) [12]: Within the dataset, there are 11 attributes present. The data is divided into 2 distinct categories that we are able to classify. In Table 3, we present the attributes count for each dataset along with the distribution of data into training and testing sets. The training set comprises approximately 70% of the data, while the remaining 30% is allocated to the testing set.

TABLE 3. Training and Testing sets of dataset.

Dataset	Attributes	Sample Train	Sample Test
Breast Cancer	11	488	211
Diabetes	9	538	230
Heart Disease	14	213	90
Parkinson	23	135	60
Indian Liver Patient Dataset (ILPD)	11	408	175

We utilized the identical set of control parameters outlined in Table 1 from Section 4.1, including a consistent number of hidden nodes ($M = 100$) and a sigmoid activation function. For each dataset specified in Table 3, we trained the model using the respective training set. The accuracy of the output data was determined through the following calculation:

$$\text{accuracy} = \frac{\text{correctly predicted data}}{\text{total data}} \times 100.$$

In Table 4, we present a comparison of the training accuracy, testing accuracy, and iteration number of Algorithm 5 with other algorithms for each dataset.

TABLE 4. The iteration number of each algorithm with the best accuracy on each dataset.

Dataset	Algorithm	Iteration no.	Accuracy train	Accuracy test
Breast Cancer	Algorithm 5	252	96.55	98.99
	BIG-SAM	1700	96.55	98.49
	iBIG-SAM	1700	96.55	98.49
	aiBIG-SAM	1700	96.55	98.49
Diabetes	Algorithm 5	98	77.11	81.98
	BIG-SAM	700	76.01	81.08
	iBIG-SAM	696	76.37	81.08
	aiBIG-SAM	1300	76.92	80.18
Heart Disease	Algorithm 5	163	87.14	82.80
	BIG-SAM	1800	86.19	82.80
	iBIG-SAM	1756	86.19	82.80
	aiBIG-SAM	2501	86.67	82.80
Parkinson	Algorithm 5	424	94.16	77.59
	BIG-SAM	659	86.13	77.59
	iBIG-SAM	660	86.13	77.59
	aiBIG-SAM	2240	87.59	77.59
Indian Liver Patient Dataset (ILPD)	Algorithm 5	376	71.46	72.25
	BIG-SAM	391	71.22	72.25
	iBIG-SAM	486	71.46	72.25
	aiBIG-SAM	790	71.46	72.25

We observe from Table 4 that Algorithm 5 has a better performance in terms of accuracy and number of iterations than BIG-SAM, iBIG-SAM, aiBIG-SAM in all experiments conducted. We also see that Algorithm 5 requires the lowest number of iterations to achieve the same accuracy as the other studied algorithms.

5. Conclusion

We have proposed and analyzed a strong convergence of a common fixed point algorithm with the inertial technique for a countable family of nonexpansive operators and then we applied it to tackle certain types of convex bilevel optimization problems. To ensure the effectiveness of our proposed algorithm, we implement our method for data classification of noncommunicable diseases and regression of a graph of cosine function. From numerical experiments, we found that our proposed algorithm has a better performance than other existing algorithms. For future work, we will employ our proposed algorithms for prediction and classification of some noncommunicable diseases data collected by Sriphat medical center, faculty of medicine, Chiang Mai University, Chiang Mai, Thailand. We are also interested to create an application for noncommunicable diseases prediction in Thailand.

Acknowledgements. This research has been financially supported by Fundamental Fund 2024, Chiang Mai University (FF030/2567). Additionally, the first author would like to thank the Science Achievement Scholarship of Thailand (SAST) for their generous financial assistance.

References

[1] Abuchu, J.-A.; Ugwunnadi, G.-C.; Narain, O.-K. Inertial proximal and contraction methods for solving monotone variational inclusion and fixed point problems. *Nonlinear Functional Analysis and Applications*, **28** (2023), no. 1 , 175–203.

[2] Akutsah, F.; Narain, O.-K.; Kim, J.-K. Improved generalized M-iteration for quasi-nonexpansive multivalued mappings with application in real Hilbert Spaces. *Nonlinear Functional Analysis and Applications*, **27** (2022), no. 1, 59–82.

[3] Aoyama, K.; Kimura, Y. Strong convergence theorems for strongly nonexpansive sequences. *Appl. Math. Comput.* **217** (2011), no. 19, 7537–7545.

[4] Aoyama, K.; Kohsaka, F.; Takahashi, W. Strong convergence theorems by shrinking and hybrid projection methods for relatively nonexpansive mappings in Banach spaces. *Nonlinear Analysis and Convex Analysis.*, (2009), 7–26.

[5] Bauschke, H.-H.; Combettes, P.-L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, USA, 2011.

[6] Beck, A.; Teboulle, M., A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences* **2** (2009), no. 1, 183–202.

[7] Boyd, S.; Vandenberghe, L. *Convex Optimization*. Cambridge University Pass, New York, USA, 2004.

[8] Burachik, RS.; Iusem, AN. *Set-valued mappings and enlargements of monotone operators*. Springer Science Business Media, New York, 2007.

[9] Bussaban, L.; Suantai, S.; Kaewkhao, A. A parallel inertial S-iteration forward-backward algorithm for regression and classification problems. *Carpathian J. Math.* **36** (2020), no. 1, 35–44.

[10] Detrano, R.; Janosi, A.; Steinbrunn, W.; Pfisterer, M.; Schmid, J.-J.; Sandhu, S.; Guppy, K.-H.; Lee, S.; Froelicher, V. International application of a new probability algorithm for the diagnosis of coronary artery disease *Am. J. Cardiol.* **64** (1989), no. 5, 304–310.

[11] Duan, P.; Zhang, Y. Alternated and multi-step inertial approximation methods for solving convex bilevel optimization problems. *Optimization* **72** (2022), no. 10, 2517–2545.

[12] Dua, D.; Graff, C. *UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]*, Irvine, CA: University of California, School of Information and Computer Science 2019.

[13] Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications *Neurocomputing* **70** (2006), no. 1-3, 489–501.

[14] Jailoka, P.; Suantai, S.; Hanjing, A. A fast viscosity forward-backward algorithm for convex minimization problems with an application in image recovery. *Carpathian J. Math.* **37** (2021), no. 3, 449–461.

[15] Lions, P.-L.; Mercier, B. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **16** (1979), no. 6, 964–979.

[16] Little, M.-A.; McSharry, P.-E.; Roberts, S.-J.; Costello, D.-A.; Moroz., I.-M. Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection. *Nature Precedings* (2007), 1–1.

[17] Mainge, P.-E.; Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization. *Set-Valued Anal.* **16** (2008), no. 7-8, 899–912.

[18] Nakajo, K.; Shimoji, K.; Takahashi, W. Strong convergence to common fixed points of families of nonexpansive mappings in Banach spaces. *J. Nonlinear Convex Anal.* **8** (2007), no. 1, 11–34.

[19] Polyak, B. Some methods of speeding up the convergence of iteration methods. *Ussr computational mathematics and mathematical physics* **4** (1964), no. 5, 1–17.

[20] Puangpee, J.; Suantai, S., A New Accelerated Viscosity Iterative Method for an Infinite Family of Nonexpansive Mappings with Applications to Image Restoration Problems. *Mathematics* **8** (2020), no. 4, 615.

[21] Sabach, S.; Shtern, S. A first order method for solving convex bilevel optimization problems. *SIAM J. Optim.* **27** (2017), no. 2, 640–660.

[22] Saejung, S.; Yotkaew, P., Approximation of zeros of inverse strongly monotone operators in Banach spaces *Nonlinear Anal.* **75** (2012), no. 2, 724–750.

[23] Shehu, Y.; Vuong, P.-T.; Zemkoho, A., An inertial extrapolation method for convex simple bilevel optimization. *Optim Methods Softw.* **36** (2019), no. 1, 1–19.

[24] Smith, J.-W.; Everhart, J.-E.; Dickson, W.-C., Knowler, W.-C.; Johannes, R.S. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. *American Medical Informatics Association* (1998).

[25] Takahashi, W. *Introduction to Nonlinear and Convex Analysis Yokohama Publishers*. Yokohama, Japan, 2009.

[26] Takahashi, W. *Nonlinear Functional Analysis*. Yokohama Publishers Yokohama, Japan, 2000.

[27] Thongsri, P.; Panyanak, B.; Suantai, S. A New Accelerated Algorithm Based on Fixed Point Method for Convex Bilevel Optimization Problems with Applications. *Mathematics* **11** (2023), no. 3, 702.

[28] Wolberg, W.-H.; Mangasarian, O.-L. Multisurface method of pattern separation for medical diagnosis applied to breast cytology *Proc. Natl. Acad. Sci. USA.* **87** (1990), no. 23, 9193–9196.

[29] Xu, H.-K. Another control condition in an iterative method for nonexpansive mappings *Bulletin of the Australian Mathematical Society* **65** (2002), no. 1, 109–113.

[30] Xu, H.-K. Viscosity approximation methods for nonexpansive mappings. *J. Math. Anal. Appl.* **298** (2004), no. 1, 279-291.

PHD DEGREE PROGRAM IN MATHEMATICS
DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE
CHIANG MAI UNIVERSITY
239 HUAYKAEW RD, 50200, CHIANG MAI, THAILAND
Email address: piti_tho@cmu.ac.th

DEPARTMENT OF MATHEMATICS, FACULTY OF SCIENCE
CHIANG MAI UNIVERSITY
239 HUAYKAEW RD, 50200, CHIANG MAI, THAILAND
Email address: Suthep.s@cmu.ac.th