Mathematics

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

AIMS Mathematics, 9(4): 8476–8496.
DOI: 10.3934/math.2024412
Received: 12 January 2024
Revised: 21 February 2024
Accepted: 23 February 2024
Published: 28 February 2024

Research article

A new two-step inertial algorithm for solving convex bilevel optimization problems with application in data classification problems

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

[1] PhD Degree Program in Mathematics, Department of Mathematics, Faculty of Science, Chiang Mai University, under the CMU Presidential Scholarship, Thailand

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

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

Abstract: In this paper, we propose a new accelerated algorithm for solving convex bilevel optimization problems using some fixed point and two-step inertial techniques. Our focus is on analyzing the convergence behavior of the proposed algorithm. We establish a strong convergence theorem for our algorithm under some control conditions. To demonstrate the effectiveness of our algorithm, we utilize it as a machine learning algorithm to solve data classification problems of some noncommunicable diseases, and compare its efficacy with BiG-SAM and iBiG-SAM.

Keywords: accelerated algorithm; two-step inertial algorithm; convex bilevel optimization problem; data classification; noncommunicable diseases
Mathematics Subject Classification: 47H10, 65K10, 90C25

1. Introduction

Data classification is an important data mining technique with a wide variety of applications to classify the different kinds of data that exist practically in all aspects of our life. It has been recognized as a critical topic in machine learning and data mining.

We begin by reviewing the history of various mathematical models and related techniques used for this purpose. Convex bilevel optimization problem plays an important role in real-world applications. It can be applied to data classification, see for example [1–4]. The convex bilevel optimization problem consists of the constrained minimization problem known as the *outer* level,

$$\min_{u\in\Lambda} \phi(u), \tag{1.1}$$

where $\mathcal{H}$ is a real Hilbert space, $\phi : \mathcal{H} \to \mathbb{R}$ is a strongly convex differentiable function, and Λ is a nonempty set of minimizers of the *inner* level given by

$$\arg\min_{u\in\mathbb{R}^m}\{\varphi(u) + \psi(u)\}, \tag{1.2}$$

where $\varphi : \mathbb{R}^m \to \mathbb{R}$ is convex differentiable function such that $\nabla\varphi$ is L_φ-Lipschitzian and $\psi \in \Gamma_0(\mathcal{H})$, the set of proper lower semicontinuous convex functions from $\mathcal{H}$ to $\mathbb{R}$. Problems (1.1) and (1.2) are labeled as a bilevel optimization problem.

Furthermore, the solution of (1.2) can be restated as the problem of finding $\hat{u} \in \Lambda$ such that

$$0 \in \nabla\varphi(\hat{u}) + \partial\psi(\hat{u}). \tag{1.3}$$

Parikh and Boyd [5] introduced the proximal gradient technique for solving (1.3), that is, $\hat{u}$ is a solution of (1.3) if and only if $\hat{u} \in \mathfrak{F}(T)$ where T is the *prox-grad mapping* defined by

$$T := \text{prox}_{t\psi}(I - t\nabla\varphi),$$

for $t > 0$, and $\mathfrak{F}(T)$ is the set of fixed points of T. It is well-known that if $t \in (0, \frac{2}{L_\varphi})$, then T is nonexpansive and $\mathfrak{F}(T) = \arg\min_{u\in\mathbb{R}^m}\{\varphi(u) + \psi(u)\}$. We also note that the set of all common fixed points of $T_n = \text{prox}_{c_n}\,\psi(I - s\nabla\varphi)$ is the set of minimizers of the inner level problem (1.2).

Furthermore, $u \in \mathfrak{F}(T)$ is a solution for problem (1.1) if u satisfies the condition

$$\langle\nabla\phi(u), v - u\rangle \geq 0,\ \forall v \in \mathfrak{F}(T).$$

Hereafter, we would like to give some background on iteration methods for finding a fixed point of the nonexpansive mapping T, that is, finding a point $u^* \in C$ such that $Tu^* = u^*$.

Let $\mathcal{H}$ be a real Hilbert space with norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$, and C be a nonempty closed convex subset of $\mathcal{H}$. One of the most popular iterative methods for finding a fixed point of a nonexpansive mapping is the Mann iteration, which was first introduced by Mann [6]. Later, Reich [7] modified it to the general version

$$u_{n+1} = \lambda_n u_n + (1 - \lambda_n)Tu_n,\ \forall n \geq 1, \tag{1.4}$$

where $u_1 \in \mathcal{H}$ and $\{\lambda_n\}$ is a real sequence in $[0, 1]$. He proved the weak convergence of (1.4) under the condition $\sum_{n=1}^{\infty} \lambda_n(1 - \lambda_n) = \infty$.

Later, Halpern [8] introduced an iterative method known as the Halpern iteration for finding a fixed point of nonexpansive mappings in real Hilbert spaces. His algorithm was given in the following form:

$$u_{n+1} = \lambda_n u_0 + (1 - \lambda_n)Tu_n,\ \forall n \geq 1, \tag{1.5}$$

where $u_0, u_1 \in C$ and $\{\lambda_n\} \subset [0, 1]$. Under some condition on $\{\lambda_n\}$, he established a strong convergence theorem of (1.5) when $u_0 = 0$. Later, Reich [9] extended the Halpern iteration (1.5) to uniformly smooth Banach spaces.

In 1974, by modifying the Mann iteration, Ishikawa [10] introduced the Ishikawa iteration process as follows:

$$\begin{cases} v_n &= (1 - \lambda_n)v_n + \lambda_n T v_n, \\ u_{n+1} &= (1 - \delta_n)u_n + \delta_n T v_n,\ \forall n \geq 1, \end{cases} \tag{1.6}$$

where $u_1 \in \mathcal{H}$ and $\{\lambda_n\}, \{\delta_n\} \subset [0, 1]$.

Moudafi [11] demonstrated a viscosity approximation method for a nonexpansive mapping in 2000, which was defined as

$$u_{n+1} = \lambda_n f(u_n) + (1-\lambda_n)Tu_n,\ \forall n \geq 1, \tag{1.7}$$

where $u_1 \in \mathcal{H}, \{\lambda_n\} \subset [0,1]$, and f is a contraction mapping. He proved that, under certain conditions, $\{u_n\}$ generated by (1.7) converges strongly to $x \in \mathfrak{F}(T)$.

By modification of the Ishikawa iteration, Agarwal et al. [12] presented the S-iteration process as follows:

$$\begin{cases} v_n &= (1-\lambda_n)u_n + \lambda_n T u_n, \\ u_{n+1} &= (1-\delta_n)Tu_n + \delta_n T v_n,\ \forall n \geq 1, \end{cases} \tag{1.8}$$

where $\{\lambda_n\}, \{\delta_n\} \subset [0,1]$ and x_1 is arbitrarily chosen. Furthemore, they demonstrated that the convergence behavior of the S-iteration is better than the iterations of Mann and Ishikawa.

Now, we would like to give some background on iteration methods to find a common fixed point of a countable family of a nonexpansive mapping $\{T_n\}$.

Aoyama et al. [13] demonstrated a Halpern type iteration

$$u_{n+1} = \lambda_n u + (1-\lambda_n)T_n u_n,\ \forall n \geq 1, \tag{1.9}$$

where $\{\lambda_n\} \subset [0,1]$ and u_1 and $u \in C$ are arbitrarily chosen. Further, they showed that, under some condition on $\{\lambda_n\}, x_n \to x \in \bigcap_{n=1}^{\infty} \mathfrak{F}(T_n)$.

Thereafter, Takahashi [14] demonstrated the iteration process

$$u_{n+1} = \lambda_n f(u_n) + (1-\lambda_n)T_n u_n,\ \forall n \geq 1, \tag{1.10}$$

where $\{\lambda_n\} \subset [0,1]$, and established a strong convergence theorem of (1.10) under some constraint on $\{\lambda_n\}$.

In 2010, Klin-eam and Suantai [15] introduced the following algorithm:

$$\begin{cases} v_n &= \lambda_n f(u_n) + (1-\delta_n)T_n u_n, \\ u_{n+1} &= (1-\delta_n)v_n + \delta_n T_n v_n,\ \forall n \geq 1, \end{cases} \tag{1.11}$$

where $\{\lambda_n\} \subset [0,1]$ and $u_1 \in C$, and showed that, under certain conditions, $\{u_n\}$ generated by (1.11) converges strongly to a common fixed point of T_n.

Polyak [16] developed an inertial methodology for improving the convergence behavior of the method. From that time on, the inertial methodology was frequently employed to accelerate the convergence behavior of methods, such as the fast iterative shrinkage-thresholding algorithm (FISTA) defined as follows:

$$\begin{cases} v_n &= Tu_n, \\ t_{n+1} &= \frac{1+\sqrt{1+4t_n^2}}{2}, \\ \theta_n &= \frac{t_n - 1}{t_{n+1}}, \\ u_{n+1} &= v_n + \theta_n(v_n - v_{n-1}), \forall n \geq 1, \end{cases} \tag{1.12}$$

where $u_1 = v_0 \in \mathbb{R}^m, t_1 = 1$, and $T = \text{prox}_{\lambda g}(I - \lambda \nabla f)$ for $\lambda > 0$. FISTA was introduced by Beck and Teboulle [17], and they applied it to solve some image restoration problems where it was shown that the performance of FISTA was better than the existing methods in the literature.

A new accelerated viscosity algorithm (NAVA) was proposed by Puangpee and Suantai [18] for finding a common fixed point of $\{T_n\}$. It was defined as follows:

$$\begin{cases} v_n &= u_n + \theta_n(u_n - u_{n-1}), \\ w_n &= (1-\sigma_n)v_n + \sigma_n T_n v_n, \\ u_{n+1} &= \lambda_n f(u_n) + \delta_n T_n v_n + \gamma_n T_n w_n, \ \forall n \geq 1, \end{cases} \tag{1.13}$$

where $u_0, u_1 \in H,$ and $\{\sigma_n\}, \{\lambda_n\}, \{\delta_n\}$, and $\{\gamma_n\} \subset (0,1)$. Moreover, they obtained a strong convergence theorem of (1.13) under certain control conditions.

Polyak [19] also highlighted how multi-step inertial methods can accelerate optimization approaches, despite the fact that neither the convergence nor the rate outcome of such multi-step inertial methods are proven in [19].

After that, Q. L. Dong et al. [20] presented the general inertial Mann algorithm as follows:

$$\begin{cases} v_n &= u_n + \theta_n(u_n - u_{n-1}), \\ w_n &= u_n + \zeta_n(u_n - u_{n-1}), \\ u_{n+1} &= (1-\gamma_n)v_n + \gamma_n T(w_n) \end{cases} \tag{1.14}$$

for each $n \geq 1$, where $\{\theta_n\} \subset [0,\theta], \{\zeta_n\} \subset [0,\zeta]$ with $\theta_1 = \zeta_1 = 0$, and $\theta, \zeta \in [0,1)$.

From here on, we would like to give a some direct methods to solve problem (1.1), namely, the *Bilevel Gradient Sequential Averaging Method* (BiG-SAM) and the *inertial Bilevel Gradient Sequential Averaging Method* (iBiG-SAM).

In 2017, Sabach and Shtern [21] presented the BiG-SAM process (Algorithm 1) as follows:

Algorithm 1 BiG-SAM

Input: $u_1 \in \mathbb{R}^m, \lambda_n \in (0,1), \iota \in (0, \frac{1}{L_\varphi})$ and $s \in (0, \frac{2}{L_\phi + \sigma})$.
For $n \geq 1$ **:**
Compute:

$$\begin{cases} v_n &= \text{prox}_{\iota g}(u_n - \iota \nabla L_\varphi(u_n)), \\ u_{n+1} &= \lambda_n(u_{n-1} - s\nabla\phi(u_n)) + (1-\lambda_n)v_n. \end{cases}$$

They showed that $u_n \to u$ where u is a solution of (1.1) and (1.2).

Later, Shehu et al. [22] introduced iBiG-SAM (Algorithm 2) by utilizing an inertial technique with BiG-SAM as follows:

Algorithm 2 iBiG-SAM

Input: $u_0, u_1 \in \mathbb{R}^m, \alpha \geq 3, \lambda_n \in (0,1), \iota \in (0, \frac{2}{L_\varphi}), s \in (0, \frac{2}{L_\phi+\sigma}]$ such that $\{\lambda_n\}$ and $\{\epsilon_n\}$ satisfying the Assumption 1.1.

For $n \geq 1$ **:**

Choose: $\theta_n \in [0, \bar{\theta_n}]$ with $\bar{\theta_n}$ defined by

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

Compute:

$$\begin{cases} v_n & = u_n + \theta_n(u_n - u_{n-1}), \\ t_n & = \text{prox}_{\iota\psi}(v_n - \iota\nabla\varphi(v_n)), \\ w_n & = v_n - s\nabla\phi(v_n), \\ u_{n+1} & = \lambda_n w_n + (1-\lambda_n)t_n. \end{cases}$$

They proved a strong convergence theorem of Algorithm 2 under Assumption 1.1 as follows:

Assumption 1.1. *Suppose* $\{\lambda_n\}_{n=1}^\infty \subset (0,1)$ *and* $\{\epsilon_n\}_{n=1}^\infty$ *are positive sequences that satisfy the following conditions:*

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

(2) $\epsilon_n = o(\lambda_n)$ *, i.e.,* $\lim_{n\to\infty}(\epsilon_n/\lambda_n) = 0$.

Motivated by ongoing research in this area, we are interested in introducing a new accelerated algorithm for solving convex bilevel optimization problems and applying it to solve data classification problems.

The following describes the way this paper is organized: Section 2 contains some fundamental definitions and helpful lemmas. The main results of this work are presented in Section 3. In this part, we provide a new accelerated algorithm for solving convex bilevel optimization problems and prove its strong convergence theorem. In Section 4, we also use our main finding to solve data classification problems. Finally, Section 5 contains a conclusion of our work.

2. Materials and methods

Throughout this paper, let C be a nonempty closed convex subset of real Hilbert space $\mathcal{H}$, and let $T : C \to C$ be a mapping. Let the strong and weak convergence of $\{u_n\}$ to $u \in \mathcal{H}$ be denoted by $u_n \to u$ and $u_n \rightharpoonup u$, respectively. A point $u \in C$ is said to be a fixed point of T if $Tu = u$, and the set of all fixed points of T is denoted by $\mathfrak{F}(T)$.

A set C is said to be convex if $\alpha u + (1-\alpha)v \in C$ for all $u, v \in C$ and $\alpha \in [0,1]$.

Definition 2.1. Let $f : \mathcal{H} \to \bar{\mathbb{R}}$. Then, the function f is *convex* on C if

$$f(\lambda u + (1-\lambda)v) \leq \lambda f(u) + (1-\lambda)f(v), \ \forall u, v \in C \text{ and } \lambda \in (0,1).$$

Definition 2.2. A function $f : \mathcal{H} \to \mathbb{R}$ is *strongly convex* with constant $\sigma > 0$ if for any $u, v \in \mathcal{H}$ and $\lambda \in [0, 1]$,

$$f(\lambda u + (1 - \lambda)v) \leq \lambda f(u) + (1 - \lambda)f(v) - \tfrac{\sigma}{2}\lambda(1 - \lambda)\|u - v\|^2.$$

Definition 2.3. For a scalar-valued function $f : \mathbb{R}^m \to \mathbb{R}$, the *derivative* of f at $\bar{u}$ is denoted by $\nabla f(\bar{u}) \in \mathbb{R}^m$ and is defined as

$$\lim_{\|h\| \to 0} \frac{f(\bar{u} + h) - f(\bar{u}) - \langle \nabla f(\bar{u}), h \rangle}{\|h\|} = 0.$$

A function f is differentiable if it is differentiable at every $u \in \mathbb{R}^m$.

Definition 2.4. Let $f : \mathbb{R}^m \to \mathbb{R}$ be convex differentiable. The *gradient* of f at u denoted by $\nabla f(u)$, is defined by

$$\nabla f(u) := \begin{bmatrix} \frac{\partial f(u)}{\partial u_1} \\ \vdots \\ \frac{\partial f(u)}{\partial u_n} \end{bmatrix}.$$

Hereafter, we will recall some important definitions, lemmas, and propositions that will be used to prove our main results.

Definition 2.5. If there exists $\tau \geq 0$ such that

$$\|Tu - Tv\| \leq \tau\|u - v\|, \ \forall u, v \in C,$$

$T : C \to C$ is said to be *Lipschitzian*.

In the above inequality, if $0 \leq \tau < 1$, T is called a *contraction*, and if $\tau = 1$, T is called *nonexpansive*. It is known that $\mathfrak{F}(T)$ is closed and convex if T is nonexpansive.

Definition 2.6. Let $u \in \mathcal{H}$. An element $u^* \in C$ is said to be a *metric projection* of u on C if

$$\|u^* - u\| \leq \|v - u\|, \ \ \forall v \in C,$$

and u^* is denoted by $P_C u$.

The function P_C is called the metric projection of $\mathcal{H}$ onto C and it is well-known that P_C is nonexpansive. Moreover,

$$\langle u - P_C u, v - P_C u \rangle \leq 0, \tag{2.1}$$

holds for all $u \in \mathcal{H}$ and $v \in C$. More information and properties of P_C can be found in [23].

For finding a common fixed point of a family of nonexpansive $\{T_n\}$, we need some important conditions, one of which is the NST- condition introduced by Nakajo et al. [24].

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

$$\lim_{n\to\infty} \|u_n - T_n u_n\| = 0 \Rightarrow \lim_{n\to\infty} \|u_n - Tu_n\| = 0, \ \forall T \in \mathfrak{T}.$$

In particular, if $\mathfrak{T} = \{T\}$, then $\{T_n\}$ is said to satisfy NST- condition (I) with T.

Definition 2.7. Let $\psi \in \Gamma_0(\mathcal{H})$ and $t > 0$. The *proximitor* of $t\psi$ at $v \in \mathcal{H}$, denoted by $\mathrm{prox}_{t\psi}(v)$, is defined as

$$\mathrm{prox}_{t\psi}(v) = \underset{u\in\mathcal{H}}{\arg\min}\left\{\psi(u) + \frac{\|u-v\|^2}{2t}\right\}.$$

The forward-backward operator T of φ and ψ with respect to t is denoted by $T := \mathrm{prox}_{t\psi}(I - t\nabla\varphi)$. Futhermore, if $t \in (0, 2/L_\varphi)$, where L_φ is the Lipschitz gradient of φ, it is generally known that T is nonexpansive.

The following lemma is required to prove our main results.

Lemma 2.8. *[25, 27] The following holds with $u, w \in \mathcal{H}$ and any arbitrary real number $\lambda \in [0, 1]$:*

(1) $\|\lambda u + (1-\lambda)w\|^2 = \lambda\|u\|^2 + (1-\lambda)\|w\|^2 - \lambda(1-\lambda)\|u-w\|^2$;

(2) $\|u \pm w\|^2 = \|u\|^2 \pm 2\langle u, w\rangle + \|w\|^2$;

(3) $\|u + w\|^2 \leq \|u\|^2 + 2\langle w, u + w\rangle$.

The following equality holds for all $u, v, w \in \mathcal{H}$ by utilizing Lemma 2.8 (*1*):

$$\|\alpha u + \beta v + \gamma w\|^2 = \alpha\|u\|^2 + \beta\|v\|^2 + \gamma\|w\|^2 - \alpha\beta\|u-v\|^2 - \beta\gamma\|v-w\|^2 - \alpha\gamma\|u-w\|^2, \tag{2.2}$$

where $\alpha, \beta, \gamma \in [0, 1]$ with $\alpha + \beta + \gamma = 1$.

Lemma 2.9. *[26] Let $\psi \in \Gamma_0(\mathcal{H})$, and $\varphi : \mathcal{H} \to \mathbb{R}$ be convex differentiable such that $\nabla\varphi$ is L_φ-Lipschitzian with $L_\varphi > 0$. Let $\{\mathfrak{c}_n\} \subset (0, 2/L_\varphi)$ and $\mathfrak{c} \in (0, 2/L_\varphi)$ such that $\mathfrak{c}_n \to \mathfrak{c}$. Define $T_n := \mathrm{prox}_{\mathfrak{c}_n\psi}(I - \mathfrak{c}_n\nabla\varphi)$, then $\{T_n\}$ satisfies NST-condition (I) with T, where $T := \mathrm{prox}_{\mathfrak{c}\psi}(I - \mathfrak{c}\nabla\varphi)$.*

Lemma 2.10. *[18] Let T be a nonexpansive mapping, and $\{T_n\}$ be a family of nonexpansive mappings such that $\emptyset \neq \mathfrak{F}(T) \subset \bigcap_{n=1}^{\infty} \mathfrak{F}(T_n)$. For any subsequences $\{k\}$ of $\{n\}$, if $\{T_n\}$ satisfies NST-condition (I) with T, then $\{T_k\}$ also satisfies NST-condition (I) with T.*

Proposition 2.11. *[21] Let ϕ be a strongly convex differentiable function from $\mathbb{R}^m$ into $\mathbb{R}$ with parameter $\sigma > 0$ such that $\nabla\phi$ is L_ϕ-Lipschitzian. Define $T_s := I - s\nabla\phi$, where I is the identity mapping. Then, T_s is a contraction for all $s \leq \frac{2}{L_\phi+\sigma}$, that is*

$$\|u - s\nabla\phi(u) - (v - s\nabla\phi(v))\| \leq \sqrt{1 - \frac{2s\sigma L_\phi}{\sigma + L_\phi}}\|u - v\|, \ \forall u, v \in \mathbb{R}^m.$$

Lemma 2.12. *[28] Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive mapping with $\mathfrak{F}(T) \neq \emptyset$. Then, $I - T$ is demiclosed at zero, that is*

$$\|u_n - Tu_n\| \to 0 \Rightarrow u \in \mathfrak{F}(T),$$

for any sequences $\{u_n\} \in \mathcal{H}$ such that $u_n \rightharpoonup u \in \mathcal{H}$.

Lemma 2.13. *[29, 30] Let $\{p_n\}, \{\xi_n\}$ be sequences of nonnegative real numbers, $\{\alpha_n\}$ a sequence in $[0, 1]$, and $\{q_n\}$ a sequence of real numbers such that*

$$p_{n+1} \leq (1 - \alpha_n)p_n + \alpha_n q_n + r_n,$$

for all $n \in \mathbb{N}$. If the following conditions hold, *(1)* $\sum_{n=1}^{\infty} \alpha_n = \infty$;

(2) $\sum_{n=1}^{\infty} r_n < \infty$;

(3) $\limsup_{n\to\infty} q_n \leq 0$;

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

Lemma 2.14. *[31] Let $\{\vartheta_n\}$ be a real sequence of numbers that does not decrease at infinity in such a way that there is a subsequence $\{\vartheta_{n_i}\}$ such that $\vartheta_{n_k} < \vartheta_{n_k+1}$ for all $k \in \mathbb{N}$. Define the sequence $\{\pi(n)\}_{n\geq n_0}$ by*

$$\pi(n) := \max\{j \leq n : \vartheta_j < \vartheta_{j+1}\},$$

where $n_0 \in \mathbb{N}$ such that $\{j \leq n_0 : \vartheta_j < \vartheta_{j+1}\} \neq \emptyset$. Then, the following hold:

(1) $\pi(n_0) \leq \pi(n_0 + 1) \leq \ldots$ *and* $\pi(n) \to \infty$;

(2) $\vartheta_{\pi(n)} \leq \vartheta_{\pi(n)+1}$ *and* $\vartheta_n \leq \vartheta_{\pi(n)+1}$ *for all* $n \geq n_0$.

3. Results

In this part, we propose a new accelerated algorithm for finding a common fixed point of a family of nonexpansive mappings in $\mathcal{H}$ by using the two-step inertial methodology with the viscosity approximation method. Second, we establish a strong convergence theorem under relevant conditions.

To do this, we start by introducing a new two-step inertial algorithm for estimating a solution for a common fixed point problem (Algorithm 3).

Throughout this section, let $\{T_n\}$ be a family of nonexpansive mappings on $\mathcal{H}$ into itself. Let f be a τ-contraction mapping on $\mathcal{H}$ with $\tau \in (0, 1)$, $\{\eta_n\} \subset (0, \infty)$, and $\{\lambda_n\}, \{\delta_n\}, \{\iota_n\} \subset (0, 1)$.

Algorithm 3 Two-step Inertial and Viscosity Algorithm

Initialize : Take $u_1, u_0, u_{-1} \in \mathcal{H}$. Let $\{\mu_n\} \subset (0, \infty)$ and $\{\rho_n\} \subset (-\infty, 0)$.

For $n \geq 1$ **:**

Set

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

$$\zeta_n = \begin{cases} \max\{\rho_n, \frac{-\eta_n\lambda_n}{\|u_n - u_{n-1}\|}\} & \text{if } u_n \neq u_{n-1}; \\ \rho_n & \text{otherwise.} \end{cases}$$

Compute

$$\begin{cases} v_n &= u_n + \theta_n(u_n - u_{n-1}) + \zeta_n(u_{n-1} - u_{n-2}), \\ w_n &= \iota_n f(v_n) + (1 - \iota_n)T_n v_n, \\ u_{n+1} &= (1 - \lambda_n - \delta_n)v_n + \lambda_n T_n w_n + \delta_n T_n v_n. \end{cases}$$

Next, we prove a strong convergence theorem of Algorithm 3.

Theorem 3.1. *Let $T : \mathcal{H} \to \mathcal{H}$ be a nonexpansive mapping with $\mathfrak{F}(T) \neq \emptyset$. Assume that $\emptyset \neq \mathfrak{F}(T) \subset \bigcap_{n=1}^{\infty} \mathfrak{F}(T_n)$ such that $\{T_n\}$ satisfies NST-condition (I) with T. Let $\{u_n\}$ be a sequence generated by Algorithm 3 such that the following additional conditions hold:*

(1) $\lim_{n\to\infty} \eta_n = 0$,

(2) $\lim_{n\to\infty} \iota_n = 0$ *and* $\sum_{n=1}^{\infty} \iota_n = \infty$,

(3) $0 < a < \lambda_n$ *for some* $a \in \mathbb{R}$,

(4) $0 < b < \delta_n < \lambda_n + \delta_n < c < 1$ *for some* $b, c \in \mathbb{R}$,

then the sequence $\{u_n\} \to u \in \mathfrak{F}(T)$ *such that* $u = P_{\mathfrak{F}(T)} f(u)$.

Proof. Let $u \in \mathfrak{F}(T)$ such that $u = P_{\mathfrak{F}(T)} f(u)$. First, we show that $\{u_n\}$ is bounded. According to the definitions of v_n and w_n, we obtain

$$\begin{aligned} \|v_n - u\| &= \|u_n + \theta_n(u_n - u_{n-1}) + \zeta_n(u_{n-1} - u_{n-2}) - u\| \\ &\leq \|u_n - u\| + \theta_n\|u_n - u_{n-1}\| + |\zeta_n|\cdot\|u_{n-1} - u_{n-2}\|, \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} \|w_n - u\| &= \|\iota_n f(v_n) + (1 - \iota_n)T_n v_n - u\| \\ &\leq \iota_n\|f(v_n) - f(u)\| + \iota_n\|f(u) - u\| + (1 - \iota_n)\|T_n v_n - u\| \\ &\leq \iota_n\tau\|v_n - u\| + \iota_n\|f(u) - u\| + (1 - \iota_n)\|v_n - u\| \\ &= (1 - (1 - \tau)\iota_n)\|v_n - u\| + \iota_n\|f(u) - u\| \\ &\leq \|v_n - u\| + \iota_n\|f(u) - u\|. \end{aligned} \tag{3.2}$$

We also know from (3.1) and (3.2) that

$$\begin{aligned} \|u_{n+1} - u\| &= \|\lambda_n T_n w_n + \delta_n T_n v_n + (1 - \lambda_n - \delta_n)v_n - u\| \\ &\leq \lambda_n\|T_n w_n - u\| + \delta_n\|T_n v_n - u\| + (1 - \lambda_n - \delta_n)\|v_n - u\| \\ &\leq \lambda_n\|w_n - u\| + \delta_n\|v_n - u\| + (1 - \lambda_n - \delta_n)\|v_n - u\| \\ &= \lambda_n\|w_n - u\| + (1 - \lambda_n)\|v_n - u\| \\ &\leq \lambda_n((1 - (1 - \tau)\iota_n)\|v_n - u\| + \iota_n\|f(u) - u\|) \\ &\quad +(1 - \lambda_n)\|v_n - u\| \\ &= (1 - (1 - \tau)\lambda_n\iota_n)\|v_n - u\| + \lambda_n\iota_n\|f(u) - u\| \\ &\leq (1 - (1 - \tau)\lambda_n\iota_n)\|u_n - u\| \\ &\quad +(1 - (1 - \tau)\lambda_n\iota_n)[\theta_n\|u_n - u_{n-1}\| + |\zeta_n|\cdot\|u_{n-1} - u_{n-2}\|] \\ &\quad +\lambda_n\iota_n\|f(u) - u\| \\ &= (1 - (1 - \tau)\lambda_n\iota_n)\|u_n - u\| \\ &\quad +(1 - \tau)\lambda_n\iota_n\tfrac{(1-(1-\tau)\lambda_n\iota_n)}{(1-\tau)\iota_n}\cdot\tfrac{\theta_n}{\lambda_n}\|u_n - u_{n-1}\| \\ &\quad +(1-\tau)\lambda_n\iota_n[\tfrac{(1-(1-\tau)\lambda_n\iota_n)}{(1-\tau)\iota_n}\cdot\tfrac{|\zeta_n|}{\lambda_n}\|u_{n-1} - u_{n-2}\| + \tfrac{\|f(u)-u\|}{1-\tau}]. \end{aligned} \tag{3.3}$$

In accordance with Assumption (1) and the definition of θ_n and ζ_n, we have

$$\frac{\theta_n}{\lambda_n}\|u_n - u_{n-1}\| \to 0 \text{ and } \frac{|\zeta_n|}{\lambda_n}\|u_{n-1} - u_{n-2}\| \to 0 \text{ as } n \to \infty.$$

Then, positive constants M_1, M_2 exist such that

$$\frac{\theta_n}{\lambda_n}\|u_n - u_{n-1}\| \leq M_1 \text{ and } \frac{|\zeta_n|}{\lambda_n}\|u_{n-1} - u_{n-2}\| \leq M_2.$$

From (3.3), we have

$$\begin{aligned}
\|u_{n+1}-u\| &\leq (1-(1-\tau)\lambda_n\iota_n)\|u_n-u\| \\
&\quad +(1-\tau)\lambda_n\iota_n\tfrac{\xi}{1-\tau}\cdot\tfrac{\theta_n}{\lambda_n}\|u_n-u_{n-1}\| \\
&\quad +(1-\tau)\lambda_n\iota_n[\tfrac{\xi}{1-\tau}\cdot\tfrac{|\zeta_n|}{\lambda_n}\|u_{n-1}-u_{n-2}\|+\tfrac{\|f(u)-u\|}{1-\tau}] \\
&\leq (1-(1-\tau)\lambda_n\iota_n)\|u_n-u\|+(1-\tau)\lambda_n\iota_n[\tfrac{\xi(M_1+M_2)+\|f(u)-u\|}{1-\tau}] \\
&\leq \max\{\|u_n-u\|,\tfrac{\xi(M_1+M_2)+\|f(u)-u\|}{1-\tau}\} \\
&\quad \vdots \\
&\leq \max\{\|u_1-u\|,\tfrac{\xi(M_1+M_2)+\|f(u)-u\|}{1-\tau}\},
\end{aligned}$$

where $\xi=\sup\{\frac{1-(1-\tau)\lambda_n\iota_n}{\iota_n}\}$. As a result, $\{u_n\}$ is bounded. Moreover, $\{v_n\}$, $\{w_n\}$, $\{f(u_n)\}$, and $\{T_nv_n\}$ are all bounded.

Using Lemma 2.8 (2), we also have

$$\begin{aligned}
\|v_n-u\|^2 &= \|u_n+\theta_n(u_n-u_{n-1})+\zeta_n(u_{n-1}-u_{n-2})-u\|^2 \\
&\leq \|u_n-u\|^2+2\theta_n\langle u_n-u,u_n-u_{n-1}\rangle+2\zeta_n\langle u_n-u,u_{n-1}-u_{n-2}\rangle \\
&\quad +\|\theta_n(u_n-u_{n-1})+\zeta_n(u_{n-1}-u_{n-2})-u\|^2 \\
&\leq \|u_n-u\|^2+2\theta_n\langle u_n-u,u_n-u_{n-1}\rangle+2\zeta_n\langle u_n-u,u_{n-1}-u_{n-2}\rangle \\
&\quad +\theta_n^2\|u_n-u_{n-1}\|^2+2\theta_n\zeta_n\langle u_n-u_{n-1},u_{n-1}-u_{n-2}\rangle \\
&\quad +\zeta_n^2\|u_{n-1}-u_{n-2}\|^2 \\
&\leq \|u_n-u\|^2+2\theta_n\|u_n-u\|\cdot\|u_n-u_{n-1}\| \\
&\quad +2|\zeta_n|\cdot\|u_n-u\|\cdot\|u_{n-1}-u_{n-2}\|+\theta_n^2\|u_n-u_{n-1}\|^2 \\
&\quad +2\theta_n|\zeta_n|\cdot\|u_n-u_{n-1}\|\cdot\|u_{n-1}-u_{n-2}\|+\zeta_n^2\|u_{n-1}-u_{n-2}\|^2.
\end{aligned} \tag{3.4}$$

Using Lemma 2.8 (3) and (3.4), we have

$$\begin{aligned}
\|u_{n+1}-u\|^2 &= \|\lambda_nT_nw_n+\delta_nT_nv_n+(1-\lambda_n-\delta_n)v_n-u\|^2 \\
&\leq \lambda_n\|T_nw_n-u\|^2+\delta_n\|T_nv_n-u\|^2+(1-\lambda_n-\delta_n)\|v_n-u\|^2 \\
&\leq \lambda_n\|w_n-u\|^2+\delta_n\|v_n-u\|^2+(1-\lambda_n-\delta_n)\|v_n-u\|^2 \\
&= \lambda_n\|w_n-u\|^2+(1-\lambda_n)\|v_n-u\|^2 \\
&= \lambda_n\|\iota_nf(v_n)+(1-\iota_n)T_nv_n-u\|^2+(1-\lambda_n)\|v_n-u\|^2 \\
&\leq \lambda_n\|\iota_n(f(v_n)-f(u))+(1-\iota_n)(T_nv_n-u)\|^2 \\
&\quad +2\lambda_n\iota_n\langle f(u)-u,w_n-u\rangle+(1-\lambda_n)\|v_n-u\|^2 \\
&\leq \lambda_n[\iota_n\|f(v_n)-f(u)\|^2+(1-\iota_n)\|T_nv_n-u\|^2] \\
&\quad +2\lambda_n\iota_n\langle f(u)-u,w_n-u\rangle+(1-\lambda_n)\|v_n-u\|^2 \\
&\leq \lambda_n\iota_n\tau\|v_n-u\|^2+\lambda_n(1-\iota_n)\|v_n-u\|^2 \\
&\quad +2\lambda_n\iota_n\langle f(u)-u,w_n-u\rangle+(1-\lambda_n)\|v_n-u\|^2 \\
&= (1-(1-\tau)\lambda_n\iota_n)\|v_n-u\|^2+2\lambda_n\iota_n\langle f(u)-u,w_n-u\rangle \\
&= (1-(1-\tau)\lambda_n\iota_n)\Big[\|u_n-u\|^2+2\theta_n\|u_n-u\|\cdot\|u_n-u_{n-1}\| \\
&\quad +2|\zeta_n|\cdot\|u_n-u\|\cdot\|u_{n-1}-u_{n-2}\|+\theta_n^2\|u_n-u_{n-1}\|^2 \\
&\quad +2\theta_n|\zeta_n|\cdot\|u_n-u_{n-1}\|\cdot\|u_{n-1}-u_{n-2}\|+\zeta_n^2\|u_{n-1}-u_{n-2}\|^2\Big] \\
&\quad +2\lambda_n\iota_n\langle f(u)-u,w_n-u\rangle.
\end{aligned} \tag{3.5}$$

Since

$$\theta_n\|u_n-u_{n-1}\| = \lambda_n\cdot\tfrac{\theta_n}{\lambda_n}\|u_n-u_{n-1}\|\to 0$$

and

$$|\zeta_n|\cdot\|u_{n-1}-u_{n-2}\| \;=\; \lambda_n\cdot\frac{|\zeta_n|}{\lambda_n}\cdot\|u_{n-1}-u_{n-2}\|\to 0$$

as $n\to\infty$, there exist positive constants M_3, M_4 such that

$$\begin{aligned}\theta_n\|u_n-u_{n-1}\| &\le M_3,\\ |\zeta_n|\cdot\|u_{n-1}-u_{n-2}\| &\le M_4.\end{aligned}$$

It follows from (3.5) that

$$\begin{aligned}\|u_{n+1}-u\|^2 \le\;& (1-(1-\tau)\lambda_n\iota_n)\|u_n-u\|^2\\ &+(1-(1-\tau)\lambda_n\iota_n)\theta_n\|u_n-u_{n-1}\|\\ &\times(2\|u_n-u\|+\theta_n\|u_n-u_{n-1}\|+|\zeta_n|\cdot\|u_{n-1}-u_{n-2}\|)\\ &+(1-(1-\tau)\lambda_n\iota_n)\,|\zeta_n|\cdot\|u_{n-1}-u_{n-2}\|\\ &\times(2\|u_n-u\|+|\zeta_n|\cdot\|u_{n-1}-u_{n-2}\|)\\ &+2\lambda_n\iota_n\langle f(u)-u, w_n-u\rangle\\ \le\;& (1-(1-\tau)\lambda_n\iota_n)\|u_n-u\|^2\\ &+(1-(1-k)\lambda_n\iota_n)\Big[5M_5\theta_n\|u_n-u_{n-1}\|+3M_5\,|\zeta_n|\cdot\|u_{n-1}-u_{n-2}\|\Big]\\ &+2\lambda_n\iota_n\langle f(u)-u, w_n-u\rangle\\ \le\;& (1-(1-\tau)\lambda_n\iota_n)\|u_n-u\|^2\\ &+(1-\tau)\lambda_n\iota_n\Big[\frac{5M_5\xi}{1-\tau}\cdot\frac{\theta_n}{\lambda_n}\|u_n-u_{n-1}\|+\frac{3M_5\xi}{1-\tau}\cdot\frac{|\zeta_n|}{\lambda_n}\|u_{n-1}-u_{n-2}\|\\ &+\frac{2}{1-\tau}\langle f(u)-u, w_n-u\rangle\Big],\end{aligned}\tag{3.6}$$

where $M_5=\max\{\sup_n\|u_n-u\|, M_3, M_4\}$. From (3.6), we set

$$p_n := \|u_n-u\|^2,\ \alpha_n := (1-\tau)\lambda_n\iota_n$$

and

$$q_n := \frac{5M_5\xi}{1-\tau}\theta_n\|u_n-u_{n-1}\|+\frac{3M_5\xi}{1-\tau}\,|\zeta_n|\cdot\|u_{n-1}-u_{n-2}\|+\frac{2}{1-\tau}\langle f(u)-u, w_n-u\rangle.$$

Hence, we obtain

$$p_{n+1}\le(1-\alpha_n)p_n+\alpha_n q_n. \tag{3.7}$$

After that, we examine the following two cases:

Case 1. Assume there is an $n_0\in\mathbb{N}$ such that the sequence $\{\|u_n-u\|\}_{n\ge n_0}$ is nonincreasing. As a result, $\{\|u_n-u\|\}$ converges since it has boundaries from below by 0. We infer that $\sum_{n=1}^{\infty}\alpha_n=\infty$, by using Assumptions (2) and (3). Then, using Lemma 2.13, we assert that

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

Indeed, by (3.2), we have

$$\begin{aligned}\|w_n-u\|^2-\|v_n-u\|^2 &\le (\|v_n-u\|+\iota_n\|f(u)-u\|)^2-\|v_n-u\|^2\\ &= 2\iota_n\|v_n-u\|\cdot\|f(u)-u\|+\iota_n^2\|f(u)-u\|^2.\end{aligned}\tag{3.8}$$

By Lemma 2.8 (1), (3.4), and (3.8), we have

$$
\begin{aligned}
\|u_{n+1}-u\|^2 &= \|\lambda_n T_n w_n + \delta_n T_n v_n + (1-\lambda_n-\delta_n)v_n - u\|^2 \\
&\leq \lambda_n\|T_n w_n - u\|^2 + \delta_n\|T_n v_n - u\|^2 + (1-\lambda_n-\delta_n)\|v_n - u\|^2 \\
&\quad -\delta_n(1-\lambda_n-\delta_n)\|v_n - T_n v_n\|^2 \\
&\leq \lambda_n\|w_n - u\|^2 + \delta_n\|v_n - u\|^2 + (1-\lambda_n-\delta_n)\|v_n - u\|^2 \\
&\quad -\delta_n(1-\lambda_n-\delta_n)\|v_n - T_n v_n\|^2 \\
&= \lambda_n\|w_n - u\|^2 + (1-\lambda_n)\|v_n - u\|^2 \\
&\quad -\delta_n(1-\lambda_n-\delta_n)\|v_n - T_n v_n\|^2 \\
&\leq \lambda_n[\|w_n - u\|^2 - \|v_n - u\|^2] \\
&\quad +\|u_n - u\|^2 + 2\theta_n\|u_n - u\|\cdot\|u_n - u_{n-1}\| \\
&\quad +2\,|\zeta_n|\cdot\|u_n - u\|\cdot\|u_{n-1} - u_{n-2}\| + \theta_n^2\|u_n - u_{n-1}\|^2 \\
&\quad +2\theta_n|\zeta_n|\cdot\|u_n - u_{n-1}\|\cdot\|u_{n-1} - u_{n-2}\| + \zeta_n^2\|u_{n-1} - u_{n-2}\|^2 \\
&\quad -\delta_n(1-\lambda_n-\delta_n)\|v_n - T_n v_n\|^2 \\
&\leq 2\lambda_n\iota_n\|v_n - u\|\cdot\|f(u) - u\| + \lambda_n\iota_n^2\|f(u) - u\|^2 \\
&\quad +\|u_n - u\|^2 + 2\theta_n\|u_n - u\|\cdot\|u_n - u_{n-1}\| \\
&\quad +2\,|\zeta_n|\cdot\|u_n - u\|\cdot\|u_{n-1} - u_{n-2}\| + \theta_n^2\|u_n - u_{n-1}\|^2 \\
&\quad +2\theta_n|\zeta_n|\cdot\|u_n - u_{n-1}\|\cdot\|u_{n-1} - u_{n-2}\| + \zeta_n^2\|u_{n-1} - u_{n-2}\|^2 \\
&\quad -\delta_n(1-\lambda_n-\delta_n)\|v_n - T_n v_n\|^2.
\end{aligned} \tag{3.9}
$$

This implies that

$$
\begin{aligned}
\delta_n(1-\lambda_n-\delta_n)\|v_n - T_n v_n\|^2 &\leq 2\lambda_n\iota_n\|v_n - u\|\cdot\|f(u) - u\| + \lambda_n\iota_n^2\|f(u) - u\|^2 \\
&\quad +\|u_n - u\|^2 - \|u_{n+1} - u\|^2 \\
&\quad +\theta_n\|u_n - u_{n-1}\| \\
&\quad \times(2\|u_n - u\| + \theta_n\|u_n - u_{n-1}\| + 2\,|\zeta_n|\cdot\|u_{n-1} - u_{n-2}\|) \\
&\quad +|\zeta_n|\cdot\|u_{n-1} - u_{n-2}\|(2\|u_n - u\| + |\zeta_n|\cdot\|u_{n-1} - u_{n-2}\|).
\end{aligned} \tag{3.10}
$$

Assumptions (2) and (4), as well as the convergence of the sequences $\{\|u_n - u\|\}$ and the fact that $\theta_n\|u_n - u_{n-1}\| \to 0$ and $|\zeta_n|\cdot\|u_{n-1} - u_{n-2}\| \to 0$, imply that

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

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

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

As a result of the definition of v_n and w_n, we have

$$
\begin{aligned}
\|v_n - w_n\| &= \|v_n - \iota_n f(v_n) - (1-\iota_n)T_n v_n\| \\
&\leq \iota_n\|f(v_n) - v_n\| + (1-\iota_n)\|T_n v_n - v_n\|.
\end{aligned} \tag{3.13}
$$

We can conclude from (3.11) and Assumption (2) that

$$\|v_n - w_n\| \to 0 \text{ as } n \to \infty. \tag{3.14}$$

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

$$\begin{aligned} \|u_{n+1} - v_n\| &\le \|u_{n+1} - T_n v_n\| + \|T_n v_n - v_n\| \\ &= \|\lambda_n T_n w_n + \delta_n T_n v_n + (1 - \lambda_n - \delta_n) v_n - T_n v_n\| \\ &\quad + \|T_n v_n - v_n\| \\ &\le \lambda_n \|T_n w_n - T_n v_n\| + (1 - \lambda_n - \delta_n)\|T_n v_n - v_n\| \\ &\quad + \|T_n v_n - v_n\| \\ &\le \lambda_n \|w_n - v_n\| + (2 - \lambda_n - \delta_n)\|T_n v_n - v_n\|, \end{aligned} \tag{3.15}$$

which implies

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

We can also conclude the following fact from the definition of v_n:

$$\|v_n - u_n\| = \theta_n \|u_n - u_{n-1}\| + |\delta_n| \cdot \|u_{n-1} - u_{n-2}\| \to 0 \text{ as } n \to \infty. \tag{3.17}$$

Hence,

$$\|u_{n+1} - u_n\| \le \|u_{n+1} - v_n\| + \|v_n - u_n\|. \tag{3.18}$$

Set

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

So, there is a subsequence $\{w_{n_k}\}$ of $\{w_n\}$ such that

$$\mathcal{V} = \lim_{k\to\infty} \langle f(u) - u, w_{n_k} - u \rangle. \tag{3.20}$$

Because $\{w_{n_k}\}$ is bounded, there must be a subsequence $\{w_{n'_k}\}$ of $\{w_{n_k}\}$ that satisfies $w_{n'_k} \rightharpoonup w \in \mathcal{H}$. We can assume $w_{n_k} \rightharpoonup w$ and (3.20) hold without losing generality.

We may conclude from (3.14) that $v_{n_k} \rightharpoonup w$, and we obtain that $w \in \mathfrak{F}(T)$ by using that fact and Lemma 2.12. Furthermore, we obtain the following fact by using $u = P_{\mathfrak{F}(T)} f(u)$ and (2.1):

$$\mathcal{V} = \lim_{k\to\infty} \langle f(u) - u, w_{n_k} - u \rangle = \langle f(u) - u, w - u \rangle \le 0. \tag{3.21}$$

Hence,

$$\mathcal{V} = \limsup_{n\to\infty} \langle f(u) - u, w_n - u \rangle \le 0, \tag{3.22}$$

which implies $\limsup_{n\to\infty} q_n \le 0$ by using $\theta_n \|u_n - u_{n-1}\| \to 0$ and $|\zeta_n| \cdot \|u_{n-1} - u_{n-2}\| \to 0$.

Using Lemma 2.13, we can conclude that $u_n \to u$.

Case 2. Assume that for any n_0, the sequence $\{\|u_n - u\|\}_{n \ge n_0}$ is not monotonically nonincreasing. We define

$$\vartheta_n := \|u_n - u\|^2.$$

So, there is a subsequence $\{\vartheta_{n_k}\}$ of $\{\vartheta_n\}$ such that $\vartheta_{n_k} < \vartheta_{n_k+1}$ for all $k \in \mathbb{N}$. We define $\pi : \{n : n \ge n_0\} \to \mathbb{N}$, by

$$\pi(n) := \max\{j \in \mathbb{N} : j \le n, \vartheta_j < \vartheta_{j+1}\}.$$

For any $n \ge n_0$, we have $\vartheta_{\pi(n)} \le \vartheta_{\pi(n)+1}$ by Lemma 2.14, that is

$$\|u_{\pi(n)} - u\| \le \|u_{\pi(n)+1} - u\|. \tag{3.23}$$

As in Case 1, by applying (3.23) we obtain

$$\begin{aligned}
&\delta_{\pi(n)}(1-\lambda_{\pi(n)}-\delta_{\pi(n)})\|v_{\pi(n)}-T_{\pi(n)}v_{\pi(n)}\|^2\\
\leq\ & 2\lambda_{\pi(n)}\iota_{\pi(n)}\|v_{\pi(n)}-u\|\cdot\|f(u)-u\|+\lambda_{\pi(n)}\iota_{\pi(n)}^2\|f(u)-u\|^2\\
&+\|u_{\pi(n)}-u\|^2-\|u_{\pi(n)+1}-u\|^2\\
&+\theta_{\pi(n)}\|u_{\pi(n)}-u_{\pi(n)-1}\|\\
&\times(2\|u_{\pi(n)}-u\|+\theta_{\pi(n)}\|u_{\pi(n)}-u_{\pi(n)-1}\|+2\left|\zeta_{\pi(n)}\right|\cdot\|u_{\pi(n)-1}-u_{\pi(n)-2}\|)\\
&+\left|\zeta_{\pi(n)}\right|\cdot\|u_{\pi(n)-1}-u_{\pi(n)-2}\|(2\|u_{\pi(n)}-u\|+\left|\zeta_{\pi(n)}\right|\cdot\|u_{\pi(n)-1}-u_{\pi(n)-2}\|)\\
\leq\ & 2\lambda_{\pi(n)}\iota_{\pi(n)}\|v_{\pi(n)}-u\|\cdot\|f(u)-u\|+\lambda_{\pi(n)}\iota_{\pi(n)}^2\|f(u)-u\|^2\\
&+\theta_{\pi(n)}\|u_{\pi(n)}-u_{\pi(n)-1}\|\\
&\times(2\|u_{\pi(n)}-u\|+\theta_{\pi(n)}\|u_{\pi(n)}-u_{\pi(n)-1}\|+2\left|\zeta_{\pi(n)}\right|\cdot\|u_{\pi(n)-1}-u_{\pi(n)-2}\|)\\
&+\left|\zeta_{\pi(n)}\right|\cdot\|u_{\pi(n)-1}-u_{\pi(n)-2}\|(2\|u_{\pi(n)}-u\|+\left|\zeta_{\pi(n)}\right|\cdot\|u_{\pi(n)-1}-u_{\pi(n)-2}\|),
\end{aligned} \tag{3.24}$$

which implies

$$\|v_{\pi(n)}-T_{\pi(n)}v_{\pi(n)}\|\to 0 \text{ as } n\to\infty. \tag{3.25}$$

Similar to the proof in Case 1, we get

$$\|v_{\pi(n)}-w_{\pi(n)}\|\to 0, \tag{3.26}$$

$$\|u_{\pi(n)+1}-v_{\pi(n)}\|\to 0, \tag{3.27}$$

and

$$\|v_{\pi(n)}-u_{\pi(n)}\|\to 0, \tag{3.28}$$

as $n\to\infty$, and so

$$\|u_{\pi(n)+1}-u_{\pi(n)}\|\to 0 \text{ as } n\to\infty. \tag{3.29}$$

As in Case 1, we then demonstrate that $\limsup_{n\to\infty}\langle f(u)-u, w_{\pi(n)}-u\rangle\leq 0$. Set

$$\mathcal{V}=\limsup_{n\to\infty}\langle f(u)-u, w_{\pi(n)}-u\rangle. \tag{3.30}$$

There exists a subsequence $\{w_{\pi(t)}\}$ of $\{w_{\pi(n)}\}$ such that $w_{\pi(t)}\rightharpoonup w\in\mathcal{H}$ and

$$\mathcal{V}=\lim_{t\to\infty}\langle f(u)-u, w_{\pi(t)}-u\rangle. \tag{3.31}$$

By Lemma 2.10, $\{T_{\pi(t)}\}$ satisfies NST-condition (I) with T. Due to inequality (3.24), $\|v_{\pi(t)}-T_{\pi(t)}v_{\pi(t)}\|\to 0$, and we obtain

$$\|v_{\pi(t)}-Tv_{\pi(t)}\|\to 0 \text{ as } t\to\infty. \tag{3.32}$$

As in Case 1, we can conclude from (3.25) that $v_{\pi(t)}\rightharpoonup w$, and $w\in\mathfrak{F}(T)$. Using $u=P_{\mathfrak{F}(T)}f(u)$ and (2.1), we obtain

$$\mathcal{V}=\lim_{t\to\infty}\langle f(u)-u, w_{\pi(t)}-u\rangle=\langle f(u)-u, w-u\rangle\leq 0. \tag{3.33}$$

Then,

$$\mathcal{V}=\limsup_{n\to\infty}\langle f(u)-u, w_{\pi(n)}-u\rangle\leq 0. \tag{3.34}$$

Since $\vartheta_{\pi(n)}\leq\vartheta_{\pi(n)+1}$, and from (3.6) along with $(1-\tau)\lambda_{\pi(n)}\iota_{\pi(n)}>0$, we obtain

$$\begin{aligned}\|u_{\pi(n)} - u\|^2 \quad \le \quad & \frac{5M_5\xi}{1-\tau}\frac{\theta_{\pi(n)}}{\lambda_{\pi(n)}}\|u_{\pi(n)} - u_{\pi(n)-1}\| \\ & + \frac{3M_5\xi}{1-\tau}\frac{|\zeta_{\pi(n)}|}{\lambda_{\pi(n)}}\|u_{\pi(n)-1} - u_{\pi(n)-2}\| \\ & + \frac{2}{1-\tau}\langle f(u) - u, w_{\pi(n)} - u\rangle. \end{aligned} \tag{3.35}$$

From $\frac{\theta_{\pi(n)}}{\lambda_{\pi(n)}}\|u_{\pi(n)} - u_{\pi(n)-1}\| \to 0$, $\frac{|\zeta_{\pi(n)}|}{\lambda_{\pi(n)}}\|u_{\pi(n)-1} - u_{\pi(n)-2}\| \to 0$, and (3.34), we obtain

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

and so $\|u_{\pi(n)} - u\| \to 0$ as $n \to \infty$.

This implies by (3.29) that $\|u_{\pi(n)+1} - u\| \to 0$ as $n \to \infty$. From Lemma 2.14 (2), we get $\vartheta_n \le \vartheta_{\pi(n)+1}$, that is,

$$\|u_n - u\| \le \|u_{\pi(n)+1} - u\| \to 0 \text{ as } n \to \infty.$$

Therefore, $u_n \to u$. □

For solving the problem (1.1), we assume the following assumptions:

Assumption 3.2. *Let Φ be the set of all solutions of problem* (1.1) *where*

(1) $\phi : \mathbb{R}^m \to \mathbb{R}$ is strongly convex with parameter $\sigma_\phi > 0$,

(2) ϕ is a continuously differentiable function such that $\nabla\phi$ is Lipschitz continuous with constant L_ϕ.

For solving the problem (1.2), we assume:

Assumption 3.3. *Let Λ be a nonempty set of minimizer of problem* (1.2).

(1) $\varphi : \mathbb{R}^m \to \mathbb{R}$ is convex and continuously differentiable, and $\nabla\varphi$ is Lipschitz continuous with constant L_φ,

(2) $\psi \in \Gamma_0(\mathbb{R}^m)$.

Next, we will present an algorithm (Algorithm 4) for solving problem (1.1).

Algorithm 4 Two-step Inertial Forward-Backward Algorithm

Input : $\mathfrak{c}_n \in (0, \frac{2}{L_\varphi})$, $s \in (0, \frac{2}{L_\phi+\sigma})$.
Initialize : Take $u_1, u_0, u_{-1} \in \mathbb{R}^m$. Let $\{\mu_n\} \subset (0, \infty)$ and $\{\rho_n\} \subset (-\infty, 0)$.
For $n \ge 1$ **:**
Set

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

$$\zeta_n = \begin{cases} \max\{\rho_n, \frac{-\eta_n\lambda_n}{\|u_n-u_{n-1}\|}\} & \text{if } u_n \neq u_{n-1}; \\ \rho_n & \text{otherwise.} \end{cases}$$

Compute

$$\begin{cases} v_n &= u_n + \theta_n(u_n - u_{n-1}) + \zeta_n(u_{n-1} - u_{n-2}), \\ w_n &= \iota_n(I - s\nabla\phi)(y_n) + (1 - \iota_n)\operatorname{prox}_{\mathfrak{c}_n\psi}(I - \mathfrak{c}_n\nabla\varphi)v_n, \\ u_{n+1} &= (1 - \lambda_n - \delta_n)v_n + \lambda_n \operatorname{prox}_{\mathfrak{c}_n\psi}(I - \mathfrak{c}_n\nabla\varphi)w_n + \delta_n \operatorname{prox}_{\mathfrak{c}_n\psi}(I - \mathfrak{c}_n\nabla\varphi)v_n. \end{cases}$$

Theorem 3.4. *Let ϕ be a function satisfying Assumption 3.2, and φ and ψ be functions satisfying Assumption 3.3. Let $\{\mathfrak{c}_n\} \subset (0, \frac{2}{L_\varphi})$ and $\mathfrak{c} \in (0, \frac{2}{L_\varphi})$ such that $\mathfrak{c}_n \to \mathfrak{c}$ as $n \to \infty$. Let $\{u_n\}$ be a sequence generated by Algorithm 4 with the same conditions as in Theorem 3.1. Then, $u_n \to u \in \Phi$.*

Proof. Set $T_n = \text{prox}_{\mathfrak{c}_n\psi}(I - \mathfrak{c}_n\nabla\varphi), T = \text{prox}_{\mathfrak{c}\psi}(I - \mathfrak{c}\nabla\varphi)$ and $f = I - s\nabla\phi$. In addition, we know that T_n and T are nonexpansive mappings. We also know from Lemma 2.9 that T_n satisfie NST-condition (I) with T. According to Proposition 2.11, f is contraction with constants $\tau = \sqrt{1 - \frac{2s\sigma L_\phi}{\sigma+L_\phi}}$ and $s \leq \frac{2}{L_\phi+\sigma}$. Theorem 3.1 clearly demonstrates that $u_n \to u \in \mathfrak{F}(T)$, where $u = P_{\mathfrak{F}(T)}f(u)$. We next claim that $u \in \Phi$. By using (2.1), we have for any $v \in \mathfrak{F}(T)$

$$\begin{aligned} \langle f(u) - u, v - u\rangle &\leq 0, \\ \langle (I - s\nabla\phi)(u) - u, v - u\rangle &= 0, \\ \langle -s\nabla\phi(u), v - u\rangle &= 0, \\ \langle \nabla\phi(u), v - u\rangle &\geq 0. \end{aligned} \tag{3.36}$$

Therefore, u is a solution of problem (1.1). □

4. Application in data classifications

In this section, we utilize our algorithm as a machine learning algorithm for data classification of Parkinson's disease and diabetes, and compare its effectiveness with BiG-SAM and iBiG-SAM.

Let $\{(x_k, t_k) \in \mathbb{R}^n \times \mathbb{R}^m : k = 1, 2, \ldots, s\}$ be a training set with s samples, with x_k representing an input and t_k representing a target. The mathematical model of single-layer feedforward neuron networks (SLFNs) is given by

$$o_k = \sum_{j=1}^{h} \alpha_j g(\langle\omega_j, x_k\rangle + b_j),\ k = 1, 2, \ldots, s,$$

where o_k is an output of ELM for SLFNs, h is the number of hidden nodes, g is an activation function, b_j is the bias, and α_j and ω_j are the weight vectors connecting the j-th hidden node with the output and input node, respectively.

The *hidden-layer output matrix* denoted by $\mathbf{H}$, is given by

$$\mathbf{H} = \begin{bmatrix} g(\langle\omega_1, x_1\rangle + b_1) & \cdots & g(\langle\omega_h, x_1\rangle + b_h) \\ \vdots & \ddots & \vdots \\ g(\langle\omega_1, x_s\rangle + b_1) & \cdots & g(\langle\omega_h, x_s\rangle + b_h) \end{bmatrix}_{s\times h}.$$

The target of standard SLFNs is to approximate these s sample with zero means, that is, $\sum_{k=1}^{s} |o_k - t_k| = 0$. Then, there exists α_j, ω_j, and b_j such that

$$t_k = \sum_{j=1}^{h} \alpha_j g(\langle\omega_j, x_k\rangle + b_j),\ k = 1, 2, \ldots, s.$$

We could derive the following simple equation from the s equations above:

$$\mathbf{H}u = \mathbf{T}, \tag{4.1}$$

where $u = [\alpha_1^T, \cdots, \alpha_h^T]^T, \mathbf{T} = [t_1^T, \cdots, t_s^T]^T$.

For solving ELM, it is necessary to calculate only the u that satisfies (4.1) with random ω_j and b_j. If there is a pseudo-inverse $\mathbf{H}^+$ of $\mathbf{H}$, $u = \mathbf{H}^+\mathbf{T}$ is the solution of (4.1). If H^+ does not exists, we can obtain a solution in terms of the least squares problem, that is,

$$\min_u \|\mathbf{H}u - \mathbf{T}\|_2^2. \tag{4.2}$$

In machine learning, model fitness plays an essential role for training set accuracy. We cannot employ an overfitting model to predict unknown data; instead, we utilize the most common technique known as the least absolute shrinkage and selection operator (LASSO). It is formulated as

$$\min_u \|\mathbf{H}u - \mathbf{T}\|_2^2 + \lambda\|u\|_1, \tag{4.3}$$

where $\|\cdot\|_1$ is the l_1-norm defined by $\|(x_1, \dots, x_n)\|_1 = \sum_{i=1}^{n} |x_i|$, and $\lambda > 0$ is a regularization parameter. We may simplify problem (4.3) to problem (1.2) by setting $\varphi(u) := \|\mathbf{H}u - \mathbf{T}\|_2^2$ and $\psi(u) := \lambda\|u\|_1$. For problem (1.1), we set $\phi := \frac{1}{2}\|u\|_2^2$ with $L_\phi = 1$ and $\sigma_\phi = 1$.

In this experiment, we aim to classify the datasets of the Parkinson's disease and diabetes from UCI and Kaggle, respectively.

Parkinson's disease dataset. [33] There are 195 examples in this dataset, all of which have 22 features. We classified two types of data in this dataset.

Diabetes dataset. [34] There are 768 examples in this set, all of which have 8 features. We classified two types of data in this dataset.

In this experiment, we establish the default settings by selecting the most advantageous choice for any parameter of each algorithm in order to reach the best level of performance, as follows:

(1) For inner level: $\nabla\varphi(u) = 2\mathbf{H}^T(\mathbf{H}u - T)$ and $L_\varphi = \lambda_{\max}(\mathbf{H}^*\mathbf{H})$, the maximum eigenvalue of $\mathbf{H}^*\mathbf{H}$.

(2) For Algorithm 1 (BiG-SAM) and Algorithm 2 (iBiG-SAM):

$$\iota = \frac{1}{L_\varphi} \text{ and } \lambda_n = \frac{1}{n}.$$

(3) For Algorithm 4 (our algorithm):

$$\lambda_n = 0.5 + \tfrac{1}{33n},\ \delta_n = 0.9 - \lambda_n,\ \iota_n = \tfrac{1}{33n},\ c_n = \tfrac{1}{L_\varphi},$$
$$\eta_n = \tfrac{33\cdot 10^{20}}{n},\ \mu_n = \tfrac{n-1}{n+\alpha-1},\ \rho_n = -0.0001.$$

(4) For all algorithms:

- Regularization parameter: $\lambda = 10^{-5}$.
- Hidden nodes: $h = 30$.
- $n = 500, \alpha = 3$, and $s = 0.01$.
- 10-fold cross-validation.

The following experiment uses the Parkinson's disease and diabetes disease datasets. We compare the effectiveness of Algorithms 1, 2, and 4 at the 500th iteration, as shown in Tables 1 and 2.

Table 1. The efficacy of each algorithm at the 500th iteration with 10-fold CV on the Parkinson's disease dataset.

	Algorithm 4		**BiG-SAM**		**iBiG-SAM**	
	acc. train	acc.test	acc. train	acc.test	acc. train	acc.test
Fold 1	86.93	94.74	85.80	94.74	85.80	94.74
Fold 2	86.29	75.00	86.29	75.00	86.29	75.00
Fold 3	86.86	85.00	86.86	85.00	86.86	85.00
Fold 4	88.57	85.00	87.43	85.00	87.43	85.00
Fold 5	84.57	95.00	84.00	95.00	84.00	95.00
Fold 6	86.86	85.00	86.86	85.00	86.86	85.00
Fold 7	88.07	78.95	87.50	78.95	87.50	78.95
Fold 8	85.23	89.47	84.66	89.47	84.66	89.47
Fold 9	87.50	84.21	85.80	84.21	85.80	84.21
Fold 10	84.66	89.47	84.66	84.21	84.66	84.21
Average acc.	86.55	86.18	85.98	85.66	85.98	85.66

Table 2. The efficacy of each algorithm at the 500th iteration with 10-fold CV on the diabetes dataset.

	Algorithm 4		**BiG-SAM**		**iBiG-SAM**	
	acc. train	acc.test	acc. train	acc.test	acc. train	acc.test
Fold 1	77.46	67.11	77.02	67.11	77.02	67.11
Fold 2	76.12	71.43	75.40	71.43	75.40	71.43
Fold 3	78.15	74.03	77.13	74.03	77.28	74.03
Fold 4	75.69	68.83	74.24	67.53	74.24	67.53
Fold 5	74.38	77.92	73.23	77.92	73.23	77.92
Fold 6	74.53	84.42	73.81	83.12	73.81	83.12
Fold 7	76.56	76.62	76.12	74.03	76.12	74.03
Fold 8	75.83	79.22	75.54	79.22	75.83	79.22
Fold 9	75.11	79.22	75.11	76.62	75.11	76.62
Fold 10	75.58	75.00	73.70	68.42	73.70	68.42
Average acc.	75.94	75.38	75.13	73.94	75.17	73.94

The results of Tables 1 and 2 reveal that Algorithm 4 provides a better accuracy for data classification than the others.

5. Conclusions

We provide a new two-step inertial accelerated algorithm in this paper. First, we analyze the convergence behavior of this algorithm and establish the strong convergence theorem under relevant conditions. Next, we utilize our algorithm as a machine learning algorithm to solve data classification problems of some noncommunicable diseases and compare its efficacy with BiG-SAM and iBiG-SAM. We find that our algorithm outperforms BiG-SAM and iBig-SAM in terms of accuracy. In our future work, we would like to employ our proposed algorithm as a machine learning algorithm for prediction and classification of some noncommunicable diseases collected from the Sriphat Medical Center, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand, and we also aim to build new innovations in the form of web applications/mobile applications/computer systems for data prediction and classification of noncommunicable diseases. These applications will have benefits for hospitals, communities, and citizens in terms of screening and preventing noncommunicable diseases.

Use of AI tools declaration

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

Acknowledgments

This work was partially supported by Chiang Mai University and Fundamental Fund 2024 (FF030/2567), Chiang Mai University.

The first author would like to thank CMU Presidential Scholarship for the financial support.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. P. Thongpaen, W. Inthakon, T. Leerapun, S. Suantai, A new accelerated algorithm for convex Bilevel optimization problems and applications in data classification, *Symmetry*, **14** (2022), 2617. https://doi.org/10.3390/sym14122617

2. K. Janngam, S. Suantai, Y. J. Cho, A. Kaewkhao, A novel inertial viscosity algorithm for Bilevel optimization problems applied to classification problems, *Mathematics*, **11** (2023), 2617. https://doi.org/10.3390/math11143241

3. P. Thongsri, B. Panyanak, S. Suantai, A new accelerated algorithm based on fixed point method for convex Bilevel optimization problems with applications, *Mathematics*, **11** (2023), 702. https://doi.org/10.3390/math11030702

4. P. Sae-jia, S. Suantai, A novel algorithm for convex Bi-level optimization problems in Hilbert spaces with applications, *Thai. J. Math.*, **21** (2023), 625–645. https://doi.org/10.3390/math11143241

5. N. Parikh, S. Boyd, Proximal algorithms, *Found. Trend. Optim.*, **1** (2014), 127–239. http://dx.doi.org/10.1561/2400000003

6. W. R. Mann, Mean value methods in iteration, *P. Am. Math. Soc.*, **4** (1953), 506–510.

7. S. Reich, Weak convergence theorems for nonexpansive mappings in Banach spaces, *J. Math. Anal. Appl.*, **67** (1979), 174–276.

8. B. Halpern, Fixed points of nonexpanding maps, *Found. Trend. Optim.*, **73** (1967), 957–961. http://dx.doi.org/10.1561/2400000003

9. S. Reich, Strong convergence theorems for resolvents of accretive operators in Banach spaces, *J. Math. Anal. Appl.*, **75** (1980), 287–292.

10. S. Ishikawa, Fixed points by a new iteration method, *P. Am. Math. Soc.*, **44** (1974), 147–150. https://doi.org/10.2307/2039245

11. A. Moudafi, Viscosity approximation methods for fixed points problems, *J. Math. Anal. Appl.*, **241** (2000), 46–55. https://doi.org/10.1006/jmaa.1999.6615

12. R. P. Agarwal, D. O. Regan, Iterative construction of fixed points of nearly asymptotically nonexpansive mappings, *J. Nonlinear Convex A.*, **8** (2007), 61.

13. K. Aoyama, Y. Kimura, W. Takahashi, M. Toyoda, Approximation of common fixed points of a countable family of nonexpansive mappings in a Banach space, *Nonlinear Anal.*, **67** (2007), 2350–2360. https://doi.org/10.1155/2010/407651

14. W. Takahashi, Viscosity approximation methods for countable family of nonexpansive mapping in Banach spaces, *Nonlinear Anal.*, **70** (2009), 719–734. https://doi.org/10.1016/j.na.2008.01.005

15. C. Klin-eam, S. Suantai, Strong convergence of composite iterative schemes for a countable family of nonexpansive mappings in Banach spaces, *Nonlinear Anal.*, **73** (2010), 431–439. https://doi.org/10.1016/j.na.2010.03.034

16. B. T. Polyak, Some methods of speeding up the convergence of iteration methods, *USSR Comput. Math. Math. Phys.*, **4** (1964), 1–17. https://doi.org/10.1016/0041-5553(64)90137-5

17. A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.*, **2** (2009), 183–202. https://doi.org/10.1137/080716542

18. J. Puangpee, S. Suantai, A new accelerated viscosity iterative method for an infinite family of nonexpansive mappings with applications to image restoration problems, *Mathematics*, **8** (2020), 615. https://doi.org/10.3390/math8040615

19. B. T. Polyak, *Introduction to optimization. optimization software*, New York: Publications Division, 1987.

20. Q. L. Dong, Y. J. Cho, T. M. Rassias, General inertial Mann algorithms and their convergence analysis for nonexpansive mappings, *Appl. Nonlinear Anal.*, 2018, 175–191. https://doi.org/10.1007/978-3-319-89815-5_7

21. S. Sabach, S. Shtern, A first order method for solving convex bilevel optimization problems, *SIAM J. Optim.*, **27** (2017), 640–660. https://doi.org/10.1137/16M105592

22. Y. Shehu, P. T. Vuong, A. Zemkoho, An inertial extrapolation method for convex simple bilevel optimization, *Optim. Method. Softw.*, **36** (2021), 1–19. https://doi.org/10.48550/arXiv.1809.06250

23. K. Goebel, S. Reich, *Uniform convexity, hyperbolic geometry, and nonexpansive mappings*, New York: Marcel Dekker, 1984.

24. K. Nakajo, K. Shimoji, W. Takahashi, Strong convergence to common fixed points of families of nonexpansive mappings in Banach spaces, *J. Nonlinear Convex A.*, **8** (2007), 11.

25. W. Takahashi, *Introduction to nonlinear and convex analysis*, Yokohama Publishers, 2009.

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

27. W. Takahashi, *Nonlinear functional analysis*, Yokohama Publishers, 2000.

28. K. Goebel, W. A. Kirk, *Topic in metric fixed point theory*, Cambridge University Press, 1990.

29. K. Aoyam, Y. Yasunori, W. Takahashi, M. Toyoda, On a strongly nonexpansive sequence in a Hilbert space, *J. Nonlinear Convex A.*, **8** (2007), 471–490.

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

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

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

33. *Little Max, Parkinsons*, UCI Machine Learning Repository, 2008. Available from: `https://archive.ics.uci.edu/dataset/174/parkinsons`.

34. *Kaggle, Diabetes dataset*, Kaggle, 1990. Available from: `https://www.kaggle.com/datasets/mathchi/diabetes-data-set/data`.