symmetry

MDPI

Article

A New Accelerated Algorithm for Convex Bilevel Optimization Problems and Applications in Data Classification

Panadda Thongpaen [1], Warunun Inthakon [2,3], Taninnit Leerapun [4] and Suthep Suantai [2,3,*]

1 Graduate Ph.D's Degree Program in Mathematics, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
2 Research Group in Mathematics and Applied Mathematics, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
3 Data Science Research Center, Department of Mathematics, Chiang Mai University, Chiang Mai 50200, Thailand
4 Sriphat Medical Center, Faculty of Medicine, Chiang Mai University, Chiang Mai 50200, Thailand
* Correspondence: suthep.s@cmu.ac.th

Abstract: In the development of algorithms for convex optimization problems, symmetry plays a very important role in the approximation of solutions in various real-world problems. In this paper, based on a fixed point algorithm with the inertial technique, we proposed and study a new accelerated algorithm for solving a convex bilevel optimization problem for which the inner level is the sum of smooth and nonsmooth convex functions and the outer level is a minimization of a smooth and strongly convex function over the set of solutions of the inner level. Then, we prove its strong convergence theorem under some conditions. As an application, we apply our proposed algorithm as a machine learning algorithm for solving some data classification problems. We also present some numerical experiments showing that our proposed algorithm has a better performance than the five other algorithms in the literature, namely BiG-SAM, iBiG-SAM, aiBiG-SAM, miBiG-SAM and amiBiG-SAM.

Keywords: convex bilevel optimization problems; accelerated algorithm; common fixed point; nonexpansive mappings; classification problems

Citation: Thongpaen, P.; Inthakon, W.; Leerapun, T.; Suantai, S. A New Accelerated Algorithm for Convex Bilevel Optimization Problems and Applications in Data Classification. *Symmetry* **2022**, *14*, 2617. https://doi.org/10.3390/sym14122617

Academic Editors: Wei-Shih Du, Marko Kostić, Radu Precup and Adrian Petrusel

Received: 3 November 2022
Accepted: 30 November 2022
Published: 10 December 2022

1. Introduction

Breast cancer is the most common type of cancer in Thai women. Anxiously, although the breast cancer can be treated, the risk of developing diseases that affect the heart or blood vessels is very high.

The three most common methods for treating breast cancer are surgery, chemotherapy and radiotherapy. However, radiotherapy often involves some incidental exposure of the heart to ionizing radiation because it was discovered, in [1], that the exposure of the heart to ionizing radiation during the therapy increases the consequent rate of ischemic heart disease which begins within a few years after exposure and continues for at least 20 years. Thus, women with preexisting cardiac risk factors have higher absolute increases in risk from this therapy than other women.

Therefore, if a patient is diagnosed with heart disease early, they will be able to prevent the risks from this type of treatment. Similarly, the malignant cells of a patient can be treated before it spreads to other parts of the body when cancer is detected at an early stage. To support the diagnosis of breast cancer and heart disease, our objective in this work is developing an algorithm for such patient prediction.

It is well known that symmetry serves as the foundation for fixed-point and optimization theory and methods. We first recall the background of some mathematical models. Consider the constrained minimization problem:

$$\min_{x \in \Gamma} \mathcal{F}(x), \tag{1}$$

when $\mathcal{H}$ is a real Hilbert space, $\mathcal{F} : \mathcal{H} \to \mathbb{R}$ is a strongly convex differentiable function with convexity parameter ρ, and Γ is the nonempty set of minimizers of the unconstrained minimization problem, as in the form:

$$\min_{x}\{\phi(x) + \psi(x)\}, \tag{2}$$

where $\psi, \phi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ are proper convex and lower semicontinuous functions and ϕ is a smooth function. Problems (1) and (2) are called outer-level and inner-level problems, respectively. In [2–5], such a problem is labeled as a simple bilevel optimization problem.

In 2017, Sabach and Shtern [6] introduced the Bilevel Gradient Sequential Averaging Method (BiG-SAM) for solving (1) and (2) as defined by Algorithm 1.

Algorithm 1 BiG-SAM: Bilevel Gradient Sequential Averaging Method

1: **Initial step.** Let $x_1 \in \mathbb{R}^n$ and $\{\alpha_k\}$ is a sequence in $(0, 1]$ satisfying the conditions assumed in [7]. Select $\lambda \in \left(0, \frac{1}{L_\phi}\right]$ and $\sigma \in \left(0, \frac{2}{L_{\mathcal{F}}+\rho}\right)$ while L_ϕ is the Lipschitz gradient of ϕ and $L_{\mathcal{F}}$ is the Lipschitz gradient of $\mathcal{F}$.

2: **Step 1.** For $k \geq 1$, compute

$$\begin{cases} y_k := \mathrm{prox}_{\lambda\psi}(x_k - \lambda\nabla\phi(x_k)), \\ u_k := x_k - \sigma\nabla\mathcal{F}(x_k), \\ x_{k+1} := \alpha_k u_k + (1-\alpha_k)y_k, \end{cases}$$

where $\nabla\phi$ and $\nabla\mathcal{F}$ are gradients of ϕ and $\mathcal{F}$, respectively.

They presented that BiG-SAM appears simpler and cheaper than the method desired in [8]. Moreover, the authors in [6] used a numerical example to show that BiG-SAM outruns the method in [8] for solving problems (1) and (2). Up to this point, the algorithm in [6] seems to be the most efficient method for convex simple bilevel optimization problems.

In 2019, Shehu et al. [9] utilized the notion of an inertial technique, which was proposed by Polyak [10], to be beneficial to accelerate the convergence rate of the BiG-SAM method, called iBiG-SAM, as defined by Algorithm 2.

Algorithm 2 iBiG-SAM: Inertial with Bilevel Gradient Sequential Averaging Method

1: **Initial step.** Let L_ϕ and $L_{\mathcal{F}}$ be Lipschitz gradients of ϕ and $\mathcal{F}$, respectively. Given $\{\alpha_k\}$ be a sequence in $(0, 1), \lambda \in \left(0, \frac{2}{L_\phi}\right)$ and $\sigma \in \left(0, \frac{2}{L_{\mathcal{F}}+\rho}\right]$. Select arbitrary points $x_1, x_0 \in \mathbb{R}^n$ and $\alpha \geq 3$.

2: **Step 1.** Choose $\mu_k \in [0, \bar{\mu}_k]$ such that for $k \geq 1$,

$$\bar{\mu}_k := \begin{cases} \min\left\{\frac{k}{k+\alpha-1}, \frac{\eta_k}{\|x_k - x_{k-1}\|}\right\} & \text{if } x_k \neq x_{k-1}, \\ \frac{k}{k+\alpha-1} & \text{otherwise.} \end{cases} \tag{3}$$

3: **Step 2.** Compute

$$\begin{cases} z_k := x_k + \mu_k(x_k - x_{k-1}), \\ y_k := \mathrm{prox}_{\lambda\psi}(z_k - \lambda\nabla\phi z_k), \\ u_k := z_k - \sigma\nabla\mathcal{F}(z_k), \\ x_{k+1} := \alpha_k u_k + (1-\alpha_k)y_k, \end{cases}$$

where $\nabla\phi$ and $\nabla\mathcal{F}$ are gradients of ϕ and $\mathcal{F}$, respectively.

They also proved that the sequence $\{x_k\}$ generated by iBiG-SAM converges to the optimal solution of problems (1) and (2) under the sequence $\{\alpha_k\}$ satisfying conditions:

(1) $\lim_{k\to\infty} \alpha_k = 0$;
(2) $\sum_{k=1}^{\infty} \alpha_k = +\infty$.

The above assumptions are derived from [7] by reducing some situations.

Recently, to accelerate the convergence of the iBiG-SAM algorithm, Duan and Zhang [11] proposed three algorithms of inertial approximation methods based on the proximal gradient algorithm as defined by Algorithms 3–5.

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

---

1: **Initial step.** Let L_ϕ and $L_{\mathcal{F}}$ be Lipschitz gradients of ϕ and $\mathcal{F}$, respectively. Given $\lambda \in \left(0, \frac{2}{L_\phi}\right), \sigma \in \left(0, \frac{2}{L_{\mathcal{F}}+\rho}\right], \epsilon > 0$. Let $\{\alpha_k\}$ be a sequence in $(0,1)$ satisfying the conditions assumed in [9]. Select arbitrary points $x_1, x_0 \in \mathcal{H}$ and $\alpha \geq 3$. Set $k = 1$.

2: **Step 1.** Compute

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

3: When k is odd, choose μ_k such that $0 \leq |\mu_k| \leq \bar{\mu_k}$ with $\bar{\mu_k}$ defined by

$$\bar{\mu_k} := \begin{cases} \min\left\{\frac{k}{k+\alpha-1}, \frac{\eta_k}{\|x_k - x_{k-1}\|}\right\} & \text{if } x_k \neq x_{k-1}, \\ \frac{k}{k+\alpha-1} & \text{if } x_k = x_{k-1}. \end{cases}$$

4: When k is even, $\mu_k = 0$.

5: **Step 2.** Compute

$$\begin{cases} y_k = \text{prox}_{\lambda\psi}(z_k - \lambda\nabla\phi(z_k)), \\ u_k = z_k - \sigma\nabla\mathcal{F}(z_k), \\ x_{k+1} = \alpha_k u_k + (1-\alpha_k)y_k, \quad k \leq 1, \end{cases}$$

where $\nabla\phi$ and $\nabla\mathcal{F}$ are gradients of ϕ and $\mathcal{F}$, respectively.

6: **Step 3.** If $\|x_k - x_{k-1}\| < \epsilon$, then stop. Otherwise, set $k = k+1$ and go to Step 1.

Algorithm 4 miBiG-SAM: The multi-step inertial Bilevel Gradient Sequential Averaging Method

---

1: **Initial step.** Let L_ϕ and $L_{\mathcal{F}}$ be Lipschitz gradients of ϕ and $\mathcal{F}$, respectively. Given $\lambda_k \in \left(0, \frac{2}{L_\phi}\right), \sigma \in \left(0, \frac{2}{L_{\mathcal{F}}+\rho}\right), \epsilon > 0$ and $\alpha \geq 3$. Let $\{\alpha_k\}$ be a sequence in $(0,1)$ satisfying the conditions assumed in [9]. Select arbitrary points $x_0, x_1, \ldots, x_{2-q} \in \mathcal{H}$ and $q \in \mathbb{N}_+$. Set $k = 1$.

2: **Step 1.** Given $x_k, x_{k-1}, \ldots, x_{k-q+1}$ and compute

$$z_k = x_k + \sum_{i\in Q} \mu_{i,k}(x_{k-i} - x_{k-1-i}),$$

where $Q = \{0, 1, \ldots, q-1\}$. Choose $\mu_{i,k}$ such that $0 \leq |\mu_{i,k}| \leq \bar{\mu_k}$ with $\bar{\mu_k}$ defined by

$$\bar{\mu_k} := \begin{cases} \min\left\{\frac{k}{k+\alpha-1}, \frac{\eta_k}{\sum_{i\in Q}\|x_{k-i} - x_{k-1-i}\|}\right\} & \text{if } \sum_{i\in Q}\|x_{k-i} - x_{k-1-i}\| \neq 0, \\ \frac{k}{k+\alpha-1} & \text{otherwise.} \end{cases}$$

3: **Step 2.** Compute

$$\begin{cases} y_k = \text{prox}_{\lambda_k\psi}(z_k - \lambda_k\nabla\phi(z_k)), \\ u_k = z_k - \sigma\nabla\mathcal{F}(z_k), \\ x_{k+1} = \alpha_k u_k + (1-\alpha_k)y_k, \quad k \leq 1, \end{cases}$$

where $\nabla\phi$ and $\nabla\mathcal{F}$ are gradients of ϕ and $\mathcal{F}$, respectively.

4: **Step 3.** If $\|x_k - x_{k-1}\| < \epsilon$, then stop. Otherwise, set $k = k+1$ and go to Step 1.

Algorithm 5 amiBiG-SAM: The multi-step alternated inertial Bilevel Gradient Sequential Averaging Method

---

1: **Initial step.** Let L_ϕ and $L_\mathcal{F}$ be Lipschitz gradients of ϕ and $\mathcal{F}$, respectively. Given $\lambda_k \in \left(0, \frac{2}{L_\phi}\right), \sigma \in \left(0, \frac{2}{L_\mathcal{F}+\rho}\right], \epsilon > 0$ and $\alpha \geq 3$. Let $\{\alpha_k\}$ be a sequence in $(0,1)$ satisfying the conditions assumed in [9]. Select arbitrary points $x_0, x_1, \ldots, x_{2-q} \in \mathcal{H}$ and $q \in \mathbb{N}_+$. Set $k = 1$.

2: **Step 1.** Given $x_k, x_{k-1}, \ldots, x_{k-q+1}$ and compute

$$z_k = \begin{cases} x_k + \sum\limits_{i \in Q} \mu_{i,k}(x_{k-i} - x_{k-1-i}), & \text{if } k = \text{odd}; \\ x_k & \text{if } k = \text{even}. \end{cases}$$

where $Q = \{0, 1, \ldots, q-1\}$. Choose $\mu_{i,k}$ such that $0 \leq |\mu_{i,k}| \leq \bar{\mu_k}$ with $\bar{\mu_k}$ defined by

$$\bar{\mu_k} := \begin{cases} \min\left\{\frac{k}{k+\alpha-1}, \frac{\eta_k}{\sum_{i\in Q}\|x_{k-i}-x_{k-1-i}\|}\right\} & \text{if } \sum_{i\in Q}\|x_{k-i} - x_{k-1-i}\| \neq 0, \\ \frac{k}{k+\alpha-1} & \text{otherwise.} \end{cases}$$

3: **Step 2.** Compute

$$\begin{cases} y_k = \mathrm{prox}_{\lambda_k\psi}(z_k - \lambda_k \nabla\phi(z_k)), \\ u_k = z_k - \sigma\nabla\mathcal{F}(z_k), \\ x_{k+1} = \alpha_k u_k + (1-\alpha_k) y_k, \quad k \leq 1, \end{cases}$$

where $\nabla\phi$ and $\nabla\mathcal{F}$ are gradients of ϕ and $\mathcal{F}$, respectively.

4: **Step 3.** If $\|x_k - x_{k-1}\| < \epsilon$, then stop. Otherwise, set $k = k+1$ and go to Step 1.

The convergence behavior of Algorithms 3–5 was shown, in [11], to be better than that of BiG-SAM and iBiG-SAM.

It is known that the following variational inequality:

$$\langle \nabla\mathcal{F}(x^\star), x - x^\star \rangle \geq 0, \quad \forall x \in \Gamma \tag{4}$$

implies $x^\star$ is a solution of convex bilevel optimization problem (1); for more details, see [12]. For recent results, see [13,14] and references therein.

It is worth noting that $x^\star \in \Gamma$ can be described by fixed-point equation:

$$\mathrm{prox}_{\lambda\psi}(x^\star - \lambda\nabla\phi(x^\star)) = x^\star, \tag{5}$$

where $\lambda > 0$ and $\mathrm{prox}_{\lambda\psi}(x) = \arg\min\limits_{u\in\mathcal{H}}\left\{\psi(u) + \frac{1}{2\lambda}\|u - x\|_2^2\right\}$, which was introduced by Moreau [15]. This means that solving the bilevel problem is equivalent to finding a fixed point of the proximal operator. It is well known that the fixed point theory plays a very crucial role in solving many real-world problems, such as problems in engineering, economics, machine learning and data science, see [16–24] for more details. For the past three decades, several fixed point algorithms were introduced and studied by many authors, see [25–34]. Some of these algorithms were applied for solving various problems in images and signal processing, data classification and regression, for example, see [19–23]. In addition, fuzzy classification is another important data classification mechanism, see [35,36].

All of the works mentioned above motivate and inspire us to establish a new accelerated algorithm to solve a convex bilevel optimization problem and apply it for solving data classification problems.

We organize the paper as follows: In Section 2, we provide some basic definitions and useful lemmas used in the later section. The main results of the paper are given in Section 3. In this section, we introduce and study a new accelerated algorithm for solving a convex bilevel optimization problem and then prove a strong convergence of our proposed algorithm. After that, we apply our main results for solving a data classification problem in Section 4. Finally, a brief conclusion of the paper is given in Section 5.

2. Preliminaries

Throughout this paper, a real Hilbert space, denoted by $\mathcal{H}$, with the inner product $\langle \cdot, \cdot \rangle$, inducing the norm $\| \cdot \|$.

A mapping $T : C \to C$ is called L-Lipschitz if there exists $L > 0$ such that

$$\|Tx - Ty\| \leq L\|x - y\|, \ \forall x, y \in C \subseteq \mathcal{H}.$$

If $L \in [0, 1)$, then T is called contraction. It is called nonexpansive if $L = 1$. We denote by $F(T)$ the set of all fixed points of T, that is, $F(T) = \{x \in C : Tx = x\}$. For a sequence $\{x_k\}$ in $\mathcal{H}$, we denote the strong convergence and the weak convergence of $\{x_k\}$ to $u \in \mathcal{H}$ by $x_k \to u$ and $x_k \rightharpoonup u$, respectively.

Let $\{T_k\}$ and $\Im$ be families of nonexpansive operators from C into itself with $\emptyset \neq F(\Im) \subset \bigcap_{k=1}^{\infty} F(T_k)$, where $F(\Im)$ is the set of all common fixed points of $\Im$ and $F(T_k)$ is the set of all fixed points of T_k.

The sequence $\{T_k\}$ is said to satisfy the NST-condition (I) with $\Im$ if for every bounded sequence $\{x_k\}$ in C,

$$\lim_{k\to\infty} \|x_k - T_k x_k\| = 0 \Longrightarrow \lim_{k\to\infty} \|x_k - Tx_k\| = 0, \ \ \forall T \in \Im,$$

see [37] for more details. In particular, if $\Im = \{T\}$, then $\{T_k\}$ is a sequence satisfying NST-condition (I) with T.

Later, NST*-condition was proposed by Nakajo et al. [38] which is a weaker condition than that of NST-condition (I). A sequence $\{T_k\}$ is said to satisfy NST*-condition if for every bounded sequence $\{x_k\}$ in C, if $\lim_{k\to\infty} \|x_k - x_{k+1}\| = 0$ and $\lim_{k\to\infty} \|x_k - T_k x_k\| = 0$ imply $\omega_w(x_k) \subset \bigcap_{k=1}^{\infty} F(T_k)$, where $\omega_w(x_k)$ is the set of all weak cluster points of $\{x_k\}$. It is easy to see that if $\{T_k\}$ satisfies the NST-condition (I), then it satisfies the NST*-condition.

In a real Hilbert space $\mathcal{H}$, these properties hold: for any $u, v \in \mathcal{H}$,

(1) $\|u + v\|^2 \leq \|u\|^2 + 2\langle v, u + v\rangle$;
(2) $\|ru + (1 - r)v\|^2 = r\|u\|^2 + (1 - r)\|v\|^2 - r(1 - r)\|u - v\|^2, \quad \forall r \in [0, 1]$.

If C is a nonempty closed convex subset of $\mathcal{H}$, then for each $x \in \mathcal{H}$, there exists a unique element in C, say $P_C x$, such that

$$\|x - P_C x\| \leq \|x - y\|, \quad \forall y \in C.$$

The mapping P_C is known as the metric projection of $\mathcal{H}$ onto C and it is also nonexpansive. Moreover,

$$\langle x - P_C x, y - P_C x\rangle \leq 0 \tag{6}$$

holds for all $x \in \mathcal{H}$ and $y \in C$.

The following results are also essential for proving our main results.

Lemma 1 ([39])**.** *Let $\{u_k\}, \{t_k\}$ be nonnegative real numbers sequences, $\{v_k\}$ a sequence in $[0, 1]$ and $\{w_k\}$ a sequence of numbers such that*

$$u_{k+1} \leq (1 - v_k)u_k + v_k w_k + t_k, \quad \forall k \in \mathbb{N}.$$

If all following conditions hold:

(1) $\sum_{k=1}^{\infty} v_k = +\infty$;
(2) $\sum_{k=1}^{\infty} t_k < +\infty$;
(3) $\limsup_{k\to\infty} w_k \leq 0$.

Then, $\limsup_{k\to\infty} u_k = 0$.

Lemma 2 ([40])**.** *Let $\mathcal{H}$ be a real Hilbert space and $T : \mathcal{H} \to \mathcal{H}$ a nonexpansive mapping with $F(T) \neq \emptyset$. Then, for any sequence $\{x_k\}$ in $\mathcal{H}$ such that $x_k \rightharpoonup u \in \mathcal{H}$ and $\lim_{k\to\infty} \|x_k - Tx_k\| = 0$ imply $u \in F(T)$.*

Lemma 3 ([41]). *Let $\{\lambda_k\}$ be a sequence of real numbers that does not decrease at infinity in the sense that there exists a subsequence $\{\lambda_{k_i}\}$ of $\{\lambda_k\}$ which satisfies $\lambda_{k_i} < \lambda_{k_i+1}$ for all $i \in \mathbb{N}$. Define $\{\varphi(k)\}_{k\geq m_0}$ of integers as follows:*

$$\varphi(k) = \max\{j \leq k : \lambda_k < \lambda_{k+1}\},$$

where $m_0 \in \mathbb{N}$ such that $\{j \leq m_0 : \lambda_k < \lambda_{k+1}\} \neq \emptyset$. Then, the following hold:
(1) $\varphi(m_0) \leq \varphi(m_0+1) \leq \ldots$ and $\varphi(k) \to \infty$;
(2) $\lambda_{\varphi(k)} \leq \lambda_{\varphi(k)+1}$ and $\lambda_k \leq \lambda_{\varphi(k)+1}$ for all $k \geq m_0$.

Proposition 1 ([6]). *Suppose $\mathcal{F} : \mathcal{H} \to \mathbb{R}$ is strongly convex with convexity parameter $\rho > 0$ and continuously differentiable function such that $\nabla\mathcal{F}$ is Lipschitz continuous with constant $L_{\mathcal{F}}$. Then, the mapping $I - \sigma\nabla\mathcal{F}$ is contraction for all $\sigma \leq \frac{2}{L_{\mathcal{F}}+\rho}$, where I is the identity operator.*

Definition 1 ([15]). *Let $\psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper convex and lower semicontinuous function. The proximity operator of parameter $\lambda > 0$ of ψ at $u \in \mathcal{H}$ is denoted by $prox_{\lambda\psi}$ and it is defined by*

$$prox_{\lambda\psi}(u) = \arg\min_{v\in\mathcal{H}}\left\{\psi(v) + \frac{1}{2\lambda}\|v-u\|^2\right\}.$$

The operator $T := \mathrm{prox}_{\lambda\psi}(I - \lambda\nabla\phi)$ is known as a forward–backward operator of ϕ and ψ with respect to λ, where $\lambda > 0$ and $\nabla\phi$ is the gradient operator of function ϕ. Moreover, T is a nonexpansive mapping whenever $\lambda \in (0, \frac{2}{L_\phi})$ where L_ϕ is a Lipschitz gradient of ϕ.

Lemma 4 ([42]). *For a real Hilbert space $\mathcal{H}$, let $\psi : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ be a proper convex and lower semicontinuous function, and $\phi : \mathcal{H} \to \mathbb{R}$ be convex differentiable with gradient $\nabla\phi$ being L_ϕ-Lipschitz gradient for some $L_\phi > 0$. If $\{T_k\}$ is the family of forward–backward operators of ϕ and ψ with respect to $c_k \in \left(0, \frac{2}{L_\phi}\right)$ such that $\{c_k\}$ converges to c, then $\{T_k\}$ satisfies NST-condition (I) with T, where T is the forward–backward operator of ϕ and ψ with respect to $c \in \left(0, \frac{2}{L_\phi}\right)$.*

3. Main Results

We start this section by introducing a new common fixed point algorithm using the inertial technique together with the modified Ishikawa iteration (see [43–45] for more details) to obtain a strong convergence theorem for two countable families of nonexpansive mappings in a real Hilbert space as seen in Algorithm 6.

Algorithm 6 IVAM (I): Inertial Viscosity Approximation Method for Two Families of Nonexpansive Mappings

1: **Input.** Let $x_0, x_1 \in \mathcal{H}$, $\{\eta_k\}$ a positive sequence and $f : \mathcal{H} \to \mathcal{H}$ a contraction with constant γ. Choose $\{\alpha_k\}, \{\beta_k\}, \{\xi_k\} \subset (0,1)$ and $\theta_k \geq 0$.

2: Select $\mu_k \in (0, \bar{\mu}_k]$ such that for $k \geq 1$,

$$\bar{\mu}_k := \begin{cases} \min\left\{\theta_k, \frac{\eta_k}{\|x_k - x_{k-1}\|}\right\} & \text{if } x_k \neq x_{k-1}, \\ \theta_k & \text{otherwise.} \end{cases} \tag{7}$$

3: Compute

$$\begin{cases} z_k = x_k + \mu_k(x_k - x_{k-1}), \\ y_k = \beta_k z_k + (1-\beta_k)T_k z_k, \\ w_k = \xi_k y_k + (1-\xi_k)S_k y_k \\ x_{k+1} = \alpha_k f(w_k) + (1-\alpha_k)w_k. \end{cases}$$

Lemma 5. *Let $\{T_k\}$ and $\{S_k\}$ be two countable families of nonexpansive mappings from $\mathcal{H}$ into itself such that $\Gamma = \cap_{k=1}^{\infty} F(T_k) \bigcap \cap_{k=1}^{\infty} F(S_k) \neq \emptyset$ and let $f : \mathcal{H} \to \mathcal{H}$ be a contraction. If $\lim_{k\to\infty} \frac{\eta_k}{\alpha_k} = 0$, then the sequence $\{x_k\}$ generated by Algorithm 6 is bounded. Furthermore, $\{f(w_k)\}$, $\{w_k\}$, $\{y_k\}$ and $\{z_k\}$ are bounded.*

Proof. Let $x^\star \in \Gamma$ be such that $x^\star = P_\Gamma f(x^\star)$. Then, by the definition of z_k and y_k in Algorithm 6, for every $k \in \mathbb{N}$, we have

$$\begin{aligned} \|z_k - x^\star\| &= \|x_k + \mu_k(x_k - x_{k-1}) - x^\star\| \\ &\leq \|x_k - x^\star\| + \mu_k\|x_k - x_{k-1}\|, \end{aligned} \tag{8}$$

and

$$\begin{aligned} \|y_k - x^\star\| &\leq \beta_k\|z_k - x^\star\| + (1-\beta_k)\|T_k z_k - x^\star\| \\ &\leq \beta_k\|z_k - x^\star\| + (1-\beta_k)\|z_k - x^\star\| \\ &= \|z_k - x^\star\|. \end{aligned} \tag{9}$$

This implies

$$\begin{aligned} \|w_k - x^\star\| &\leq \xi_k\|y_k - x^\star\| + (1-\xi_k)\|S_k y_k - x^\star\| \\ &\leq \xi_k\|y_k - x^\star\| + (1-\xi_k)\|y_k - x^\star\| \\ &= \|y_k - x^\star\| \qquad (10) \\ &\leq \|z_k - x^\star\|. \qquad (11) \end{aligned}$$

It follows from (8) and (11) that

$$\begin{aligned} \|x_{k+1} - x^\star\| &= \|\alpha_k(f(w_k) - x^\star) + (1-\alpha_k)(w_k - x^\star)\| \\ &\leq \alpha_k\|f(w_k) - x^\star\| + (1-\alpha_k)\|w_k - x^\star\| \\ &= \alpha_k\|f(w_k) - f(x^\star) + f(x^\star) - x^\star\| + (1-\alpha_k)\|w_k - x^\star\| \\ &\leq \alpha_k\|f(w_k) - f(x^\star)\| + \alpha_k\|f(x^\star) - x^\star\| + (1-\alpha_k)\|w_k - x^\star\| \\ &\leq \alpha_k\gamma\|w_k - x^\star\| + \alpha_k\|f(x^\star) - x^\star\| + (1-\alpha_k)\|w_k - x^\star\| \\ &= [1-\alpha_k(1-\gamma)]\|w_k - x^\star\| + \alpha_k\|f(x^\star) - x^\star\| \\ &\leq [1-\alpha_k(1-\gamma)]\|z_k - x^\star\| + \alpha_k\|f(x^\star) - x^\star\| \\ &\leq [1-\alpha_k(1-\gamma)](\|x_k - x^\star\| + \mu_k\|x_k - x_{k-1}\|) + \alpha_k\|f(x^\star) - x^\star\| \\ &= [1-\alpha_k(1-\gamma)]\|x_k - x^\star\| + \mu_k\|x_k - x_{k-1}\| - \alpha_k(1-\gamma)\mu_k\|x_k - x_{k-1}\| \\ &\quad + \alpha_k\|f(x^\star) - x^\star\| \\ &\leq [1-\alpha_k(1-\gamma)]\|x_k - x^\star\| + \mu_k\|x_k - x_{k-1}\| + \alpha_k\|f(x^\star) - x^\star\| \\ &= [1-\alpha_k(1-\gamma)]\|x_k - x^\star\| + \alpha_k(1-\gamma)\left[\frac{\frac{\mu_k}{\alpha_k}\|x_k - x_{k-1}\| + \|f(x^\star) - x^\star\|}{1-\gamma}\right] \\ &\leq \max\left\{\|x_k - x^\star\|, \frac{\frac{\mu_k}{\alpha_k}\|x_k - x_{k-1}\| + \|f(x^\star) - x^\star\|}{1-\gamma}\right\}. \end{aligned}$$

Using $\lim\limits_{k\to\infty} \dfrac{\eta_k}{\alpha_k} = 0$ and (7), we obtain

$$\lim_{k\to\infty} \frac{\mu_k}{\alpha_k}\|x_k - x_{k-1}\| = \lim_{k\to\infty} \frac{\eta_k}{\alpha_k\|x_k - x_{k-1}\|}\|x_k - x_{k-1}\| = \lim_{k\to\infty} \frac{\eta_k}{\alpha_k} = 0.$$

Thus, there exists $0 < \overline{M}$ such that $\frac{\mu_k}{\alpha_k}\|x_k - x_{k-1}\| < \overline{M}$ for all $k \in \mathbb{N}$, which implies

$$\|x_{k+1} - x^\star\| \le \max\left\{\|x_k - x^\star\|, \frac{\overline{M} + \|f(x^\star) - x^\star\|}{1-\gamma}\right\}. \tag{12}$$

By mathematical induction, we conclude that $\|x_k - x^\star\| \le M$ for all $k \in \mathbb{N}$, where $M = \max\left\{\|x_1 - x^\star\|, \frac{\overline{M} + \|f(x^\star) - x^\star\|}{1-\gamma}\right\}$. It follows that $\{x_k\}$ is bounded. This implies that the sequences $\{f(w_k)\}, \{w_k\}, \{y_k\}$ and $\{z_k\}$ are bounded. □

We now prove a strong convergence theorem of the sequence $\{x_k\}$ generated by Algorithm 6 to solve a common fixed point problem as follows.

Theorem 1. *Let $\{T_k\}$ and $\{S_k\}$ be two countable families of nonexpansive mappings from $\mathcal{H}$ into $\mathcal{H}$ such that $\Gamma = \cap_{k=1}^\infty F(T_k) \bigcap \cap_{k=1}^\infty F(S_k) \neq \emptyset$. Let $\{x_k\}$ be a sequence generated by Algorithm 6. Suppose $\{T_k\}$ and $\{S_k\}$ satisfy NST*-conditions and the following conditions hold:*

(1) $0 < a < \alpha_k < \hat{a} < 1$;
(2) $0 < b < \beta_k < \hat{b} < 1$;
(3) $0 < c < \xi_k < \hat{c} < 1$;
(4) $\lim_{k\to\infty} \alpha_k = 0$ *and* $\sum_{k=1}^\infty \alpha_k = +\infty$;
(5) $\lim_{k\to\infty} \frac{\eta_k}{\alpha_k} = 0$,

where $a, b, c, \hat{a}, \hat{b}$ and $\hat{c}$ are real positive numbers. Then, $\{x_k\}$ converges strongly to $x^\star \in \Gamma$, where $x^\star = P_\Gamma f(x^\star)$.

Proof. Let $x^\star \in \Gamma$ be such that $x^\star = P_\Gamma f(x^\star)$. It follows from (11) that

$$\begin{aligned}
\|x_{k+1} - x^\star\|^2 &= \|\alpha_k[f(w_k) - f(x^\star)] + (1-\alpha_k)(w_k - x^\star) + \alpha_k(f(x^\star) - x^\star)\|^2 \\
&\le \|\alpha_k[f(w_k) - f(x^\star)] + (1-\alpha_k)(w_k - x^\star)\|^2 + 2\alpha_k\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle \\
&= \alpha_k\|f(w_k) - f(x^\star)\|^2 + (1-\alpha_k)\|w_k - x^\star\|^2 \\
&\quad - \alpha_k(1-\alpha_k)\|(f(w_k) - f(x^\star)) - (w_k - x^\star)\|^2 + 2\alpha_k\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle \\
&\le \alpha_k\|f(w_k) - f(x^\star)\|^2 + (1-\alpha_k)\|w_k - x^\star\|^2 + 2\alpha_k\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle \\
&\le \alpha_k\gamma^2\|w_k - x^\star\|^2 + (1-\alpha_k)\|w_k - x^\star\|^2 + 2\alpha_k\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle \\
&= [1 - \alpha_k(1-\gamma^2)]\|w_k - x^\star\|^2 + 2\alpha_k\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle \\
&\le [1 - \alpha_k(1-\gamma^2)]\|z_k - x^\star\|^2 + 2\alpha_k\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle.
\end{aligned}$$

This together with $z_k - x^\star = (x_k - x^\star) + \mu_k(x_k - x_{k-1})$ and $0 \le \gamma < 1$ give us that

$$\begin{aligned}
\|x_{k+1} - x^\star\|^2 &\le [1-\alpha_k(1-\gamma)]\|(x_k - x^\star) + \mu_k(x_k - x_{k-1})\|^2 + 2\alpha_k\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle \\
&\le [1-\alpha_k(1-\gamma)]\left(\|x_k - x^\star\|^2 + 2\mu_k\|x_k - x^\star\|\|x_k - x_{k-1}\| + \mu_k^2\|x_k - x_{k-1}\|^2\right) \\
&\quad + 2\alpha_k\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle \\
&= [1 - \alpha_k(1-\gamma)]\|x_k - x^\star\|^2 + 2\alpha_k\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle \\
&\quad + [1 - \alpha_k(1-\gamma)]\mu_k\|x_k - x_{k-1}\|(2\|x_k - x^\star\| + \mu_k\|x_k - x_{k-1}\|).
\end{aligned} \tag{13}$$

Because $\lim\limits_{k\to\infty} \mu_k\|x_k - x_{k-1}\| = \lim\limits_{k\to\infty} \alpha_k \dfrac{\mu_k}{\alpha_k}\|x_k - x_{k-1}\| = 0$, there exists $0 < M_1$ such that

$$\mu_k\|x_k - x_{k-1}\| \le M_1 \tag{14}$$

for all $k \in \mathbb{N}$.

Put $M_2 := \sup_{k\in\mathbb{N}}\{\|x_k - x^\star\|, M_1\}$. This together with (13) and (14) yields

$$\begin{aligned}
\|x_{k+1} - x^\star\|^2 &\leq [1-\alpha_k(1-\gamma)]\|x_k - x^\star\|^2 + 2\alpha_k\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle \\
&\quad + \mu_k\|x_k - x_{k-1}\|(2\|x_k - x^\star\| + M_1) \\
&\leq [1-\alpha_k(1-\gamma)]\|x_k - x^\star\|^2 + 2\alpha_k\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle \\
&\quad + \mu_k\|x_k - x_{k-1}\|(2M_2 + M_2) \\
&= [1-\alpha_k(1-\gamma)]\|x_k - x^\star\|^2 + 2\alpha_k\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle \\
&\quad + 3M_2\mu_k\|x_k - x_{k-1}\| \\
&= [1-\alpha_k(1-\gamma)]\|x_k - x^\star\|^2 \\
&\quad + \alpha_k(1-\gamma)\left[\frac{3M_2\frac{\mu_k}{\alpha_k}\|x_k - x_{k-1}\| + 2\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle}{1-\gamma}\right].
\end{aligned} \tag{15}$$

We now set u_k, v_k and s_k as the following:

$$u_k := \|x_k - x^\star\|^2, \quad v_k := \alpha_k(1-\gamma)$$

and

$$s_k := \frac{3M_2\mu_k}{\alpha_k(1-\gamma)}\|x_k - x_{k-1}\| + \frac{2}{1-\gamma}\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle.$$

So, we have from (15) that

$$u_{k+1} \leq (1-v_k)u_k + v_k s_k, \quad \forall k \in \mathbb{N}. \tag{16}$$

Next, we analyze the convergence of sequence $\{x_k\}$ by considering the following two cases:

Case 1. Suppose $\{\|x_k - x^\star\|\}_{k\geq m_o}$ is nonincreasing for some $m_0 \in \mathbb{N}$. Because $\{\|x_k - x^\star\|\}$ is bounded from below by zero, we obtain $\lim_{k\to\infty}\|x_k - x^\star\|$ exists. It follows from $\lim_{k\to\infty}\alpha_k = 0$ and $\sum_{k=1}^{\infty}\alpha_k = +\infty$ that

$$\sum_{k=1}^{\infty} v_k = \sum_{k=1}^{\infty}\alpha_k(1-\gamma) = (1-\gamma)\sum_{k=1}^{\infty}\alpha_k = +\infty.$$

To apply Lemma 1, we need to claim that $\limsup_{k\to\infty}\langle f(x^\star) - x^\star, x_{k+1} - x^\star\rangle \leq 0$. Indeed, by definition of y_k, we have

$$\begin{aligned}
\|y_k - x^\star\|^2 &= \|\beta_k(z_k - x^\star) + (1-\beta_k)(T_k z_k - x^\star)\|^2 \\
&= \beta_k\|z_k - x^\star\|^2 + (1-\beta_k)\|T_k z_k - x^\star\|^2 - \beta_k(1-\beta_k)\|z_k - T_k z_k\|^2 \\
&\leq \beta_k\|z_k - x^\star\|^2 + (1-\beta_k)\|z_k - x^\star\|^2 - \beta_k(1-\beta_k)\|z_k - T_k z_k\|^2 \\
&= \|z_k - x^\star\|^2 - \beta_k(1-\beta_k)\|z_k - T_k z_k\|^2.
\end{aligned} \tag{17}$$

By Algorithm 6, (10) and (17), we obtain

$$
\begin{aligned}
\|x_{k+1}-x^\star\|^2 &= \|\alpha_k(f(w_k)-x^\star)+(1-\alpha_k)(w_k-x^\star)\|^2 \\
&= \alpha_k\|f(w_k)-x^\star\|^2+(1-\alpha_k)\|w_k-x^\star\|^2-\alpha_k(1-\alpha_k)\|f(w_k)-w_k\|^2 \\
&\le \alpha_k\|f(w_k)-x^\star\|^2+(1-\alpha_k)\|w_k-x^\star\|^2 \\
&\le \alpha_k\|f(w_k)-x^\star\|^2+(1-\alpha_k)\|y_k-x^\star\|^2 \\
&\le \alpha_k\|f(w_k)-x^\star\|^2+(1-\alpha_k)\Big(\|z_k-x^\star\|^2-\beta_k(1-\beta_k)\|z_k-T_kz_k\|^2\Big) \\
&= \alpha_k\|f(w_k)-x^\star\|^2+(1-\alpha_k)\|(x_k-x^\star)+\mu_k(x_k-x_{k-1})\|^2 \\
&\quad -(1-\alpha_k)\beta_k(1-\beta_k)\|z_k-T_kz_k\|^2 \\
&\le \alpha_k\|f(w_k)-x^\star\|^2-(1-\alpha_k)\beta_k(1-\beta_k)\|z_k-T_kz_k\|^2 \\
&\quad +(1-\alpha_k)\Big(\|x_k-x^\star\|^2+2\mu_k\|x_k-x^\star\|\|x_k-x_{k-1}\|+\mu_k^2\|x_k-x_{k-1}\|^2\Big),
\end{aligned} \tag{18}
$$

which implies that for any $k \in \mathbb{N}$,

$$
\begin{aligned}
(1-\alpha_k)\beta_k(1-\beta_k)\|z_k-T_kz_k\|^2 &\le \alpha_k\|f(w_k)-x^\star\|^2+(1-\alpha_k)\|x_k-x^\star\|^2-\|x_{k+1}-x^\star\|^2 \\
&\quad +2\mu_k(1-\alpha_k)\|x_k-x^\star\|\|x_k-x_{k-1}\| \\
&\quad +\mu_k^2(1-\alpha_k)\|x_k-x_{k-1}\|^2 \\
&= \alpha_k(\|(f(w_k)-f(x_k))+(f(x_k)-x^\star)\|^2) \\
&\quad +(1-\alpha_k)\|x_k-x^\star\|^2-\|x_{k+1}-x^\star\|^2 \\
&\quad +2\mu_k(1-\alpha_k)\|x_k-x^\star\|\|x_k-x_{k-1}\| \\
&\quad +\mu_k^2(1-\alpha_k)\|x_k-x_{k-1}\|^2 \\
&\le \alpha_k\Big(\|f(w_k)-f(x_k)\|^2+2\|f(w_k)-f(x_k)\|\|f(x_k)-x^\star\|\Big) \\
&\quad +\alpha_k\|f(x_k)-x^\star\|^2+(1-\alpha_k)\|x_k-x^\star\|^2 \\
&\quad -\|x_{k+1}-x^\star\|^2+2\mu_k(1-\alpha_k)\|x_k-x^\star\|\|x_k-x_{k-1}\| \\
&\quad +\mu_k^2(1-\alpha_k)\|x_k-x_{k-1}\|^2.
\end{aligned}
$$

Taking $k \to \infty$, we obtain

$$\lim_{k\to\infty}\|z_k-T_kz_k\|=0. \tag{19}$$

This implies

$$
\begin{aligned}
\lim_{k\to\infty}\|y_k-z_k\| &= \lim_{k\to\infty}(1-\beta_k)\|T_kz_k-z_k\| \\
&\le \lim_{k\to\infty}\|T_kz_k-z_k\| \\
&= 0.
\end{aligned} \tag{20}
$$

Because $\|z_k-x_k\|=\mu_k\|x_k-x_{k-1}\|$ and $\lim_{k\to\infty}\mu_k\|x_k-x_{k-1}\|=0$, we derive

$$\lim_{k\to\infty}\|z_k-x_k\|=0. \tag{21}$$

From $\|y_k-x_k\|\le\|y_k-z_k\|+\|z_k-x_k\|$, (20) and (21), we obtain

$$\lim_{k\to\infty}\|y_k-x_k\|=0. \tag{22}$$

Moreover, we have from (9), (18) and nonexpansiveness of S_k that

$$\begin{aligned}
\|x_{k+1}-x^\star\|^2 &\le \alpha_k\|f(w_k)-x^\star\|^2+(1-\alpha_k)\|w_k-x^\star\|^2\\
&= \alpha_k\|f(w_k)-x^\star\|^2+\|w_k-x^\star\|^2-\alpha_k\|w_k-x^\star\|^2\\
&\le \alpha_k\|f(w_k)-x^\star\|^2+\|w_k-x^\star\|^2\\
&= \alpha_k\|f(w_k)-x^\star\|^2+\|\xi_k(y_k-x^\star)+(1-\xi_k)(S_ky_k-x^\star)\|^2\\
&= \alpha_k\|f(w_k)-x^\star\|^2+\xi_k\|y_k-x^\star\|^2+(1-\xi_k)\|S_ky_k-x^\star\|^2\\
&\quad -\xi_k(1-\xi_k)\|y_k-S_ky_k\|^2\\
&\le \alpha_k\|f(w_k)-x^\star\|^2+\xi_k\|y_k-x^\star\|^2+(1-\xi_k)\|y_k-x^\star\|^2\\
&\quad -\xi_k(1-\xi_k)\|y_k-S_ky_k\|^2\\
&= \alpha_k\|f(w_k)-x^\star\|^2+\|y_k-x^\star\|^2-\xi_k(1-\xi_k)\|y_k-S_ky_k\|^2\\
&\le \alpha_k\|f(w_k)-x^\star\|^2+\|z_k-x^\star\|^2-\xi_k(1-\xi_k)\|y_k-S_ky_k\|^2\\
&= \alpha_k\|f(w_k)-x^\star\|^2+\|(x_k-x^\star)+\mu_k(x_k-x_{k-1})\|^2-\xi_k(1-\xi_k)\|y_k-S_ky_k\|^2\\
&\le \alpha_k\|f(w_k)-x^\star\|^2+\|x_k-x^\star\|^2+2\mu_k\|x_k-x^\star\|\|x_k-x_{k-1}\|\\
&\quad +\mu_k^2\|x_k-x_{k-1}\|^2-\xi_k(1-\xi_k)\|y_k-S_ky_k\|^2.
\end{aligned}$$

The above inequality implies

$$\begin{aligned}
\xi_k(1-\xi_k)\|y_k-S_ky_k\|^2 \le{}& \alpha_k\|f(w_k)-x^\star\|^2+\|x_k-x^\star\|^2+2\mu_k\|x_k-x^\star\|^2\|x_k-x_{k-1}\|^2\\
&+\mu_k^2\|x_k-x_{k-1}\|^2-\|x_{k+1}-x^\star\|^2.
\end{aligned}$$

By assumptions (3), (4) and $\lim_{k\to\infty}\|x_k-x^\star\|$ exists together with $\lim\limits_{k\to\infty}\mu_k\|x_k-x_{k-1}\|=0$, we obtain

$$\lim_{k\to\infty}\|y_k-S_ky_k\|=0. \tag{23}$$

From the definition of w_k and assumption (3), we have

$$\begin{aligned}
\|w_k-x_k\| &\le \xi_k\|y_k-x_k\|+(1-\xi_k)\|S_ky_k-x_k\|\\
&\le \xi_k\|y_k-x_k\|+(1-\xi_k)(\|S_ky_k-y_k\|+\|y_k-x_k\|)\\
&= \|y_k-x_k\|+(1-\xi_k)\|S_ky_k-y_k\|\\
&\le \|y_k-x_k\|+\|S_ky_k-y_k\|.
\end{aligned}$$

It follows from (22) and (23) that

$$\lim_{k\to\infty}\|w_k-x_k\|=0. \tag{24}$$

Using the definition of x_{k+1}, we have

$$\begin{aligned}
\|x_{k+1}-x_k\| &\le \alpha_k\|f(w_k)-x_k\|+(1-\alpha_k)\|w_k-x_k\|\\
&\le \alpha_k\|f(w_k)-f(x^\star)\|+\alpha_k\|f(x^\star)-x_k\|+(1-\alpha_k)\|w_k-x_k\|\\
&\le \alpha_k\gamma\|w_k-x^\star\|+\alpha_k\|f(x^\star)-x_k\|+\|w_k-x_k\|.
\end{aligned}$$

Due to $\lim_{k\to\infty}\alpha_k=0$, (24) and the boundedness of $\{x_k\}$ and $\{w_k\}$, we obtain

$$\lim_{k\to\infty}\|x_{k+1}-x_k\|=0. \tag{25}$$

Let $\zeta=\limsup_{k\to\infty}\langle f(x^\star)-x^\star,x_{k+1}-x^\star\rangle$. The boundedness of $\{x_k\}$ implies that there exists a subsequence $\{x_{k_j}\}$ such that

$$\lim_{j\to\infty}\langle f(x^\star)-x^\star,x_{k_j+1}-x^\star\rangle=\limsup_{k\to\infty}\langle f(x^\star)-x^\star,x_{k+1}-x^\star\rangle=\zeta$$

and $x_{k_j} \rightharpoonup x \in H$. It derives from the nonexpansiveness of T_k that

$$\begin{aligned}\|x_k - T_k x_k\| &\leq \|x_k - z_k\| + \|z_k - T_k z_k\| + \|T_k z_k - T_k x_k\| \\ &\leq \|x_k - z_k\| + \|z_k - T_k z_k\| + \|z_k - x_k\| \\ &= 2\|x_k - z_k\| + \|z_k - T_k z_k\|. \end{aligned} \tag{26}$$

It follows from (19) and (21) that $\lim_{k\to\infty} \|x_k - T_k x_k\| = 0$.

Using Lemma 2, we obtain $x \in \cap_{k=1}^{\infty} F(T_k)$. Due to S_k being nonexpansive, we have for any $k \in \mathbb{N}$,

$$\begin{aligned}\|x_k - S_k x_k\| &\leq \|x_k - y_k\| + \|y_k - S_k y_k\| + \|S_k y_k - S_k x_k\| \\ &\leq \|x_k - y_k\| + \|y_k - S_k y_k\| + \|y_k - x_k\| \\ &= 2\|x_k - y_k\| + \|y_k - S_k y_k\|, \end{aligned} \tag{27}$$

which implies $\lim_{k\to\infty} \|x_k - S_k x_k\| = 0$ by employing (22) and (23). By Lemma 2, we obtain $x \in \cap_{k=1}^{\infty} F(S_k)$. Because $\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0$, it follows that x_{k_j+1} converges weakly to x.

In addition, utilizing $x^\star = P_\Gamma f(x^\star)$ together with (6) gives us that

$$\zeta = \lim_{j\to\infty} \langle f(x^\star) - x^\star, x_{k_j+1} - x^\star \rangle = \langle f(x^\star) - x^\star, x - x^\star \rangle \leq 0.$$

Therefore,

$$\limsup_{k\to\infty} \langle f(x^\star) - x^\star, x_k - x^\star \rangle = 0. \tag{28}$$

Invoking $\lim_{k\to\infty} \frac{\mu_k}{\alpha_k} \|x_k - x_{k-1}\| = 0$ and (28), we obtain

$$\limsup_{k\to\infty} s_k = \limsup_{k\to\infty} \left[3M_2 \frac{\mu_k}{\alpha_k(1-\gamma)} \|x_k - x_{k-1}\| + \frac{2}{1-\gamma} \langle f(x^\star) - x^\star, x_{k+1} - x^\star \rangle \right] \leq 0. \tag{29}$$

Coming back to (16), by Lemma 1, we can conclude that $x_k \to x^\star$.

Case 2. Suppose that $\{\|x_k - x^\star\|\}$ is not a monotonically decreasing sequence. To apply Lemma 3, put $\lambda_k := \|x_k - x^\star\|$. Then, there exists a subsequence $\{\lambda_{k_i}\}$ of $\{\lambda_k\}$ such that

$$\lambda_{k_i} < \lambda_{k_i+1}, \quad \forall i \in \mathbb{N}.$$

In this case, let $\varphi : \mathbb{N} \to \mathbb{N}$ be defined by

$$\varphi(k) := \max\{j \in \mathbb{N} : j \leq k,\ \lambda_{k_j} < \lambda_{k_j+1}\}.$$

Therefore, $\varphi(k)$ satisfies the condition in Lemma 3. Hence, we have $\lambda_{\varphi(k)} \leq \lambda_{\varphi(k)+1}$ for all k. This means that

$$\|x_{\varphi(k)} - x^\star\| \leq \|x_{\varphi(k)+1} - x^\star\|, \quad \forall k.$$

As the proof in Case 1, we also have that for any k,

$$\begin{aligned}&\beta_{\varphi(k)}(1 - \beta_{\varphi(k)})(1 - \alpha_{\varphi(k)})\|z_{\varphi(k)} - T_{\varphi(k)} z_{\varphi(k)}\|^2 \\ &\quad \leq \alpha_{\varphi(k)} \|f(w_{\varphi(k)}) - f(x_{\varphi(k)})\|^2 + \alpha_{\varphi(k)} \|f(x_{\varphi(k)}) - x^\star\|^2 \\ &\qquad + 2\alpha_{\varphi(k)} \|f(w_{\varphi(k)}) - f(x_{\varphi(k)})\| \|f(x_{\varphi(k)}) - x^\star\| - \alpha_{\varphi(k)} \|x_{\varphi(k)} - x^\star\|^2 \\ &\qquad + \|x_{\varphi(k)} - x^\star\|^2 - \|x_{\varphi(k)+1} - x^\star\|^2 \\ &\qquad + \mu_{\varphi(k)}(1 - \alpha_{\varphi(k)})\|x_{\varphi(k)} - x_{\varphi(k)-1}\| \Big(2\|x_{\varphi(k)} - x^\star\| + \mu_{\varphi(k)} \|x_{\varphi(k)} - x_{\varphi(k)-1}\| \Big). \end{aligned}$$

Because $\|x_{\varphi(k)} - x^\star\| \le \|x_{\varphi(k)+1} - x^\star\|$ for all k, the above inequality leads to

$$\begin{aligned}
\beta_{\varphi(k)}(1-\beta_{\varphi(k)})&(1-\alpha_{\varphi(k)})\|z_{\varphi(k)} - T_{\varphi(k)}z_{\varphi(k)}\|^2 \\
&\le \alpha_{\varphi(k)}\|f(w_{\varphi(k)}) - f(x_{\varphi(k)})\|^2 + \alpha_{\varphi(k)}\|f(x_{\varphi(k)}) - x^\star\|^2 \\
&\quad + 2\alpha_{\varphi(k)}\|f(w_{\varphi(k)}) - f(x_{\varphi(k)})\|\|f(x_{\varphi(k)}) - x^\star\| - \alpha_{\varphi(k)}\|x_{\varphi(k)} - x^\star\|^2 \\
&\quad + \mu_{\varphi(k)}(1-\alpha_{\varphi(k)})\|x_{\varphi(k)} - x_{\varphi(k)-1}\|\Big(2\|x_{\varphi(k)} - x^\star\| + \mu_{\varphi(k)}\|x_{\varphi(k)} - x_{\varphi(k)-1}\|\Big).
\end{aligned}$$

Using $\lim_{k\to\infty} \alpha_{\varphi(k)} = 0$ and $\lim_{k\to\infty} \mu_{\varphi(k)}\|x_{\varphi(k)} - x_{\varphi(k)-1}\| = 0$, we obtain

$$\lim_{k\to\infty} \|z_{\varphi(k)} - T_{\varphi(k)}z_{\varphi(k)}\| = 0. \tag{30}$$

Similar to the proof of Case 1, we conclude

$$\lim_{k\to\infty} \|z_{\varphi(k)} - x_{\varphi(k)}\| = 0, \tag{31}$$

$$\lim_{k\to\infty} \|y_{\varphi(k)} - x_{\varphi(k)}\| = 0, \tag{32}$$

$$\lim_{k\to\infty} \|y_{\varphi(k)} - S_{\varphi(k)}y_{\varphi(k)}\| = 0, \tag{33}$$

and so

$$\lim_{k\to\infty} \|x_{\varphi(k)+1} - x_{\varphi(k)}\| = 0. \tag{34}$$

Put $\delta := \limsup_{k\to\infty}\langle f(x^\star) - x^\star, x_{\varphi(k)+1} - x^\star\rangle$. Due to $\{x_{\varphi(k)}\}$ being bounded, there exists a subsequence $\{x_{\varphi(k_j)}\}$ of $\{x_{\varphi(k)}\}$ such that

$$\delta := \limsup_{k\to\infty}\langle f(x^\star) - x^\star, x_{\varphi(k)+1} - x^\star\rangle = \delta := \lim_{j\to\infty}\langle f(x^\star) - x^\star, x_{\varphi(k_j)+1} - x^\star\rangle$$

and $x_{\varphi(k_j)} \rightharpoonup \nu$ for some $\nu \in \mathcal{H}$. The nonexpansiveness of $T_{\varphi(k)}$ and $S_{\varphi(k)}$ implies

$$\begin{aligned}
\|x_{\varphi(k)} - T_{\varphi(k)}x_{\varphi(k)}\| &\le \|x_{\varphi(k)} - z_{\varphi(k)}\| + \|z_{\varphi(k)} - T_{\varphi(k)}z_{\varphi(k)}\| + \|T_{\varphi(k)}z_{\varphi(k)} - T_{\varphi(k)}x_{\varphi(k)}\| \\
&\le \|x_{\varphi(k)} - z_{\varphi(k)}\| + \|z_{\varphi(k)} - T_{\varphi(k)}z_{\varphi(k)}\| + \|z_{\varphi(k)} - x_{\varphi(k)}\|
\end{aligned} \tag{35}$$

and

$$\begin{aligned}
\|x_{\varphi(k)} - S_{\varphi(k)}x_{\varphi(k)}\| &\le \|x_{\varphi(k)} - y_{\varphi(k)}\| + \|y_{\varphi(k)} - S_{\varphi(k)}y_{\varphi(k)}\| + \|S_{\varphi(k)}y_{\varphi(k)} - S_{\varphi(k)}x_{\varphi(k)}\| \\
&\le \|x_{\varphi(k)} - y_{\varphi(k)}\| + \|y_{\varphi(k)} - S_{\varphi(k)}y_{\varphi(k)}\| + \|y_{\varphi(k)} - x_{\varphi(k)}\|.
\end{aligned} \tag{36}$$

Taking $k \to \infty$ in (35) and (36), we derive from (30)–(33) that

$$\lim_{k\to\infty} \|x_{\varphi(k)} - T_{\varphi(k)}x_{\varphi(k)}\| = 0 \tag{37}$$

and

$$\lim_{k\to\infty} \|x_{\varphi(k)} - S_{\varphi(k)}x_{\varphi(k)}\| = 0. \tag{38}$$

By Lemma 2, we obtain $\nu \in \Gamma$. Due to $\lim_{j\to\infty} \|x_{\varphi(k_j)+1} - x_{\varphi(k_j)}\| = 0$, we obtain $x_{\varphi(k_j)+1} \rightharpoonup \nu$. Furthermore, it follows from $x^\star := P_\Gamma f(x^\star)$ and (6) that

$$\delta := \lim_{j\to\infty}\langle f(x^\star) - x^\star, x_{\varphi(k_j)+1} - x^\star\rangle = \langle f(x^\star) - P_\Gamma f(x^\star), \nu - P_\Gamma f(x^\star)\rangle \le 0,$$

and thus

$$\limsup_{k\to\infty}\langle f(x^\star)-x^\star, x_{\varphi(k)+1}-x^\star\rangle=\delta\le 0. \tag{39}$$

Because $\lambda_{\varphi(k)}\le\lambda_{\varphi(k)+1}$, as in the proof of Case 1, we have that for every k,

$$\begin{aligned}
&\|x_{\varphi(k)}-x^\star\|^2\\
&\quad\le\|x_{\varphi(k)+1}-x^\star\|^2\\
&\quad\le\Big(1-\alpha_{\varphi(k)}(1-\gamma)\Big)\|x_{\varphi(k)}-x^\star\|^2\\
&\qquad+\alpha_{\varphi(k)}(1-\gamma)\left[\frac{3M_2\frac{\mu_{\varphi(k)}}{\alpha_{\varphi(k)}}\|x_{\varphi(k)}-x_{\varphi(k)-1}\|+2\langle f(x^\star)-x^\star,x_{\varphi(k)+1}-x^\star\rangle}{1-\gamma}\right].
\end{aligned}\tag{40}$$

Therefore,

$$\begin{aligned}
&\alpha_{\varphi(k)}(1-\gamma)\|x_{\varphi(k)}-x^\star\|^2\\
&\quad\le\alpha_{\varphi(k)}(1-\gamma)\left[\frac{3M_2\frac{\mu_{\varphi(k)}}{\alpha_{\varphi(k)}}\|x_{\varphi(k)}-x_{\varphi(k)-1}\|+2\langle f(x^\star)-x^\star,x_{\varphi(k)+1}-x^\star\rangle}{1-\gamma}\right].
\end{aligned}\tag{41}$$

From $\alpha_{\varphi(k)}\in(0,1)$ and $\gamma\in[0,1)$, we obtain $\alpha_{\varphi(k)}(1-\gamma)>0$, which implies

$$\|x_{\varphi(k)}-x^\star\|^2\le\frac{3M_2\frac{\mu_{\varphi(k)}}{\alpha_{\varphi(k)}}\|x_{\varphi(k)}-x_{\varphi(k)-1}\|+2\langle f(x^\star)-x^\star,x_{\varphi(k)+1}-x^\star\rangle}{1-\gamma}. \tag{42}$$

Invoking $\lim\limits_{k\to\infty}\frac{\mu_k}{\alpha_k}\|x_k-x_{k-1}\|=0$ and (39), we obtain

$$\limsup_{k\to\infty}\|x_{\varphi(k)}-x^\star\|=0,$$

and hence

$$\lim_{k\to\infty}\|x_{\varphi(k)}-x^\star\|=0.$$

It follows from (34) that $\lim\limits_{k\to\infty}\|x_{\varphi(k)+1}-x^\star\|=0$. By Lemma 3, we obtain

$$0\le\lim_{k\to\infty}\|x_k-x^\star\|\le\lim_{k\to\infty}\|x_{\varphi(k)+1}-x^\star\|=0.$$

Therefore, $\{x_k\}$ converges strongly to $x^\star$. □

We observe that Algorithm 6 can be reduced to Algorithm 7 by setting $S_k=T_k$ for finding a common fixed point of a countable family of nonexpansive mappings of $\{T_k\}$.

Corollary 1. *Let $\{T_k\}$ be a countable family of nonexpansive mappings from $\mathcal{H}$ into itself such that $\Gamma=\cap_{k=1}^{\infty}F(T_k)\neq\emptyset$. Suppose $\{T_k\}$ satisfies NST*-conditions and the following conditions hold:*

(1) $0<a<\alpha_k<\hat{a}<1$;
(2) $0<b<\beta_k<\hat{b}<1$;
(3) $0<c<\xi_k<\hat{c}<1$;
(4) $\lim_{k\to\infty}\alpha_k=0$ *and* $\sum_{k=1}^{\infty}\alpha_k=+\infty$;
(5) $\lim_{k\to\infty}\frac{\eta_k}{\alpha_k}=0$,

where $a,b,c,\hat{a},\hat{b}$ and $\hat{c}$ are real positive numbers. Then, the sequence $\{x_k\}$ generated by Algorithm 7 converges strongly to $x^\star\in\Gamma$, where $x^\star=P_\Gamma f(x^\star)$.

Algorithm 7 IVMIA (II): Inertial Viscosity Approximation Method for a family of Nonexpansive Mappings

1: **Input.** Let $x_0, x_1 \in \mathcal{H}$, $\{\eta_k\}$ a positive sequence and $f : \mathcal{H} \to \mathcal{H}$ a γ-contraction. Choose $\{\alpha_k\}, \{\beta_k\}, \xi_k \subset (0,1)$ and $\theta_k \geq 0$.

2: Select $\mu_k \in (0, \bar{\mu_k}]$ such that for $k \geq 1$,

$$\bar{\mu_k} := \begin{cases} \min\left\{\theta_k, \frac{\eta_k}{\|x_k - x_{k-1}\|}\right\} & \text{if } x_k \neq x_{k-1}, \\ \theta_k & \text{otherwise.} \end{cases} \tag{43}$$

3: Compute

$$\begin{cases} z_k = x_k + \mu_k(x_k - x_{k-1}), \\ y_k = \beta_k z_k + (1-\beta_k)T_k z_k, \\ w_k = \xi_k y_k + (1-\xi_k)T_k y_k \\ x_{k+1} = \alpha_k f(w_k) + (1-\alpha_k)w_k. \end{cases}$$

4. Application to Convex Bilevel Optimization Problems

The aim of this section is to apply our proposed algorithm for solving the following convex bilevel optimization problem:

$$\min_{x \in \Gamma} \mathcal{F}(x), \tag{44}$$

where $\mathcal{F} : \mathcal{H} \to \mathbb{R}$ is strongly convex differentiable with $\nabla \mathcal{F}$ being $L_{\mathcal{F}}$-Lipschitz continuous and Γ is the set of all common minimizers of the following unconstrained minimization problems:

$$\min_x \{\phi_1(x) + \psi_1(x)\} \quad \text{and} \quad \min_x \{\phi_2(x) + \psi_2(x)\}, \tag{45}$$

where $\psi_i, \phi_i : \mathcal{H} \to (-\infty, +\infty]$, $i = 1, 2$, are proper convex and lower semicontinuous functions and ϕ_1, ϕ_2 are differentiable functions. Problem (45) can be reduced to (2) if $\phi_1 = \phi_2$ and $\psi_1 = \psi_2$. As in the literature, we know that $x^\star \in \Gamma$ if and only if

$$x^\star = \text{prox}_{\lambda_k \psi_1}(I - \lambda_k \nabla \phi_1) \quad \text{and} \quad x^\star = \text{prox}_{\varepsilon_k \psi_2}(I - \varepsilon_k \nabla \phi_2),$$

where $\lambda_k \in \left(0, \frac{2}{L_{\phi_1}}\right), \varepsilon_k \in \left(0, \frac{2}{L_{\phi 2}}\right)$ while L_{ϕ_1} and L_{ϕ_2} are Lipschitz gradients of $\nabla \phi_1$ and $\nabla \phi_2$, respectively. In addition, $x^\star \in \Gamma$ also is a solution of problem (44) if it satisfies the following form:

$$\langle \nabla \mathcal{F}(x^\star), x - x^\star \rangle \geq 0, \quad \forall x \in \Gamma. \tag{46}$$

Therefore, we solve convex bilevel optimization problems (44) and (45) by finding a common fixed point $x^\star$ of $\text{prox}_{\lambda_k \psi_1}(I - \lambda_k \nabla \phi_1)$ and $\text{prox}_{\varepsilon_k \psi_2}(I - \varepsilon_k \nabla \phi_2)$, which satisfies the formulation of (46).

Next, we present the algorithm derived from our main result for solving the convex bilevel optimization problem as defined by Algorithm 8.

In order to solve (44) and (45), we suppose the following conditions hold:

(1) $f : \mathcal{H} \to \mathcal{H}$ is a γ-contraction with $\gamma \in [0, 1)$;

(2) $\{\lambda_k\} \subset \left(0, \frac{2}{L_{\phi_1}}\right)$ and $\lambda \in \left(0, \frac{2}{L_{\phi_1}}\right)$ with $\lambda_k \to \lambda$;

(3) $\{\varepsilon_k\} \subset \left(0, \frac{2}{L_{\phi_2}}\right)$ and $\varepsilon \in \left(0, \frac{2}{L_{\phi_2}}\right)$ with $\varepsilon_k \to \varepsilon$;

(4) $\{\alpha_k\}, \{\beta_k\}$ and $\{\xi_k\}$ are sequences in $(0, 1)$;

(5) ψ_i, $i = 1,2$, be two lower semicontinuous functions and convex from $\mathcal{H}$ into $\mathbb{R} \cup +\infty$;
(6) $\phi_i, i = 1,2$, be two smooth convex loss functions and differentiable with L_{ϕ_i}-Lipschitz continuous gradients of $\nabla\phi_i, i = 1,2$, respectively;
(7) $\mathcal{F} : \mathcal{H} \to \mathbb{R}$ is strongly convex differentiable with $\nabla\mathcal{F}$ being $L_{\mathcal{F}}$-Lipschitz constant and $\sigma \in \left(0, \frac{2}{L_{\mathcal{F}}+\rho}\right)$ where ρ is a parameter such that $\mathcal{F}$ is strongly convex.

Theorem 2. *Let $\{x_k\}$ be a sequence generated by Algorithm 8 such that all conditions as in Theorem 1 hold. Let Ω be the set of all solutions of (44). Then, $\{x_k\}$ converges strongly to $x^\star \in \Omega$ which satisfies $x^\star = P_\Gamma f(x^\star)$.*

Algorithm 8 iVMBi(I): Inertial Viscosity Method for Bilevel Optimization Problem (I)

1: **Input.** Let $x_0, x_1 \in \mathcal{H}$, $\{\eta_k\}$ a positive sequence. Choose $\{\alpha_k\}, \{\beta_k\}, \{\xi_k\} \subset (0,1)$ and $\theta_k \geq 0$.

2: **Step 1.** Select $\mu_k \in (0, \bar{\mu_k}]$ such that for $k \geq 1$,

$$\bar{\mu_k} := \begin{cases} \min\left\{\theta_k, \frac{\eta_k}{\|x_k - x_{k-1}\|}\right\} & \text{if } x_k \neq x_{k-1}, \\ \theta_k & \text{otherwise.} \end{cases} \tag{47}$$

3: **Step 2.** Compute

$$\begin{cases} z_k = x_k + \mu_k(x_k - x_{k-1}), \\ y_k = \beta_k z_k + (1-\beta_k)\text{prox}_{\lambda_k\psi_1}(I - \lambda_k\nabla\phi_1)z_k, \\ w_k = \xi_k y_k + (1-\xi_k)\text{prox}_{\varepsilon_k\psi_2}(I - \varepsilon_k\nabla\phi_2)y_k \\ u_k = (I - \sigma\nabla\mathcal{F})(w_k), \\ x_{k+1} = \alpha_k u_k + (1-\alpha_k)w_k. \end{cases}$$

Proof. Let $T_k := \text{prox}_{\lambda_k\psi_1}(I - \lambda_k\nabla\phi_1)$ and $S_k := \text{prox}_{\varepsilon_k\psi_2}(I - \varepsilon_k\nabla\phi_2)$ as in Algorithm 6, where $\lambda_k \in \left(0, \frac{2}{L_{\phi_1}}\right), \varepsilon_k \in \left(0, \frac{2}{L_{\phi_2}}\right)$ while L_{ϕ_i}, $i = 1,2$, are Lipschitz gradients of $\nabla\phi_i$, $i = 1,2$, respectively. Using Proposition 1, we get that $I - \sigma\nabla\mathcal{F}$ is a contraction mapping. By Theorem 1 and setting $f := I - \sigma\nabla\mathcal{F}$, we obtain that $\{x_k\}$ converges strongly to $x^\star \in \Gamma$, where $x^\star = P_\Gamma f(x^\star)$. Observe that $f(x^\star) = x^\star - \sigma\nabla\mathcal{F}(x^\star)$. It is derived from (6) that for any $x \in \Gamma$,

$$\begin{aligned} 0 &\leq \langle P_\Gamma f(x^\star) - f(x^\star), x - P_\Gamma f(x^\star)\rangle \\ &= \langle x^\star - f(x^\star), x - x^\star\rangle \\ &= \langle x^\star - (x^\star - \sigma\nabla\mathcal{F}(x^\star)), x - x^\star\rangle \\ &= \sigma\langle\nabla\mathcal{F}(x^\star), x - x^\star\rangle. \end{aligned}$$

Because $0 < \sigma$, we conclude $0 \leq \langle\nabla\mathcal{F}(x^\star), x - x^\star\rangle$ for all $x \in \Gamma$, that is, $x^\star$ is an optimal solution of problem (44). Hence, we obtain the desired result. □

Furthermore, our algorithm can be applied to solving convex bilevel optimization problems (1) and (2) by using the same proximity operator in step 2 and 3 as seen in Algorithm 9.

Algorithm 9 iVMBi(II): Inertial Viscosity Method for Bilevel Optimization Problem (II)

1: **Input.** Let $x_0, x_1 \in \mathcal{H}$, $\{\eta_k\}$ a positive sequence. Choose $\{\alpha_k\}, \{\beta_k\}, \{\xi_k\} \subset (0,1)$ and $\theta_k \geq 0$.

2: **Step 1.** Select $\mu_k \in (0, \bar{\mu_k}]$ such that for $k \geq 1$,

$$\bar{\mu_k} := \begin{cases} \min\left\{\theta_k, \frac{\eta_k}{\|x_k - x_{k-1}\|}\right\} & \text{if } x_k \neq x_{k-1}, \\ \theta_k & \text{otherwise.} \end{cases} \tag{48}$$

3: **Step 2.** Compute

$$\begin{cases} z_k = x_k + \mu_k(x_k - x_{k-1}), \\ y_k = \beta_k z_k + (1 - \beta_k)\mathrm{prox}_{\lambda_k \psi}(I - \lambda_k \nabla \phi) z_k, \\ w_k = \xi_k y_k + (1 - \xi_k)\mathrm{prox}_{\lambda_k \psi}(I - \lambda_k \nabla \phi) y_k \\ u_k = (I - \sigma \nabla \mathcal{F})(w_k), \\ x_{k+1} = \alpha_k u_k + (1 - \alpha_k) w_k. \end{cases}$$

The following result is immediately obtained by Theorem 2.

Theorem 3. *Let $\{x_k\}$ be a sequence generated by Algorithm 9 such that all conditions as in Corollary 1 hold. Then, $\{x_k\}$ converges strongly to $x^\star \in \arg\min(\phi + \psi)$ which satisfies*

$$x^\star = P_\Gamma f(x^\star) \text{ and } \langle \nabla \mathcal{F}(x^\star), x - x^\star \rangle \geq 0, \quad \forall x \in \Gamma,$$

that is, $x_k \to x^\star \in \Omega$, where Ω is the set of all solutions of problems (1) and (2).

Next, we use Algorithm 9 as a machine learning algorithm for solving some data classification problems applying on UCI-datasets of breast cancer and heart disease. Moreover, we compare the performance of Algorithm 9 with BiG-SAM, iBiG-SAM, aiBiG-SAM, miBiG-SAM and amiBiG-SAM.

In order to employ Algorithm 9 for solving data classification, we need to know what is the objective function of the inner level. To obtain this, we use a single-layer feedback neuron network (SLFNs) model and the concept of extreme learning machine (ELM) introduced by Huang et al. [46].

In supervised learning, we start with the training set of N samples $S := \{(p_k, q_k) : p_k \in \mathbb{R}^n, q_k \in \mathbb{R}^m, k = 1, 2, \ldots, N\}$, where p_k is input data and q_k is a target. The mathematical model of ELM for SLFNs with M hidden nodes and activate function $\mathcal{G}$ is given by

$$o_j = \sum_{i=1}^{M} m_i \mathcal{G}(\langle w_i, p_j \rangle + r_i), \ \ j = 1, 2, \ldots, N$$

where m_i is the weight vector connecting the i-th hidden node and the output node, r_i is a bias and w_i is the weight vector connecting the i-th hidden node and the input node.

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

$$\mathbf{A} = \begin{bmatrix} \mathcal{G}(\langle w_1, p_1 \rangle + r_1) & \cdots & \mathcal{G}(\langle w_M, p_1 \rangle + r_M) \\ \vdots & \ddots & \vdots \\ \mathcal{G}(\langle w_1, p_N \rangle + r_1) & \cdots & \mathcal{G}(\langle w_M, p_N \rangle + r_M) \end{bmatrix}.$$

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

For prediction or classification problem by using ELM model, we need a zero mean, that is, $\sum_{j=1}^{N} |o_j - q_j| = 0$. Hence,

$$q_j = \sum_{i=1}^{M} m_i \mathcal{G}(\langle w_i, x_j \rangle + r_i), \;\; i = 1, 2, \ldots, N.$$

We can write the above system of linear equations of M variable and N equations as a matrix equation as follows:

$$\mathbf{A}m = \mathbf{Q}, \tag{49}$$

where $m = [m_1^T, \ldots, m_M^T]^T$ and $\mathbf{Q} = [q_1^T, \ldots, q_N^T]^T$ is the training data. To solve ELM, it is to find a weight m satisfies (49). If the Moore–Penrose generalized inverse $\mathbf{A}^\dagger$ of $\mathbf{A}$ exists, then $m = \mathbf{A}^\dagger\mathbf{Q}$. However, in the case that $\mathbf{A}^\dagger$ does not exist, we can find m as the minimizer of the following convex minimization problem:

$$\min_m \|\mathbf{A}m - \mathbf{Q}\|_2^2. \tag{50}$$

Using a least squares model (50) may cause the over fitting problem. In order to prevent this problem, the regularization methods were proposed. The classical one is the Tikhonov regularization [47], which was employed to solve the following minimization problem:

$$\text{Minimize:} \;\; \|\mathbf{A}m - \mathbf{Q}\|_2^2 + \beta\|Km\|_2^2, \tag{51}$$

where β is the regularized parameter and K is the Tikhonov matrix. In the standard form, K is set to be the identity.

Another regularization method is the least absolute shrinkage and selection operator (LASSO), which was proposed by Tibshirani [48] for solving the following convex minimization problem:

$$\text{Minimize:} \;\; \|\mathbf{A}m - \mathbf{Q}\|_2^2 + \beta\|m\|_1, \tag{52}$$

where β is the regularized parameter and $\|(x_1, x_2, \ldots, x_p)\|_1 = \sum_{i=1}^{p} |x_i|$.

In this work, we set $\psi(m) = \beta\|m\|_1$ and $\phi(m) = \|\mathbf{A}m - \mathbf{Q}\|_2^2$. Based on model (52), we can apply Algorithm 9 for solving the convex bilevel optimization problems (1) and (2) while the objective function of the outer level $\mathcal{F}(m) = \frac{1}{2}\|m\|_2^2$. We now conduct some numerical experiments for classifications of the following datasets.

In these experiments, we aim to classify the datasets of breast cancer and heart disease from https://archive.ics.uci.edu, accessed on 12 June 2022.

Breast cancer dataset [49]. This dataset contains 699 samples, each of which has 11 attributes. In this dataset, we classify two classes of data.

Heart disease dataset [50]. This dataset contains 303 samples, each of which has 13 attributes. In this dataset, we classify two classes of data.

Throughout these experiments, all the results are performed under MATLAB 9.6 (R2019a) running on a MacBook Air 13.3-inch, 2020, with Apple M1 chip processor and 8-core GPU, configured with 8 GB of RAM.

In all the experiments, sigmoid is used as an activation function, and we set the number of hidden node $M = 30$. The following formula for the accuracy of the data classification is given by

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

where TP is the model successfully predicting the patient as positive, TN denotes the model successfully predicting the patient as negative, FN represents the prediction of the diseased patient as healthy by negative test results and FP means the prediction of a healthy patient as diseased by a positive test result.

We also compute the success probability of making a correct positive class classification as the following form:

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

In addition, we measure the sensitivity of the model toward identifying the positive class as the following form:

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

The Lipschitz gradient L_ϕ of $\nabla\phi$ is computed by $2\|\mathbf{A}\|^2$. When the dimension of $\mathbf{A}$ is so large, it is hard to compute such L_ϕ. All parameters for each algorithm of our experiments are given in Table 1.

Table 1. Chosen parameters of each algorithm.

Parameters	Algorithm 9	Algorithm 1	Algorithm 2	Algorithm 3	Algorithm 4	Algorithm 5
σ	$\frac{2}{L_{\mathcal{F}}+\rho}$	$\frac{2}{L_{\mathcal{F}}+\rho}$	$\frac{2}{L_{\mathcal{F}}+\rho}$	$\frac{2}{L_{\mathcal{F}}+\rho}$	$\frac{2}{L_{\mathcal{F}}+\rho}$	$\frac{2}{L_{\mathcal{F}}+\rho}$
λ	-	$\frac{1}{L_\phi}$	$\frac{1}{L_\phi}$	$\frac{1}{L_\phi}$	-	-
λ_k	$\frac{1}{L_\phi}$	-	-	-	$\frac{1}{L_\phi}$	$\frac{1}{L_\phi}$
α	-	-	3	3	3	3
α_k	$\frac{1}{50k}$	$\frac{1}{k+2}$	$\frac{1}{50k}$	$\frac{1}{k+2}$	$\frac{1}{k+2}$	$\frac{1}{k+2}$
θ_k	$\frac{k}{k+1}$	-	-	-	-	-
η_k	$\frac{10^{50}}{k^2}$	-	$\frac{10^{50}}{k^2}$	$\frac{\alpha_k}{k^{0.01}}$	$\frac{\alpha_k}{k^{0.01}}$	$\frac{\alpha_k}{k^{0.01}}$
q	-	-	-	-	4	4

From Table 1, we select the best choice of parameter for each algorithm in order to achieve the highest performance. It is worth noting that all parameters satisfy the assumptions of each convergence theorem, see [6,9,11] for more details. In addition, we set $\beta = 0.00001$ which is a regularized parameter of problem (52). In Algorithm 9, we choose $\xi_k, \beta_k = \frac{1}{k+2}$ for experimentation on the breast cancer dataset, while the classification of heart disease uses $\xi_k = 0.5$ together with $\beta_k = 0.1$.

We compare the performance of each method at the 100th and 500th iterations and obtain the following results, as seen in Tables 2 and 3, respectively.

Table 2. The performance of each algorithm at 100th iteration on each dataset.

Dataset	Algorithm	Pre Train	Rec Train	Pre Test	Rec Test	Acc Train	Acc Test
Breast Cancer	Algorithm 1	0.8845	0.9812	0.9718	1.0000	90.4082	98.0861
	Algorithm 2	0.9686	0.9625	0.9857	1.0000	95.5102	99.0431
	Algorithm 3	0.8845	0.9812	0.9718	1.0000	90.4082	98.0861
	Algorithm 4	0.8966	0.9750	0.9718	1.0000	91.0204	98.0861
	Algorithm 5	0.8845	0.9812	0.9718	1.0000	90.4082	98.0861
	Algorithm 9	0.9747	0.9625	0.9857	1.0000	95.9184	99.0431
Heart Disease	Algorithm 1	0.7656	0.8522	0.7647	0.7800	77.6190	75.2688
	Algorithm 2	0.8306	0.8957	0.7719	0.8800	84.2857	79.5699
	Algorithm 3	0.7656	0.8522	0.7647	0.7800	77.6190	75.2688
	Algorithm 4	0.8049	0.8609	0.7593	0.8200	80.9524	76.3441
	Algorithm 5	0.7656	0.8522	0.7647	0.7800	77.6190	75.2688
	Algorithm 9	0.8268	0.9130	0.7667	0.9200	84.7619	80.6452

Table 3. The performance of each algorithm at 500th iteration on each dataset.

Dataset	Algorithm	Pre Train	Rec Train	Pre Test	Rec Test	Acc Train	Acc Test
Breast Cancer	Algorithm 1	0.9506	0.9625	0.9857	1.0000	94.2857	99.0431
	Algorithm 2	0.9778	0.9625	0.9928	1.0000	96.1224	99.5215
	Algorithm 3	0.9506	0.9625	0.9857	1.0000	94.2857	99.0431
	Algorithm 4	0.9536	0.9625	0.9857	1.0000	94.4898	99.0431
	Algorithm 5	0.9506	0.9625	0.9857	1.0000	94.2857	99.0431
	Algorithm 9	0.9778	0.9625	0.9928	1.0000	96.1224	99.5215
Heart Disease	Algorithm 1	0.8065	0.8696	0.7679	0.8600	81.4286	78.4946
	Algorithm 2	0.8455	0.9043	0.7797	0.9200	85.7143	81.7204
	Algorithm 3	0.8065	0.8696	0.7679	0.8600	81.4286	78.4946
	Algorithm 4	0.8115	0.8609	0.7544	0.8600	81.4286	77.4194
	Algorithm 5	0.8065	0.8696	0.7679	0.8600	81.4286	78.4946
	Algorithm 9	0.8455	0.9043	0.7833	0.9400	85.7143	82.7957

Table 2 shows that our algorithm performs the best accuracy at the 100th iteration. Moreover, Table 3 shows the performance of each algorithm at the 500th iteration. We found that Algorithm 9 has a better accuracy than the others.

Next, we show the performance for the prediction of each algorithm in terms of the number of iterations and training times for which each algorithm achieves the highest accuracy.

From Table 4, comparing with Algorithm 1 (BiG-SAM), Algorithm 2 (iBiG-SAM), Algorithm 3 (aiBiG-SAM), Algorithm 4 (miBiG-SAM) and Algorithm 5 (amiBiG-SAM), Algorithm 9 provides a higher value of accuracy for training. In the testing case, we found that the accuracy of Algorithm 2 (iBiG-SAM) is better than our algorithm on the breast cancer experimentation. However, our method has the lowest number of iterations and training times compared with the others.

Table 4. The iteration number and training time of each algorithm with the highest accuracy on each dataset.

Dataset	Algorithm	Iteration No.	Training Time	Acc Train	Acc Test
Breast Cancer	Algorithm 1	819	0.0272	95.1020	99.0431
	Algorithm 2	264	0.0095	96.1224	99.5215
	Algorithm 3	819	0.0267	95.1020	99.0431
	Algorithm 4	531	0.0320	95.1020	99.0431
	Algorithm 5	819	0.0330	95.1020	99.0431
	Algorithm 9	78	0.0054	96.1224	99.0431
Heart Disease	Algorithm 1	2024	0.0293	86.1905	79.5699
	Algorithm 2	556	0.0096	86.1905	81.7204
	Algorithm 3	2024	0.0452	86.1905	79.5699
	Algorithm 4	1226	0.0517	86.1905	79.5699
	Algorithm 5	1398	0.0317	85.7143	78.4946
	Algorithm 9	192	0.0064	86.1905	82.7957

We also construct a 10-fold cross validation to appraise the performance of each algorithm and use Average accuracy as the appraising tool. It is defined as follows:

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

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

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

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

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

We split the data into training sets and testing sets by using the 10-fold cross validation, as seen in Table 5.

Table 5. Number of samples in each fold for all datasets.

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

In Table 6, we show the average of the accuracy of each algorithm with the 500th iteration.

Table 6. Average accuracy of each algorithm at 500th iteration with 10-fold cross validation.

Algorithm	Breast Cancer			Heart Disease		
	Acc Train	Acc Test	$\text{Error}_{\%}$	Acc Train	Acc Test	$\text{Error}_{\%}$
Algorithm 1	95.8989	95.9876	4.0534	79.8319	78.5484	20.8104
Algorithm 2	96.7094	96.9876	3.1474	85.0318	82.8387	16.0616
Algorithm 3	95.8989	95.9876	4.0534	79.8319	78.5484	20.8104
Algorithm 4	96.0420	96.1304	3.9103	80.9683	80.5269	19.2519
Algorithm 5	95.8989	95.9876	4.0534	79.8319	78.5484	20.8104
Algorithm 9	96.7889	97.4182	2.8930	85.8148	82.8387	15.9883

Table 6 demonstrates that Algorithm 9 performs better than Algorithm 1 (BiG-SAM), Algorithm 2 (iBiG-SAM), Algorithm 3 (aiBiG-SAM), Algorithm 4 (miBiG-SAM) and Algorithm 5 (amiBiG-SAM) in terms of the accuracy in all the experiments conducted.

5. Conclusions

We propose a novel iterative method based on a fixed-point approach with an inertial technique for approximating a common fixed point of two countable families of nonexpansive mappings in a Hilbert space and also present strong convergence theorems. Our algorithm leads to a sequence converging strongly to a solution for convex bilevel optimization problems for which the inner level consists of the minimization of the sum of smooth and nonsmooth functions. Furthermore, we apply the proposed algorithm to the data classification of breast cancer and heart disease datasets and then their performances are assessed and compared with the other algorithms. We derive from the experiment that our algorithm provides a higher value of accuracy of training and testing on various datasets. We can conclude the advantages of our proposed algorithm from our experiments in that it requires a lower number of iterations and less training time compared with the others. It is worth mentioning that our proposed algorithm is intelligent machine learning for the prediction and classification of big data. It is an efficient algorithm that can be developed to software/applications for prediction and classifications in future works. Furthermore, we aim to employ our proposed algorithm for real datasets of the patients at Sriphat Medical Center, Faculty of Medicine, Chiang Mai University, Thailand.

Author Contributions: Conceptualization, S.S.; formal analysis, P.T. and W.I.; investigation, P.T.; methodology, S.S.; supervision, S.S.; validation, T.L. and S.S.; writing—original draft, P.T.; writing—review and editing, S.S. and W.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research has received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B05F640183] and Chiang Mai University. The first and fourth authors were also supported by Thailand Science Research and Innovation under the project IRN62W0007.

Data Availability Statement: All data in this work is obtainable at https://archive.ics.uci.edu, accessed on 12 June 2022.

Acknowledgments: The authors would like to thank the referees for the valuable suggestions. This research has received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B05F640183] and Chiang Mai University. The first and fourth authors were also supported by Thailand Science Research and Innovation under the project IRN62W0007.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Darby, S.C.; Ewertz, M.; McGale, P.; Bennet, A.M.; Blom-Goldman, U.; Brønnum, D.; Correa, C.; Cutter, D.; Gagliardi, G.; Gigante, B.; et al. Risk of ischemic heart disease in women after radiotherapy for breast cancer. *N. Engl. J. Med.* **2013**, *368*, 987–998. [CrossRef] [PubMed]
2. Solodov, M. An explicit descent method for bilevel convex optimization. *J. Convex Anal.* **2007**, *4*, 227–237.
3. Cabot, A. Proximal point algorithm controlled by a slowly vanishing term: Applications to hierarchical minimization. *SIAM J. Optim.* **2005**, *15*, 555–572. [CrossRef]
4. Helou, E.S.; Simões, L.E. ϵ-subgradient algorithms for bilevel convex optimization. *Inverse Probl.* **2017**, *33*, 5. [CrossRef]
5. Dempe, S.; Kalashnikov, V.; Pérez-Valdés, G.A.; Kalashnykova, N. Bilevel programming problems. In *Energy Systems*; Springer: Berlin, Germany, 2015.
6. Sabach, S.; Shtern, S. A first order method for solving convex bilevel optimization problems. *SIAM J. Optim.* **2017**, *27*, 640–660. [CrossRef]
7. Xu, H.K. Viscosity approximation methods for nonexpansive mappings. *J. Math. Anal. Appl.* **2004**, *298*, 279–291. [CrossRef]
8. Beck, A.; Sabach, S. A first order method for finding minimal norm-like solutions of convex optimization problems. *Math. Program.* **2014**, *147*, 25–46. [CrossRef]
9. Shehu, Y.; Vuong, P.T.; Zemkoho, A. An inertial extrapolation method for convex simple bilevel optimization. *Optim. Methods Softw.* **2019**, *2019*, 1–20. [CrossRef]
10. Polyak, B.T. Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.* **1964**, *4*, 1–7. [CrossRef]
11. Duan, P.; Zhang, Y. Alternated and multi-step inertial approximation methods for solving convex bilevel optimization problems. *Optimization* **2022**. [CrossRef]
12. Anulekha, D.; Joydeep, D. *Optimality Conditions in Convex Optimization: A Finite-Dimensional View*; CRC Press: Boca Raton, FL, USA, 2011.
13. Yao, Y.; Iyiola, O.S.; Shehu, Y. Subgradient extragradient method with double inertial steps for variational inequalities. *J. Sci. Comput.* **2022**, *90*, 1–29. [CrossRef]
14. Zhao, X.; Yao, J.C.; Yao, Y. A nonmonotone gradient method for constrained multiobjective optimization problems. *J. Nonlinear Var. Anal.* **2022**, *6*, 693–706.
15. Moreau, J.J. Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. Fr.* **1965**, *93*, 273–299. [CrossRef]
16. Combettes, P.L.; Wajs, V.R. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **2005**, *4*, 1168–1200. [CrossRef]
17. Byrne, C. Iterative oblique projection onto convex subsets and the split feasibility problem. *Inverse Probl.* **2002**, *18*, 441–453. [CrossRef]
18. Byrne, C. Aunified treatment of some iterative algorithms in signal processing and image reconstruction. *Inverse Probl.* **2004**, *20*, 103–120. [CrossRef]
19. Cholamjiak, P.; Shehu, Y. Inertial forward-backward splitting method in Banach spaces with application to compressed sensing. *Appl. Math.* **2019**, *64*, 409–435. [CrossRef]
20. Kunrada, K.; Pholasa, N.; Cholamjiak, P. On convergence and complexity of the modified forward-backward method involving new linesearches for convex minimization. *Math. Meth. Appl. Sci.* **2019**, *42*, 1352–1362.
21. Suantai, S.; Eiamniran, N.; Pholasa, N.; Cholamjiak, P. Three-step projective methods for solving the split feasibility problems. *Mathematics* **2019**, *7*, 712. [CrossRef]
22. Suantai, S.; Kesornprom, S.; Cholamjiak, P. Modified proximal algorithms for finding solutions of the split variational inclusions. *Mathematics* **2019**, *7*, 708. [CrossRef]
23. Thong, D.V.; Cholamjiak, P. Strong convergence of a forward-backward splitting method with a new step size for solving monotone inclusions. *Comput. Appl. Math.* **2019**, *38*, 94. [CrossRef]

24. Thongpaen, P.; Wattanaweekul, R. A fast fixed-point algorithm for convex minimization problems and its application in image restoration problems. *Mathematics* **2021**, *9*, 2619. [CrossRef]
25. Mann, W.R. Mean value methods in iteration. *Proc. Am. Math. Soc.* **1953**, *4*, 506–510. [CrossRef]
26. Halpern, B. Fixed points of nonexpansive maps. *Bull. Am. Math. Soc.* **1967**, *73*, 957–961. [CrossRef]
27. Ishikawa, S. Fixed points by a new iteration method. *Proc. Am. Math. Soc.* **1974**, *44*, 147–150. [CrossRef]
28. Phuengrattana, W.; Suantai, S. On the rate of convergence of Mann, Ishikawa, Noor and SP-iterations for continuous functions on an arbitrary interval. *J. Comput. Appl. Math.* **2011**, *235*, 3006–3014. [CrossRef]
29. Wongyai, S.; Suantai, S. Convergence Theorem and Rate of Convergence of a New Iterative Method for Continuous Functions on Closed Interval. In Proceedings of the AMM and APAM Conference Proceedings, Bangkok, Thailand, 23–25 May 2016; pp. 111–118.
30. De la Sen, M.; Agarwal, R.P. Common fixed points and best proximity points of two cyclic self-mappings. *Fixed Point Theory Appl.* **2012**, *2012*, 1–17. [CrossRef]
31. Shoaib, A. Common fixed point for generalized contraction in b-multiplicative metric spaces with applications. *Bull. Math. Anal. Appl.* **2020**, *12*, 46–59.
32. Kim, K.S. A Constructive scheme for a common coupled fixed point problems in Hilbert space. *Mathematics* **2020**, *8*, 1717. [CrossRef]
33. Haghi, R.H.; Bakhshi, N. Some coupled fixed point results without mixed monotone property. *J. Adv. Math. Stud.* **2022**, *15*, 456–463.
34. 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.* **2021**, *37*, 449–461. [CrossRef]
35. Tang, Y.M.; Zhang, L.; Bao, G.Q.; Ren, F.J.; Pedrycz, W. Symmetric implicational algorithm derived from intuitionistic fuzzy entropy. *Iran. J. Fuzzy Syst.* **2022**, *19*, 27–44.
36. Mokeddem, S.A. A fuzzy classification model for myocardial infarction risk assessment. *Appl. Intell.* **2018**, *48*, 1233–1250. [CrossRef]
37. Nakajo, K.; Shimoji, K.; Takahashi, W. Strong convergence to common fixed points of families of nonexpansive mappings in Banach spaces. *J. Nonlinear Convex Anal.* **2007**, *8*, 11–34.
38. Nakajo, K; Shimoji, K.; Takahashi, W. Strong convergence theorems by the hybrid method for families of nonexpansive mappings in Hilbert spaces. *Taiwan J. Math.* **2006**, *10*, 339–360. [CrossRef]
39. Xu, H.K. Another control condition in an iterative method for nonexpansive mappings. *Bull. Aust. Math. Soc.* **2002**, *65*, 109–113. [CrossRef]
40. Goebel, K.; Kirk, W.A. *Topic in Metric Fixed Point Theory*; Cambridge Studies in Advanced Mathematics; Cambridge University Press: Cambridge, UK, 1990; Volume 28.
41. Mainge, P.E. Strong convergence of projected subgradient methods for nonsmooth and nostrictly convex minimization. *Set-Valued Anal.* **2008**, *16*, 899–912. [CrossRef]
42. Bussaban, L.; Suantai, S.; Kaewkhao, A. A parallel inertial S-iteration forward-backward algorithm for regression and classification problems. *Carpathian J. Math.* **2020**, *36*, 21–30. [CrossRef]
43. Das, G.; Debata, J.P. Fixed point of quasi-non-expansive mappings. *Indian J. Pure Appl. Math.* **1968**, *17*, 1263–1269.
44. Takahashi, W.; Tamura, T. Convergence theorems for a pair of non-expansive mappings. *J. Convex Anal.* **1998**, *5*, 45–58.
45. Thongpaen, P.; Inthakon, W. Common attractive points of widely more generalized hybrid mappings in Hilbert spaces. *Thai J. Math.* **2020**, *18*, 861–869.
46. Huang, G.B.; Zhu Q.Y.; Siew, C.K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [CrossRef]
47. Tikhonov, A.N.; Arsenin, A.Y. *Solution of Ill-Posed Problems*; John Wiley & Sons: Washington, DC, USA, 1997.
48. Tibshirani, R. Regression shrinkage abd selection via lasso. *J. R. Stat. Soc. Ser. B (Method)* **1996**, *58*, 267–288.
49. Street, W.N.; Wolberg, W.H.; Mangasarian, O.L. Nuclear feature extraction for breast tumor diagnosis. *Biomed. Image Process. Biomed. Vis.* **1993**, *1905*, 861–870.
50. 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.* **1989**, *64*, 304–310. [CrossRef]