

## Article

# An Inertial Modified S-Algorithm for Convex Minimization Problems with Directed Graphs and Its Applications in Classification Problems

Kobkoon Janngam <sup>1</sup> and Suthep Suantai <sup>2,3,\*</sup>

<sup>1</sup> Graduate Ph.D. Degree Program in Mathematics, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

<sup>2</sup> Data Science Research Center, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

<sup>3</sup> Research Center in Mathematics and Applied Mathematics, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

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

**Abstract:** In this paper, we propose a new accelerated common fixed-point algorithm for two countable families of  $G$ -nonexpansive mappings. Weak convergence results are obtained in the context of directed graphs in real Hilbert spaces. As applications, we apply the obtained results to solving some convex minimization problems and employ our proposed algorithm to solve the data classification of Breast Cancer, Heart Diseases and Ionosphere. Moreover, we also compare the performance of our proposed algorithm with other algorithms in the literature and it is shown that our algorithm has a better convergence behavior than the others.

**Keywords:** classification problems; convex minimization; coordinate affine; forward-backward algorithm;  $G$ -nonexpansive

**MSC:** 47H09; 90C25; 65K10



**Citation:** Janngam, K.; Suantai, S. An Inertial Modified S-Algorithm for Convex Minimization Problems with Directed Graphs and Its Applications in Classification Problems.

*Mathematics* **2022**, *10*, 4442. <https://doi.org/10.3390/math10234442>

Academic Editors: Wei-Shih Du, Marko Kostić, Vladimir E. Fedorov and Manuel Pinto

Received: 4 November 2022

Accepted: 21 November 2022

Published: 24 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The Banach contraction mapping principle [1] unquestionably plays a significant role in the literature on fixed-point theory, despite the fact that it is just one of many cornerstone results that are presented. In fact, the metric fixed-point theory is thought to have its roots in this idea, which is one of the fundamental outcomes of mathematical analysis. However, this fact is a strong motivation for creating other mappings that satisfy specific contractive conditions; see [2–4]. In 2004, Ran et al. [5] introduced a new concept of Banach's fixed-point theorem in partially ordered sets and applied this result to solve linear and nonlinear matrix equations. In 2007, Jachymski [6] presented the notion of single-valued  $G$ -contraction on complete metric spaces with graphs and proved a fixed-point theorem which extends the results of [5]. He called such mappings a Banach  $G$ -contraction. The Banach  $G$ -contraction was subsequently extended in various ways by many authors; see [7–10]. In the past decade, many researchers introduced algorithms for finding the fixed points of  $G$ -nonexpansive mappings; see [11–14]. Recently, Janngam et al. [15–17] introduced fixed-point algorithms in Hilbert spaces with directed graphs and applied these results to classification and image recovery.

At present, fixed-point theory was applied to solve various problems in sciences, engineering, economics, physics, and data science such as signal/image processing; see [18–22], and intensity-modulated radiation therapy treatment planning; see [23,24].

In the field of image processing, the image restoration problem is an interesting and important topic. The least absolute shrinkage and selection operator (LASSO) model can be used to convert this problem into an optimization problem. For this problem, there are

several optimizations and fixed-point methods; see [25–29] for more detail. A fast iterative shrinkage-thresholding algorithm (FISTA) is one of the most widely used approaches for resolving image restoration problems. Beck et al. [30] demonstrated that FISTA with the inertial step technique has a faster convergence rate than previous methods in the literature.

From this perspective, the primary purpose of this study is to construct an accelerated algorithm for finding the common fixed points of two countable families of  $G$ -nonexpansive mappings in real Hilbert spaces with graphs based on the idea of the inertial technique. The applications of this result are to solve convex minimization and data classification problems. Moreover, we compared our algorithm's performance with those of other algorithms.

The structure of the paper is as follows. In Section 2, we provide fundamental ideas about fixed-point theorems. In Section 3, we present an inertial modified  $S$ -algorithm and prove a weak-convergence theorem. In Section 4, convex minimization and classification problems are discussed. Moreover, some numerical experiments on classification problems are also given in Section 5. Finally, we provide the conclusions and discussions.

## 2. Preliminaries

Let  $\mathcal{H}$  be a real Hilbert space with the norm  $\|\cdot\|$  and let  $\mathcal{C}$  be a nonempty closed convex subset of  $\mathcal{H}$ . A mapping  $T$  of  $\mathcal{C}$  into itself is called *nonexpansive* if  $\|Tx - Ty\| \leq \|x - y\|$  for all  $x, y \in \mathcal{C}$ . For a mapping  $T$  of  $\mathcal{C}$  into itself, we denote by  $F(T)$  the set all fixed points of  $T$ , that is,  $F(T) := \{x \in \mathcal{C} : Tx = x\}$ .

Let  $G$  be a directed graph such that the set  $V(G)$  of its vertices corresponds to  $\mathcal{C}$  and  $\Delta$  is the diagonal of  $\mathcal{C} \times \mathcal{C}$  such that  $\Delta \subseteq E(G)$ , where  $E(G)$  is a set of its edges. When two or more edges in a directed graph  $G$  connect the same ordered pair of vertices, the edges are said to be parallel.

Assume that  $G$  has no parallel edges. Consequently,  $G$  can be identified by the pair  $(V(G), E(G))$ . The graph is obtained from  $G$  by reversing the direction of the edges, which is represented by  $G^{-1}$ . That is

$$E(G^{-1}) = \{(u, v) \in \mathcal{C} \times \mathcal{C} : (v, u) \in E(G)\}.$$

Here, we will give a basic knowledge of the definitions of the graph properties that will be used in this work; see [31].

**Definition 1.** A graph  $G = (V(G), E(G))$  is said to be

- (i) *Connected* if there is a path between every pair of vertices;
- (ii) *Symmetric* if  $(u, v) \in E(G)$ , then  $(v, u) \in E(G)$  for all  $u, v \in V(G)$ ;
- (iii) *Transitive* if  $(u, v)$  and  $(v, w) \in E(G)$  then,  $(u, w) \in E(G)$  for all  $u, v, w \in V(G)$ .

**Definition 2.** Let  $G = (V(G), E(G))$  be a directed graph. A mapping  $T : \mathcal{C} \rightarrow \mathcal{C}$  is said to be

- (i)  *$G$ -contraction* [6] if
  - (a)  $T$  preserves edges of  $G$ , that is, if  $(u, v) \in E(G)$ , then  $(Tu, Tv) \in E(G)$ ;
  - (b) There exists  $c \in (0, 1)$  such that for any  $u, v \in V(G)$  if  $(u, v) \in E(G)$ , then  $\|Tu - Tv\| \leq c\|u - v\|$ , where  $c$  is a contraction factor;
- (ii)  *$G$ -nonexpansive* [13] if
  - (a)  $T$  preserves edges of  $G$ ;
  - (b)  $\|Tu - Tv\| \leq \|u - v\|$ , whenever  $(u, v) \in E(G)$  for all  $u, v \in V(G)$ .

**Example 1** ([11]). Let  $\mathcal{C} = [0, 2] \subset \mathbb{R}$  and  $G = (V(G), E(G))$  be a directed graph such that  $V(G) = \mathcal{C}$  and  $(u, v) \in E(G)$  if and only if  $0.5 \leq u \leq v \leq 1.7$ , where  $S$  and  $T$  are mappings of  $\mathcal{C}$  into itself and given by

$$Su = 1 + \frac{2}{3}\arcsin(u - 1) \text{ and } Tu = 1 + \frac{1}{3}\tan(u - 1),$$

for all  $u \in C$ . It is shown in [11] that both  $S$  and  $T$  are  $G$ -nonexpansive but not nonexpansive.

We write  $\rightharpoonup$  and  $\rightarrow$  denote the weak and strong convergences, respectively. A mapping  $T : C \rightarrow C$  is said to be  $G$ -demiclosed at 0 if, for any  $\{u_n\} \subseteq C$  with  $(u_n, u_{n+1})$  and such that  $u_n \rightharpoonup C$  and  $Tu_n \rightarrow 0$  imply  $Tu = 0$ .

The following definition is necessary for our algorithm to be well defined.

**Definition 3** ([17]). Assume that  $Y := \bigcap_{n=1}^{\infty} F(T_n) \neq \emptyset$  and  $Y \times Y \subseteq E(G)$ . Then,  $E(G)$  is called

- (i) Right coordinate affine if for any  $(p, q), (p, n) \in E(G)$ , then  $\gamma(p, q) + \xi(p, n) \in E(G)$  for all  $\gamma, \xi \in \mathbb{R}$  with  $\gamma + \xi = 1$ ;
- (ii) Left coordinate affine if for any  $(p, q), (m, q) \in E(G)$ , then  $\gamma(p, q) + \xi(m, q) \in E(G)$  for all  $\gamma, \xi \in \mathbb{R}$  with  $\gamma + \xi = 1$ .

If  $E(G)$  is right and left coordinate affine, then  $E(G)$  is coordinate affine.

Our main result will be proved using the following lemma.

**Lemma 1** ([32]). Let  $\{x_n\}, \{y_n\}$  and  $\{\zeta_n\}$  be sequences of non-negative real numbers satisfying the inequality

$$x_{n+1} \leq (1 + \zeta_n)x_n + y_n$$

for all  $n \geq 1$ . If  $\sum_{n=1}^{\infty} \zeta_n < \infty$  and  $\sum_{n=1}^{\infty} y_n < \infty$ , then  $\lim_{n \rightarrow \infty} x_n$  exists.

**Lemma 2** ([33]). Let  $m, n \in \mathcal{H}$  and  $\xi \in [0, 1]$ . Then,

- (i)  $\|\xi m + (1 - \xi)n\|^2 = \xi\|m\|^2 + (1 - \xi)\|n\|^2 - \xi(1 - \xi)\|m - n\|^2$ ;
- (ii)  $\|m \pm n\|^2 = \|m\|^2 \pm 2\langle m, n \rangle + \|n\|^2$ .

**Lemma 3** ([34]). Let  $\{u_n\}$  and  $\{\mu_n\}$  be sequences of non-negative real numbers satisfying the inequality

$$u_{n+1} \leq (1 + \mu_n)u_n + \mu_n u_{n-1}$$

for all  $n \geq 1$ . Then, the following inequality holds:

$$u_{n+1} \leq M \cdot \prod_{j=1}^n (1 + 2\mu_j),$$

where  $M = \max\{u_1, u_2\}$ . Moreover, if  $\sum_{n=1}^{\infty} \mu_n < \infty$ , then  $\{u_n\}$  is bounded.

We say that  $v \in C$  is a weak cluster point of  $\{u_n\}$  if there is a subsequence  $\{u_{n_k}\}$  of  $\{u_n\}$  such that  $u_{n_k} \rightharpoonup v$  and the set of all weak cluster points of  $\{u_n\}$  is denoted by  $\omega_w(u_n)$ .

To prove our main convergence result, we need the following Opial's lemma.

**Lemma 4** ([35]). Let  $\{u_n\}$  be a sequence in  $\mathcal{H}$  such that there exists  $\emptyset \neq Y \subset \mathcal{H}$ . If for any  $p \in Y$ ,  $\lim_{n \rightarrow \infty} \|u_n - p\|$  exists and  $\omega_w(u_n) \in Y$ , then there exists  $v \in Y$  such that  $\{u_n\}$  weakly converges to  $v$ .

**Definition 4** ([36]). Let  $\{S_n\}$  and  $\varphi$  be two families of nonexpansive mappings of  $C$  into itself. Suppose that  $\emptyset \neq F(\varphi) \subset \bigcap_{n=1}^{\infty} F(S_n)$ , where  $F(\varphi)$  stands for the set of all common fixed points of each  $S \in \varphi$ . The sequence  $\{S_n\}$  satisfies the NST-condition (I) with  $\varphi$  if

$$\lim_{n \rightarrow \infty} \|S_n u_n - u_n\| = 0 \implies \lim_{n \rightarrow \infty} \|S u_n - u_n\| = 0$$

for all bounded sequences  $\{u_n\} \subset \mathcal{C}$  and  $S \in \varphi$ . A sequence  $\{S_n\}$  satisfies the NST-condition (I) with  $S$  if  $\varphi = \{S\}$ .

**Example 2** ([37]). Define  $T_n = \beta_n I + (1 - \beta_n)T$ , where  $T \in \psi$  and  $0 < a \leq \beta_n \leq b < 1$  for all  $n \geq 1$ . Then,  $T_n$  is  $G$ -nonexpansive and  $\{T_n\}$  satisfies the NST-condition (I) with  $\psi$ ; see [37] for more details.

**Definition 5** ([20,38]). Let  $f, g : \mathbb{R}^n \rightarrow (-\infty, +\infty]$  be the forward-backward operator of lower semi-continuous and convex functions. A forward-backward operator  $T$  is defined by

$$T := \text{prox}_{\mu g}(I - \mu \nabla f),$$

where  $\mu > 0$  and

$$\text{prox}_{\mu g}x := \operatorname{argmin}_{y \in \mathcal{H}} \left\{ g(y) + \frac{1}{2\mu} \|y - x\|^2 \right\}.$$

This operator was introduced by Moreau [39] and it is known as the proximity operator with respect to  $\mu$  and function  $g$ . If  $\mu \in (0, \frac{2}{L})$ , then  $T$  is a nonexpansive mapping, where  $L$  is a Lipschitz constant of  $\nabla f$ .

For the definition of the proximity operator, we have the following remark; see [40].

**Remark 1** ([40]). Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be such that  $f(x) = \mu \|x\|_1$ . The proximity operator of  $f$  is evaluated by

$$\text{prox}_{\mu \|\cdot\|_1}(x) = (\text{sign}(x_i) \max(|x_i| - \mu, 0))_{i=1}^n.$$

Bussaban et al. [41] proved the following lemma.

**Lemma 5.** Let  $f$  be a convex differentiable function from  $\mathcal{H}$  into  $\mathbb{R}$  with gradient  $\nabla f$  being  $L$ -Lipschitz constant for some  $L > 0$  and  $g$  be a proper lower semi-continuous convex function from  $\mathcal{H}$  into  $\mathbb{R} \cup \{\infty\}$ . Let  $T$  be the forward-backward operator of  $f$  and  $g$ . Then,  $\{T_n\}$  satisfies the NST-condition (I) with  $T$  if  $\{T_n\}$  is the forward-backward operator of  $f$  and  $g$  such that  $a_n \rightarrow a$  with  $a, a_n \in (0, 2/L)$ .

### 3. Main Results

In this section, we introduce a new modified S-algorithm (Algorithm 1) with the inertial technical term and then we prove a weak convergence theorem of the sequence  $\{x_n\}$  which is defined by Algorithm 1 as a common fixed point of two families for  $G$ -nonexpansive mappings in Hilbert spaces with graphs.

Throughout this section, let  $\mathcal{C}$  be a nonempty closed and convex subset of a real Hilbert space  $\mathcal{H}$  and let  $G = (V(G), E(G))$  where  $V(G) = \mathcal{C}$  and  $E(G)$  is convex, right coordinate affine, symmetric, and transitive. Let  $T, S : \mathcal{C} \rightarrow \mathcal{C}$  be  $G$ -nonexpansive mappings with  $F(T) \cap F(S) \neq \emptyset$ . Let  $\{T_n\}$  and  $\{S_n\}$  be families of  $G$ -nonexpansive mappings of  $\mathcal{C}$  into itself such that  $F(T) \subset \bigcap_{n=1}^{\infty} F(T_n)$  and  $F(S) \subset \bigcap_{n=1}^{\infty} F(S_n)$ . We also let  $\mathfrak{F} = \bigcap_{n=1}^{\infty} F(T_n) \cap \bigcap_{n=1}^{\infty} F(S_n)$ .

To prove the weak convergence result of Algorithm 1, the following tools are needed.

**Proposition 1.** Let  $\check{v} \in \mathfrak{F}$  and  $y_0, x_1 \in \mathcal{C}$  be such that  $(\check{v}, y_0), (\check{v}, x_1) \in E(G)$ . Suppose that  $E(G)$  is right coordinate affine, symmetric, and transitive. Let a sequence  $\{x_n\}$  be generated by Algorithm 1. Then,  $(\check{v}, z_n), (\check{v}, y_n), (\check{v}, x_n)$  and  $(x_n, x_{n+1})$  are in  $E(G)$  for all  $n \geq 1$ .

**Proof.** We shall use strong mathematical induction to prove our result. In order to prove this, we use Algorithm 1 to obtain

$$\begin{aligned} (\check{v}, z_1) &= (\check{v}, (1 - \beta_1)x_1 + \beta_1 T_1 x_1) \\ &= (1 - \beta_1)(\check{v}, x_1) + \beta_1(\check{v}, T_1 x_1). \end{aligned}$$

Since  $T_n$  is edge-preserving and  $(\check{v}, x_1) \in E(G)$ , we have  $(\check{v}, z_1) \in E(G)$ . Using Algorithm 1, we obtain

$$\begin{aligned}(\check{v}, y_1) &= (\check{v}, (1 - \alpha_1)T_1x_1 + \alpha_1S_1z_1) \\ &= (1 - \alpha_1)(\check{v}, T_1x_1) + \alpha_1(\check{v}, S_1z_1).\end{aligned}$$

Since  $T_n$  and  $S_n$  are edge-preserving and  $(\check{v}, z_1) \in E(G)$ , we have  $(\check{v}, y_1) \in E(G)$ .

---

**Algorithm 1** (IMSA) An inertial modified S-algorithm.

---

- 1: **Initial.** Take arbitrary  $y_0, x_1 \in \mathcal{C}$  and  $n = 1$ ,  $\beta_n \in [a, b] \subset (0, 1)$ , and  $q_n \geq 0$  such that  $\sum_{n=1}^{\infty} q_n < \infty$  and  $\alpha_n \rightarrow 1$ .
- 2: **Step 1.** Compute  $y_n$  and  $z_n$ :

$$\begin{aligned}z_n &= (1 - \beta_n)x_n + \beta_nT_nx_n, \\ y_n &= (1 - \alpha_n)T_nx_n + \alpha_nS_nz_n.\end{aligned}$$

**Step 2.** Compute the inertial step:

$$x_{n+1} = y_n + q_n(y_n - y_{n-1}).$$

Then,  $n := n + 1$  and back to the first step.

---

For all  $k < n$ , we assume that  $(\check{v}, z_k), (\check{v}, y_k)$  and  $(\check{v}, x_k) \in E(G)$ . We obtain from Algorithm 1 that

$$\begin{aligned}(\check{v}, z_{k+1}) &= (\check{v}, (1 - \beta_{k+1})x_{k+1} + \beta_{k+1}T_{k+1}x_{k+1}) \\ &= (1 - \beta_{k+1})(\check{v}, x_{k+1}) + \beta_{k+1}(\check{v}, T_{k+1}x_{k+1}),\end{aligned}\tag{1}$$

$$\begin{aligned}(\check{v}, y_{k+1}) &= (\check{v}, (1 - \alpha_{k+1})T_{k+1}x_{k+1} + \alpha_{k+1}S_{k+1}z_{k+1}) \\ &= (1 - \alpha_{k+1})(\check{v}, T_{k+1}x_{k+1}) + \alpha_{k+1}(\check{v}, S_{k+1}z_{k+1})\end{aligned}\tag{2}$$

and

$$\begin{aligned}(\check{v}, x_{k+1}) &= (\check{v}, y_k + q_k(y_k - y_{k-1})) \\ &= (\check{v}, (1 + q_k)y_k - q_ky_{k-1}) \\ &= (1 + q_k)(\check{v}, y_k) - q_k(\check{v}, y_{k-1}).\end{aligned}\tag{3}$$

Since (1)–(3) and  $T_n, S_n$  preserve edges, it follows from the fact that  $E(G)$  is the right coordinate affine that  $(\check{v}, x_{k+1}), (\check{v}, z_{k+1})$  and  $(\check{v}, y_{k+1}) \in E(G)$ . Using strong mathematical induction, we have  $(\check{v}, x_n), (\check{v}, z_n), (\check{v}, y_n) \in E(G)$  for all  $n \in \mathbb{N}$ . It easy to see that  $(x_n, \check{v}) \in E(G)$ . Since  $E(G)$  is transitive and  $(x_n, \check{v}), (\check{v}, x_{n+1}) \in E(G)$ , we obtain  $(x_n, x_{n+1}) \in E(G)$ , as required.  $\square$

**Lemma 6.** Let  $\mathcal{H}$  be a real Hilbert space and  $\mathcal{C}$  be a nonempty closed convex subset of  $\mathcal{H}$ . Let  $\{y_n\}$  be a sequence generated by Algorithm 1 and  $(\check{v}, y_0), (\check{v}, x_1) \in E(G)$  for arbitrary  $y_0, x_1 \in \mathcal{C}$  and  $\check{v} \in \mathfrak{F}$ . Then,  $\lim_{n \rightarrow \infty} \|\check{v} - y_n\|$  exists.

**Proof.** Let  $\check{v} \in \mathfrak{F}$ . By Proposition 1, we have  $(\check{v}, z_n), (\check{v}, x_n), (\check{v}, y_n) \in E(G)$ . Then

$$\begin{aligned}\|\check{v} - z_n\| &= \|\check{v} - \beta_n T_n x_n - (1 - \beta_n)x_n\| \\ &\leq (1 - \beta_n)\|\check{v} - x_n\| + \beta_n\|\check{v} - T_n x_n\| \\ &\leq (1 - \beta_n)\|\check{v} - x_n\| + \beta_n\|\check{v} - x_n\| \\ &= \|\check{v} - x_n\|,\end{aligned}\quad (4)$$

and

$$\begin{aligned}\|\check{v} - y_n\| &= \|\check{v} - \alpha_n S_n z_n - (1 - \alpha_n)T_n x_n\| \\ &\leq (1 - \alpha_n)\|\check{v} - T_n x_n\| + \alpha_n\|\check{v} - S_n z_n\| \\ &\leq (1 - \alpha_n)\|\check{v} - x_n\| + \alpha_n\|\check{v} - z_n\|\end{aligned}\quad (5)$$

$$\begin{aligned}&\leq (1 - \alpha_n)\|\check{v} - x_n\| + \alpha_n\|\check{v} - x_n\| \\ &= \|\check{v} - x_n\|.\end{aligned}\quad (6)$$

We obtain from (6) that

$$\begin{aligned}\|\check{v} - y_n\| &\leq \|\check{v} - x_n\| \\ &= \|\check{v} - y_{n-1} - \varrho_{n-1}(y_{n-1} - y_{n-2})\| \\ &\leq \|\check{v} - y_{n-1}\| + \varrho_{n-1}\|y_{n-2} - y_{n-1}\| \\ &\leq (1 + \varrho_{n-1})\|\check{v} - y_{n-1}\| + \varrho_{n-1}\|\check{v} - y_{n-2}\|.\end{aligned}\quad (7)$$

It follows from Lemma 3 that  $\|\check{v} - y_n\| \leq K \cdot \prod_{j=1}^n (1 + 2\varrho_j)$ , where  $K = \max\{\|\check{v} - y_1\|, \|\check{v} - y_2\|\}$ . Hence,  $\{y_n\}$  is bounded sequence. Moreover,  $\{x_n\}$  and  $\{z_n\}$  are bounded. Therefore,

$$\sum_{n=1}^{\infty} \varrho_n \|y_n - y_{n-1}\| < \infty. \quad (8)$$

Applying Lemma 1 and (7), the conclusion of Lemma 6 holds.  $\square$

**Lemma 7.** Let  $\mathcal{H}$  be a real Hilbert space and  $\mathcal{C}$  be a nonempty closed convex subset of  $\mathcal{H}$ . Let  $\{y_n\}$  be a sequence generated by Algorithm 1 and  $(\check{v}, y_0), (\check{v}, x_1) \in E(G)$  for arbitrary  $y_0, x_1 \in \mathcal{C}$  and  $\check{v} \in \mathfrak{F}$ . Then,  $\lim_{n \rightarrow \infty} \|T_n x_n - x_n\| = \lim_{n \rightarrow \infty} \|S_n x_n - x_n\| = 0$ .

**Proof.** Let  $\check{v} \in \mathfrak{F}$ . Applying Lemma 2 together with  $G$ -nonexpansiveness of  $T_n$ , we have

$$\begin{aligned}\|\check{v} - z_n\|^2 &= \|\check{v} - \beta_n T_n x_n - (1 - \beta_n)x_n\|^2 \\ &= \|\beta_n(\check{v} - T_n x_n) + (1 - \beta_n)(\check{v} - x_n)\|^2 \\ &= \beta_n\|\check{v} - T_n x_n\|^2 + (1 - \beta_n)\|\check{v} - x_n\|^2 - \beta_n(1 - \beta_n)\|T_n x_n - x_n\|^2 \\ &\leq \|\check{v} - x_n\|^2 - \beta_n(1 - \beta_n)\|T_n x_n - x_n\|^2.\end{aligned}$$

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

$$\beta_n(1 - \beta_n)\|T_n x_n - x_n\|^2 \leq \|\check{v} - x_n\|^2 - \|\check{v} - z_n\|^2. \quad (9)$$

Next, we shall show that

$$\lim_{n \rightarrow \infty} \|T_n x_n - x_n\| = 0.$$

In order to do this, we know from Lemma 6 that  $\lim_{n \rightarrow \infty} \|\check{v} - y_n\|$  exists. Call it  $a$ . From (6), we have

$$\|\check{v} - y_n\| \leq \|\check{v} - x_n\|.$$

Taking the  $\liminf$  yields

$$a \leq \liminf_{n \rightarrow \infty} \|\check{v} - x_n\|. \quad (10)$$

It follows from (8) and

$$\|\check{v} - x_{n+1}\| \leq \|\check{v} - y_n\| + \varrho_n \|y_{n-1} - y_n\|$$

that

$$\limsup_{n \rightarrow \infty} \|\check{v} - x_n\| \leq a. \quad (11)$$

Using (10) and (11), we have

$$\lim_{n \rightarrow \infty} \|\check{v} - x_n\| = a. \quad (12)$$

Since  $\|\check{v} - z_n\| \leq \|\check{v} - x_n\|$ , we obtain

$$\limsup_{n \rightarrow \infty} \|\check{v} - z_n\| \leq \limsup_{n \rightarrow \infty} \|\check{v} - x_n\| = a.$$

Then

$$\limsup_{n \rightarrow \infty} \|\check{v} - z_n\| \leq a. \quad (13)$$

Since  $\alpha_n \rightarrow 1$  as  $n \rightarrow \infty$  and (5), we obtain

$$a \leq \liminf_{n \rightarrow \infty} \|\check{v} - z_n\|. \quad (14)$$

This together with (14) yields

$$\lim_{n \rightarrow \infty} \|\check{v} - z_n\| = a. \quad (15)$$

Combining expressions (9), (12) and (15), we obtain

$$\lim_{n \rightarrow \infty} \|T_n x_n - x_n\| = 0. \quad (16)$$

Finally, we shall show that

$$\lim_{n \rightarrow \infty} \|S_n x_n - x_n\| = 0.$$

In order to show this, we consider the following

$$\begin{aligned} \|\check{v} - y_n\|^2 &= \|\alpha_n(\check{v} - S_n z_n) + (1 - \alpha_n)(\check{v} - T_n x_n)\|^2 \\ &= \alpha_n \|\check{v} - S_n z_n\|^2 + (1 - \alpha_n) \|\check{v} - T_n x_n\|^2 - \alpha_n(1 - \alpha_n) \|T_n x_n - S_n z_n\|^2 \\ &\leq \|\check{v} - x_n\|^2 - \alpha_n(1 - \alpha_n) \|T_n x_n - S_n z_n\|^2. \end{aligned}$$

Since  $\lim_{n \rightarrow \infty} \|\check{v} - y_n\| = a$  and (12), the above inequality leads to

$$\lim_{n \rightarrow \infty} \|T_n x_n - S_n z_n\| = 0. \quad (17)$$

Now

$$\|x_n - z_n\| \leq \beta_n \|T_n x_n - x_n\|$$

implies by (16) that

$$\lim_{n \rightarrow \infty} \|x_n - z_n\| = 0. \quad (18)$$

Using (16), (17) and (18), we have

$$\|S_n x_n - x_n\| = \|S_n x_n - S_n z_n\| + \|S_n z_n - T_n x_n\| + \|T_n x_n - x_n\| \quad (19)$$

and so

$$\lim_{n \rightarrow \infty} \|S_n x_n - x_n\| = 0, \quad (20)$$

as required.  $\square$

We now prove the weak convergence of Algorithm 1 to a common fixed point of two families for  $G$ -nonexpansive mappings in Hilbert spaces.

**Theorem 1.** Let  $\mathcal{H}$  be a real Hilbert space and  $\mathcal{C}$  be a nonempty closed convex subset of  $\mathcal{H}$ . Let  $\{y_n\}$  be a sequence generated by Algorithm 1 and  $(\check{v}, y_0), (\check{v}, x_1) \in E(G)$  for arbitrary  $y_0, x_1 \in \mathcal{C}$  and  $\check{v} \in \mathfrak{F}$ . Suppose that  $\{T_n\}$  and  $\{S_n\}$  satisfy the NST-condition (I) with  $T$  and  $S$ , respectively. Then,  $\{x_n\}$  converge weakly to a point in  $\mathfrak{F}$ .

**Proof.** Let  $\check{v} \in \mathfrak{F}$  be such that  $(\check{v}, y_0), (\check{v}, x_1) \in E(G)$ . Then,  $\lim_{n \rightarrow \infty} \|\check{v} - y_n\|$  exists as proven in Lemma 6. By Lemma 7 and  $\{T_n\}$  and  $\{S_n\}$  satisfy the NST-condition (I) with  $T$  and  $S$ , respectively, therefore

$$\lim_{n \rightarrow \infty} \|T x_n - x_n\| = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \|S x_n - x_n\| = 0. \quad (21)$$

Since  $I - T$  and  $I - S$  are  $G$ -demiclosed at 0, we obtain  $\omega_w(x_n) \subset F(T) \cap F(S)$ . We conclude from Lemma 4 that  $\{x_n\}$  converges weakly to  $\check{v} \in F(T) \cap F(S)$ , as required.  $\square$

#### 4. Applications

In 2004, Huang et al. [42] proposed the extreme learning machine (ELM) as a feedforward neural network-based machine learning technique. The single hidden layer feedforward neural network algorithm can be more effectively used if standard ELM employs the structure of a single-layer feedforward neural network (SLFN); see [43] for more detail. Only the weight vector between the hidden and output nodes needs to be determined in the initial ELM because hidden nodes can be random [42]. The training can be completed considerably more quickly because there are a lot fewer parameters that need to be updated than with traditional SLFNs. Fast learning times, easy implementation, and little human involvement are some of the benefits of ELM; see [44]. On the other hand, unstable results necessitate many experiments to identify the best ELM design; see [45] for more details. ELM is employed in a variety of areas, including computational intelligence and pattern rearrangement.

Let us give some basic knowledge of ELM for data classification problems. After that, we use our obtained results to the convex minimization problem.

Let  $\{(x_k, t_k) : x_k \in \mathbb{R}^n, t_k \in \mathbb{R}^m, k = 1, 2, \dots, N\}$  be a set of training of  $N$  distinct samples, where  $x_k = [x_{k1}, x_{k2}, \dots, x_{kn}]$  is an input data and  $t_k = [t_{k1}, t_{k2}, \dots, t_{km}]$  is a target.  $\mathfrak{W}(x)$  that represents the activation function, and ELM with  $M$  hidden nodes can be represented as the following mathematical model:

$$\sum_{j=1}^M \rho_j \mathfrak{W}(\langle w_j, x_i \rangle + d_j) = o_i, \quad i = 1, \dots, N,$$



where  $\rho_j = [\rho_{j1}, \rho_{j2}, \dots, \rho_{jm}]^T$  is the weight vector that connects the hidden node and the  $j$ -th output node,  $w_j = [w_{j1}, w_{j2}, \dots, w_{jn}]^T$  is the weight vector that connects the hidden node and the  $j$ -th input nodes and  $d_j$  is the  $j$ -th hidden node's threshold.

The standard of SLFNs with  $M$  hidden nodes can be taken as samples of  $N$  without error. In other words,  $\sum_{i=1}^N \|t_i - o_i\| = 0$ , that is, there exist  $\rho_j, w_j, d_j$  such that

$$\sum_{j=1}^M \rho_j \mathfrak{W}(\langle w_j, x_i \rangle + d_j) = t_i, \quad i = 1, \dots, N.$$

From the above equations, it can be written as follows:

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

$$\mathbf{H} = \begin{bmatrix} \mathfrak{W}(\langle w_1, x_1 \rangle + d_1) & \cdots & \mathfrak{W}(\langle w_M, x_1 \rangle + d_M) \\ \vdots & \ddots & \vdots \\ \mathfrak{W}(\langle w_1, x_N \rangle + d_1) & \cdots & \mathfrak{W}(\langle w_M, x_N \rangle + d_M) \end{bmatrix},$$

$$\rho = [\rho_1^T, \dots, \rho_M^T]_{m \times M}^T, \quad \mathbf{T} = [t_1^T, \dots, t_N^T]_{m \times N}^T.$$

For the model  $\mathbf{H}\rho = \mathbf{T}$ , we aim to estimate the parameter  $\rho$  for solving the minimization problem known as ordinary least square (OLS),

$$\min_{\rho} \|\mathbf{H}\rho - \mathbf{T}\|_2^2, \quad (22)$$

where  $\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$ ,  $\mathbf{T} \in \mathbb{R}^{N \times m}$  is the target data,  $\rho \in \mathbb{R}^{M \times m}$  is an output weight,  $\mathbf{H} \in \mathbb{R}^{N \times M}$  is the hidden layer output matrix,  $N$  is the number of training data, and  $M$  is the number of unknown variables.

There are several ways to estimate the solution of Equation (22) using mathematical models. The output weight  $\rho$  can be obtained in different ways; see [42,46–48]. The solution  $\rho$  is obtained from  $\rho = \mathbf{H}^\dagger \mathbf{T}$  when the Moore–Penrose generalized inverse  $\mathbf{H}^\dagger$  of  $\mathbf{H}$  exists. However, the number of unknown variables  $M$  in a realistic situation is substantially more than the quantity of training data  $N$ , which might cause the network to become overfitted. The accuracy is low, whereas there are few  $M$  hidden nodes. Subset selection and ridge regression are the two classical methods for improving (22); see [49] for more detail. One well-known model for estimation of the output weight  $\rho$ , called least absolute shrinkage and selection operator (LASSO) [50],

$$\min_{\rho} \|\mathbf{H}\rho - \mathbf{T}\|_2^2 + \lambda \|\rho\|_1, \quad (23)$$

where  $\lambda$  is a regularization parameter. The LASSO maintains the beneficial features of both ridge regression and subset selection, that is, regression analysis using LASSO improves the predictability and interpretability of the statistical model by performing both variable selection and regularization. Five years after, the regularization techniques and the original ELM were established to enhance OLS performance. In more general, we can rewrite (23) as a minimization of the sum of the following form:

$$\min_{x \in \mathcal{H}} (f(x) + g(x)) \quad (24)$$

where  $f, g : \mathcal{H} \rightarrow (-\infty, \infty]$  are proper lower semi-continuous functions such that  $f$  is differentiable. Let  $\mathfrak{S} := \operatorname{argmin}(f + g)$  be the set of all solutions of the problem (24).

We consider the convex minimization problem (24). We also know that  $\check{v}$  is the solution of problem (24) if and only if  $\check{v} = T\check{v}$ , where  $T = \operatorname{prox}_{\mu g_1}(I - \mu \nabla f_1)$  and  $\mu > 0$ ; see [20] for more detail.

Several methods have been proposed to solve the convex minimization problem (24). Polyak [51] was the first to present a method for accelerating algorithms and providing an improved convergence behavior by including an inertial step. Since then, numerous authors have employed the inertial technique to speed up the convergence rate of their algorithms to solve various problems; see [30,34,41,52–56].

The fast iterative shrinkage–thresholding algorithm (FISTA) [30] which performs an inertial step, is one of the most well-known forward–backward-type algorithms. It is defined by

$$\begin{cases} y_n = Tx_n, \\ t_{n+1} = \frac{1 + \sqrt{1 + 4t_n^2}}{2}, \\ \theta_n = \frac{t_n - 1}{t_{n+1}}, \\ x_{n+1} = y_n + \theta_n(y_n - y_{n-1}), \end{cases} \quad (25)$$

where  $n \geq 1$ ,  $T := \text{prox}_{\frac{1}{L}g}(I - \frac{1}{L}\nabla f)$ ,  $x_1 = y_0 \in \mathbb{R}^n$ ,  $t_1 = 1$ , and  $\theta_n$  is the inertial step size, which was introduced by Nesterov [57]. Beck et al. [30] introduced FISTA and proved the convergence rate of this algorithm. They also applied these results to the image restoration problem.

Recently, Bussaban et al. [41] introduced parallel inertial S-iteration forward–backward algorithm (PISFBA) [41]. It is defined by

$$\begin{cases} y_n = x_n + \theta_n(x_n - x_{n-1}), \\ z_n = (1 - \beta_n)x_n + \beta_n T_n x_n, \\ x_{n+1} = (1 - \alpha_n)T_n y_n + \alpha_n T_n z_n, \end{cases} \quad (26)$$

where  $n \geq 1$ ,  $x_0 = x_1 \in \mathcal{H}$ ,  $0 < q < \alpha_n \leq 1$ ,  $0 < s < \beta_n < r < 1$  and  $\sum_{n=1}^{\infty} \theta_n \|x_n - x_{n-1}\| < \infty$ . They proved the weak convergence theorem of PISFBA and applied this method to solve regression and data classification problems.

Finally, we constructed Algorithm 2 to solve the convex minimization problem (24) by applying Algorithm 1. Let  $T_n = \text{prox}_{\mu_n g_1}(I - \mu_n \nabla f_1)$  and  $S_n = \text{prox}_{\kappa_n g_2}(I - \kappa_n \nabla f_2)$ , where  $\mu_n \in (0, 2/L_1)$ ,  $\kappa_n \in (0, 2/L_2)$  and  $f_i, g_i : \mathcal{H} \rightarrow (-\infty, \infty]$ ,  $i = 1, 2$  are proper lower semi-continuous functions such that  $f_i$  are differentiable and that  $\nabla f_i$  are a Lipschitz continuity with constant  $L_i > 0$ .

---

**Algorithm 2** (FBIMSA) A forward–backward inertial modified S-algorithm.

---

- 1: **Initial.** Take arbitrary  $y_0, x_1 \in \mathcal{C}$  and  $n = 1$  when  $\beta_n, \alpha_n$  and  $q_n$  are the same as in Algorithm 1.
- 2: **Step 1.** Compute  $y_n$  and  $z_n$ :

$$\begin{aligned} z_n &= (1 - \beta_n)x_n + \beta_n \text{prox}_{\mu_n g_1}(I - \mu_n \nabla f_1)x_n, \\ y_n &= (1 - \alpha_n)\text{prox}_{\mu_n g_1}(I - \mu_n \nabla f_1)x_n + \alpha_n \text{prox}_{\kappa_n g_2}(I - \kappa_n \nabla f_2)z_n. \end{aligned}$$

**Step 2.** Compute the inertial step:

$$x_{n+1} = y_n + q_n(y_n - y_{n-1}).$$

Then,  $n := n + 1$  and back to the first step.

---

In the next theorem, we use the result of the convergence theorem of Algorithm 1 to obtain the convergence theorem of Algorithm 2.

**Theorem 2.** Let a sequence  $\{x_n\}$  be generated by Algorithm 2. Then,  $x_n \rightarrow \check{v} \in \mathfrak{S}$ , where  $\mathfrak{S} := \text{argmin}(f_1 + g_1) \cap \text{argmin}(f_2 + g_2)$ .

**Proof.** Let  $T_n = \text{prox}_{\mu_n g_1}(I - \mu_n \nabla f_1)$  and  $S_n = \text{prox}_{\kappa_n g_2}(I - \kappa_n \nabla f_2)$ , where  $\mu_n \in (0, 2/L_1)$  and  $\kappa_n \in (0, 2/L_2)$ . Then,  $T_n$  and  $S_n$  are nonexpansive operators for all  $n$ . Similarly, we set  $T$  and  $S$  to be forward-backward operators of  $f_1$  and  $f_2$  with respect to  $\mu$  and  $\kappa$ , respectively, where  $\mu \in (0, 2/L_1)$  and  $\kappa \in (0, 2/L_2)$ . Then,  $T$  and  $S$  are nonexpansive operators. Thus,  $T = \text{prox}_{\mu g_1}(I - \mu \nabla f_1)$  and  $S = \text{prox}_{\kappa g_2}(I - \kappa \nabla f_2)$ . By Proposition 26.1 in [38], we know that  $\cap_{n=1}^{\infty} F(T_n) = \text{argmin}(f_1 + g_1)$  and  $\cap_{n=1}^{\infty} F(S_n) = \text{argmin}(f_2 + g_2)$ . It is derived from Lemma 5 that  $\{T_n\}$  and  $\{S_n\}$  satisfy the NST-condition (I) with  $T$  and  $S$ , respectively. Applying Theorem 1, we obtain the required result directly by setting the complete graph  $G = \mathbb{R}^n \times \mathbb{R}^n$  on  $\mathbb{R}^n$ .  $\square$

## 5. Numerical Experiments

This section will present the basic ELM model and its fundamental supervised classification versions. We also give the result of data classification using each method.

For solving the convex minimization problem (24), we use the model of LASSO when  $\mathcal{W}(x)$  is sigmoid. We set  $f_1(x) = f_2(x) = \|\mathbf{H}\rho - \mathbf{T}\|_2^2$  and  $g_1(x) = g_2(x) = \lambda \|\rho\|_1$  for our algorithm. For other algorithms, we set  $f(x) = \|\mathbf{H}\rho - \mathbf{T}\|_2^2$ ,  $g(x) = \lambda \|\rho\|_1$ .

The values shown in Table 1 are set for all control parameters,  $L = 2\|\mathbf{H}_1\|^2$ , where  $\mathbf{H}_1$  is a hidden layer output matrix of a training matrix, and  $I$  is an iterations number. We use the output data's accuracy to measure the performance of each method which is calculated by

$$\text{accuracy} = \frac{\text{correct predicted data}}{\text{all data}} \times 100$$

**Table 1.** Selected parameters of each method.

Methods	Setting
Algorithm 2	$\alpha_n = \frac{n}{n+1}, \beta_n = 0.99, c = 1/L, q_n = \frac{n}{n+1}$ if $1 \leq n \leq I$ , and $1/2^n$ otherwise
PISFBA	$\alpha_n = \beta_n = \frac{0.9n}{n+1}, c = 1/L, \theta_n = \frac{1}{2^n \ x_n - x_{n-1}\ }$ if $x_n \neq x_{n-1}$ , and 0 otherwise
FISTA	$t_1 = 1, t_{n+1} = (1 + \sqrt{1 + 4t_n^2})/2,$ $\theta_n = (t_n - 1)/t_{n+1}$

Next, we use the Breast Cancer, Heart Disease UCI and Ionosphere data sets for classifying which are detailed as follows:

Wisconsin Breast Cancer data set [58]: W.H. Wolberg created this data set, at the General Surgery Department, University of Wisconsin, Clinical Sciences Center, W.N. Street, and O.L. Mangasarian, Computer Sciences Department, University of Wisconsin. It contains 2 classes, 569 observations, and 30 attributes.

Heart Disease UCI [59]: This data set contains 76 attributes. However, all published studies use only a subset of 14 of them. This data set shows the patient's presence of heart disease. Our goal is to divide the data into two categories.

Ionosphere data set [60]: This radar data set, from the Ionosphere collection, was gathered by a system near Goose Bay, Labrador. This data set consists of 351 observations and 34 attributes. Radar results indicating signs of an ionosphere structure are considered "good". Bad returns are those whose transmissions do not penetrate the ionosphere.

We set up the training and testing data on Table 2.

We performed the experiments in order to compare the performance of each studied algorithm, namely Algorithm 2, PISFBA, and FISTA. In each data set, we use the number of hidden nodes  $M$  and the number of iterations  $I$  as follows:

The number of hidden nodes  $M$  depends on the characteristic of each data set and the number of iterations for each data set is selected to achieve the highest performance for each studied algorithm.

**Table 2.** Data sets of Breast Cancer, Heart Disease UCI, and Ionosphere, 70% of training and 30% of testing of each data set.

Data Set	Features	Sample	
		Training Set	Testing Set
Breast Cancer	14	478	205
Heart Disease UCI	14	213	90
Ionosphere	34	205	146

The following numerical experiments are obtained by each algorithm and each data set under the control sequences in Table 1 and the selected parameters for each data set in Table 3.

**Table 3.** Number of hidden nodes and iterations for each data set.

Data Sets	Number of Hidden Nodes ( $M$ )	Number of Iterations ( $I$ )
Breast Cancer	100	400
Heart Disease UCI	350	500
Ionosphere	50	100

In Table 4, we use acc.Train and acc.Test to represent the accuracy of training and testing, respectively.

**Table 4.** Performance comparison using different methods.

Data Set	Algorithm 2		PISFBA		FISTA	
	acc.Train	acc.Test	acc.Train	acc.Test	acc.Train	acc.Test
Breast Cancer	97.11	97.46	96.11	95.89	92.64	93.25
Heart Disease UCI	78.34	79.01	72.54	73.84	69.64	68.52
Ionosphere	93.98	94.09	90.54	90.71	91.33	91.71

We observe from Table 4 that our proposed algorithm, Algorithm 2, has a higher performance than PISFBA and FISTA in terms of the accuracy of training and testing of each data set. So, we can conclude from our experiments that Algorithm 2 can be used for data classifications of the selected data sets with higher accuracy compared to PISFBA and FISTA.

**Remark 2.** Limitations of the proposed algorithm and its applications.

Our proposed algorithm, Algorithm 2, guarantees weak convergence in a setting of real Hilbert spaces under the control sequences  $\{\alpha_n\}$  and  $\{\beta_n\}$  together with the inertial parameter  $q_n$  such that the conditions  $\alpha_n \rightarrow 1$ ,  $\beta_n \in [a, b] \subset (0, 1)$  and  $q_n \geq 0$ ,  $\sum_{n=1}^{\infty} q_n < \infty$ . For applications of Algorithm 2, we have to choose  $\{\alpha_n\}$ ,  $\{\beta_n\}$  and  $\{q_n\}$  under above restrictions. However, in finite-dimensional Euclidean spaces, we obtain a strong convergence of Algorithm 2. Another limitation of Algorithm 2 is computation technique for Lipschitzian constant of  $\nabla f$  when  $f(x) = \|\mathbf{H}\rho - \mathbf{T}\|_2^2$ . In

the case of big data sets, it may cause difficulty in such computation because of the large dimension of the matrix  $H$ .

## 6. Discussions

In this work, we propose a new accelerated common fixed-point algorithm, Algorithm 2, and employ it to solve data classifications of Breast Cancer, Heart Diseases and Ionosphere. A convergence theorem of the proposed algorithm is established under some control conditions  $\alpha_n \rightarrow 1$ ,  $\beta_n \in [a, b] \subset (0, 1)$  and  $\varrho_n \geq 0$ ,  $\sum_{n=1}^{\infty} \varrho_n < \infty$ . From our experiments, Algorithm 2 has a higher performance than PISFBA and FISTA. The performance of our proposed algorithm depends on the inertial parameter  $\varrho_n$ . We note that if we choose  $\varrho_n$  which is closed to 1, then we obtain a higher performance of Algorithm 2. We also observe that the performance Algorithm 2 depends on the number of hidden nodes and characteristics of data sets. However, future research will focus on finding new methods or techniques that increase the performance of algorithms for the classification of big real data sets of NCDs of patients from the Sriphat medical center, the faculty of medicine, Chiang Mai University, Thailand.

## 7. Conclusions

We introduce and prove the weak convergence theorem of an inertial modified S-algorithm (IMSA) for finding a common fixed point of two countable families of G-nonexpansive mappings. Firstly, we proved the weak convergence of IMSA. Secondly, we proposed a new forward-backward inertial modified S-algorithm (FBIMSA) for solving the convex minimization problem. Finally, we applied the proposed algorithm to solve the data classification of Breast Cancer, Heart Diseases and Ionosphere. The numerical results demonstrated the advantages of the proposed algorithm.

**Author Contributions:** Conceptualization, S.S.; Formal analysis, K.J. and S.S.; Investigation, K.J.; Methodology, S.S.; Supervision, S.S.; Validation, S.S.; Writing—original draft, K.J.; Writing—review and editing, S.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** NSRF via the program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B05F640183].

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All data were obtained from <https://archive.ics.uci.edu> (accessed on 23 September 2022).

**Acknowledgments:** This research has also 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.

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

## References

1. Banach, S. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fundam. Math.* **1922**, *3*, 133–181.
2. Branciari, A. A fixed point theorem for mappings satisfying a general contractive condition of integral type. *Int. J. Math. Math. Sci.* **2002**, *29*, 531–536.
3. Suzuki, T. A generalized Banach contraction principle that characterizes metric completeness. *Proc. Am. Math. Soc.* **2008**, *136*, 1861–1869.
4. Zhang, X. Common fixed point theorems for some new generalized contractive type mappings. *J. Math. Anal. Appl.* **2007**, *333*, 780–786.
5. Ran, A.C.M.; Reurings, M.C.B. A fixed point theorem in partially ordered sets and some applications to matrix equations. *Proc. Am. Math. Soc.* **2004**, *132*, 1435–1443.
6. Jachymski, J. The contraction principle for mappings on a metric space with a graph. *Proc. Am. Math. Soc.* **2008**, *136*, 1359–1373.
7. Thianwan, T.; Yambangwai, D. Convergence analysis for a new two-step iteration process for G-nonexpansive mappings with directed graphs. *Fixed Point Theory Appl.* **2019**, *2019*, 44.

8. Bojor, F. Fixed point of  $\psi$ -contraction in metric spaces endowed with a graph. *Anna. Univ. Crai. Math. Comp. Sci. Ser.* **2010**, *37*, 85–92.
9. Aleomraninejad, S.M.A.; Rezapour, S.; Shahzad, N. Some fixed point result on a metric space with a graph. *Topol. Appl.* **2012**, *159*, 659–663.
10. Tiammee, J.; Suantai, S. Coincidence point theorems for graph-preserving multi-valued mappings. *Fixed Point Theory Appl.* **2014**, *2014*, 70. <https://doi.org/10.1186/1687-1812-2014-70>.
11. Sridarat, P.; Supaturatorn, R.; Suantai, S.; Cho, Y.J. Convergence analysis of SP-iteration for G-nonexpansive mappings with directed graphs. *Bull. Malays. Math. Sci. Soc.* **2019**, *42*, 2361–2380.
12. Tripak, O. Common fixed points of G-nonexpansive mappings on Banach spaces with a graph. *Fixed Point Theory Appl.* **2016**, *2016*, 87.
13. Tiammee, J.; Kaewkhao, A.; Suantai, S. On Browder's convergence theorem and Halpern iteration process for G-nonexpansive mappings in Hilbert spaces endowed with graphs. *Fixed Point Theory Appl.* **2015**, *2015*, 187.
14. Suantai, S.; Kankam, K.; Chalamjiak, P.; Chalamjiak, W. A parallel monotone hybrid algorithm for a finite family of G-nonexpansive mappings in Hilbert spaces endowed with a graph applicable in signal recovery. *Comput. Appl. Math.* **2021**, *40*, 145.
15. Janngam, K.; Wattanataweekul, R. A new accelerated fixed-point algorithm for classification and convex minimization problems in Hilbert spaces with directed graphs. *Symmetry* **2022**, *14*, 1059. <https://doi.org/10.3390/sym14051059>.
16. Janngam, K.; Wattanataweekul, R. An accelerated fixed-point algorithm with an inertial technique for a countable family of G-nonexpansive mappings applied to image recovery. *Symmetry* **2022**, *14*, 662. <https://doi.org/10.3390/sym14040662>.
17. Wattanataweekul, R.; Janngam, K. An accelerated common fixed point algorithm for a countable family of G-nonexpansive mappings with applications to image recovery. *J. Inequal. Appl.* **2022**, *2022*, 68. <https://doi.org/10.1186/s13660-022-02796-y>.
18. Byrne, C. A unified treatment of some iterative algorithms in signal processing and image reconstruction. *Inverse Probl.* **2004**, *20*, 103–120.
19. Chalamjiak, P.; Shehu, Y. Inertial forward-backward splitting method in Banach spaces with application to compressed sensing. *Appl. Math.* **2019**, *64*, 409–435.
20. Combettes, P.L.; Wajs V.R. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **2005**, *4*, 1168–1200.
21. Suantai, S.; Eiamniran, N.; Pholasa, N.; Chalamjiak, P. Three-step projective methods for solving the split feasibility problems. *Mathematics* **2019**, *7*, 712. [doi:10.3390/math7080712](https://doi.org/10.3390/math7080712).
22. Suantai, S.; Kesornprom, S.; Chalamjiak, P. Modified proximal algorithms for finding solutions of the split variational inclusions. *Mathematics* **2019**, *7*, 708. [doi:10.3390/math7080708](https://doi.org/10.3390/math7080708).
23. Censor, Y.; Bortfeld, T.; Martin, B.; Trofimov, A. A unified approach for inversion problems in intensity-modulated radiation therapy. *Phys. Med. Biol.* **2006**, *51*, 2353–2365.
24. Censor, Y.; Elfving, T.; Kopf, N.; Bortfeld, T. The multiple set split feasibility problem and its applications. *Inverse Probl.* **2005**, *21*, 2071–2084.
25. Ben-Tal, A.; Nemirovski, A. *Lectures on Modern Convex Optimization, Analysis, Algorithms, and Engineering Applications*; MPS/SIAM Ser. Optim.; SIAM: Philadelphia, PA, USA, 2001.
26. Bioucas-Dias, J.; Figueiredo, M. A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.* **2007**, *16*, 2992–3004.
27. Chen, S.S.; Donoho, D.L.; Saunders, M.A. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **1998**, *20*, 33–61.
28. Donoho, D.L.; Johnstone, I.M. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Assoc.* **1995**, *90*, 1200–1224.
29. Figueiredo, M.A.T.; Nowak, R.D.; Wright, S.J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Signal Process.* **2007**, *1*, 586–597.
30. Beck, A.; Teboulle, M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202.
31. Johnsonbaugh, R. *Discrete Mathematics*; Pearson: Hoboken, NJ, USA, 1997.
32. Tan, K.; Xu, H.K. Approximating fixed points of nonexpansive mappings by the ishikawa iteration process. *J. Math. Anal. Appl.* **1993**, *178*, 301–308.
33. Takahashi, W. *Introduction to Nonlinear and Convex Analysis*; Yokohama Publishers: Yokohama, Japan, 2009.
34. Hanjing A.; Suantai, S. A fast image restoration algorithm based on a fixed point and optimization method. *Mathematics* **2020**, *8*, 378.
35. Moudafi, A.; Al-Shemas, E. Simultaneous iterative methods for split equality problem. *Trans. Math. Program. Appl.* **2013**, *1*, 1–11.
36. Nakajo, K.; Shimoji, K.; Takahashi, W. Strong convergence to a common fixed point of families of nonexpansive mappings in Banach spaces. *J. Nonlinear Convex Anal.* **2007**, *8*, 11–34.
37. Suantai, S.; Donganont, M.; Chalamjiak, W. Hybrid methods for a countable family of G-nonexpansive mappings in Hilbert spaces endowed with graphs. *Mathematics* **2019**, *7*, 936. <https://doi.org/10.3390/math7100936>.
38. Bauschke, H.H.; Combettes, P.L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2017.
39. Moreau, J.J. Fonctions convexes duales et points proximaux dans un espace hilbertien. *Comptes Rendus Acad. Sci. Paris Ser. A Math.* **1962**, *255*, 2897–2899.

40. Beck, A. *First-Order Methods in Optimization*; Tel-Aviv University: Tel Aviv-Yafo, Israel, 2017; pp. 129–177, ISBN 978-1-61197-498-0.
41. 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.
42. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501.
43. Ding, S.; Xu, X.; Nie, R. Extreme learning machine and its applications. *Neural Comput. Appl.* **2014**, *25*, 549–556, <https://doi.org/10.1007/s00521-013-1522-8>.
44. Wang, Z.; Xin, J.; Sun, P.; Lin, Z.; Yao, Y.; Gao, X. Improved lung nodule diagnosis accuracy using lung CT images with uncertain class, *Comput. Methods Programs Biomed.* **2018**, *162*, 197–209, <https://doi.org/10.1016/j.cmpb.2018.05.028>.
45. Silitonga, A.S.; Shamsuddin, A.H.; Mahlia, T.M.I.; Milano, J.; Kusumo, F. Siswanto, J.; Dharma, S.; Sebayang, A.H.; Masjuki, H.H.; Ong, H.C. Biodiesel synthesis from Ceiba pentandra oil by microwave irradiation-assisted transesterification: ELM modeling and optimization. *Renew. Energy* **2020**, *146*, 1278–1291. <https://doi.org/10.1016/j.renene.2019.07.065>.
46. Huang, G.-B.; Chen, L.; Siew, C.-K. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *Trans. Neural Netw.* **2006**, *17*, 879–892.
47. Widrow, B.; Greenblatt, A.; Kim, Y.; Park, D. The no-prop algorithm: A new learning algorithm for multilayer neural networks. *J. Comput. Graph. Stat.* **2013**, *17*, 182–188.
48. Brunton, S. L.; Kutz, J. N. Brunton, S. L.; Kutz, J. N. Singular Value Decomposition (SVD). In *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*; Cambridge University Press: Cambridge, UK, 2019; pp. 3–46.
49. Tikhonov, A.N.; Arsenin, V.Y. *Solutions of Ill-Posed Problems*; V.H. Winston: Washington, DC, USA, 1977.
50. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B Methodol.* **1996**, *58*, 267–288.
51. Polyak, B. Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.* **1964**, *4*, 1–17.
52. Lions, P.L.; Mercier, B. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **1979**, *16*, 964–979.
53. Janngam, K.; Suantai, S. An accelerated forward-backward algorithm with applications to image restoration problems. *Thai. J. Math.* **2021**, *19*, 325–339.
54. Alakoya, T.O.; Jolaoso, L.O.; Mewomo, O.T. Two modifications of the inertial Tseng extra gradient method with self-adaptive step size for solving monotone variational inequality problems. *Demonstr. Math.* **2020**, *53*, 208–224.
55. Gebrie, A.G.; Wangkeeree, R. Strong convergence of an inertial extrapolation method for a split system of minimization problems. *Demonstr. Math.* **2020**, *53*, 332–351.
56. Yatakoat, P.; Suantai, S.; Hanjing, A. On some accelerated optimization algorithms based on fixed point and linesearch techniques for convex minimization problems with applications. *Adv. Contin. Discrete Models* **2022**, *2022*, 25. <https://doi.org/10.1186/s13662-022-03698-5>.
57. Nesterov, Y. A method for solving the convex programming problem with convergence rate  $O(1/k^2)$ . *Dokl. Akad. Nauk SSSR* **1983**, *269*, 543–547.
58. Mangasarian, O. L.; Wolberg W. H. Cancer diagnosis via linear programming. *SIAM News* **1990**, *23*, 1–18.
59. Lichman, M. UCI Machine Learning Repository. Available online: <https://archive.ics.uci.edu/ml/datasets/Heart+Disease> (accessed on 20 April 2020).
60. Dua, D.; Graff, C. UCI Machine Learning Repository. Available online: <https://archive.ics.uci.edu/ml/datasets/ionosphere> (accessed on 23 September 2022).