

Article

A New Accelerated Fixed-Point Algorithm for Classification and Convex Minimization Problems in Hilbert Spaces with Directed Graphs

Kobkoon Janngam ¹ and Rattanakorn Wattanataweekul ^{2,*}

¹ Graduate Ph.D. Degree Program in Mathematics, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand; kobkoon_jan@cmu.ac.th

² Department of Mathematics, Statistics and Computer, Faculty of Science, Ubon Ratchathani University, Ubon Ratchathani 34190, Thailand

* Correspondence: rattanakorn.w@ubu.ac.th

Abstract: A new accelerated algorithm for approximating the common fixed points of a countable family of G -nonexpansive mappings is proposed, and the weak convergence theorem based on our main results is established in the setting of Hilbert spaces with a symmetric directed graph G . As an application, we apply our results for solving classification and convex minimization problems. We also apply our proposed algorithm to estimate the weight connecting the hidden layer and output layer in a regularized extreme learning machine. For numerical experiments, the proposed algorithm gives a higher performance of accuracy of the testing set than that of FISTA-S, FISTA, and nAGA.

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



Citation: 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>

Academic Editor: Lorentz JÄNTSCHI

Received: 29 April 2022

Accepted: 19 May 2022

Published: 21 May 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

Let H be a real Hilbert space with the norm $\|\cdot\|$ and C be a nonempty closed convex subset of H . A mapping $T : C \rightarrow C$ is said to be *nonexpansive* if it satisfies the following symmetric contractive-type condition:

$$\|Tx - Ty\| \leq \|x - y\|,$$

for all $x, y \in C$; see [1].

The element $x \in C$ is a fixed point of T if $Tx = x$ and $F(T) := \{x \in C : x = Tx\}$ stands for the set of all fixed points of T .

Fixed point theory, i.e., the study of the conditions under which a map admits a fixed point, is an extensive area of research due to its numerous applications in many fields. It started with Banach's work, which established the existence of a unique fixed point for a contraction using a classical theorem known as the Banach contraction principle; see [2]. The contraction principle of Banach has been expanded and generalized in various directions due to its applications in mathematics and other fields. One of the more recent generalizations is due to Jachymski.

Jachymski [3] introduced the structure of the graph on metric spaces using fixed point theory and obtained certain conditions for self-mapping to be a Picard operator. Several authors [4–7] proved fixed point theorems for a new type of contraction on a metric space endowed with graphs. Aleomraninejad et al. [8] used the idea of Reich et al. [9] and proved a strong convergence theorem for G -contractive and G -nonexpansive mappings. On hyperbolic metric spaces, Alfuraidan and Khamsi [10] gave a definition of G -monotone nonexpansive multivalued mappings and proved the existence of a fixed point for multivalued contraction and monotone single-valued mappings. Later on, Alfuraidan [11]

presented and study the existence of fixed points for G -monotone nonexpansive and extended the results of Jachymski [3]. For approximating common fixed points of a finite family of G -nonexpansive mappings, Suantai et al. [12] used the shrinking projection with the parallel monotone hybrid method. They also used a graph to prove a strong convergence theorem in Hilbert spaces under specific conditions, and they then applied their iterative scheme to signal recovery.

In the past decade, algorithms for approximating fixed points of G -nonexpansive mappings without inertial techniques have been proposed by many researchers; see [3,8,10–18]. We require more efficient algorithms for solving such problems. As a result, some accelerated fixed-point algorithms using inertial techniques have been proposed to improve convergence behavior; see [19–27]. Recently, Janngam et al. [28] proved the weak convergence theorem for a countable family of G -nonexpansive mappings in a Hilbert space by using a coordinate affine structure with an inertial technique. They also applied their method to image recovery.

Inspired by previous research described above, we introduce a new accelerated algorithm based on the concept of the inertial technique for finding a common fixed point of a family of G -nonexpansive mappings in Hilbert spaces. We employ our result to solve data classification and convex minimization problems and also compare our algorithm efficiency to that of FISTA-S, FISTA, and nAGA.

This paper is classified as follows: in Section 2, we give certain terminology as well as some facts that will be useful in later sections. We investigate and prove our algorithm's weak convergence in Section 3. For application, we apply our method for solving convex minimization and data classification problems in Section 4 and provide numerical experiments of classification problems in Section 5. The last section of our paper, Section 6, is a summary.

2. Preliminaries

Let C be a nonempty subset of a real Banach space X . Let $\Delta = \{(u, u) : u \in C\}$, where Δ stands for the diagonal of the Cartesian product $C \times C$. Consider a directed graph G in which the set $V(G)$ of its vertices corresponds to C , and the set $E(G)$ of its edges contains all loops. A directed graph G is said to have parallel edges if two or more edges with both the same tail vertex and the same head vertex.

Assume that G does not have parallel edges. Then, $G = (V(G), E(G))$. The conversion of a graph G is denoted by G^{-1} . Thus, we have

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

Let us give some definitions of basic graph properties which are used in this paper (see [29] for more details).

Definition 1. A graph G is said to be

- (i) *symmetric* if for all $(x, y) \in E(G)$; we have $(y, x) \in E(G)$;
- (ii) *transitive* if for any $u, v, w \in V(G)$ with $(u, v), (v, w) \in E(G)$; then, $(u, w) \in E(G)$;
- (iii) *connected* if there is a path between any two vertices of the graph G .

The definition of G -contraction [3] and G -nonexpansive mappings [13] are given as follows.

Definition 2. A mapping $T : C \rightarrow C$ is said to be

- (i) *G -contraction* if
 - (a) T is edge-preserving, i.e., $(Tu, Tv) \in E(G)$ for all $(u, v) \in E(G)$.
 - (b) There exists $\rho \in [0, 1)$ such that $\|Tu - Tv\| \leq \rho\|u - v\|$ for all $(u, v) \in E(G)$, where ρ is called a contraction factor.
- (ii) *G -nonexpansive* if

- (a) T is edge-preserving.
- (b) $\|Tu - Tv\| \leq \|u - v\|$ for all $(u, v) \in E(G)$.

Example 1. Let $C = [0, 2] \subset \mathbb{R}$. Suppose that $(u, v) \in E(G)$ if and only if $0.4 \leq u, v \leq 1.6$ or $u = v$, where $u, v \in \mathbb{R}$. Let $S, T : C \rightarrow C$ be defined by

$$Su = \cos(\tan(u - 1))\sin\left(\frac{\pi}{2}\right) \text{ and } Tu = 1 + \frac{\ln(u)}{3},$$

for all $u \in C$. Then, both S and T are G -nonexpansive but not nonexpansive (see [30] for more details).

Definition 3. A mapping $T : C \rightarrow C$ is called G -demiclosed at 0 if $u_n \rightharpoonup u$ and $Tu_n \rightarrow 0$, then $Tu = 0$ for all sequence $\{u_n\} \subseteq C$ with $(u_n, u_{n+1}) \in E(G)$.

To prove our main result, we need the definition of the coordinate affine of the graph G as follows.

Definition 4 ([28]). Assume that $\Lambda := \bigcap_{n=1}^{\infty} F(T_n) \neq \emptyset$ and $\Lambda \times \Lambda \subseteq E(G)$. Then, $E(G)$ is said to be

- (i) left coordinate affine if $\alpha(x, y) + \beta(u, y) \in E(G)$ for all $(x, y), (u, y) \in E(G)$ and all $\alpha, \beta \in \mathbb{R}$ with $\alpha + \beta = 1$.
- (ii) right coordinate affine if $\alpha(x, y) + \beta(x, z) \in E(G)$ for all $(x, y), (x, z) \in E(G)$ and all $\alpha, \beta \in \mathbb{R}$ with $\alpha + \beta = 1$.

We say that $E(G)$ is coordinate affine if $E(G)$ is both left and right coordinate affine.

The results of the following lemmas can be used to prove our main theorem; see also [19,31,32].

Lemma 1 ([31]). Let $\{\eta_n\}, \{v_n\}$ and $\{\vartheta_n\}$ be sequences of nonnegative real numbers such that $\sum_{n=1}^{\infty} \vartheta_n < \infty$ and $\sum_{n=1}^{\infty} v_n < \infty$. Suppose that

$$\eta_{n+1} \leq (1 + \vartheta_n)\eta_n + v_n \text{ for all } n \geq 1.$$

Then, $\lim_{n \rightarrow \infty} \eta_n$ exists.

Lemma 2 ([32]). For a real Hilbert space H , the following results hold:

- (i) For any $u, v \in H$ and $\gamma \in [0, 1]$,

$$\|\gamma u + (1 - \gamma)v\|^2 = \gamma\|u\|^2 + (1 - \gamma)\|v\|^2 - \gamma(1 - \gamma)\|u - v\|^2.$$

- (ii) For any $u, v \in H$,

$$\|u \pm v\|^2 = \|u\|^2 \pm 2\langle u, v \rangle + \|v\|^2.$$

Lemma 3 ([19]). Let $\{v_n\}$ and $\{\mu_n\}$ be sequences of nonnegative real numbers such that

$$v_{n+1} \leq (1 + \mu_n)v_n + \mu_n v_{n-1} \text{ for all } n \geq 1.$$

Then,

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

where $M = \max\{v_1, v_2\}$. If $\sum_{n=1}^{\infty} \mu_n < \infty$, then $\{v_n\}$ is bounded.

Let $\{u_n\}$ be a sequence in X . We write $u_n \rightharpoonup u$ to indicate that a sequence $\{u_n\}$ converges weakly to a point $u \in H$. Similarly, $u_n \rightarrow u$ will symbolize the strong convergence.

For $v \in C$, if there is a subsequence $\{u_{n_k}\}$ of $\{u_n\}$ such that $u_{n_k} \rightharpoonup v$, then v is called a weak cluster point of $\{u_n\}$. The set of all weak cluster points of $\{u_n\}$ is denoted by $\omega_w(u_n)$.

The following lemma was proved by Moudafi and Al-Shemas; see [33].

Lemma 4 ([33]). Let $\{u_n\}$ be a sequence in a real Hilbert space H such that there exists $\emptyset \neq \Lambda \subset H$ satisfying:

(i) For any $p \in \Lambda$, $\lim_{n \rightarrow \infty} \|u_n - p\|$ exists.

(ii) Any weak cluster point of $\{u_n\} \in \Lambda$.

Then, there exists $x^* \in \Lambda$ such that $u_n \rightharpoonup x^*$.

Let $\{T_n\}$ and ψ be families of nonexpansive mappings of C into itself such that $\emptyset \neq F(\psi) \subset \bigcap_{n=1}^{\infty} F(T_n)$, where $F(\psi)$ is the set of all common fixed points of each $T \in \psi$. A sequence $\{T_n\}$ satisfies the NST-condition (I) with ψ if, for any bounded sequence $\{u_n\}$ in C ,

$$\lim_{n \rightarrow \infty} \|T_n u_n - u_n\| = 0 \text{ implies } \lim_{n \rightarrow \infty} \|T u_n - u_n\| = 0,$$

for all $T \in \psi$; see [34]. If $\psi = \{T\}$, then $\{T_n\}$ satisfies the NST-condition (I) with T .

Example 2 ([30]). Let $T \in \psi$. Define $T_n = \beta_n I + (1 - \beta_n)T$, where $0 < s \leq \beta_n \leq t < 1$ for all $n \in \mathbb{N}$. Therefore, $\{T_n\}$ is a family of G -nonexpansive mappings and satisfies the NST-condition.

Let $A : H \rightarrow 2^H$ be a maximal monotone operator and $c > 0$. The resolvent of A is defined by $J_{cA} = (I + cA)^{-1}$, where I is an identity operator. If $A = \partial f$ for some $f \in \Gamma_0(H)$, where $\Gamma_0(H)$ stands for the set of proper lower semicontinuous convex functions from $H \rightarrow (-\infty, +\infty]$, then $J_{cA} = \text{prox}_{cf}$. The forward-backward operator of lower semicontinuous and convex functions of $f, g : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ has the following definition:

A forward-backward operator T is defined by $T := \text{prox}_{\lambda g}(I - \lambda \nabla f)$ for $\lambda > 0$, where ∇f is the gradient operator of function f and $\text{prox}_{\lambda g} x := \argmin_{y \in H} \left\{ g(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\}$ (see [35,36]). Moreau [37] defined the operator $\text{prox}_{\lambda g}$ as the proximity operator with respect to λ and g . If $\lambda \in (0, 2/L)$, then T is nonexpansive and L is a Lipschitz constant of ∇f .

We have the following remark for the definition of the proximity operator; see [38].

Remark 1. Let $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be given by $g(x) = \lambda \|x\|_1$. The proximity operator of g is evaluated by the following formula

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

where $x = (x_1, x_2, \dots, x_n)$ and $\|x\|_1 = \sum_{i=1}^n |x_i|$.

The following lemma was proved by Bussaban et al.; see [20].

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

3. Main Results

Let C be a nonempty closed and convex subset of a real Hilbert space H with a directed graph $G = (V(G), E(G))$ such that $V(G) = C$. Let $\{T_n\}$ be a family of G -nonexpansive mappings of C into itself such that $\emptyset \neq \bigcap_{n=1}^{\infty} F(T_n)$.

The following proposition is useful for our main theorem.

Proposition 1. Let $x^* \in \cap_{n=1}^{\infty} F(T_n)$ and $y_0, x_1 \in C$ be such that $(x^*, y_0), (x^*, x_1) \in E(G)$. Let $\{x_n\}$ be a sequence generated by Algorithm 1. Suppose $E(G)$ is symmetric, transitive, and a right coordinate affine. Then, $(x^*, x_n), (x^*, y_n), (x^*, z_n)$ and $(x_n, x_{n+1}) \in E(G)$ for all $n \in \mathbb{N}$.

Algorithm 1 (ASA) An Accelerated S-algorithm

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

$$\begin{aligned} z_n &= (1 - \beta_n)x_n + \beta_n T_n x_n, \\ y_n &= (1 - \alpha_n)T_n x_n + \alpha_n T_n z_n, \\ x_{n+1} &= y_n + \theta_n(y_n - y_{n-1}). \end{aligned}$$

Then, $n := n + 1$ and go to Step 1.

Proof. We shall prove the results by using strong mathematical induction. From Algorithm 1, we obtain

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

Since $(x^*, x_1) \in E(G)$ and T_n is edge preserving, we obtain $(x^*, z_1) \in E(G)$. Again, by Algorithm 1, we obtain

$$\begin{aligned} (x^*, y_1) &= (x^*, (1 - \alpha_1)T_1 x_1 + \alpha_1 T_1 z_1) \\ &= (1 - \alpha_1)(x^*, T_1 x_1) + \alpha_1(x^*, T_1 z_1). \end{aligned}$$

Since $(x^*, z_1) \in E(G)$ and T_n is edge preserving, we obtain $(x^*, y_1) \in E(G)$. Next, we assume that $(x^*, z_k), (x^*, y_k)$ and $(x^*, x_k) \in E(G)$ for all $k < n$. By Algorithm 1, we obtain

$$\begin{aligned} (x^*, z_{k+1}) &= (x^*, (1 - \beta_{k+1})x_{k+1} + \beta_{k+1}(T_{k+1}x_{k+1})) \\ &= (1 - \beta_{k+1})(x^*, x_{k+1}) + \beta_{k+1}(x^*, T_{k+1}x_{k+1}), \end{aligned} \quad (1)$$

$$\begin{aligned} (x^*, y_{k+1}) &= (x^*, (1 - \alpha_{k+1})T_{k+1}y_{k+1} + \alpha_{k+1}T_{k+1}z_{k+1}) \\ &= (1 - \alpha_{k+1})(x^*, T_{k+1}y_{k+1}) + \alpha_{k+1}(x^*, T_{k+1}z_{k+1}) \end{aligned} \quad (2)$$

and

$$\begin{aligned} (x^*, x_{k+1}) &= (x^*, y_k + \theta_k(y_k - y_{k-1})) \\ &= (x^*, (1 + \theta_k)y_k - \theta_k y_{k-1}) \\ &= (1 + \theta_k)(x^*, y_k) - \theta_k(x^*, y_{k-1}). \end{aligned} \quad (3)$$

By (1)–(3) and since $E(G)$ is right coordinate affine and T_n is edge preserving, we obtain $(x^*, x_{k+1}), (x^*, y_{k+1})$ and (x^*, z_{k+1}) are in $E(G)$. By strong mathematical induction, we conclude that $(x^*, x_n), (x^*, y_n), (x^*, z_n) \in E(G)$ for all $n \in \mathbb{N}$. Since $E(G)$ is symmetric, we obtain $(x_n, x^*) \in E(G)$. Since $(x_n, x^*), (x^*, x_{n+1}) \in E(G)$ and $E(G)$ is transitive, we obtain $(x_n, x_{n+1}) \in E(G)$ as required. \square

We now prove the weak convergence of G -nonexpansive mapping in a real Hilbert space by using Algorithm 1.

Theorem 1. Let C be a nonempty closed and convex subset of a real Hilbert space H with a directed graph $G = (V(G), E(G))$ with $V(G) = C$ and $E(G)$ is symmetric, transitive, and right coordinate affine. Let $y_0, x_1 \in C$ and $\{x_n\}$ be a sequence in H defined by Algorithm 1. Suppose $\{T_n\}$ satisfies

NST-condition (I) with T such that $\emptyset \neq F(T) \subset \cap_{n=1}^{\infty} F(T_n)$ and $(x^*, y_0), (x^*, x_1) \in E(G)$ for all $x^* \in \cap_{n=1}^{\infty} F(T_n)$. Then, $x_n \rightarrow x^* \in F(T)$.

Proof. Let $x^* \in \cap_{n=1}^{\infty} F(T_n)$. By the definition of z_n and y_n , we obtain

$$\begin{aligned}\|x^* - z_n\| &\leq (1 - \beta_n)\|x^* - x_n\| + \beta_n\|x^* - T_n x_n\| \\ &= \|x^* - x_n\|\end{aligned}\quad (4)$$

and

$$\begin{aligned}\|x^* - y_n\| &\leq (1 - \alpha_n)\|x^* - T_n x_n\| + \alpha_n\|x^* - T_n z_n\| \\ &\leq (1 - \alpha_n)\|x^* - x_n\| + \alpha_n\|x^* - z_n\|\end{aligned}\quad (5)$$

which implies that

$$\|x^* - y_n\| \leq \|x^* - x_n\|. \quad (6)$$

By the definition of x_n , we obtain

$$\|x^* - x_n\| \leq \|x^* - y_{n-1}\| + \theta_{n-1}\|y_{n-2} - y_{n-1}\|. \quad (7)$$

From (6) and (7), we obtain

$$\|x^* - y_n\| \leq \|x^* - y_{n-1}\| + \theta_{n-1}\|y_{n-2} - y_{n-1}\|. \quad (8)$$

It follows from (8) that

$$\|x^* - y_n\| \leq (1 + \theta_{n-1})\|x^* - y_{n-1}\| + \theta_{n-1}\|x^* - y_{n-2}\|. \quad (9)$$

Applying Lemma 3, we obtain $\|x^* - y_n\| \leq M \cdot \prod_{j=1}^n (1 + 2\theta_j)$, where $M = \max\{\|x^* - y_1\|, \|x^* - y_2\|\}$. Since $\sum_{n=1}^{\infty} \theta_n < \infty$, we obtain that $\{y_n\}$ is bounded and so are $\{z_n\}$ and $\{x_n\}$. Thus,

$$\sum_{n=1}^{\infty} \theta_n \|y_n - y_{n-1}\| < \infty \quad (10)$$

By Lemma 1 and (9), we obtain $\lim_{n \rightarrow \infty} \|x^* - y_n\|$ exists. By Lemma 2(i) and the definition of z_n , we obtain

$$\begin{aligned}\|x^* - z_n\|^2 &= \|(1 - \beta_n)(x^* - x_n) + \beta_n(x^* - T_n x_n)\|^2 \\ &= (1 - \beta_n)\|x^* - x_n\|^2 + \beta_n\|x^* - T_n x_n\|^2 - (1 - \beta_n)\beta_n\|T_n x_n - x_n\|^2 \\ &\leq \|x^* - x_n\|^2 - (1 - \beta_n)\beta_n\|T_n x_n - x_n\|^2.\end{aligned}\quad (11)$$

Let $\lim_{n \rightarrow \infty} \|x^* - y_n\| = a$. From the boundedness of $\{x_n\}$ and (6), we obtain

$$\liminf_{n \rightarrow \infty} \|x^* - x_n\| \geq a. \quad (12)$$

Since $\|x^* - x_{n-1}\| \leq \|x^* - y_n\| + \theta_n\|y_n - y_{n-1}\|$ and (10), we obtain

$$\limsup_{n \rightarrow \infty} \|x^* - x_n\| \leq a. \quad (13)$$

It follows from (12) and (13) that

$$\lim_{n \rightarrow \infty} \|x^* - x_n\| = a. \quad (14)$$

From (4), one can easily see that $\limsup_{n \rightarrow \infty} \|x^* - z_n\| \leq a$. By (5) with $\alpha_n \rightarrow 1$, we obtain that $a \leq \liminf_{n \rightarrow \infty} \|x^* - z_n\|$. Thus,

$$\lim_{n \rightarrow \infty} \|x^* - z_n\| = a. \quad (15)$$

Use the facts that (11), (14), and (15) yield

$$\|T_n x_n - x_n\| \rightarrow 0.$$

According to $\{T_n\}$ satisfying the NST-condition (I) with T , we obtain that $\|T x_n - x_n\| \rightarrow 0$ as $n \rightarrow \infty$. Let $\omega_w(x_n)$ be the set of all weak cluster point of $\{x_n\}$. Thus, $\omega_w(x_n) \subset F(T)$ by demicloseness of $I - T$ at 0. From Lemma 4, we conclude that $x_n \rightharpoonup x^*$ with $x^* \in F(T)$ as required. \square

Corollary 1. *Let C be a nonempty closed and convex subset of a real Hilbert space H and let $\{T_n\}$ be a family of nonexpansive mappings of C into itself. Let $y_0, x_1 \in C$, and $\{x_n\}$ be a sequence in H defined by Algorithm 1. Suppose that $\{T_n\}$ satisfies NST-condition (I) with T such that $\emptyset \neq F(T) \subset \cap_{n=1}^{\infty} F(T_n)$. Then, $\{x_n\}$ converges weakly to a point in $F(T)$.*

4. Applications

In the past decade, extreme learning machine (ELM) [39], a new learning algorithm for single-hidden layer feedforward networks (SLFNs), has been extensively studied in various research topics for machine learning and artificial intelligence such as face classification, image segmentation, regression, and data classification problems. ELM was theoretically proven to have extremely fast learning speed and good performance better than the gradient-based learning such as backpropagation in most of the cases. The target of this model is to find the parameter β that solves the following minimization problem, called ordinary least square (OLS),

$$\min_{\beta} \|\mathbf{H}\beta - \mathbf{T}\|, \quad (16)$$

where $\|\cdot\|$ is the l_2 -norm defined by $\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$, $\mathbf{T} \in \mathbb{R}^{N \times m}$ is the target of data, $\beta \in \mathbb{R}^{M \times m}$ is a weight which connects hidden layer and output layer, and $\mathbf{H} \in \mathbb{R}^{N \times M}$ is the hidden layer output matrix. In general mathematical modeling, there are several methods to estimate the solution of (16); in this case, the solution β obtained by $\beta = \mathbf{H}^{\dagger} \mathbf{T}$, where \mathbf{H}^{\dagger} is the Moore–Penrose generalized inverse of \mathbf{H} . However, in a real situation, the number of unknown variable M is much more than the number of training data N , which causes the network to possibly lead to overfitting. On the other hand, the accuracy is low while the number of hidden nodes M is small. Thus, in order to improve (16), several regularization methods were introduced. The classical two standard techniques for improving (16) are subset selection and ridge regression (sometimes called Tikhonov regularization) [40].

In this paper, we focus on the following problem, called least absolute shrinkage and selection operator (LASSO) [41],

$$\min_{\beta} \|\mathbf{H}\beta - \mathbf{T}\|_2^2 + \lambda \|\beta\|_1, \quad (17)$$

where λ is a regularization parameter. LASSO tries to retain the good features of both subset selection and ridge regression [41]. After the regularization methods and the original ELM were introduced for improving performance of OLS, five years later, the regularized extreme learning machine [42] was proposed and applied to solve regression problems. In general, the (17) can be rewritten as minimization of $f + g$, that is,

$$\min_x F(x) := f(x) + g(x), \quad (18)$$

where f is a smooth convex function with gradient having Lipschitz constant L and g is a convex smooth (or possible non-smooth) function. The solution of (18) can be rewritten into \tilde{x} is a minimizer of $(f + g)$ if and only if $0 \in \nabla f(\tilde{x}) + \partial g(\tilde{x})$, where $\nabla f(\tilde{x})$ is the gradient of f and $\partial g(\tilde{x})$ is a subdifferential of g by using Fermat's rule (see [35] for more details). In fixed point theory, Parikh et al. [43] characterized (18) as follows: \tilde{x} is a minimizer of $f + g$ if and only if

$$\tilde{x} = \text{prox}_{\lambda g}(I - \lambda \nabla f)(\tilde{x}) = J_{\lambda \partial g}(I - \lambda \nabla f)(\tilde{x}), \quad (19)$$

where $\text{prox}_{\lambda g}$ is the proximity operator of λg , $\lambda > 0$ and $J_{\partial g}$ is defined by $J_{\partial g} = (I + \partial g)^{-1}$, $J_{\partial g}$ is the resolvent of ∂g and I is an identity operator. The problem (18) can be rewritten into a general problem, called a zero of sum of two operators problem, by finding \tilde{x} such that

$$\tilde{x} \in \text{zer}(A + B),$$

where $A, B : H \rightarrow 2^H$ are two set-valued operators and $\text{zer}(A + B) := \{x : 0 \in Ax + Bx\}$. In this case, we assume that $A : H \rightarrow 2^H$ is a maximal monotone operator and $B : H \rightarrow H$ is an L -Lipschitz operator. For convenience, (19) also can be rewritten as:

$$\tilde{x} = T\tilde{x},$$

where $T = \text{prox}_{\lambda g}(I - \lambda \nabla f)$. It is also known that T is nonexpansive if $\lambda \in (0, 2/L)$ when L is a Lipschitz constant of ∇f .

We are interested in applying our proposed method for solving a convex minimization problem and compared the convergence behavior of our proposed algorithm with the others and give some applications to solve classification problems. Our proposed method will be used to solve (18). Over the past two decades, several algorithms have been introduced for solving the problem (18). A simple and classical algorithm is the forward-backward algorithm (FBA), which was introduced by Lions and Mercier [21].

The forward-backward algorithm (FBA) is defined by

$$\begin{cases} y_n = x_n - \gamma \nabla f x_n, \\ x_{n+1} = x_n + \rho_n (J_{\gamma \partial g} y_n - x_n), \end{cases} \quad (20)$$

where $n \geq 1$, $x_0 \in H$ and L is a Lipschitz constant of ∇f , $\gamma \in (0, 2/L)$, $\delta = 2 - (\gamma L/2)$ and $\{\rho_n\}$ is a sequence in $[0, \delta]$ such that $\sum_{n \in \mathbb{N}} \rho_n (\delta - \rho_n) = +\infty$. A technique for improving speed and giving a better convergence behavior of the algorithms was introduced firstly by Polyak [44] by adding an inertial step called inertial techniques. Since then, many authors have employed the inertial technique to accelerate their algorithms for various kinds of problems; see [19,20,22–26]. The performance of FBA can be improved using an iterative method with the inertial steps described below.

A fast iterative shrinkage-thresholding algorithm (FISTA) [25] is defined by

$$\begin{cases} y_n = T x_n, \\ s_{n+1} = \frac{1 + \sqrt{1 + 4s_n^2}}{2}, \\ \mu_n = \frac{s_n - 1}{s_{n+1}}, \\ x_{n+1} = y_n + \mu_n (y_n - y_{n-1}), \end{cases} \quad (21)$$

where $n \geq 1$, $s_1 = 1$, $x_1 = y_0 \in \mathbb{R}^n$, $T := \text{prox}_{\frac{1}{L}g}(I - \frac{1}{L}\nabla f)$ and μ_n is the inertial step size. Beck and Teboulle [25] solved the image recovery and proved the convergence rate using

FISTA. The inertial step size μ_n of the FISTA was firstly introduced by Nesterov [45].

A fast iterative shrinkage-thresholding algorithm-Iteration (FISTA-S) [27] is defined by

$$\begin{cases} s_{n+1} = \frac{1+\sqrt{1+4s_n^2}}{2}, \mu_n = \frac{s_n-1}{s_{n+1}}, \\ y_n = x_n + \mu_n(x_n - x_{n-1}), \\ z_n = (1 - \beta_n)y_n + \beta_n T_n y_n, \\ x_{n+1} = (1 - \alpha_n)T_n y_n + \alpha_n T_n z_n, \end{cases} \quad (22)$$

where $x_0, x_1 \in H$, $\alpha_n, \beta_n \in [a, b] \subset (0, 1)$, $T_n = \text{prox}_{c_n g}(I - c_n \nabla f)$ and $T = \text{prox}_{cg}(I - c \nabla f)$. Bussaban et al. [27] solved the image recovery and proved the weak convergence theorem using FISTA-S.

A new accelerated proximal gradient algorithm (nAGA) [26] is defined by

$$\begin{cases} y_n = x_n + \mu_n(x_n - x_{n-1}), \\ x_{n+1} = T_n[(1 - \rho_n)y_n + \rho_n T_n y_n], \end{cases} \quad (23)$$

where $n \geq 1$, $T_n = \text{prox}_{a_n g}(I - a_n \nabla f)$ with $a_n \in (0, 2/L)$ and $\{\mu_n\}, \{\rho_n\}$ are sequences in $(0, 1)$ and $\frac{\|x_n - x_{n-1}\|_2}{\mu_n} \rightarrow 0$. The nAGA was introduced for proving a convergence theorem by Verma and Shukla [26]. The nonsmooth convex minimization problem with sparsity, including regularizers, was solved using this method for the multitask learning framework.

Theorem 2. Let H be a Hilbert space, $A : H \rightarrow 2^H$ be maximal monotone operator and $B : H \rightarrow H$ be an L -Lipschitz operator. Let $a \in (0, 2/L)$ and $\{a_n\} \subset (0, 2/L)$ such that $a_n \rightarrow a$. Define $T_n = J_{c_n A}(I - c_n B)$ and $T = J_{cA}(I - cB)$. Suppose that $\text{zer}(A + B) \neq \emptyset$. Let $\{x_n\}$ be a sequence in H defined by Algorithm 1. Then, $\{x_n\}$ converges weakly to a point in $\text{zer}(A + B)$.

Proof. Using Proposition 26.1(iv) (see [35]), we have $\{T_n\}$ and T are nonexpansive mappings such that $F(T) = F(T_n) = \text{zer}(A + B)$. Then, the proof is completed by Theorem 1 and Lemma 5. \square

The convergence of Algorithm 2 is obtained by using our main result.

Algorithm 2 (FBASA) A forward-backward accelerated S-algorithm

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

$$\begin{aligned} z_n &= (1 - \beta_n)x_n + \beta_n \text{prox}_{a_n g}(I - a_n \nabla f)y_n, \\ y_n &= (1 - \alpha_n) \text{prox}_{a_n g}(I - a_n \nabla f) + \alpha_n \text{prox}_{a_n g}(I - a_n \nabla f)z_n, \\ x_{n+1} &= y_n + \theta_n(y_n - y_{n-1}). \end{aligned}$$

Then, $n := n + 1$ and go to Step 1.

Theorem 3. For $f, g : \mathbb{R}^n \rightarrow (-\infty, \infty]$, f is a smooth convex function with a gradient having a Lipschitz constant L and g is a convex function. Let $a_n \in (0, 2/L)$ be such that $\{a_n\}$ converges to a and let $T := \text{prox}_{ag}(I - a \nabla f)$ and $T_n := \text{prox}_{a_n g}(I - a_n \nabla f)$ and let $\{x_n\}$ be a sequence generated by Algorithm 2, where β_n, α_n and θ_n are the same as in Algorithm 1. Then,

- (i) $\|x^* - x_{n+1}\| \leq M \cdot \prod_{j=1}^n (2\theta_j + 1)$, where $M = \max\{\|x^* - x_1\|, \|x^* - x_2\|\}$ and $x^* \in \text{Argmin}(f + g)$;
- (ii) $\{x_n\}$ converges weakly to a point in $\text{Argmin}(f + g)$.

Proof. We know that T and $\{T_n\}$ are nonexpansive operators, and $F(T) = \bigcap_{n=1}^{\infty} F(T_n) = \text{Argmin}(f + g)$ for all n ; see [35]. Then, $\{T_n\}$ satisfies the NST-condition (I) with T by using

Lemma 5. We get the desired result immediately from Theorem 1 by putting $G = \mathbb{R}^n \times \mathbb{R}^n$, the complete graph, on \mathbb{R}^n . \square

5. Numerical Experiments

The classification problem is one of the most important problems in the convex minimization problem. We illustrate the process of reformulating the data classification problem in machine learning.

We first present a basic idea of extreme learning machines for data classification problem, and use our algorithm to find this problem through numerical experiments. Moreover, the performance of Algorithm 2, FISTA-S, FISTA, and nAGA are compared.

Extreme learning machine (ELM). Let $R := \{(x_k, t_k) : x_k \in \mathbb{R}^n, t_k \in \mathbb{R}^m, k = 1, 2, \dots, N\}$ be a training set of different samples N , where x_k is input data and t_k is a target. A standard SLFNs with activation function $\Psi(x)$ (for instance sigmoid), and M hidden nodes can be rewritten as

$$\sum_{j=1}^M \beta_j \Psi(\langle w_j, x_i \rangle + c_j) = o_i, i = 1, \dots, N,$$

where β_j is the weight vector connecting the j -th the output node and hidden node, w_j is the weight vector connecting the j -th the input node and hidden node, and c_j is the threshold of the j -th hidden node. The objective of standard SLFNs is to estimate these N different samples with $\sum_{i=1}^N \|t_i - o_i\| = 0$, that is, there exist β_j, w_j, c_j such that

$$\sum_{j=1}^M \beta_j \Psi(\langle w_j, x_i \rangle + c_j) = t_i, i = 1, \dots, N.$$

We can derive a simple equation from the above N equations as follows:

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

$$\mathbf{H} = \begin{bmatrix} \Psi(\langle w_1, x_1 \rangle + c_1) & \cdots & \Psi(\langle w_M, x_1 \rangle + c_M) \\ \vdots & \ddots & \vdots \\ \Psi(\langle w_j, x_N \rangle + c_1) & \cdots & \Psi(\langle w_N, x_N \rangle + c_M) \end{bmatrix},$$

$$\beta = [\beta_1^T, \dots, \beta_{m \times M}^T]_{m \times M}^T, \mathbf{T} = [t_1^T, \dots, t_N^T]_{m \times N}^T.$$

A standard SLFN goal is to estimate β_j, w_j , and c_j to solve (18), whereas an ELM goal is to find only β_j with w_j and c_j chosen at random.

In an experiment on classification problems, we employ the model (17) to solve the convex minimization problem. We set $f(x) = \|\mathbf{H}\beta - \mathbf{T}\|_2^2$ and $g(x) = \lambda \|\beta\|_1$. Next, we use the Iris dataset to classify iris plant types, and the Heart Disease UCI dataset to identify heart patients which are detailed as follows:

Iris dataset [46]. This dataset has three classes of 50 examples, each of which represents a different variety of iris plant. The purpose is to identify each iris plant species based on the length of its sepals and petals.

Heart Disease UCI dataset [47]. Although there are 76 attributes in the original dataset, all published experiments only use 14 of them. Data on patients with heart disease are provided in this dataset. We divide the data into two classes based on the predicted attributes.

The dataset was graciously provided by <https://archive.ics.uci.edu> (accessed on April 2020).

All control parameters are set to the values shown in Table 1, $L = 2\|\mathbf{H}_1\|^2$, where \mathbf{H}_1 is a hidden layer output matrix of a training matrix, $M = 100$, and $\Psi(x)$ is sigmoid.

Each dataset is given a training set, as indicated in Table 2. We evaluated the output data's accuracy by

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

Table 1. Parameters of each methods.

Methods	Setting
Algorithm 2	$\alpha_n = \beta_n = 0.5, c = 1/L, \theta_n = 0.9$ if $1 \leq n \leq 400$, and $1/2^n$ otherwise
FISTA-S	$\alpha_n = \beta_n = 0.5, c = 1/L, s_1 = 1,$ $s_{n+1} = (1 + \sqrt{1 + 4s_n^2})/2, q_n = (s_n - 1)/s_{n+1}$ if $1 \leq n \leq 400$, and $1/2^n$ otherwise
FISTA	$s_1 = 1, s_{n+1} = (1 + \sqrt{1 + 4s_n^2})/2,$ $q_n = (s_n - 1)/s_{n+1}$
nAGA	$\rho_n = 0.9, \gamma = 1/L, s_1 = 1,$ $s_{n+1} = (1 + \sqrt{1 + 4s_n^2})/2, q_n = (s_n - 1)/s_{n+1}$

Table 2. Training and testing sets of the Iris and Heart Disease UCI datasets.

Dataset	Attributes	Sample	
		Train	Test
Heart Disease UCI	14	213	90
Iris	4	105	45

From the results in Table 3, we conclude that the proposed learning algorithm under selection with the identical number of hidden nodes M has high performance in terms of the accuracy. The weight computed by Algorithm 2 converges faster to the optimal weight and performs accuracy better than those computed by FISTA-S, FISTA, and nAGA.

Table 3. Performance comparison using different methods at the 400th iteration.

Dataset	Algorithm 2		FISTA-S		FISTA		nAGA	
	Train	Test	Train	Test	Train	Test	Train	Test
Heart Disease UCI	68.26	62.37	39.96	37.63	53.76	50.50	67.83	60.22
Iris	100	98.10	95.56	94.29	95.56	94.29	94.29	93.33

6. Conclusions

We generate and study an algorithm for approximating the common fixed points of a countable family of G -nonexpansive mappings and prove the weak convergence of our algorithm. In addition, we give an application of our result for solving data classification and convex minimization problems. Finally, our numerical experiments assert that our proposed algorithm provides more accuracy than FISTA-S, FISTA, and nAGA.

In future work, we aim to find new models and methods for data prediction and classification of real datasets in medical science and create new innovations for health care service.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The first author was supported by Fundamental Fund 2022, Chiang Mai University, Thailand, under the supervision of Suthep Suantai. The second author would like to thank Ubon Ratchathani University, Thailand.

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

References

- Berinde, V. A Modified Krasnosel'skiĭ–Mann Iterative Algorithm for Approximating Fixed Points of Enriched Nonexpansive Mappings. *Symmetry* **2022**, *14*, 123. [\[CrossRef\]](#)
- Banach, S. Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. *Fund. Math.* **1922**, *3*, 133–181. [\[CrossRef\]](#)
- Jachymski, J. The contraction principle for mappings on a metric space with a graph. *Proc. Am. Math. Soc.* **2008**, *136*, 1359–1373. [\[CrossRef\]](#)
- Bojor, F. Fixed point of ψ -contraction in metric spaces endowed with a graph. *Anna. Univ. Crai. Math. Comp. Sci. Ser.* **2010**, *37*, 85–92.
- Chifu, C.; Petrușel, G. Generalized contractions in metric spaces endowed with a graph. *Fixed Point Theory Appl.* **2012**, *2012*, 161. [\[CrossRef\]](#)
- Acar, Ö.; Altun, I. Multivalued F-contractive mappings with a graph and some fixed point results. *Publ. Math. Debr.* **2016**, *88*, 305–317. [\[CrossRef\]](#)
- Acar, Ö.; Aydi, H.; De la Sen, M. New Fixed Point Results via a Graph Structure. *Mathematics* **2021**, *9*, 1013. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Riech, S.; Zaslavski, A.J. Convergence of inexact iterative schemes for nonexpansive set-valued mappings. *Fixed Point Theory Appl.* **2010**, *2010*, 518243. [\[CrossRef\]](#)
- Alfuraidan, M.R.; Khamsi, M.A. Fixed points of monotone nonexpansive mappings on a hyperbolic metric space with a graph. *Fixed Point Theory Appl.* **2015**, *2015*, 44. [\[CrossRef\]](#)
- Alfuraidan, M.R. Fixed points of monotone nonexpansive mappings with a graph. *Fixed Point Theory Appl.* **2015**, *2015*, 49. [\[CrossRef\]](#)
- 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. *Comp. Appl. Math.* **2021**, *40*, 145. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Tripak, O. Common fixed points of G-nonexpansive mappings on Banach spaces with a graph. *Fixed Point Theory Appl.* **2016**, *2016*, 87. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Glowinski, R.; Tallec, P.L. *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*; SIAM: Philadelphia, PA, USA, 1989.
- Haubridge, S.; Nguyen, V.H.; Strodiot, J.J. Convergence analysis and applications of the Glowinski Le Tallec splitting method for finding a zero of the sum of two maximal monotone operators. *J. Optim. Theory Appl.* **1998**, *97*, 645–673. [\[CrossRef\]](#)
- Noor, M.A. New approximation schemes for general variational inequalities. *J. Math. Anal. Appl.* **2000**, *251*, 217–229. [\[CrossRef\]](#)
- Hanjing A.; Suantai, S. A fast image restoration algorithm based on a fixed point and optimization method. *Mathematics* **2020**, *8*, 378. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Lions, P.L.; Mercier, B. Splitting Algorithms for the Sum of Two Nonlinear Operators. *SIAM J. Numer. Anal.* **1979**, *16*, 964–979. [\[CrossRef\]](#)
- Janngam, K.; Suantai, S. An accelerated forward-backward algorithm with applications to image restoration problems. *Thai J. Math.* **2021**, *19*, 325–339.
- Alakoya, T.O.; Jolaoso, L.O.; Mewomo, O.T. Two modifications of the inertial Tseng extragradient method with self-adaptive step size for solving monotone variational inequality problems. *Demonstr. Math.* **2020**, *53*, 208–224. [\[CrossRef\]](#)
- 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. [\[CrossRef\]](#)
- Beck, A.; Teboulle, M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [\[CrossRef\]](#)

26. Verma, M.; Shukla, K. A new accelerated proximal gradient technique for regularized multitask learning framework. *Pattern Recogn. Lett.* **2018**, *95*, 98–103. [[CrossRef](#)]
27. Bussaban, L.; Kaewkhao, A.; Suantai, S. Inertial s-iteration forward-backward algorithm for a family of nonexpansive operators with applications to image restoration problems. *Filomat* **2021**, *35*, 771–882. [[CrossRef](#)]
28. 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. [[CrossRef](#)]
29. Johnsonbaugh, R. *Discrete Mathematics*; Pearson: Hoboken, NJ, USA, 1997.
30. 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. [[CrossRef](#)]
31. Tan, K.; Xu, H.K. Approximating fixed points of nonexpansive mappings by the ishikawa iteration process. *J. Math. Anal. Appl.* **1993**, *178*, 301–308. [[CrossRef](#)]
32. Takahashi, W. *Introduction to Nonlinear and Convex Analysis*; Yokohama Publishers: Yokohama, Japan, 2009.
33. Moudafi, A.; Al-Shemas, E. Simultaneous iterative methods for split equality problem. *Trans. Math. Program. Appl.* **2013**, *1*, 1–11.
34. 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.
35. Bauschke, H.H.; Combettes, P.L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed.; Springer: Berlin, Germany, 2017.
36. Combettes, P.L.; Wajs V.R. Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **2005**, *4*, 168–1200. [[CrossRef](#)]
37. Moreau, J.J. Fonctions convexes duales et points proximaux dans un espace hilbertien. *C. R. Acad. Sci. Paris Ser. A Math.* **1962**, *255*, 2897–2899.
38. Beck, A. *First-Order Methods in Optimization*; Tel-Aviv University: Tel Aviv-Yafo, Israel, 2017; pp. 129–177. ISBN 978-1-61197-498-0.
39. Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme learning machine: Theory and applications. *Neurocomputing* **2006**, *70*, 489–501. [[CrossRef](#)]
40. Tikhonov, A.N.; Arsenin, V.Y. *Solutions of Ill-Posed Problems*; V.H. Winston: Washington, DC, USA, 1977.
41. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B Methodol.* **1996**, *58*, 267–288. [[CrossRef](#)]
42. Martinez-Martinez, J.M.; Escandell-Montero, P.; Soria-Olivas, E.; Martn-Guerrero, J.D.; MagdalenaBenedito, R.; Gmez-Sanchis, J. Regularized extreme learning machine for regression problems. *Neurocomputing* **2011**, *74*, 3716–3721. [[CrossRef](#)]
43. Parikh, N.; Boyd, S. Proximal Algorithms. *Found. Trends Optim.* **2014**, *1*, 127–239. [[CrossRef](#)]
44. Polyak, B. Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.* **1964**, *4*, 1–17. [[CrossRef](#)]
45. 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.
46. Dua, D.; Karra, E. *Taniskidou*, UCI Machine Learning Repository; University of California Irvine: Irvine, CA, USA, 2017.
47. Lichman, M. *UCI Machine Learning Repository*; University of California Irvine: Irvine, CA, USA, 2013.