Article

The Modified Viscosity Approximation Method with Inertial Technique and Forward–Backward Algorithm for Convex Optimization Model

Adisak Hanjing [1]**, Limpapat Bussaban** [2] **and Suthep Suantai** [3,4,*]

[1] Department of Science and Mathematics, Rajamangala University of Technology Isan Surin Campus, Surin 32000, Thailand; adisak.ha@rmuti.ac.th
[2] Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand; lim.bussaban@gmail.com
[3] Data Science Research Center, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand
[4] Research Group 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 algorithm for finding a common fixed point of nonexpansive operators, and then, a strong convergence result of the proposed method is discussed and analyzed in real Hilbert spaces. As an application, we create a new accelerated viscosity forward–backward method (AVFBM) for solving nonsmooth optimization problems of the sum of two objective functions in real Hilbert spaces, and the strong convergence of AVFBM to a minimizer of the sum of two convex functions is established. We also present the application and simulated results of AVFBM for image restoration and data classification problems.

Keywords: Hilbert space; common fixed points; viscosity forward–backward algorithm; convergence theorems; convex optimization model

MSC: 47H10; 47J25; 65K05; 90C30

1. Introduction

Image restoration is a fundamental problem in image processing. The image restoration (image deblurring or image deconvolution) is concerned with the reconstruction or estimation of the uncorrupted image from a blurred and noisy image [1,2]. Thus, the main objective of the image restoration algorithms is to reduce the blurring effects and the noise that degraded the image by minimizing the noise of the degraded image to produce an estimate image that approaches the original image.

The image restoration problem can be modeled by a linear inverse problem, which is formulated by:

$$u = Bv + e, \tag{1}$$

where $B \in \mathbb{R}^{m \times n}$ is the blurring matrix, $v \in \mathbb{R}^n$ is an original image, $u \in \mathbb{R}^m$ is the observed image, and $e \in \mathbb{R}^m$ is a noise. One of the most popular models to solve Problem (1) is the least absolute shrinkage and selection operator (LASSO) [3], which can be considered in the following form:

$$\min_v \|Bv - u\|_2^2 + \tau \|v\|_1, \tag{2}$$

where $\tau > 0$ is a regularization parameter, $\|\cdot\|_1$ is l_1-norm, and $\|\cdot\|_2$ is l_2-norm. Moreover, Problem (2) can be applied to solving many areas of science and applied science such as astronomical imaging [4], microscopy [5], and signal recovery problems [6].

The nonsmooth convex optimization model which includes (2) as a particular case has the following form:

$$\min_{x \in \mathcal{H}} \phi_1(x) + \phi_2(x), \tag{3}$$

where $\mathcal{H}$ is a Hilbert space with norm $\| \cdot \|$, and inner product $\langle \cdot, \cdot \rangle$, $\phi_2 : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ is a proper convex and lower semi-continuous function, and $\phi_1 : \mathcal{H} \to \mathbb{R}$ is convex differentiable with a Lipschitz continuous gradient constant $L > 0$. The solution set of Problem (3) will be denoted by $\Omega := Argmin(\phi_1 + \phi_2)$. Furthermore, x is a solution of Problem (3) if and only if x satisfies the fixed point equation:

$$x = prox_{c\phi_2}(I - c\nabla\phi_1)(x), \tag{4}$$

where $c > 0$, I is an identity operator, ∇f is the gradient of f, $prox_{\phi_2} = (I + \partial\phi_2)^{-1}$, and $\partial\phi_2$ is the subdifferential of ϕ_2 defined by:

$$\partial\phi_2(a^*) := \{u \in \mathcal{H} : \phi_2(a) \geq \langle u, a - a^* \rangle + \phi_2(a^*), \ a \in \mathcal{H}\},$$

see [7–9] for more details. For solving (3), the Forward–Backward splitting (FBS) algorithm [10] has been using the following form:

$$x_{k+1} = \underbrace{prox_{c_k\phi_2}}_{\text{backward step}} \underbrace{(I - c_k\nabla\phi_1)}_{\text{forward step}}(x_k), \ k \in \mathbb{N}, \tag{5}$$

where $x_1 \in \mathcal{H}$ and $0 < c_k < 2/L$. To accelerate the proximal gradient algorithm, the inertial technique or extrapolation technique was proposed by Nesterov in 1983 [11] for solving a class of convex optimization problems (3), where $F := \phi_1 + \phi_2$ is a smooth and convex function. A typical algorithm takes the following form:

$$\begin{cases} y_k = x_k + \theta_k(x_k - x_{k-1}), \\ x_{k+1} = y_k + c\nabla F(y_k), \ k \in \mathbb{N}, \end{cases} \tag{6}$$

where $c > 0$ is the step size depending on the Lipschitz continuity modulus of ∇F and the inertial parameter $\theta_k \in (0, 1)$ for all k. He also showed that by choosing $\{\theta_k\}$ such that $\sup_k \theta_k = 1$, this algorithm has a faster convergence rate than the general gradient algorithm; see [11]. In 2009, Beck and Teboulle [12] improved FBS by using the inertial techniques; this algorithm is known as the fast iterative shrinkage-thresholding algorithm (FISTA), which is defined as follows:

$$\begin{cases} y_k = prox_{\frac{1}{L}\phi_2}(x_k - \frac{1}{L}\nabla\phi_1(x_k)), \\ t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}, \quad \theta_k = \frac{t_k-1}{t_{k+1}}, \\ x_{k+1} = y_k + \theta_k(y_k - y_{k-1}), \ k \in \mathbb{N}, \end{cases} \tag{7}$$

where $x_1 = y_0 \in \mathcal{H}, t_1 = 1$ and θ_k is the inertial parameter. The FISTA has been recognized as a fast method. It is noted that the inertial parameter $\{\theta_k\}$ in (7) satisfies $\sup_k \theta_k = 1$. So, the sequence generated by FISTA has a rate of convergence that is proven to be significantly better, both theoretically and practically. Recently, Liang and Schonlieb [13] modified FISTA, called "FISTA-Mod", for the short and proved weak convergence theorem of FISTA-Mod. Moreover, they proved $\|x_k - x_{k-1}\| = O(1/k)$. The FBS and FISTA are only weak convergence in Hilbert spaces. For strong convergence, the viscosity approximation method (VAM) of the fixed point of a nonexpansive operator T was proposed by Moudafi [14], who proved the strong convergence of the methods (8) in real Hilbert spaces.

$$x_{k+1} = \gamma_k g(x_k) + (1 - \gamma_k)Tx_k, \ k \in \mathbb{N}, \tag{8}$$

where $x_1 \in \mathcal{H}$, γ_k is a sequence in $(0, 1)$ and g is a contraction operator. In 2008, Takahashi [15] modified the viscosity approximation method of Moudafi [14] for finding a common fixed point of a countable family of nonexpansive operators $\{T_k\}$. His algorithm takes the following form:

$$x_{k+1} = \gamma_k g(x_k) + (1 - \gamma_k)T_k x_k, \ k \in \mathbb{N}, \tag{9}$$

where $x_1 \in \mathcal{H}$, $\{\gamma_k\} \subset (0,1)$, and g is a contraction operator. He proved a strong convergence theorem of (9) under some conditions on $\{T_k\}$ and $\{\gamma_k\}$.

In 2012, He and Guo [16] introduced the following modified viscosity approximation method for a countable family of nonexpansive operators:

$$x_{k+1} = \gamma_k g(x_k) + (1 - \gamma_k) L_k x_k, \quad k \in \mathbb{N}, \tag{10}$$

where $\{\gamma_k\} \subset (0,1)$, $L_k = \sum_{i=1}^{k} \left(\frac{w_i}{s_k} \right) T_i$, $s_k = \sum_{i=1}^{k} w_i, w_i > 0$ with $\sum_{i=k}^{\infty} w_k = 1$. They proved strongly the convergence of (10) under the condition on $\{\gamma_k\}$ without any other condition on $\{T_k\}$. However, this algorithm needs larger computational work than that of (9). After that, several algorithms for the common fixed points of a countable family of nonexpansive operators were introduced and discussed; see [16–20].

Inspired by [10,12,15], in this paper, we propose a simple method with the inertial technique for solving a common fixed point problem of a countable family of nonexpansive operators in a real Hilbert space. We then prove a strong convergence of the proposed method under some suitable conditions. Finally, we apply our proposed method to solving the image restoration and classification problems.

The rest of this paper is organized as follows: In Section 2, we present some notation and useful lemmas that will be used in this paper. The strong convergence of the accelerated viscosity fixed point method and the accelerated viscosity forward–backward method are analyzed in Section 3. Applications and simulated results for image restoration and data classification problems are given in Section 4. Finally, we give a conclusion remark for further study in Section 5.

2. Preliminaries

In this section, we present some definitions and useful lemmas for proving our main results in the next section. Throughout this paper, we adopt the following notations:

- $\mathcal{H}$ denotes a real Hilbert space with norm $\| \cdot \|$ and inner product $\langle \cdot, \cdot \rangle$;
- C denotes a nonempty closed convex subset of $\mathcal{H}$;
- $Fix(T)$ denotes the set of all fixed points of T;
- $\rightharpoonup$ and $\rightarrow$ denote the weak convergence and strong convergence, respectively;
- $prox_{c\phi_2}(I - c\nabla\phi_1)$ denotes the forward–backward operator of ϕ_1 and ϕ_2 with respect to c.

A mapping $T : C \rightarrow C$ is said to be an L-Lipschitz operator if there exists $L > 0$ such that $\|Ta - Tb\| \leq L\|a - b\|$ for all $a, b \in C$. An L-Lipschitz operator is called a nonexpansive operator and contraction operator if $L = 1$ and $L \in (0,1)$, respectively. If $T : C \rightarrow C$ is a nonexpansive operator with $Fix(T) \neq \varnothing$, then $Fix(T)$ is closed and convex, and the mapping $I - T$ is demiclosed at zero, that is for any sequence $\{x_k\} \subset C$ such that $x_k \rightharpoonup a$ and $\|x_k - Tx_k\| \rightarrow 0$ imply $a \in Fix(T)$. A mapping P_C is said to be a metric projection of $\mathcal{H}$ onto C, if for every $a \in \mathcal{H}$, there exists a unique nearest point in C denoted by $P_C a$ such that:

$$\|a - P_C a\| \leq \|a - b\|, \quad \forall b \in C.$$

Moreover, P_C is firmly nonexpansive mapping and P_C satisfying $\langle a - P_C a, b - P_C a \rangle \leq 0$, $\forall a \in \mathcal{H}, b \in C$. Let $\{T_k\}$ and Λ be families of nonexpansive operators of C into itself such that $\varnothing \neq Fix(\Lambda) \subset \Gamma := \cap_{k=1}^{\infty} Fix(T_k)$, where $Fix(\Lambda)$ is the set of all common fixed points of Λ. A sequence $\{T_k\}$ is said to satisfy the NST-condition (I) with Λ [21] if for every bounded sequence $\{x_k\}$ in C,

$$\lim_{k \rightarrow \infty} \|x_k - T_k x_k\| = 0 \quad \text{implies} \quad \lim_{k \rightarrow \infty} \|x_k - Tx_k\| = 0 \quad \text{for all} \ \ T \in \Lambda.$$

If Λ is singleton, i.e., $\Lambda = \{T\}$, then $\{T_k\}$ is said to satisfy the NST-condition (I) with T. After that, Aoyama, Kohsaka and Takahashi [22] introduced the condition (Z), which is more general than that of NST-condition (I). A sequence $\{T_k\}$ is said to satisfy the condition

(Z) if whenever $\{x_k\}$ is a bounded sequence in C such that $\lim_{k \to \infty} \|x_k - T_k x_k\| = 0$, it follows that every weak cluster point of $\{x_k\}$ belongs to Γ.

It is also known that $prox_{c\phi_2}(I - c\nabla\phi_1)$ is a nonexpansive mapping when $0 < c < 2/L$. The following lemmas are useful for proving our main results.

Lemma 1 ([23]). *Let $\phi_1 : \mathcal{H} \to \mathbb{R}$ be a convex and differentiable function with an L-Lipschitz continuous gradient of ϕ_1 and let $\phi_2 : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ be a proper lower semi-continuous and convex function. Let $T_k := prox_{c_k\phi_2}(I - c_k\nabla\phi_1)$ and $T := prox_{c\phi_2}(I - c\nabla\phi_1)$, where $c_k, c \in (0, 2/L)$ with $c_k \to c$ as $k \to \infty$. Then, $\{T_k\}$ satisfies NST-condition (I) with T.*

Lemma 2 ([24]). *For all $a, b \in \mathcal{H}$, and $t \in [0, 1]$ the following hold:*

(i) $\|ta + (1-t)b\|^2 = t\|a\|^2 + (1-t)\|b\|^2 - t(1-t)\|a-b\|^2$;
(ii) $\|a \pm b\|^2 = \|a\|^2 \pm 2\langle a, b \rangle + \|b\|^2$;
(iii) $\|a + b\|^2 = \|a\|^2 + 2\langle b, a+b \rangle$.

Lemma 3 ([25]). *Let $\{a_i, i = 1, 2, \ldots, k\} \subset \mathcal{H}$. For $b_i \in (0, 1), i = 1, 2, \ldots, k$ such that $\sum_{i=1}^{k} b_i = 1$. Then, the following identity holds:*

$$\left\| \sum_{i=1}^{k} b_i a_i \right\|^2 = \sum_{i=1}^{k} b_i \|a_i\|^2 - \sum_{i,j=1, i \neq j}^{k} b_i b_j \|a_i - a_j\|^2.$$

Lemma 4 ([26]). *Let $\{a_k\}$ be a sequence of non-negative real numbers, $\{b_k\}$ be a sequence of real numbers, and $\{t_k\}$ be a sequence of real numbers in $(0, 1)$ such that $\sum_{n=1}^{\infty} t_k = \infty$. Assume that:*

$$a_{k+1} \leq (1 - t_k)a_k + t_k b_k, \quad k \in \mathbb{N}.$$

If $\limsup_{i \to \infty} b_{k_i} \leq 0$ for every subsequence $\{a_{k_i}\}$ of $\{a_k\}$ satisfying the condition:

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

then $\lim_{k \to \infty} a_k = 0$.

3. Main Results

In this section, we propose a new accelerated viscosity fixed point method, which is called "AVFPM" for solving a common fixed point of nonexpansive operators in a real Hilbert space. In order to introduce AVFPM, we assume the following:

- $g : \mathcal{H} \to \mathcal{H}$ is a contraction with constant $\eta \in (0, 1)$;
- $\{T_k : \mathcal{H} \to \mathcal{H}\}$ is a family of nonexpansive operators;
- $\{T_k\}$ satisfies condition (Z);
- $\Gamma := \cap_{k=1}^{\infty} Fix(T_k) \neq \emptyset$.

Theorem 5. *Let $\{x_k\}$ be a sequence generated by Algorithm 1 (AVFPM). Then, $\{x_k\}$ converges strongly to an element $a^* \in \Gamma$, where $a^* = P_\Gamma g(a^*)$.*

Now, we prove the strong convergence of Algorithm 1 (AVFPM).

Algorithm 1: An accelerated viscosity fixed point method (AVFPM).

Initialization: Take $x_0, x_1 \in \mathcal{H}$ arbitrarily and positive sequences $\{\lambda_k\}, \{\sigma_k\}, \{\gamma_k\}, \{\beta_k\}$, and $\{\alpha_k\}$ satisfy the following conditions:

$$\{\alpha_k\} \subset (0,1), \quad \lim_{k \to \infty} \alpha_k = 0 \text{ and } \sum_{k=1}^{\infty} \alpha_k = \infty,$$

$$\{\beta_k\} \subset (0,1), \quad 0 < a_5 \leq \gamma_k < 1, \quad \alpha_k + \beta_k + \gamma_k = 1,$$

$$0 < a_1 \leq \lambda_k \leq a_2 < 1, \quad 0 < a_3 \leq \sigma_k \leq a_4 < 1.$$

for some positive real numbers a_1, a_2, a_3, and a_4.

Iterative steps: Calculate x_{k+1} as follows:

Step 1. Choose a bounded sequence of non-negative real numbers $\{\mu_k\}$. For $k \geq 1$, set

$$\theta_k = \begin{cases} \min\left\{\mu_k, \dfrac{\tau_k}{\|x_k - x_{k-1}\|}\right\} & \text{if } x_k \neq x_{k-1}, \\ \mu_k & \text{otherwise,} \end{cases}$$

where $\{\tau_k\}$ is a sequence of positive real numbers such that $\lim_{k \to \infty} \tau_k / \alpha_k = 0$.

Step 2. Compute

$$\begin{cases} w_k = x_k + \theta_k(x_k - x_{k-1}), \\ z_k = (1 - \lambda_k)w_k + \lambda_k T_k w_k, \\ y_k = (1 - \sigma_k)T_k w_k + \sigma_k T_k z_k, \\ x_{k+1} = \alpha_k g(w_k) + \beta_k T_k z_k + \gamma_k T_k y_k. \end{cases}$$

Update $k := k + 1$ and return to Step 1.

Proof. By the Banach contraction principle, there exists a unique $a^* \in \Gamma$ such that $a^* = P_\Gamma g(a^*)$. By definitions of x_{k+1}, we have:

$$\|w_k - a^*\| \leq \|x_k - a^*\| + \theta_k \|x_k - x_{k-1}\|, \tag{11}$$

and:

$$\|z_k - a^*\| \leq (1 - \lambda_k)\|w_k - a^*\| + \lambda_k \|T_k w_k - a^*\| \leq \|w_k - a^*\|. \tag{12}$$

From (12), we get:

$$\|y_k - a^*\| \leq (1 - \sigma_k)\|T_k w_k - a^*\| + \sigma_k \|T_k z_k - a^*\| \leq \|w_k - a^*\|. \tag{13}$$

From (11)–(13), we obtain:

$$\begin{aligned} \|x_{k+1} - a^*\| &\leq \alpha_k \|g(w_k) - g(a^*)\| + \alpha_k \|g(a^*) - a^*\| \\ &\quad + \beta_k \|T_k z_k - a^*\| + \gamma_k \|T_k y_k - a^*\| \\ &\leq \alpha_k \eta \|w_k - a^*\| + \alpha_k \|g(a^*) - a^*\| \\ &\quad + \beta_k \|z_k - a^*\| + \gamma_k \|y_k - a^*\| \\ &\leq (1 - \alpha_k(1 - \eta))\|w_k - a^*\| + \alpha_k \|g(a^*) - a^*\| \\ &\leq (1 - \alpha_k(1 - \eta))\|x_k - a^*\| \\ &\quad + \alpha_k \left[\frac{\theta_k}{\alpha_k}\|x_k - x_{k-1}\| + \|g(a^*) - a^*\|\right]. \end{aligned}$$

By the condition of θ_k, we have $\lim_{k \to \infty} \dfrac{\theta_k}{\alpha_k}\|x_k - x_{k-1}\| = 0$, and so there exists a constant $M > 0$ such that $\dfrac{\theta_k}{\alpha_k}\|x_k - x_{k-1}\| \leq M \ \forall k \geq 1$. Thus:

$$\|x_{k+1} - a^*\| \le (1 - \alpha_k(1 - \eta))\|x_k - a^*\| + \alpha_k(M + \|g(a^*) - a^*\|).$$

By mathematical induction, we get:

$$\|x_{k+1} - a^*\| \le \max\left\{ \|x_1 - a^*\|, \frac{M + \|g(a^*) - a^*\|}{1 - \eta} \right\} \quad \forall k \ge 1.$$

This implies that $\{x_k\}$ is bounded and $\{w_k\}, \{z_k\}, \{y_k\}, \{T_k w_k\}, \{T_k z_k\}$, $\{T_k y_k\}, and \{g(w_k)\}$ are also bounded. By Lemma 2, we obtain:

$$\begin{aligned}
\|w_k - a^*\|^2 &\le \|x_k - a^*\|^2 + \theta_k^2 \|x_k - x_{k-1}\|^2 \\
&\quad + 2\theta_k \|x_k - a^*\| \|x_k - x_{k-1}\|,
\end{aligned} \tag{14}$$

and:

$$\|z_k - a^*\|^2 \le \|w_k - a^*\|^2 - \lambda_k(1 - \lambda_k)\|w_k - T_k w_k\|^2. \tag{15}$$

By Lemma 2(i) and (15), we obtain:

$$\begin{aligned}
\|y_k - a^*\|^2 &\le (1 - \sigma_k)\|T_k w_k - a^*\|^2 + \sigma_k \|T_k z_k - a^*\|^2 \\
&\quad - \sigma_k(1 - \sigma_k)\|T_k w_k - T_k z_k\|^2 \\
&\le \|w_k - a^*\|^2 - \sigma_k \lambda_k(1 - \lambda_k)\|w_k - T_k w_k\|^2 \\
&\quad - \sigma_k(1 - \sigma_k)\|T_k w_k - T_k z_k\|^2
\end{aligned} \tag{16}$$

From (12), (14), (16), Lemmas 2(iii) and 3, we have:

$$\begin{aligned}
\|x_{k+1} - a^*\|^2 &= \|\alpha_k(g(w_k) - g(a^*)) + \beta_k(T_k z_k - a^*) + \gamma_k(T_k y_k - a^*)\|^2 \\
&\quad + 2\alpha_k \langle g(a^*) - a^*, x_{k+1} - a^* \rangle \\
&\le \alpha_k \|g(w_k) - g(a^*)\|^2 + \beta_k \|T_k z_k - a^*\|^2 + \gamma_k \|T_k y_k - a^*\|^2 \\
&\quad + 2\alpha_k \langle g(a^*) - a^*, x_{k+1} - a^* \rangle \\
&\le \alpha_k \eta \|w_k - a^*\|^2 + \beta_k \|w_k - a^*\|^2 + \gamma_k \|w_k - a^*\|^2 \\
&\quad - \gamma_k \sigma_k \lambda_k(1 - \lambda_k)\|w_k - T_k w_k\|^2 \\
&\quad - \gamma_k \sigma_k(1 - \sigma_k)\|T_k w_k - T_k z_k\|^2 + 2\alpha_k \langle g(a^*) - a^*, x_{k+1} - a^* \rangle \\
&\le (1 - \alpha_k(1 - \eta))\|x_k - a^*\|^2 \\
&\quad + \alpha_k \left[2\|x_k - a^*\| \left(\frac{\theta_k}{\alpha_k} \|x_k - x_{k-1}\| \right) \right. \\
&\quad + \left(\frac{\theta_k}{\alpha_k} \|x_k - x_{k-1}\| \right) \theta_k \|x_k - x_{k-1}\| \\
&\quad \left. + 2\langle g(a^*) - a^*, x_{k+1} - a^* \rangle \right] - \gamma_k \sigma_k \lambda_k(1 - \lambda_k)\|w_k - T_k w_k\|^2 \\
&\quad - \gamma_k \sigma_k(1 - \sigma_k)\|T_k w_k - T_k z_k\|^2.
\end{aligned} \tag{17}$$

So, we get:

$$\gamma_k \sigma_k \lambda_k(1 - \lambda_k)\|w_k - T_k w_k\|^2 \le \|x_k - a^*\|^2 - \|x_{k+1} - a^*\|^2 + \alpha_k M', \tag{18}$$

and:

$$\gamma_k \sigma_k(1 - \sigma_k)\|T_k w_k - T_k z_k\|^2 \le \|x_k - a^*\|^2 - \|x_{k+1} - a^*\|^2 + \alpha_k M', \tag{19}$$

where:

$$M' = \sup_{n \geq 1} \left\{ 2\|x_k - a^*\| \left(\frac{\theta_k}{\alpha_k} \|x_k - x_{k-1}\| \right) + \left(\frac{\theta_k}{\alpha_k} \|x_k - x_{k-1}\| \right) \theta_k \|x_k - x_{k-1}\| \right.$$
$$\left. + 2\langle g(a^*) - a^*, x_{k+1} - a^* \rangle \right\}.$$

Now, we show that $\{x_k\}$ converges strongly to a^*. Let $a_k = \|x_k - a^*\|^2$. Suppose that $\{a_{k_i}\}$ is a subsequence of $\{a_k\}$ such that $\liminf_{i \to \infty}(a_{k_i+1} - a_{k_i}) \geq 0$. By (19) and conditions of $\{\lambda_k\}, \{\sigma_k\}, \{\gamma_k\}, \{\beta_k\}$, and $\{\alpha_k\}$, we have:

$$\begin{aligned}
\limsup_{i \to \infty} \gamma_{k_i} \sigma_{k_i} (1 - \sigma_{k_i}) \|T_{k_i} w_{k_i} - T_{k_i} z_{k_i}\|^2 &\leq \limsup_{i \to \infty} (a_{k_i} - a_{k_i+1} + \alpha_{k_i} M') \\
&\leq \limsup_{i \to \infty} (a_{k_i} - a_{k_i+1}) \\
&\quad + \limsup_{i \to \infty} \alpha_{k_i} M' \\
&= -\liminf_{i \to \infty} (a_{k_i+1} - a_{k_i}) \\
&\leq 0.
\end{aligned} \tag{20}$$

This implies that:

$$\lim_{i \to \infty} \|T_{k_i} w_{k_i} - T_{k_i} z_{k_i}\| = 0. \tag{21}$$

Similarly, we have $\lim_{i \to \infty} \|w_{k_i} - T_{k_i} w_{k_i}\| = 0$. By definitions of θ_k and x_{k+1}, we get:

$$\lim_{i \to \infty} \|w_{k_i} - x_{k_i}\| = 0. \tag{22}$$

So, we obtain:

$$\begin{aligned}
\|x_{k_i} - T_{k_i} x_{k_i}\| &\leq \|x_{k_i} - T_{k_i} w_{k_i}\| + \|T_{k_i} w_{k_i} - T_{k_i} x_{k_i}\| \\
&\leq 2\|w_{k_i} - x_{k_i}\| + \|w_{k_i} - T_{k_i} w_{k_i}\| \to 0.
\end{aligned} \tag{23}$$

From definitions of x_{k+1}, we have:

$$\begin{aligned}
\|z_{k_i} - x_{k_i}\| &\leq \|w_{k_i} - x_{k_i}\| + \lambda_{k_i} \|w_{k_i} - T_{k_i} w_{k_i}\|, \\
\|y_{k_i} - T_{k_i} w_{k_i}\| &= \sigma_{k_i} \|T_{k_i} w_{k_i} - T_{k_i} z_{k_i}\|,
\end{aligned} \tag{24}$$

and:

$$\|y_{k_i} - x_{k_i}\| \leq \|y_{k_i} - T_{k_i} w_{k_i}\| + \|T_{k_i} w_{k_i} - w_{k_i}\| + \|w_{k_i} - x_{k_i}\|. \tag{25}$$

This implies:

$$\lim_{i \to \infty} \|z_{k_i} - x_{k_i}\| = \lim_{i \to \infty} \|y_{k_i} - x_{k_i}\| = 0. \tag{26}$$

Moreover,

$$\begin{aligned}
\|x_{k_i+1} - x_{k_i}\| &\leq \|x_{k_i+1} - T_{k_i} x_{k_i}\| + \|T_{k_i} x_{k_i} - x_{k_i}\| \\
&\leq \alpha_{k_i} \|g(w_{k_i}) - T_{k_i} x_{k_i}\| + \beta_{k_i} \|z_{k_i} - x_{k_i}\| \\
&\quad + \gamma_{k_i} \|y_{k_i} - x_{k_i}\| + \|T_{k_i} x_{k_i} - x_{k_i}\|,
\end{aligned} \tag{27}$$

which implies $\lim_{i \to \infty} \|x_{k_i+1} - x_{k_i}\| = 0$. Now, we claim:

$$\limsup_{i \to \infty} \langle g(a^*) - a^*, x_{k_i+1} - a^* \rangle \leq 0.$$

Indeed, choose a subsequence $\{x_{k_{i_j}}\}$ of $\{x_{k_i}\}$ such that:

$$\limsup_{i \to \infty} \langle g(a^*) - a^*, x_{k_i} - a^* \rangle = \lim_{j \to \infty} \langle g(a^*) - a^*, x_{k_{i_j}} - a^* \rangle.$$

Since $\{x_{k_{i_j}}\}$ is bounded, there exists a subsequence $\{x_{k_{i_{j_p}}}\}$ of $\{x_{k_{i_j}}\}$ such that $x_{k_{i_{j_p}}} \rightharpoonup u \in H$. Without loss of generality, we may assume that $x_{k_{i_j}} \rightharpoonup u \in \mathcal{H}$. Since $\{T_k\}$ satisfies condition (Z), we have $u \in \Gamma$. As $\lim_{i \to \infty} \|x_{k_i+1} - x_{k_i}\| = 0$ and $a^* = P_\Gamma g(a^*)$, we obtain:

$$\limsup_{i \to \infty} \langle g(a^*) - a^*, x_{k_i+1} - a^* \rangle = \langle g(a^*) - a^*, u - a^* \rangle \leq 0. \tag{28}$$

By (17), (28), and $\lim_{k \to \infty} \frac{\theta_k}{\alpha_k} \|x_k - x_{k-1}\| = 0$, we can apply Lemma 4 to obtain $\lim_{k \to \infty} \|x_k - a^*\| = 0$; that is, $\{x_k\}$ converges strongly to $a^* = P_\Gamma g(a^*)$. This completes the proof. $\square$

Finally, we will apply the Algorithm 1 (AVFPM) for solving the nonsmooth convex optimization problems (3) of the sum of two objective functions ϕ_1 and ϕ_2 by assuming the following:

- $g : \mathcal{H} \to \mathcal{H}$ is a contraction with constant $\eta \in (0,1)$;
- $\phi_1 : \mathcal{H} \to \mathbb{R}$ is convex differentiable with Lipschitz continuous gradient constant $L > 0$;
- $\phi_2 : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ is a proper convex and lower semi-continuous function;
- $\Omega := Argmin(\phi_1 + \phi_2) \neq \emptyset$.

By setting $T_k = prox_{c_k \phi_2}(I - c_k \nabla \phi_1)$, which is the forward–backward operator of ϕ_1 and ϕ_2 with respect to $c_k \in (0, 2/L)$ and $c_k \to c$, we have an accelerated viscosity forward–backward method for solving the problems (3) as follows:

Next, we prove the strong convergence of Algorithm 2 (AVFBM) by using Theorem 5.

Algorithm 2: An accelerated viscosity forward–backward method (AVFBM).

Initialization: Take $x_0, x_1 \in \mathcal{H}$ arbitrarily and positive sequences $\{\lambda_k\}, \{\sigma_k\}, \{\gamma_k\}, \{\beta_k\}$, and $\{\alpha_k\}$ satisfy the following conditions:

$$\{\alpha_k\} \subset (0,1), \quad \lim_{k \to \infty} \alpha_k = 0 \text{ and } \sum_{k=1}^{\infty} \alpha_k = \infty,$$

$$\{\beta_k\} \subset (0,1), \quad 0 < a_5 \leq \gamma_k < 1, \quad \alpha_k + \beta_k + \gamma_k = 1,$$

$$0 < a_1 \leq \lambda_k \leq a_2 < 1, \quad 0 < a_3 \leq \sigma_k \leq a_4 < 1.$$

for some positive real numbers a_1, a_2, a_3, and a_4.
Iterative steps: Calculate x_{k+1} as follows:
Step 1. Choose a bounded sequence of non-negative real numbers $\{\mu_k\}$. For $k \geq 1$, defined θ_k by the same as Algorithm 1.
Step 2. Compute

$$\begin{cases} w_k = x_k + \theta_k(x_k - x_{k-1}), \\ z_k = (1 - \lambda_k)w_k + \lambda_k prox_{c_k \phi_2}(I - c_k \nabla \phi_1)w_k, \\ y_k = (1 - \sigma_k)prox_{c_k \phi_2}(I - c_k \nabla \phi_1)w_k + \sigma_k prox_{c_k \phi_2}(I - c_k \nabla \phi_1)z_k, \\ x_{k+1} = \alpha_k g(w_k) + \beta_k prox_{c_k \phi_2}(I - c_k \nabla \phi_1)z_k \\ \qquad\quad + \gamma_k prox_{c_k \phi_2}(I - c_k \nabla \phi_1)y_k. \end{cases}$$

Update $k := k + 1$ and return to Step 1.

Theorem 6. *Let $\{x_k\}$ be a sequence generated by Algorithm 2 (AVFBM). Then, $\{x_k\}$ converges strongly to an element $a^* \in \Omega$, where $a^* = P_\Omega g(a^*)$.*

Proof. Let $T := prox_{c\phi_2}(I - c\nabla\phi_1)$ and $T_k := prox_{c_k\phi_2}(I - c_k\nabla\phi_1)$. Then, T and $\{T_k\}$ are nonexpansive operators for all k, and $Fix(T) = \cap_{k=1}^{\infty} Fix(T_k) = Argmin(\phi_1 + \phi_2)$. By Lemma 1, we have $\{T_k\}$, which satisfies condition (Z). Therefore, we obtain the result directly by Theorem 5. $\square$

## 4. Application and Simulated Results

4.1. Image Restoration

In this example, we apply Algorithm 2 (AVFBM) to solving an image restoration problem (2) and compare the deblurring efficiency of AVFBM, FBS [10], and FISTA [12]. Our programs are written in MATLAB and run on a laptop with an Intel core i5, 4.00 GB RAM, and windows 8 (64-bit). All algorithms applied to the l_1-regularization problem (2); that is, $\phi_1(x) = \|Bx - b\|_2^2$ and $\phi_2(x) = \tau\|x\|_1$, where B is the blurring operator, b is the observed image, and τ is the regularization parameter. The maximum iteration number for all methods was fixed at 500.

In these experiments, we consider four gray-scale images (Cameraman, Lenna, Woman, and Boy) with size of 256×256 as the original images and consider Gaussian blur of filter size 9×9 with a standard deviation $\sigma = 4$ with noise 10^{-4}. We have measured the performance of AVFBM, FBS, and FISTA by means of the Signal-to-Noise Ratio (SNR) [27] and Peak Signal-to-Noise Ratio (PSNR) [28]. The SNR and PSNR at x_k of the restored images are defined as:

$$SNR(x, x_k) = 10\log_{10}\left\{\frac{\|x - \bar{x}\|_2}{\|x_k - x\|_2}\right\},$$

$$PSNR(x_k) = 10\log_{10}\left(\frac{255^2}{MSE}\right),$$

where $MSE = \frac{1}{256^2}\|x_k - x\|_2^2$, x is the original image, and $\bar{x}$ is the mean of the original image. The regularization parameter was chosen to be $\tau = 10^{-4}$, and the initial image was the blurred image. The Lipschitz constant L of the gradient ∇f is $L = 2\lambda_{max}(B^T B)$ [12]. The parameters of the algorithms are chosen as follows: $\lambda_k = \frac{0.5k}{k+1}$, $\sigma_k = \frac{0.99k}{k+1}$, $\alpha_k = \frac{1}{50k}$, $\beta_k = \frac{1}{300k+1}$, $\gamma_k = 1 - \alpha_k - \beta_k$, $c_k = \frac{k}{L(k+1)}$, $c = \frac{1}{L}$, $\tau_k = \frac{10^{15}}{k^2}$ and $\mu_k = \frac{t_k-1}{t_{k+1}}$, where t_k is a sequence defined by $t_1 = 1$ and $t_{k+1} = \frac{1+\sqrt{1+4t_k^2}}{2}$. The contraction mapping is defined by $g(a) = 0.95a$ for all $a \in \mathbb{R}^n$. The comparison of the performance of AVFBM, FISTA, and FBS by means of SNR and PSNR is shown in Figure 1. The plot of SNR and PSNR at x_k of the restored images is shown in Figure 2. We see from Figures 1 and 2 that AVFBM gives a higher performance of SNR and PSNR than the other methods. The comparison results for deblurring of the three methods of the four images are shown in Figure 3.

(a) **Signal-to-Noise Ratio (SNR)**

	Blurred Image	FBS	FISTA	AVFBM
Cameraman	3.7432	6.6511	9.9796	10.9632
Lenna	3.8312	6.7045	9.901	10.7501
Woman	4.5189	8.0734	11.5319	12.3696
Boy	3.3431	6.1435	9.4715	10.1616

(b) **Peak Signal-to-Noise Ratio (PSNR)**

	Blurred Image	FBS	FISTA	AVFBM
Cameraman	20.9709	26.7867	33.4438	35.411
Lenna	23.3723	29.1191	35.5119	37.2102
Woman	20.3108	27.4197	34.3367	36.0121
Boy	20.2437	25.8446	32.5005	33.8808
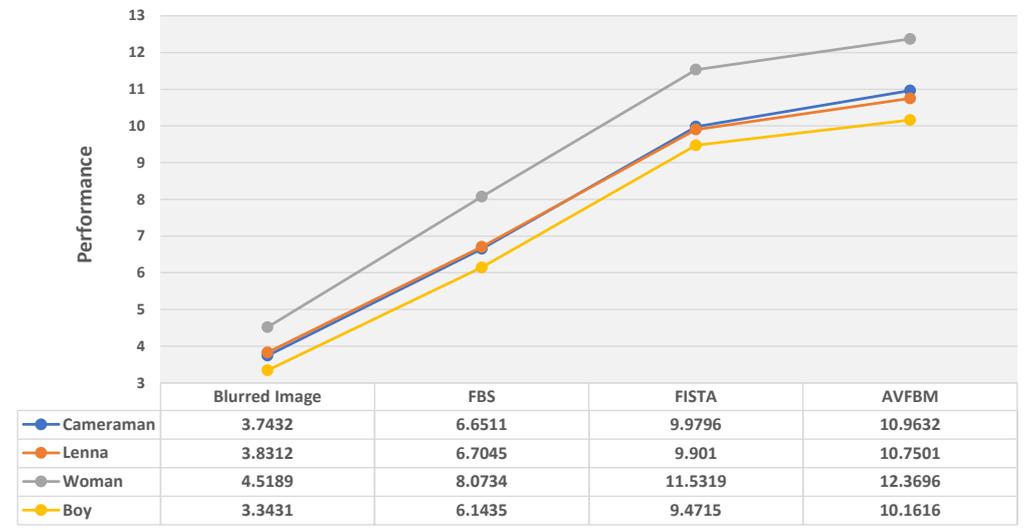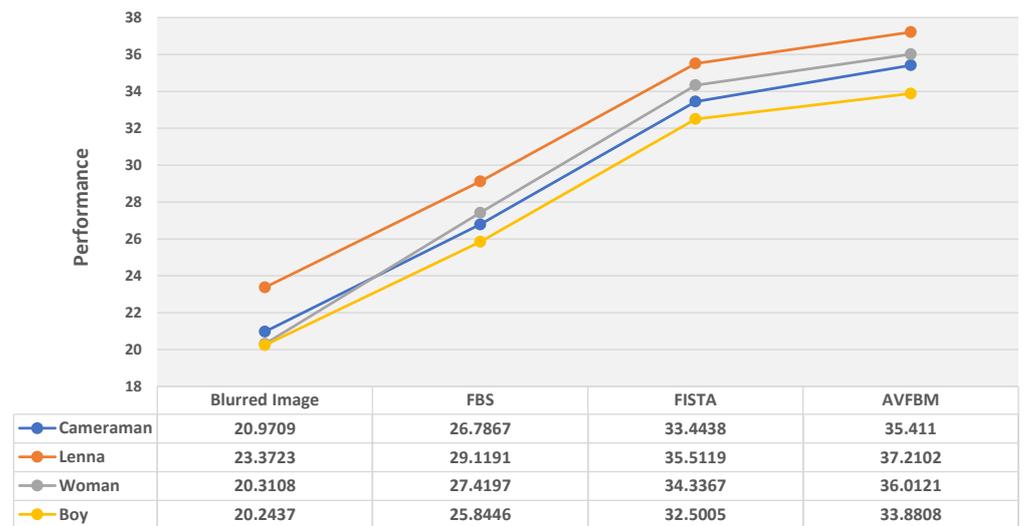
Figure 1. Comparison of SNR and PSNR by FBS, FISTA, AVFBM.

(a) Cameraman (b) Lenna (c) Woman (d) Boy

(e) Cameraman (f) Lenna (g) Woman (h) Boy
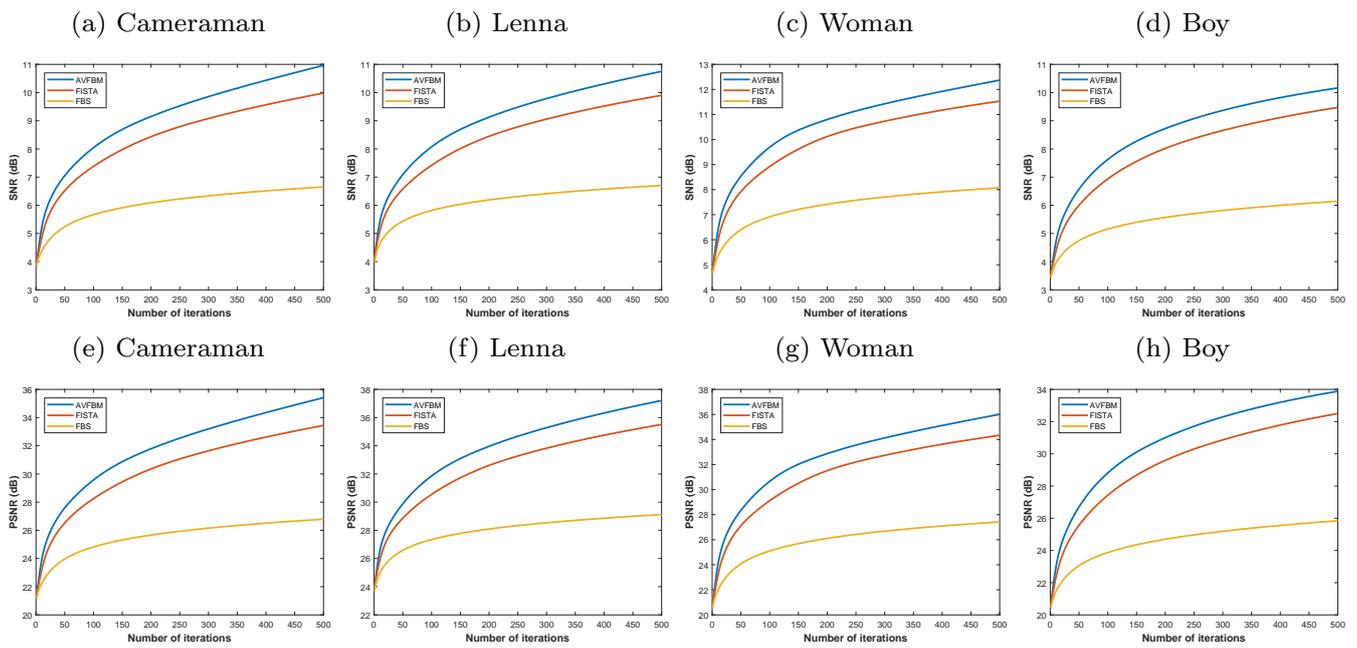
Figure 2. Plot of SNR and PSNR for the images.

(a) Original Image (b) Blurred Image (c) FBS (d) FISTA (e) AVFBM

(f) Original Image (g) Blurred Image (h) FBS (i) FISTA (j) AVFBM

(k) Original Image (l) Blurred Image (m) FBS (n) FISTA (o) AVFBM

(p) Original Image (q) Blurred Image (r) FBS (s) FISTA (t) AVFBM

Figure 3. Original images, blurred images, and Deblurring images by FBS, FISTA, AVFBM.

4.2. Data Classification

In this section, a learning algorithm named extreme learning machine (ELM) [29] will be investigated. ELM is a learning algorithm for single-hidden layer feedforward neural networks (SLFNs). Let $D = \{(\mathbf{x}_i, \mathbf{t}_i) : \mathbf{x}_i \in \mathbb{R}^n, \mathbf{t}_i \in \mathbb{R}^m, i = 1, 2, \ldots, N\}$ be a training dataset with N distinct training data $\mathbf{x}_i$ and label $\mathbf{t}_i$. For a given M nodes in the hidden layer, the SLFNs output for the jth pattern, $\mathbf{o}_j \in \mathbb{R}^m$, is given by:

$$\mathbf{o}_j = \sum_{i=1}^{M} \mathbf{w}_i f(\langle \mathbf{h}_i, \mathbf{x}_j \rangle + b_i), j = 1, 2, \ldots, N, \tag{29}$$

where f is the activation function, $\mathbf{h}_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ for $i = 1, 2, \ldots, M$ are the weight vector and bias connecting the input layer to the ith hidden node, respectively, and $\mathbf{w}_i \in \mathbb{R}^m$ for $i = 1, 2, \ldots, M$ is the weight vector connecting the ith hidden layer to the output layer. The target of SLFNs is to approximate the parameters $\mathbf{w}_i, \mathbf{h}_i, b_i$ for all $i = 1, 2, \ldots, M$ such that:

$$\mathbf{t}_j = \sum_{i=1}^{M} \mathbf{w}_i f(\langle \mathbf{h}_i, \mathbf{x}_j \rangle + b_i), j = 1, 2, \ldots, N, \tag{30}$$

which means that zero error $\sum_{i=1}^{N} \|\mathbf{o}_i - \mathbf{t}_i\|$ is close to 0 while ELM is used to find only parameter $\mathbf{w}_i$ with random $\mathbf{h}_i$ and b_i. As the above N equations, Equation (30) can be rewritten as:

$$\mathbf{Hw} = \mathbf{T} \tag{31}$$

where:

$$\mathbf{H} = \begin{bmatrix} f(\langle \mathbf{h}_1, \mathbf{x}_1 \rangle + b_1) & \cdots & f(\langle \mathbf{h}_M, \mathbf{x}_1 \rangle + b_M) \\ \vdots & \ddots & \vdots \\ f(\langle \mathbf{h}_1, \mathbf{x}_N \rangle + b_1) & \cdots & f(\langle \mathbf{h}_M, \mathbf{x}_N \rangle + b_M) \end{bmatrix}_{N \times M},$$

$\mathbf{w} = \left[\mathbf{w}_1^T, \ldots, \mathbf{w}_M^T\right]_{m \times M}^T$ and $\mathbf{T} = \left[\mathbf{t}_1^T, \ldots, \mathbf{t}_N^T\right]_{m \times N}^T$. From Equation (31), the ELM learning algorithm estimates the weight $\mathbf{w}$ by $\mathbf{w} = \mathbf{H}^\dagger \mathbf{T}$ where $\mathbf{H}^\dagger = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$ is the pseudo-inverse matrix of $\mathbf{H}$. Note that the linear system (31) can be represented by a least squared method. As shown in [29], ELM has an extremely fast training speed and good generalization performance. Nevertheless, its solutions also have some drawbacks [30]. To overcome these drawbacks, regularized extreme learning machine (RegELM) [30] replacing the least square method by the regularization method, i.e., ridge regression, for the training model was proposed, and the mathematical model of the RegELM algorithm can be described as:

$$\min_{\mathbf{w} \in \mathbb{R}^{M \times m}} \frac{1}{2} \|\mathbf{Hw} - \mathbf{T}\|_2^2 + \frac{\lambda}{2} \|\mathbf{w}\|_2^2, \tag{32}$$

where $\lambda > 0$ is called the regularization parameter. The RegELM's output weight can be calculated by $\mathbf{w} = (\lambda \mathbf{I} + \mathbf{H}^T \mathbf{H})^\dagger \mathbf{H}^T \mathbf{T}$, where $\mathbf{I}$ is the identity matrix. Although RegELM can be expected to provide better generalization ability than ELM and its running time is extremely fast similarly to ELM, we can define a greater generalization of RegELM as in [31] by replacing Equation (32) in a generalized way as follows:

$$\min_{\mathbf{w} \in \mathbb{R}^{M \times m}} \frac{1}{2} \|\mathbf{Hw} - \mathbf{T}\|_2^2 + \lambda \left[(1 - \alpha) \frac{1}{2} \|\mathbf{w}\|_2^2 + \alpha \|\mathbf{w}\|_1\right], \tag{33}$$

where $0 \leq \alpha \leq 1$. Equation (33), called elastic net, trades off between the ride regression ($\alpha = 0$) and the LASSO ($\alpha = 1$). In this paper, we present a new algorithm for RegELM and employ our results to data classification problems with benchmark datasets. For this case, we set $\alpha = 1$, and Problem (33) becomes a LASSO problem. From our result (Theorem 6) in Section 3, we can apply AVFBM (Algorithm 2) to solve the LASSO problem and define a learning algorithm for RegELM as follows:

RegELM-AVFBM: Given a training set $D = \{(\mathbf{x}_i, \mathbf{t}_i) : \mathbf{x}_i \in \mathbb{R}^n, \mathbf{t}_i \in \mathbb{R}^m, i = 1, 2, \ldots, N\}$, activation function f,

> Step 1: Select regularization parameter λ and hidden node number M.
> Step 2: Randomly $\mathbf{h}_i$ and $b_i, i = 1, \ldots, M$.
> Step 3: Calculate the hidden layer output matrix $\mathbf{H}$.
> Step 4: Obtain the output weight $\mathbf{w}$ by using AVFBM (Algorithm 2).

Several benchmark problems were chosen for experiments. All datasets were downloaded from https://archive.ics.uci.edu/ (accessed on 6 April 2020). The information of each dataset viz name of datasets, the number of attributes (number of input nodes), the number of classes (number of output nodes), and the number of (sample) data are summarized in Table 1. Each dataset was normalized to zero mean and unit variance; 70% of the data were sampled for training, and the remaining 30% were used for testing. For each method, we tested a different number of hidden nodes M in order to see which architecture provided the best results. The number of nodes in the hidden layer varied from 1 to 200 for the abalone dataset and from 1 to 100 for the other five datasets. For each method, we set the sigmoid function as the activation function f and the regularization parameter $\lambda = 1 \times 10^{-5}$ for regularized methods (RegELM and RegELM-AVFBM). However, for approximation methods (AVFBM, FISTA), we use relative error criteria, $\frac{\|x_k - x_{k-1}\|}{\|x_k\|} < \epsilon$, for the stopping algorithm and set all control sequences $(\lambda_k, \sigma_k, \alpha_k, \beta_k, \gamma_k, c_k, \tau_k, \mu_k)$ as in Section 4.1. For evaluating the performance of each method, an accuracy is defined as the total accuracy rate of classifying each case correctly. Accuracy is a value that represents the power of a model to correctly predict, and it is described as follows.

$$Accuracy = (TP + TN)/(TP + FP + TN + FN).$$

In experimental results, the accuracy of training and testing in percentage and the suitable number of hidden nodes of our method compared with direct methods viz standard ELM [29] and RegELM [30] are described in Table 2. RegELM-AVFBM has a good behavior in terms of accuracy of prediction and fit for testing datasets compared with the two direct methods. However, it is hard to compare the computational time, since approximation methods take time to iterate for convergence to the solution. Thus, to evaluate in the same way, we use two approximation methods (FISTA and AVFBM) for training RegELM and train the model with five different stopping errors ϵ under a maximum of 100,000 iterations. Table 3 shows the performance viz accuracy of training and testing (in percentage), computational time (in second), number of computed iterations, and number of suitable nodes in the hidden layer.

Table 1. Information of benchmark datasets.

Datasets	# Attributes	# Classes	# Observations	
			# Train (≈70%)	# Test (≈30%)
Zoo	16	7	70	31
Iris	4	3	105	45
Wine	13	3	128	50
Parkinsons	23	2	135	60
Heart Disease UCI	14	2	213	90
Abalone	8	3	2924	1253

Note that the cardinal number of set A is denoted by #A. For example, # Attributes is the number of attributes of the data.

Table 2. Comparison of the accuracy of training and testing as well as the number of hidden nodes for ELM, RegELM, and RegELM-AVFBM.

Datasets	ELM			RegELM			RegELM-AVFBM		
	Training (%)	Testing (%)	# Nodes	Training (%)	Testing (%)	# Nodes	Training (%)	Testing (%)	# Nodes
Zoo	97.1429	93.5484	13	97.1429	93.5484	13	100	96.7742	93
Iris	99.0476	100	60	98.0952	100	42	98.0952	100	54
Wine	98.4375	100	36	98.4375	100	36	100	100	40
Parkinsons	94.8148	75	31	94.8148	75	31	96.2963	81.6667	78
Heart Disease UCI	86.385	84.4444	25	86.385	84.4444	25	88.7324	85.5556	33
Abalone	69.0492	67.4381	89	69.0492	67.4381	89	68.4337	67.518	111

Table 3. Comparison of the accuracy of training and testing, computation time, number of iterations, and number of hidden nodes for RegELM-FISTA and RegELM-AVFBM. The sign ∞ in the column # Iters means that the model was computed over the maximum iterations (100,000 iterations for this case).

Datasets	ϵ	RegELM-FISTA					RegELM-AVFBM				
		Training (%)	Testing (%)	Time(s)	# Iters	# Nodes	Training (%)	Testing (%)	Time(s)	# Iters	# Nodes
Zoo	0.1	90	90.3226	0.0005994	11	33	98.5714	93.5484	0.0008407	9	82
	0.01	98.5714	93.5484	0.0021261	40	72	100	96.7742	0.0079208	81	93
	0.001	100	93.54839	0.0085405	455	42	98.57143	93.54839	0.005203	157	13
	0.0001	97.142857	93.548387	0.0152656	1330	13	97.14286	93.54839	0.0142225	384	13
	0.00001	97.142857	93.548387	0.0353528	2609	13	97.14286	93.54839	0.0193836	694	13
	0.000001	97.142857	93.548387	0.0561337	4193	13	97.142857	93.548387	0.0317928	1354	13
Iris	0.1	79.0476	91.1111	0.0006506	11	39	80	91.1111	0.0003463	7	20
	0.01	80	91.1111	0.0008544	27	9	96.19048	100	0.0033075	110	56
	0.001	92.38095	97.77778	0.0141314	658	22	98.09524	100	0.0217391	804	54
	0.0001	96.190476	100	0.051715	4232	38	98.095238	100	0.1359273	4891	53
	0.00001	98.0952381	100	0.6242778	41,584	56	98.0952381	100	0.9981745	47,695	42
	0.000001	-	-	-	∞	-	-	-	-	∞	-
Wine	0.1	97.6563	96	0.0005292	11	31	99.2188	98	0.0008743	8	65
	0.01	99.2188	98	0.0015979	30	64	100	100	0.0047111	98	40
	0.001	99.21875	100	0.0106758	364	45	98.4375	100	0.006298	271	36
	0.0001	98.4375	100	0.0536025	4374	36	98.4375	100	0.0234622	1146	36
	0.00001	98.4375	100	0.2150904	18,406	36	98.4375	100	0.0710794	3135	36
	0.000001	98.4375	100	0.4733094	39,108	36	98.4375	100	0.1426151	7342	36
Parkinsons	0.1	80.7407	75	0.0005362	11	5	80.7407	73.3333	0.0007843	4	5
	0.01	80.7407	75	0.000316	11	5	96.2963	81.66667	0.0034927	111	78
	0.001	96.2963	78.33333	0.0143303	649	83	95.55556	76.66667	0.0092722	252	31
	0.0001	98.518519	85	0.1072512	4702	95	100	76.666667	0.0401135	1533	60
	0.00001	99.2592593	78.3333333	0.3693551	31,488	60	95.555556	75	0.0395779	2266	31
	0.000001	95.5555556	75	0.3067227	32,185	31	94.814815	75	0.0887627	5421	31
Heart Disease UCI	0.1	82.6291	84.4444	0.000561	11	52	86.8545	85.5556	0.0006593	8	72
	0.01	84.9765	84.4444	0.0008177	31	57	85.9155	84.4444	0.0027626	73	25
	0.001	87.79343	86.66667	0.0115745	600	61	88.73239	85.55556	0.0051466	240	33
	0.0001	90.140845	85.555556	0.1300254	5231	58	86.38498	84.44444	0.0131317	644	25
	0.00001	86.384977	84.444444	0.0629385	6507	25	86.384977	84.444444	0.042063	1707	25
	0.000001	86.3849765	84.4444444	0.1245222	12,505	25	86.384977	84.444444	0.054982	3366	25
Abalone	0.1	57.2845	56.3448	0.0008332	11	9	57.0109	56.664	0.0007203	7	16
	0.01	59.13133	57.86113	0.0116477	47	147	66.72367	66.0016	0.0199067	111	96
	0.001	64.74008	64.08619	0.07772	445	111	68.63885	67.11891	0.2978755	817	175
	0.0001	66.792066	66.400638	0.8201147	5560	96	68.433653	67.517957	0.9515634	4480	111
	0.00001	68.5362517	67.1987231	11.9269803	51,900	149	68.7414501	67.6775738	3.4826104	21,877	89
	0.000001	-	-	-	∞	-	68.9466484	67.6775738	13.0566531	81,392	89

5. Conclusions

In this work, by using the inertial technique together with the viscosity approximation method, we propose a new accelerated algorithm for finding a common fixed point of a countable family of nonexpansive operates in a real Hilbert space. The strong convergence of the proposed method is established under some suitable conditions. As a special case, we obtain a new accelerated algorithm, called the accelerated viscosity forward–backward

method (AVFBM), for solving nonsmooth convex optimization problems. We also apply our algorithm, AVFBM, to solving image restoration and classification problems. By our experiments, for image restoration problem, they show that our algorithm, AVFBM, has a better performance for SNR and PSNR than that of FBS and FISTA, which are the most popular methods for solving such problems. Moreover, for the classification problems of six datasets—Zoo, Iris, Wine, Parkinsons, Heart Disease UCI, and Abalone (https://archive.ics.uci.edu/, accessed on 6 April 2020)—we use our algorithm, AVFBM, as a learning algorithm for finding the optimal output weight **w** in the mathematical model (32) of the classification problems. We compare the efficiency of our method with ELM, RegELM, and RegELM-FISTA by using the measurement of accuracy of training and testing. We found that our algorithm outperforms the other methods, as seen from Tables 2 and 3.

Author Contributions: Formal analysis, writing—original draft preparation, A.H.; methodology, writing—review and editing, software, L.B.; Conceptualization, revised the manuscript, S.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare that they have no competing interest.

References

1. Maurya, A.; Tiwari, R. A Novel Method of Image Restoration by using Different Types of Filtering Techniques. *Int. J. Eng. Sci. Innov. Technol.* **2014**, *3*, 124–129.
2. Suseela, G.; Basha, S.A.; Babu, K.P. Image Restoration Using Lucy Richardson Algorithm For X-Ray Images. *IJISET Int. J. Innov. Sci. Eng. Technol.* **2016**, *3*, 280–285.
3. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B Methodol.* **1996**, 58, 267–288.
4. Vogel, C. *Computational Methods for Inverse Problems*; SIAM: Philadelphia, PA, USA, 2002.
5. Sluder, G.; Wolf, D.E. *Digital Microscopy*, 3rd ed.; Elsevier: New York, NY, USA, 2007.
6. Suantai, S.; Kankam, K.; Cholamjiak, P. A Novel Forward-Backward Algorithm for Solving Convex Minimization Problem in Hilbert Spaces. *Mathematics* **2020**, *8*, 42. [CrossRef]
7. Bauschke, H.H.; Combettes, P.L. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*; Springer: New York, NY, USA, 2011.
8. Burachik, R.S.; Iusem, A.N. *Set-Valued Mappings and Enlargements of Monotone Operators*; Springer Science Business Media: New York, NY, USA, 2007.
9. Moreau, J.J. Proximité et dualité dans un espace hilbertien. *B. Soc. Math. Fr.* **1965**, *93*, 273–299.
10. Lions, P.L.; Mercier, B. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **1979**, *16*, 964–979.
11. Nesterov, Y.E. A method for solving the convex programming problem with convergence rate O($1/k^2$). *Sov. Math. Dokl.* **1983**, *27*, 372–376.
12. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202.
13. Liang, J.; Schonlieb, C.B. Improving fista: Faster, smarter and greedier. *arXiv* **2018**, arXiv:1811.01430.
14. Moudafi, A. Viscosity approximation method for fixed-points problems. *J. Math. Anal. Appl.* **2000**, 241, 46–55.
15. Takahashi, W. Viscosity approximation methods for countable families of nonexpansive mappings in Banach spaces. *Nonlinear Anal.* **2009**, *70*, 719–734.
16. He, S.; Guo, J. Iterative algorithm for common fixed points of infinite family of nonexpansive mappings in Banach spaces. *J. Appl. Math.* **2012**, *2012*, 787419.

17. Shimoji, K.; Takahashi, W. Strong convergence to common fixed points of infinite nonexpansive mappings and applications. *Taiwan. J. Math.* **2001**, *5*, 387–404.
18. Aoyama, K.; Kimura, Y.; Takahashi, W.; Toyoda, M. Finding common fixed points of a countable family of nonexpansive mappings in a Banach space. *Sci. Math. Jpn.* **2007**, *66*, 325–335.
19. Takahashi, W.; Takeuchi, Y.; Kubota, R. Strong convergence theorems by hybrid methods for families of nonexpansive mappings in Hilbert spaces. *J. Math. Anal. Appl.* **2008**, *341*, 276–286.
20. Takahashi, W.; Yao, J.-C. Strong convergence theorems by hybrid methods for countable families of nonlinear operators in Banach spaces. *J. Fixed Point Theory Appl.* **2012**, *11*, 333–353.
21. Nakajo, K.; Shimoji, K.; Takahashi, W. Strong convergence to common fixed points of families of nonexpansive mappings in Banach spaces. *J. Nonlinear Convex Anal.* **2007**, *8*, 11–34.
22. Aoyama, K.; Kohsaka, F.; Takahashi, W. Strong convergence theorems by shrinking and hybrid projection methods for relatively nonexpansive mappings in Banach spaces. In Proceedings of the 5th International Conference on Nonlinear Analysis and Convex Analysis, Hakodate, Japan, 26–31 August 2009; pp. 7–26.
23. Bussaban, L.; Suantai, S.; Kaewkhao, A. A parallel inertial S-iteration forward-backward algorithm for regression and classification problems. *Carpathian J. Math.* **2020**, *36*, 35–44.
24. Takahashi, W. *Introduction to Nonlinear and Convex Analysis*; Yokohama Publishers: Yokohama, Japan, 2009.
25. Chidume, C.E.; Ezeora, J.N. Krasnoselskii-type algorithm for family of multi-valued strictly pseudo-contractive mappings. *Fixed Point Theory Appl.* **2014**, *2014*, 111. [CrossRef]
26. Saejung, S.; Yotkaew, P. Approximation of zeros of inverse strongly monotone operators in Banach spaces. *Nonlinear Anal.* **2012**, *75*, 724–750.
27. Chen, D.Q.; Zhang, H.; Cheng, L.Z. A fast fixed point algorithm for total variation deblurring and segmentation. *J. Math. Imaging Vis.* **2012**, *43*, 167–179.
28. Thung, K.; Raveendran, P. A survey of image quality measures. In Proceedings of the 2009 International Conference for Technical Postgraduates (TECHPOS), Kuala Lumpur, Malaysia, 14–15 December 2009; pp. 1–4.
29. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine. *Neurocomputing* **2006**, *70*, 489–501.
30. Deng, W.; Zheng, Q.; Chen, L. Regularized extreme learning machine. In Proceedings of the 2009 IEEE Symposium on Computational Intelligence and Data Mining, Nashville, TN, USA, 30 March–2 April 2009; pp. 389–395.
31. Martínez-Martínez, J.M.; Escandell-Montero, P.; Soria-Olivas, E.; Martín-Guerrero, J.D.; Magdalena-Benedito, R.; Gómez-Sanchis, J. Regularized extreme learning machine for regression problems. *Neurocomputing* **2011**, *74*, 3716–3721.