Thai Journal of Mathematics
Volume 18 Number 4 (2020)
Pages 2001–2011

http://thaijmath.in.cmu.ac.th

ISSN 1686-0209

New Accelerated Fixed Point Algorithms with Applications to Regression and Classification Problems

Piti Thongsri[1] **and Suthep Suantai**[2,*]

[1] *Graduate Master's Degree Program in Mathematics Department of Mathematics, Faculty of Science Chiang Mai University, Chiang Mai 50200, Thailand*
e-mail : piti_thongsri@cmu.ac.th

[2] *Data Science Research Center, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand*
e-mail : suthep.s@cmu.ac.th

Abstract In this paper, we introduce and study a novel algorithm for finding a common fixed point of a countable family of nonexpansive mappings in Hilbert spaces. The main results obtained in this paper can be applied to solving regression and classification problems by using machine learning model.

MSC: 49K35; 47H10; 20M12

Keywords: Hilbert space; fixed point algorithm; forward-backward algorithm; regression and classification problems

Submission date: 09.07.2020 / Acceptance date: 23.07.2020

1. Introduction

Fixed point theory plays very important role in science, applied science, medical science, data science, engineering and economic. Moreover, it has different applications to many areas of science, such as existence of nonlinear equations, optimization problems [1], equilibrium problems [2], image restoration [3], regression and data classification problems [4], variational inequality problems [5] and split feasibility problems [6]. Many real life problems can be converted into equations in the form of fixed point problems meaning that we have to find a fixed point of some operators. Recently, many fixed point algorithms have been proposed and studied to solving various kinds of real world problems, such as Picard iteration, Mann iteration [7], Halpern iteration [8], etc.

In this work, we are dealing with the convex minimization problem which can be formulated as

$$\min_{x\in\mathbb{R}^n}\{F(x)=f(x)+g(x)\}, \tag{1.1}$$

*Corresponding author.

Published by The Mathematical Association of Thailand.

where $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is proper convex and lower semi-continuous function, and $f : \mathbb{R}^n \to \mathbb{R}$ is convex differentiable with gradient ∇f being L-Lipschitz constant for some $L > 0$. The set of minimizers of F is denoted by $Argmin(F)$. The classical Forward-Backward splitting (FBS) algorithm [9] for problem (1.1) is given by the following iterative formula:

$$x_{n+1} = prox_{c_n g}(I - c_n \nabla f)(x_n), \ c_n \in (0, 2/L), \tag{1.2}$$

where c_n is the step-size, I is an identity operator and $prox_g$ is the proximity operator of g defined by

$$prox_g(x) = \arg\min_y \left\{ g(y) + \frac{\|x-y\|^2}{2} \right\}. \tag{1.3}$$

This method (1.2) has been widely used due to its simplicity, as a result it has been improved by many works such as the method that has improved the convergence rate of (1.2) signicantly. It is a method known as the fast iterative shrinkage-threshodling algorithm or FISTA. It was proposed by Beck and Teboulle [10] as follows:

$$\begin{aligned} & x_1 = y_0 \in \mathbb{R}^n, t_1 = 1 \\ & y_n = Tx_n \\ & t_{n+1} = \frac{1+\sqrt{1+4t_n^2}}{2}, \ \theta_n = \frac{t_n - 1}{t_{n+1}}, \\ & x_{n+1} = y_n + \theta_n(y_n - y_{n-1}). \end{aligned} \tag{1.4}$$

Very recently, Laing and Schonlieb [11] modified FISTA by replacing $t_{n+1} = \frac{p+\sqrt{q+rt_n^2}}{2}$ where $p, q > 0$ and $0 < r \leq 4$, and proved weak convergence theorem of FISTA.

Later, the new accelerated proximal gradient algorithm (NAGA) was proposed by Verma and Shukla in [12] as follows:

$$\begin{aligned} & x_0, x_1 \in C \\ & y_n = x_n + \theta_n(x_n - x_{n-1}), \\ & x_{n+1} = T_n[(1-\alpha_n)y_n + \alpha_n T_n y_n], \end{aligned} \tag{1.5}$$

where T_n is the forward-backward operator of f and g with respect to $c_n \in (0, 2/L)$. They proved a convergence theorem of NAGA and applied this method for solving the non-smooth convex minimization problem.

Motivated by those works mentioned above, in this paper, a new accelerated fixed point algorithms for sloving (1.1) is proposed by employing the concepts of Aoyama et al. iteration process together with the inertial step for a countable family of nonexpansive mappings. We also prove the convergence of our algorithm under some conditions and apply it to solving regression and classification problems. The organization of this paper is as follows. In Section 2, we describes some notation and useful Lemmas for the later section. In Section 3, we introduce our proposed algorithm for common fixed point problem, give the theoretical proofs of its convergence under particular condition. In Section 4, we apply our algorithm to solving regression and classification problems.

2. Preliminaries

Let H be a real Hilbert space with norm $\|\cdot\|$ and inner product $\langle\cdot,\cdot\rangle$, and C a nonempty closed convex subset of H. A mapping $T : C \to C$ is said to be *L-Lipschitz operator* if there exists $L > 0$ such that $\|Tx - Ty\| \le \|x - y\|$ $\forall x, y \in C$. An L-Lipschitz operator is called *nonexpansive operator* if $L = 1$. A point $x \in C$ is a fixed point of T if $Tx = x$. Let $F(T) := \{x \in C : Tx = x\}$, the fixed point set of T. Let $\{T_n\}$ and Ω be families of nonexpansive operator of C into itself such that $\emptyset \ne F(\Omega) \subset \Gamma := \bigcap_{n=1}^{\infty} F(T_n)$, where $F(\Omega)$ is the set of all common fixed points of Ω, and let $\omega_w(x_n)$ denote the set of all weak-cluster point of a bounded sequence $\{x_n\}$ in C. A sequence T_n is said to satisfy the *NST-condition(I)* with Ω [13], if for every bounded sequence $\{x_n\}$ in C,

$$\lim_{n\to\infty} \|x_n - T_n x_n\| = 0 \text{ implies } \lim_{n\to\infty} \|x_n - Tx_n\| = 0 \text{ for all } T \in \Omega.$$

If Ω is singleton, i.e., $\Omega = T$, then $\{T_n\}$ is siad to satisfy the *NST-condition(I)* with T. After that, Nakajo et al. [14] introduced the NST*-condition which is more general than that of NST-condition. A sequence $\{T_n\}$ is siad to satisfy the *NST*-condition* if for every bounded sequence $\{x_n\}$ in C,

$$\lim_{n\to\infty} \|x_n - T_n x_n\| = \lim_{n\to\infty} \|x_n - x_{n+1}\| = 0 \text{ implies } \omega_w(x_n) \subset \Gamma.$$

It follows directly from above definition that if $\{T_n\}$ satisfies the NST-condition(I), then $\{T_n\}$ satisfies the NST*-condition. Observe that if $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is a proper convex and lower semi-continuous function, then for all $x \in \mathbb{R}^n$ the $prox_g(x)$ exists and unique [15]. The solution of (1.1) can be characterized by Theorem 16.3 of Bauschke and combettes [16] as follows:

$$x^* \text{ is a minimizer of } (f + g) \text{ if and only if } 0 \in \partial g(x^*) + \nabla f(x^*)$$

where ∂g is subdifferential of g and ∇f is the gradient of f. The subdifferential of g at x^*, denoted by $\partial g(x^*)$, is defined by

$$\partial g(x^*) := \{u : g(x) \ge \langle u, x - x^*\rangle + g(x^*) \text{ for all } x\},$$

It is well-known that the subdifferential operator ∂g is maximal monotone, see [17] for more details. For solving (1.1) is characterized by the following fixed point problem:

$$\begin{aligned} x^* \text{ is a minimizer of } (f + g) \text{ if and only if } x^* &= prox_{cg}(I - c\nabla f)(x^*) \\ &= J_{c\partial g}(I - c\nabla f)(x^*), \end{aligned}$$

where $c > 0$ and $J_{\partial g}$ is resolvent of ∂g defined by $J_{\partial g} = (I + \partial g)^{-1}$. It is also known that $prox_{cg}(I - c\nabla f)$ is a nonexpansive mapping when $c \in (0, 2/L)$. The operator $prox_{cg}(I - c\nabla f)$ is called the forward-backward operator of f and g with respect to c. We end this part with the following Lemmas which will be used to prove our main results.

Lemma 2.1. [4] *For a real Hilbert speace H, let $g : H \to \mathbb{R} \cup \{\infty\}$ be proper convex and lower semi-continuous function, and $f : H \to \mathbb{R}$ be convex differentiable with gradient ∇f being L-Lipschitz constant for some $L > 0$. If $\{T_n\}$ is the forward-backward operator of f and h with respect to $c_n \in (0, 2/L)$ such that c_n converges to c, then $\{T_n\}$ satisfies NST-condition(I) with T, where T is the forward-backward operator of f and h with respect to $c \in (0, 2/L)$.*

Lemma 2.2. [18] *Let H be a real Hilbert space. Then the following results hold:*

(i) for all $t \in [0,1]$ *and* $x, y \in H$,

$$\|tx + (1-t)y\|^2 = t\|x\|^2 + (1-t)\|y\|^2 - t(1-t)\|x-y\|^2;$$

(ii) $\|x \pm y\|^2 = \|x\|^2 \pm 2\langle x, y\rangle + \|y\|^2 \ \forall x, y \in H.$

Lemma 2.3. [19] *Let* $\{a_n\}, \{b_n\}$ *and* $\{\gamma_n\}$ *be sequences of nonnegative real numbers such that*

$$a_{n+1} \le (1+\gamma_n)a_n + b_n, \ n \in \mathbb{N}.$$

If $\sum_{n=1}^{\infty} \gamma_n < \infty$ *and* $\sum_{n=1}^{\infty} b_n < \infty$, *then* $\lim_{n\to\infty} a_n$ *exists.*

Lemma 2.4. [20] *Let* H *be a Hilbert space and* $\{x_n\}$ *be a sequence in* H *such that there exists a nonempty set* $\Gamma \subset H$ *satisfying*

(i) for every $p \in \Gamma$, $\lim_{n\to\infty} \|x_n - p\|$ *exists;*
(ii) each weak-cluster point of the sequence $\{x_n\}$ *is in* Γ.

Then there exists $x^* \in \Gamma$ *such that* $\{x_n\}$ *weakly converges to* x^*.

Lemma 2.5. [3] *Let* $\{a_n\}$ *and* $\{\theta_n\}$ *be sequences of nonnegative real numbers such that*

$$a_{n+1} \le (1+\theta_n)a_n + \theta_n a_{n-1}, \ n \in \mathbb{N}.$$

Then the following holds

$$a_{n+1} \le K \cdot \prod_{j=1}^{n}(1+2\theta_j), \ \text{where } K = \max\{a_1, a_2\}.$$

Moreover, if $\sum_{n=1}^{\infty} \theta_n < \infty$, *then* $\{a_n\}$ *is bounded.*

3. Main Results

In this section, we propose a new accelerated fixed point algorithms for finding a common fixed point of a countable family of nonexpansive operators in a real Hilbert space. We now ready to introduce this algorithm by the theorem as follows:

Theorem 3.1. *Let* H *be a real Hilbert space and* $\{T_n : H \to H\}$ *be a family of nonexpansive operators. Suppose* $\{T_n\}$ *satisfies* NST^**-condition and* $\Gamma := \bigcap_{n=1}^{\infty} F(T_n) \neq \emptyset$. *Let* $\{x_n\}$ *be a sequence generated by* ***Algorithm 1*** *as follows:*

$$\begin{aligned}
&x_0, x_1 \in H\\
&w_n = x_n + \theta_n(x_n - x_{n-1})\\
&z_n = (1-\gamma_n)w_n + \gamma_n T_n w_n\\
&x_{n+1} = \alpha_n x_n + \beta_n T_n w_n + (1-\alpha_n-\beta_n)T_n z_n
\end{aligned} \tag{3.1}$$

where $\gamma_n \in [a_1, b_1] \subset (0,1)$, $\beta_n \in [0,1]$, $\alpha_n \in [a_1, b_1) \subset [0,1)$, $\alpha_n + \beta_n \in [a_1, b_1] \subset (0,1)$, $\theta_n \ge 0$ *and* $\sum_{n=1}^{\infty} \theta_n < \infty$. *Then the following hold:*

(i) $\|x_{n+1} - x^*\| \le K \cdot \prod_{j=1}^{n}(1+2\theta_j)$, *where* $K = \max\{\|x_1 - x^*\|, \|x_2 - x^*\|\}$ *and* $x^* \in \bigcap_{n=1}^{\infty} F(T_n)$,
(ii) $x_n \rightharpoonup x^* \in \bigcap_{n=1}^{\infty} F(T_n)$.

Proof. (i) Let $x^* \in \Gamma$. By (3.1) and nonexpansiveness of T_n, we have

$$\begin{aligned}\|w_n - x^*\| &= \|x_n + \theta_n(x_n - x_{n-1}) - x^*\| \\ &\le \|x_n - x^*\| + \theta_n\|x_n - x_{n-1}\|, \qquad (3.2)\\ \|z_n - x^*\| &= \|(1-\gamma_n)w_n + \gamma_n T_n w_n - x^*\| \\ &= \|(1-\gamma_n)(w_n - x^*) + \gamma_n(T_n w_n - x^*)\| \\ &\le (1-\gamma_n)\|w_n - x^*\| + \gamma_n\|T_n w_n - x^*\| \\ &\le \|w_n - x^*\|, \qquad (3.3)\end{aligned}$$

and

$$\begin{aligned}\|x_{n+1} - x^*\| &= \|\alpha_n x_n + \beta_n T_n w_n + (1-\alpha_n-\beta_n)T_n z_n - x^*\| \\ &= \|\alpha_n(x_n - x^*) + \beta_n(T_n w_n - x^*) + (1-\alpha_n-\beta_n)(T_n z_n - x^*)\| \\ &\le \alpha_n\|x_n - x^*\| + \beta_n\|T_n w_n - x^*\| + (1-\alpha_n-\beta_n)\|T_n z_n - x^*\| \\ &\le \alpha_n\|x_n - x^*\| + \beta_n\|w_n - x^*\| + (1-\alpha_n-\beta_n)\|z_n - x^*\| \\ &\le \alpha_n\|x_n - x^*\| + \beta_n(\|x_n - x^*\| + \theta_n\|x_n - x_{n-1}\|) \\ &\quad + (1-\alpha_n-\beta_n)(\|x_n - x^*\| + \theta_n\|x_n - x_{n-1}\|) \\ &= \|x_n - x^*\| + (1-\alpha_n)\theta_n\|x_n - x_{n-1}\|. \qquad (3.4)\end{aligned}$$

This implies

$$\begin{aligned}\|x_{n+1} - x^*\| &\le \|x_n - x^*\| + \theta_n\|x_n - x_{n-1} + x^* - x^*\| \qquad (3.5)\\ &\le (1+\theta_n)\|x_n - x^*\| + \theta_n\|x_{n-1} - x^*\|\end{aligned}$$

Apply Lemma 2.5, we get $\|x_{n+1} - x^*\| \le K \cdot \prod_{j=1}^{n}(1+2\theta_j)$, where $K = \max\{\|x_1 - x^*\|, \|x_2 - x^*\|\}$. Since $\sum_{n=1}^{\infty}\theta_n < \infty$, it follows that $\{x_n\}$ is bounded. This implies $\sum_{n=1}^{\infty}\theta_n\|x_n - x_{n-1}\| < \infty$.

(ii) By (3.4) and Lemma 2.3, we obtain $\lim_{n\to\infty}\|x_n - x^*\|$ exists. By Lemma 2.2(ii), we obtain

$$\begin{aligned}\|w_n - x^*\|^2 &\le (\|x_n - x^*\| + \theta_n\|x_n - x_{n-1}\|)^2 \\ &= \|x_n - x^*\|^2 + \theta_n^2\|x_n - x_{n-1}\|^2 + 2\theta_n\|x_n - x^*\|\|x_n - x_{n-1}\|. \qquad (3.6)\end{aligned}$$

By Lemma 2.2(i), we have

$$\begin{aligned}\|z_n - x^*\|^2 &= \|(1-\gamma_n)(w_n - x^*) + \gamma_n(T_n w_n - x^*)\|^2 \\ &= (1-\gamma_n)\|w_n - x^*\|^2 + \gamma_n\|T_n w_n - x^*\|^2 - \gamma_n(1-\gamma_n)\|w_n - T_n w_n\|^2 \\ &\le \|w_n - x^*\|^2 - \gamma_n(1-\gamma_n)\|w_n - T_n w_n\|^2. \qquad (3.7)\end{aligned}$$

Using Lemma 2.2(i) together with (3.6) and (3.7), we get

$$\begin{aligned}
\|x_{n+1}-x^*\|^2 &= \|\alpha_n x_n + \beta_n T_n w_n + (1-\alpha_n-\beta_n)T_n z_n - x^*\|^2 \\
&= \|\alpha_n(x_n-x^*) + \beta_n(T_n w_n - x^*) + (1-\alpha_n-\beta_n)(T_n z_n - x^*)\|^2 \\
&\leq \alpha_n\|x_n-x^*\|^2 + \beta_n\|T_n w_n - x^*\|^2 + (1-\alpha_n-\beta_n)\|T_n z_n - x^*\|^2 \\
&\quad - \alpha_n\beta_n\|T_n w_n - x_n\|^2 \\
&\leq \alpha_n\|x_n-x^*\|^2 + \beta_n\|w_n-x^*\|^2 + (1-\alpha_n-\beta_n)\|z_n-x^*\|^2 \\
&\leq \alpha_n\|x_n-x^*\|^2 + \beta_n\|x_n-x^*\|^2 + \beta_n\theta_n^2\|x_n-x_{n-1}\|^2 \\
&\quad + 2\theta_n\beta_n\|x_n-x^*\|\|x_n-x_{n-1}\| + (1-\alpha_n-\beta_n)(\|x_n-x^*\|^2 \\
&\quad + \theta_n^2\|x_n-x_{n-1}\|^2 + 2\theta_n\|x_n-x^*\|\|x_n-x_{n-1}\| \\
&\quad - \gamma_n(1-\gamma_n)\|w_n-T_n w_n\|^2).
\end{aligned}$$

Thus

$$\begin{aligned}
(1-\alpha_n-\beta_n)\gamma_n(1-\gamma_n)\|w_n-T_n w_n\|^2 &\leq \|x_n-x^*\|^2 + (1-\alpha_n)\theta_n^2\|x_n-x_{n-1}\|^2 \\
&\quad + (1-\alpha_n)2\theta_n\|x_n-x^*\|\|x_n-x_{n-1}\| \\
&\quad - \|x_{n+1}-x^*\|^2. \qquad (3.8)
\end{aligned}$$

Since $\sum_{n=1}^{\infty}\theta_n\|x_n-x_{n-1}\| < \infty$ and $\lim_{n\to\infty}\|x_n-x^*\|$ exists, it follows that $\lim_{n\to\infty}\|w_n-T_n w_n\| = 0$. Note that

$$\begin{aligned}
\|x_n-T_n x_n\| &\leq \|x_n-w_n\| + \|w_n-T_n w_n\| + \|T_n w_n - T_n x_n\| \\
&\leq 2\|x_n-w_n\| + \|w_n-T_n w_n\|, \qquad (3.9)
\end{aligned}$$

and

$$\begin{aligned}
\|z_n-w_n\| &= \|(1-\gamma_n)w_n + \gamma_n T_n w_n - w_n\| \\
&= \gamma_n\|T_n w_n - w_n\|. \qquad (3.10)
\end{aligned}$$

These imply $\lim_{n\to\infty}\|x_n-T_n x_n\| = 0$ and $\lim_{n\to\infty}\|z_n-w_n\| = 0$. By (3.1) and nonexpansiveness of T_n, we have

$$\begin{aligned}
\|x_{n+1}-x_n\| &= \|\alpha_n x_n + \beta_n T_n w_n + (1-\alpha_n-\beta_n)T_n z_n - x_n\| \\
&= \|\beta_n(T_n w_n - T_n z_n) + (1-\alpha_n)(T_n z_n - x_n)\| \\
&\leq \beta_n\|T_n w_n - T_n z_n\| + (1-\alpha_n)\|T_n z_n - x_n\| \\
&\leq \beta_n\|w_n-z_n\| + (1-\alpha_n)(\|T_n z_n - T_n x_n\| + \|T_n x_n - x_n\|) \\
&\leq \beta_n\|w_n-z_n\| + (1-\alpha_n)\|z_n-x_n\| + (1-\alpha_n)\|T_n x_n - x_n\| \\
&\leq \beta_n\|w_n-z_n\| + (1-\alpha_n)\|z_n-x_n\| + (1-\alpha_n)\|w_n-x_n\| \\
&\quad + (1-\alpha_n)\|T_n x_n - x_n\|,
\end{aligned} \qquad (3.11)$$

$$\|w_n-x_n\| = \theta_n\|x_n-x_{n+1}\| \to 0, \text{ and } \|w_n-z_n\| = \gamma_n\|T_n w_n - w_n\| \to 0.$$

These imply $\lim_{n\to\infty}\|x_n-x_{n+1}\| = 0$. Since $\{T_n\}$ satisfies NST*-condition, we get $\omega_w(x_n) \subset \Gamma := \bigcap_{n=1}^{\infty} F(T_n)$. Therefore, by Lemma 2.4, we obtain that $\{x_n\}$ converges weakly to a point $x^* \in \Gamma$. This completes the proof. ∎

Finally, we apply our proposed algorithm, for solving the minimization problem (1.1) by setting $T_n = prox_{c_n g}(I - c_n \nabla f)$, the forward-backward operator of f and g with respect to c_n, where $g : H \to \mathbb{R} \cup \{\infty\}$ is proper covex and lower semi-continuous function and $f : H \to \mathbb{R}$ is convex differentiable with gradient ∇f being L-Lipschitz constant for some $L > 0$ with $Argmin(f + g) \neq \emptyset$. The following result is directly obtained by Theorem 3.1.

4. Simulated Results for Regression and Classification Problems

In this section, we predict a sine function and classify datasets by our proposed learning algorithm. All results are performed on Intel Core-i5 gen 8th with 8.00 GB RAM, windows 10, under MATLAB computing environment.

Extreme learning machine(ELM). Let $D = \{(x_i, t_i) : x_i \in \mathbb{R}^n, t_i \in \mathbb{R}^m, i = 1, 2, ..., N\}$ be a training set with N distinct samples, x_i and t_i are called input data and target, respectively. A standard SLFNs with M hidden nodes and activation function $\Phi(x)$, e.g. sigmoid, are mathematically modeled as

$$\sum_{j=1}^{M} \beta_j \Phi(\langle w_j, x_i \rangle + b_j) = o_i,\ i = 1, ..., N,$$

where w_j is the weight vector connecting the jth hidden node and the input node, β_j is the weight vector connecting the jth hidden node and the output node, and b_j is the threshold of the jth hidden node. The target of standard SLFNs is to approximate these N samples with zero error means that $\sum_{i=1}^{N} \|o_i - t_i\| = 0$, i.e., there exist β_j, w_j, b_j such that

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

From above N equations, we can formulate a simple equation as

$$H\beta = T,$$

where

$$H = \begin{bmatrix} \Phi(\langle w_1, x_1 \rangle + b_1) & \cdots & \Phi(\langle w_M, x_1 \rangle + b_M) \\ \vdots & \ddots & \vdots \\ \Phi(\langle w_1, x_N \rangle + b_1) & \cdots & \Phi(\langle w_M, x_N \rangle + b_M) \end{bmatrix}_{N \times M}$$

$$\beta = [\beta_1^T, ..., \beta_M^T]^T_{m \times M},\ T = [t_1^T, ..., t_N^T]^T_{m \times N}.$$

The goal of a standard SLFNs is estimate β_j, w_j and b_j for solving (1.1) while ELM aim to find only β_j with randomly w_j and b_j.

We conduct some experiments on regression and classication problems, the problem is formulated as the following convex minimization problem:

$$\min_{\beta} \|H\beta - T\|_2^2 + \lambda \|\beta\|_1,$$

where $\|\cdot\|_1$ is l_1-norm defined by $\|x\|_1 = \sum_{i=1}^{n} |x_i|$ and $\lambda > 0$ is called regularization parameter. This problem is called the least absolute shrinkage and selection operator (LASSO) [21].

4.1. Regression for a Sine Function

In order to regress a sine function, we create a training set by randomly 10 distinct data,our activation function is sigmoid, number of hidden nodes $M = 100$, and regularization parameter $\lambda = 1 \times 10^{-5}$. In Algorithm 1, we set $T_n = prox_{c_n g}(I - c_n \nabla f)$, $\alpha_n = 0.1$, $\beta_n = 0.1$, $\gamma_n = 0.9$ and

$$\theta_n = \begin{cases} \frac{1}{2^n} & ,\text{if } n > 1700, \\ 0.9 & ,\text{if } n \leq 1700. \end{cases}$$

We then get results compared to FISTA and NAGA as in Figure 1 and Table 1.

Method	MSE	Computational time
Algorithm 1	2.016534×10^{-3}	4.23423×10^{-2}
FISTA	1.982393×10^{-1}	2.11288×10^{-2}
NAGA	9.596212×10^{-2}	2.49963×10^{-2}

Table 1. Numerical results of regression of a function sine.

Table1 and Figure1 show that Algorithm 1 gives a better performance to predict a sine function than FISTA and NAGA while a computational time have a few difference.

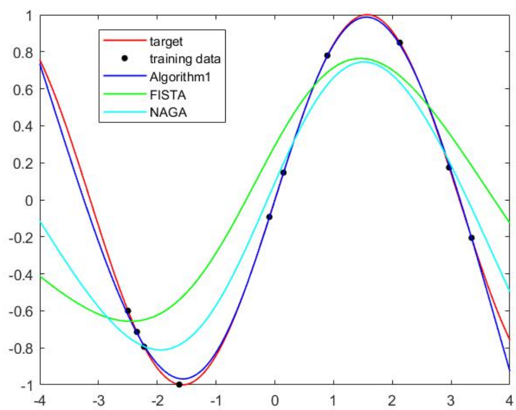

Figure 1. A regression of the sine function at 1700th step

4.2. Data Classification

In order to classify datasets, we classify the type of iris plants from Iris dataset and identify heart patient from Heart Disease UCI dataset. We would to thanks https://www.kaggle.com/ and https://archive.ics.uci.edu/ for supporting database website.

- Iris dataset [22] This dataset contains 3 classes of 50 instances where each class refers to a type of iris plant. The aim is to separate each type of iris plant (iris setosa, iris versicolour and iris virginica) from sepal and petal length.
- Heart Disease UCI dataset [23] The original dataset contains 76 attributes, but all published experiments refer to using a subset of 14 of them. This dataset refers to the presence of heart disease in the patient. The predicted attribute is aim to classify the data into 2 classes.

Table 2 shows information about the datasets,number of attributes and number of samples for training (around 70% of data) and testing (remainder 30% of data) sets.

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

TABLE 2. Information about the datasets.

Dataset	RegularizedELM					
	Algorithm 1		NAGA		FISTA	
	test	train	test	train	test	train
Heart Disease UCI	68.57	62.37	67.39	61.29	51.74	54.84
Iris	98.10	100.00	95.24	97.78	94.29	95.56

TABLE 3. Information about the datasets.

We set all control parameters λ, α_n, β_n, γ_n, θ_n as in Section 4.1, activation function is sigmoid, and the nmber of hidden nodes $M = 100$. Given a training set for each dataset as mentioned in Table 2. An accuracy of the output data is calculated by

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

Table 3 shows the performance in term of accuracy of training set and accuracy of testing set of each methods.
The results presented in Table 3 are obtained as follows:

- The proposed learning algorithm have a high performance of accuracy of training set with a few difference.
- The optimal weight computed by proposed algorithm gives a performance of accuracy better than those computed by NAGA and FISTA
- The proposed learning algorithm has a high performance of accuracy of testing set.

ACKNOWLEDGEMENTS

The author would like to thank the referees for valueable comments and suggestions for improving this work. This work was supported by Science Achievement Scholarship of Thailand (SAST) and Chiang Mai University.

References

[1] S.-S. Chang, H.W.J. Lee, C.K. Chan, A new method for solving equilibrium problem fixed point problem and variational inequality problem with application to optimization, Nonlinear Analysis 70 (2009) 3307–3319.

[2] A. Hanjing, S. Suanta, Hybrid inertial accelerated algorithms for split fixed point problems of demicontractive mappings and equilibrium problems, Numerical Algorithms (2020).

[3] A. Hanjing, S. Suantai, A fast image restoration algorithm based on a fixed point and optimization method, Mathematics 8 (2020) 378.

[4] L. Bussaban, S. Suantai, A. Kaewkhao, A parallel inertial S-iteration forward-backward algorithm for regression and classification problems, Carpathian Jornal of Mathematics 36 (2020) 35–44.

[5] L.C. Ceng, P. Cubiotti, J.C. Yao, An implicit iterative scheme for monotone variational inequalities and fixed point problems, Nonlinear Analysis 69 (2008) 2445–2457

[6] S. Suantai, Y. Shehu, P. Cholamjiak, O.S. lyiola, Strong convergence of a self-adaptive method for the split feasibility problem in Banach spaces, J. Fixed Point Theory Appl. (2018).

[7] W.R. Mann, Mean value methods in iteration, Proc. Amer. Math. Soc. 4 (1953) 506–510.

[8] B. Halpern, Fixed points of nonexpanding maps. Bull. Amer. Math. Soc. 73 (1967) 957–961.

[9] P.L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM Journal on Numerical Anal. 16 (6) (1979) 964–979.

[10] A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM Journal on Imaging Sciences 2 (1) (2009) 183–202.

[11] J. Liang, C.B. Schonlieb, Improving fista: Faster, smarter and greedier, (2018) Available at abs/1811.01430.

[12] M. Vermaand, K. Shukla, A new accelerated proximal gradient technique for regularized multitask learning framework, Pattern Recognition Letters 95 (2017) 98–103.

[13] K. Nakajo, K. Shimoji, W. Takahashi, Strong convergence to common fixed points of families of nonexpansive mappings in Banach spaces, J. Nonlinear Convex Anal. 8 (2007) 11–34.

[14] K. Nakajo, K. Shimoji, W. Takahashi, On strong convergencde by thehybrid method for families of mappings in Hilbert spaces, Nonlinear Analysis, Theory, Methods and Applications 71 (1–2) (2009) 112–119.

[15] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Pass, New York, 2004.

[16] H.H. Bauschke, P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, Springer, New York, 2011.

[17] R.S. Burachik, A.N. Iusem, Set-valued mappings and enlargements of monotone operator, 8, Springer Science Business Media, New York, 2007.

[18] W. Takahashi, Introduction to Nonlinear and Convex Analysis, Yokohama Publishers, Yokohama, 2009.

[19] K. Tan, H.K. Xu, Approximating fixed points of nonexpansive mappings by the Ishikawa iteration process, J. Math. Anal. Appl. 178 (2) (1993) 301–308.

[20] A. Moudafi, E. Al-Shemas, Simultaneous iterative methods for split equality problem, Trans. Math. Program. Appl. 1 (2013) 1–11.

[21] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society, Series B (Methodological) 58 (1) (1996) 267–288.

[22] D. Dua, E. Karra Taniskidou, UCI Machine Learning Repository, 2017.

[23] M. Lichman, UCI Machine Learning Repository, 2013.