CARPATHIAN J. MATH.
Volume **39** (2023), No. 1,
Pages 125 - 138

Online version at https://semnul.com/carpathian/
Print Edition: ISSN 1584 - 2851; Online Edition: ISSN 1843 - 4401
DOI: https://doi.org/10.37193/CJM.2023.01.08

In memoriam Professor Charles E. Chidume (1947- 2021)

A New Accelerated Viscosity Forward-backward Algorithm with a Linesearch for Some Convex Minimization Problems and its Applications to Data Classification

DAWAN CHUMPUNGAM[1], PANITARN SARNMETA[2] and SUTHEP SUANTAI[1]

ABSTRACT. In this paper, we focus on solving *convex minimization problem* in the form of a summation of two convex functions in which one of them is Frecét differentiable. In order to solve this problem, we introduce a new accelerated viscosity forward-backward algorithm with a new linesearch technique. The proposed algorithm converges strongly to a solution of the problem without assuming that a gradient of the objective function is L-Lipschitz continuous. As applications, we apply the proposed algorithm to classification problems and compare its performance with other algorithms mentioned in the literature.

1. INTRODUCTION

Many problems in computer science, economics, engineering, statistics, physics and medical science, such as signal processing, compressed sensing, medical image reconstruction, digital image processing, data prediction and classification, can be formulated as *convex minimization problems* in the form of the sum of two convex functions in which one of these functions is Frecét differentiable. The problem is defined as follows:

$$\min_{x\in H}\{f(x)+g(x)\}, \tag{1.1}$$

where $f : H \to \mathbb{R}\cup\{+\infty\}$ is a proper, convex differentiable function and $g : H \to \mathbb{R}\cup\{+\infty\}$ is proper, lower semicontinuous convex function defined on Hilbert space H. We refer to [4, 5, 11, 12, 17, 25] for more information on its applications.

A solution of (1.1) is in fact a fixed point of a mapping $prox_{\alpha g}(I-\alpha\nabla f)$. To be more precise, x^* solves (1.1) if and only if

$$x^* = prox_{\alpha g}(I-\alpha\nabla f)(x^*), \tag{1.2}$$

where $\alpha>0$, and $prox_{\alpha g}(x) = \arg\min_{y\in H}\{g(y)+\frac{1}{2\alpha}\|x-y\|^2\}$.

The *forward-backward algorithm* [13] was proposed to solve (1.1). This algorithm is defined by

$$x_{n+1} = \underbrace{prox_{\alpha_n g}}_{\text{backward}}\underbrace{(I-\alpha_n\nabla f)}_{\text{forward}}(x_n), \text{ for all } n\in\mathbb{N}, \tag{1.3}$$

where α_n is a positive step size. This algorithm converges to a solution of (1.1) under the assumption that ∇f is L-Lipschitz continuous and $\alpha_n \in (0,\frac{2}{L})$. Many authors have studied and introduced forward-backward type algorithms to solve (1.1), see for instance

Received: 22.09.2021. In revised form: 12.04.2022. Accepted: 23.04.2022
2020 *Mathematics Subject Classification.* 65K05, 90C25, 90C30.
Key words and phrases. *convex minimization problems, machine learning, forward-backward algorithm, linesearch, accelerated algorithm, data classification.*
Corresponding author: Suthep Suantai; suthep.s@cmu.ac.th

[1, 5, 15, 22]. There are also several works which focus on the nonconvex cases that can be reduced to (1.1), we refer to the work of Wu et al. [23] and the references therein, for more in-depth discussion on this topic. However, most of these works assume the Lipschitz assumption on $\triangledown f$ which might not be an easy task to verify in general. So, in this paper, we study another approach where $\triangledown f$ is not necessary L-Lipschitz continuous.

In 2016, Cruz and Nghia [2] introduced a linesearch technique as the following:

Linesearch 1 Given $x \in domg$, $\delta > 0$, $\sigma > 0$ and $\theta \in (0, 1)$.
Input Set $\alpha = \sigma$.
While $\alpha\|\triangledown f(prox_{\alpha g}(x - \alpha\triangledown fx)) - \triangledown fx\| > \delta\|prox_{\alpha g}(x - \alpha\triangledown fx) - x\|$
Set $\alpha = \theta\alpha$
End While
Output α.

They asserted that **Linesearch 1** stops after finitely many steps and proposed the following algorithm:

Algorithm 1 Given $x_0 \in domg$, $\delta \in (0, \frac{1}{2})$, $\sigma > 0$, and $\theta \in (0, 1)$, for all $n \in \mathbb{N}$, define

$$x_{n+1} = prox_{\gamma_n g}(I - \gamma_n \triangledown f)(x_n), \tag{1.4}$$

where $\gamma_n :=$ **Linesearch 1**$(x_n, \delta, \sigma, \theta)$. They also showed that a sequence generated by Algorithm 1 converges weakly to a solution of (1.1) under assumptions A1 and A2 as follows:

A1. f, g are proper lower semi-continuous convex functions with $domg \subseteq domf$,
A2. f is differentiable on an open set containing $domg$, and $\triangledown f$ is uniformly continuous on any bounded subset of $domg$ and maps any bounded subset of $domg$ to a bounded set in H.

As we can see, the L-Lipschitz continuity of $\triangledown f$ is not necessary. Moreover, if $\triangledown f$ is L-Lipschitz continuous, then A2 is satisfied.

In 2019, Kankam et. al. [11] proposed the following linesearch:

Linesearch 2 Given $x \in domg$, $\delta > 0$, $\sigma > 0$ and $\theta \in (0, 1)$. Set

$$L(x, \alpha) = prox_{\alpha g}(x - \alpha\triangledown f(x)), \text{ and}$$

$$S(x, \alpha) = prox_{\alpha g}(L(x, \alpha) - \alpha\triangledown f(L(x, \alpha))).$$

Input Set $\alpha = \sigma$.
While $\alpha \max\{(\|\triangledown f(S(x, \alpha)) - \triangledown f(L(x, \alpha))\|, \|\triangledown f(L(x, \alpha)) - \triangledown f(x)\|)\}$
$> \delta(\|S(x, \alpha) - L(x, \alpha)\| + \|L(x, \alpha) - x\|),$
Set $\alpha = \theta\alpha$
End While
Output α.

They also asserted that Linesearch 2 stops at finitely many steps, and proposed the following algorithm:

Algorithm 2 Given $x_0 \in domg$, $\delta \in (0, \frac{1}{8})$, $\sigma > 0$ and $\theta \in (0, 1)$, for all $n \in \mathbb{N}$, define

$$\begin{cases} y_n = prox_{\gamma_n g}(x_n - \gamma_n\triangledown f(x_n)), \\ x_{n+1} = prox_{\gamma_n g}(y_n - \gamma_n\triangledown f(y_n)), \end{cases}$$

where $\gamma_n :=$ **Linesearch 2**$(x_n, \delta, \sigma, \theta)$.

A weak convergence result of this algorithm was obtained under assumptions A1 and A2. Although, Algorithm 1 and 2 obtained weak convergence results without the Lipschitz assumption on $\triangledown f$, some improvements are still welcome, specifically, to improve its convergence behavior by using inertial step and viscosity technique.

Recently, in order to accelerate the convergence behavior of an algorithm, many authors has utilized the inertial technique. Polyak [16] was the first to introduce and investigate this technique. In his work, he focused on solving smooth convex minimization problems.

There are other approaches to solving the convex minimization problem (1.1) without the Lipschitz assumption on the gradient of the objective function, for instance Shehu et al. [18] considered the split feasibility problems which can be converted to convex minimization problem (1.1), we refer to [18, 19] for more details.

Inspired by the works mentioned earlier, we focus our attention on linesearch algorithms which relax L-Lipschitz continuity assumption on $\triangledown f$. Algorithm 1 and 2 only obtain weak convergence results which might not be enough for infinite dimension spaces. Moreover, accelerated techniques are also not used in these algorithms. So, our main objective is to introduce a modified linesearch technique and propose a new algorithm which utilizes a viscosity step, an inertial step and a modified linesearch to improve its convergence behavior and obtain its strong convergence theorem to a solution of (1.1) without the Lipschitz assumption on $\triangledown f$. Furthermore, we apply this new algorithm to solving classification problems and compare its performance with Algorithm 1 and 2 to show that the proposed algorithm has better performance.

2. PRELIMINARIES

In this section, we recall some definitions and lemmas which will be used in the main results.

Let $\{x_n\}$ be a sequence in H. A strong and weak convergence of $\{x_n\}$ to x are denoted by $x_n \to x$ and $x_n \rightharpoonup x$, respectively. Let $x \in H$ and $h : H \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous convex function. We denote $domh = \{x \in H : h(x) < +\infty\}$.

A *subdifferential* of h at x is defined as follows:

$$\partial h(x) := \{u \in H : \langle u, y - x\rangle \ + h(x) \leq h(y), \ y \in H\}.$$

A *proximal operator* $prox_{\alpha h} : H \to domh$ is defined as follows:

$$prox_{\alpha h}(x) = (I + \alpha\partial h)^{-1}(x),$$

where I is an identity and α is a positive number. This operator is single-valued and the following is satisfied:

$$\frac{x - prox_{\alpha h}(x)}{\alpha} \in \partial h(prox_{\alpha h}(x)), \text{ for all } x \in H \text{ and } \alpha > 0. \tag{2.5}$$

Next, we present some important lemmas for this work.

Lemma 2.1 ([3]). *A subdifferential ∂h is maximal monotone. Furthermore, a graph, $Gph(\partial h) := \{(x, y) \in H \times H : y \in \partial h(x)\}$, is demiclosed, i.e., for any sequence $\{(x_n, y_n)\} \subseteq Gph(\partial h)$ such that $\{x_n\} \rightharpoonup x$ and $\{y_n\} \to y$, then $(x, y) \in Gph(\partial h)$.*

Lemma 2.2 ([10]). *Let $f, g : H \to \mathbb{R} \cup \{+\infty\}$ be proper lower semicontinuous convex functions with $dom\ g \subseteq dom\ f$ and $J(x, \alpha) = prox_{\alpha g}(x - \alpha\triangledown f(x))$. Then for any $x \in dom\ g$ and*

$\alpha_2 \geq \alpha_1 > 0$, *we have*

$$\frac{\alpha_2}{\alpha_1}\|x - J(x,\alpha_1)\| \geq \|x - J(x,\alpha_2)\| \geq \|x - J(x,\alpha_1)\|.$$

Lemma 2.3 ([20]). *Let H be a real Hilbert space. Then, for all $a, b, c \in H$ and $\zeta \in [0,1]$, the following hold,*

(i) $\|a \pm b\|^2 = \|a\|^2 \pm 2\langle a, b\rangle + \|b\|^2$,
(ii) $\|\zeta a + (1-\zeta)b\|^2 = \zeta\|a\|^2 + (1-\zeta)\|b\|^2 - \zeta(1-\zeta)\|a-b\|^2$,
(iii) $\|a+b\|^2 \leq \|a\|^2 + 2\langle b, a+b\rangle$,
(iv) $\langle a-b, b-c\rangle = \frac{1}{2}(\|a-c\|^2 - \|a-b\|^2 - \|b-c\|^2)$.

Lemma 2.4 ([14]). *Let $\{a_n\}$ be a sequence of real numbers such that there exists a subsequence $\{a_{m_j}\}$ of $\{a_n\}$ such that $a_{m_j} < a_{m_j+1}$, for all $j \in \mathbb{N}$. Then there exists a nondecreasing sequence $\{n_k\}$ of $\mathbb{N}$ such that $\lim_{k\to+\infty} n_k = +\infty$ and for all sufficiently large $k \in \mathbb{N}$ the following holds:*

$$a_{n_k} \leq a_{n_k+1} \quad \text{and} \quad a_k \leq a_{n_k+1}.$$

Lemma 2.5 ([24]). *Let $\{a_n\}$ be a sequence of nonnegative real numbers, $\{\zeta_n\}$ a sequence in $(0,1)$ with $\sum_{n=1}^{+\infty}\zeta_n = +\infty$, $\{c_n\}$ a sequence of nonnegative real numbers with $\sum_{n=1}^{+\infty} c_n < +\infty$ and $\{b_n\}$ a sequence of real numbers with $\limsup_{n\to+\infty} b_n \leq 0$. Suppose that the following holds*

$$a_{n+1} \leq (1-\zeta_n)a_n + \zeta_n b_n + c_n,$$

for all $n \in \mathbb{N}$, then $\lim_{n\to\infty} a_n = 0$.

3. Main results

In this section, we denote S_* the set of all solutions of (1.1). We suppose that $f, g : H \to \mathbb{R} \cup \{+\infty\}$ are two proper lower semicontinuous convex functions which satisfy assumptions A1 and A2. Furthermore, we also suppose that $S_* \neq \emptyset$.

We introduce the following linesearch technique.

Linesearch 3 Given $x \in domg$, $\delta > 0$, $\sigma > 0$ and $\theta \in (0,1)$.
Set

$$L(x,\alpha) = prox_{\alpha g}(x - \alpha\nabla f(x)), \text{ and}$$

$$S(x,\alpha) = prox_{\alpha g}(L(x,\alpha) - \alpha\nabla f(L(x,\alpha))).$$

Input Set $\alpha = \sigma$.
While

$$\frac{\alpha}{2}(\|\nabla f(S(x,\alpha)) - \nabla f(L(x,\alpha))\| + \|\nabla f(L(x,\alpha)) - \nabla f(x)\|) > \delta(\|S(x,\alpha) - L(x,\alpha)\| + \|L(x,\alpha) - x\|),$$

$$\textbf{or} \qquad \alpha\|\nabla f(L(x,\alpha)) - \nabla f(x)\| > 4\delta\|L(x,\alpha) - x\|.$$

Set $\alpha = \theta\alpha$
End While
Output α.

We first show that Linesearch 3 terminates at finitely many steps.

Lemma 3.6. *Linesearch 3 stops at finitely many steps.*

Proof. If $x \in S_*$, then $x = L(x,\sigma) = S(x,\sigma)$, so Linsearch 3 stops with zero step. If $x \notin S_*$, suppose by contradiction that, for all $n \in \mathbb{N}$, the following hold

$$\frac{\sigma\theta^n}{2}(\|\nabla f(S(x,\sigma\theta^n)) - \nabla f(L(x,\sigma\theta^n))\| + \|\nabla f(L(x,\sigma\theta^n)) - \nabla f(x)\|)$$
$$> \delta(\|S(x,\sigma\theta^n) - L(x,\sigma\theta^n)\| + \|L(x,\sigma\theta^n) - x\|), \tag{3.6}$$

or

$$\sigma\theta^n\|\nabla f(L(x,\sigma\theta^n)) - \nabla f(x)\| > 4\delta\|L(x,\sigma\theta^n) - x\|. \tag{3.7}$$

Then, from these assumptions, we can find a subsequence $\{\sigma\theta^{n_k}\}$ of $\{\sigma\theta^n\}$ such that (3.6) or (3.7) holds. Then, it is sufficient to consider the following 2 cases.

Case 1, suppose that (3.6) holds, then it follows from assumption A2 and Lemma 2.2 that
$\|S(x,\sigma\theta^{n_k}) - L(x,\sigma\theta^{n_k})\| \to 0$ and $\|L(x,\sigma\theta^{n_k}) - x\| \to 0$, as $k \to +\infty$. Since ∇f is uniformly continuous, we get

$$\|\nabla f(S(x,\sigma\theta^{n_k})) - \nabla f(L(x,\sigma\theta^{n_k}))\| \to 0 \text{ and } \|\nabla f(L(x,\sigma\theta^{n_k})) - \nabla f(x)\| \to 0,$$

as $k \to +\infty$. So, it follows from (3.6) that $\frac{\|L(x,\sigma\theta^{n_k})-x\|}{\sigma\theta^{n_k}} \to 0$, as $k \to +\infty$. By (2.5), we obtain

$$\frac{x - \sigma\theta^{n_k}\nabla f(x) - L(x,\sigma\theta^{n_k})}{\sigma\theta^{n_k}} \in \partial g(L(x,\sigma\theta^{n_k})).$$

Thus, $\frac{L(x,\sigma\theta^{n_k})-x}{\sigma\theta^{n_k}} - \nabla f(x) \in \partial g(L(x,\sigma\theta^{n_k}))$. Since $L(x,\sigma\theta^{n_k}) \to x$, as $k \to +\infty$, we obtain from Lemma 2.1 that $0 \in \nabla f(x) + \partial g(x) \subseteq \partial(f+g)(x)$. Hence, $x \in S_*$ which is a contradiction.

Case 2, suppose that (3.7) holds, for all $\sigma\theta^{n_k}$. Then, from A2 and Lemma 2.2, we have $\|L(x,\sigma\theta^{n_k}) - x\| \to 0$, as $k \to +\infty$. Again, from the uniform continuity of ∇f, we have

$$\|\nabla f(L(x,\sigma\theta^{n_k})) - \nabla f(x)\| \to +\infty,$$

as $k \to +\infty$. From (3.7), we conclude that

$$\frac{\|L(x,\sigma\theta^{n_k}) - x\|}{\sigma\theta^{n_k}} \to 0,$$

as $k \to +\infty$. By the same argument as in Case 1, we can show that $0 \in \partial(f+g)(x)$, and hence $x \in S_*$, a contradiction. Therefore, we conclude that Linesearch 3 stops with finite steps, and the proof is complete. □

Let $F : dom g \to dom g$ be a contractive mapping. We define a new algorithm as follows:

Algorithm 3 Given $x_0, x_1 \in dom g, \delta \in (0, \frac{1}{8}), \sigma > 0, \theta \in (0,1), \mu \geq 0, \tau_n \in (0,1), \zeta_n \in (0,1)$ and $\alpha_n \in [0,1]$, for all $n \in \mathbb{N}$, define

$$\begin{cases} \hat{x}_n = x_n + \beta_n(x_n - x_{n-1}), \\ y_n = P_{dom g}\hat{x}_n, \\ z_n = (1-\zeta_n)y_n + \zeta_n F(y_n), \\ w_n = prox_{\gamma_n g}(z_n - \gamma_n \nabla f(z_n)), \\ x_{n+1} = (1-\alpha_n)w_n + \alpha_n prox_{\gamma_n g}(w_n - \gamma_n \nabla f(w_n)), \end{cases}$$

where $\gamma_n :=$ **Linesearch 3**$(z_n, \delta, \sigma, \theta)$, $\beta_n = \begin{cases} \min\{\mu, \frac{\tau_n}{\|x_n - x_{n-1}\|}\}, & \text{if } x_n \neq x_{n-1}, \\ \mu, & \text{otherwise,} \end{cases}$

and $P_{dom g}$ is a metric projection onto $dom g$.

Lemma 3.7. *Let* $\gamma_n :=$ ***Linesearch 3***$(z_n, \delta, \sigma, \theta)$*. Then, for all* $n \in \mathbb{N}$ *and* $x \in domg$*, the following hold:*

(I) $\|z_n - x\|^2 - \|w_n - x\|^2 \geq 2\gamma_n[(f+g)(w_n) - (f+g)(x)] + (1-8\delta)\|w_n - z_n\|^2$,

(II) $\|z_n - x\|^2 - \|v_n - x\|^2 \geq 2\gamma_n[(f+g)(w_n) + (f+g)(v_n) - 2(f+g)(x)]$
$+ (1-8\delta)(\|w_n - z_n\|^2 + \|v_n - w_n\|^2)$,

where $v_n = prox_{\gamma_n g}(w_n - \gamma_n \nabla f(w_n))$.

Proof. We first show that (I) is true. From (2.5), we know that

$$\frac{z_n - w_n}{\gamma_n} - \nabla f(z_n) \in \partial g(w_n), \quad \text{for all } n \in \mathbb{N}.$$

It follows from the definitions of $\partial g(w_n)$, $\nabla f(z_n)$ and $\nabla f(w_n)$ that

$$g(x) - g(w_n) \geq \langle \frac{z_n - w_n}{\gamma_n} - \nabla f(z_n), x - w_n\rangle,$$

$$f(x) - f(z_n) \geq \langle \nabla f(z_n), x - z_n\rangle \text{ and } f(z_n) - f(w_n) \geq \langle \nabla f(w_n), z_n - w_n\rangle,$$

for all $n \in \mathbb{N}$. Consequently,

$$\begin{aligned}
f(x) - f(z_n) + g(x) - g(w_n) &\geq \frac{1}{\gamma_n}\langle z_n - w_n, x - w_n\rangle + \langle \nabla f(z_n), w_n - z_n\rangle \\
&= \frac{1}{\gamma_n}\langle z_n - w_n, x - w_n\rangle + \langle \nabla f(z_n) - \nabla f(w_n), w_n - z_n\rangle \\
&\quad + \langle \nabla f(w_n), w_n - z_n\rangle \\
&\geq \frac{1}{\gamma_n}\langle z_n - w_n, x - w_n\rangle - \|\nabla f(z_n) - \nabla f(w_n)\|\|w_n - z_n\| \\
&\quad + \langle \nabla f(w_n), w_n - z_n\rangle \\
&\geq \frac{1}{\gamma_n}\langle z_n - w_n, x - w_n\rangle - \frac{4\delta}{\gamma_n}\|w_n - z_n\|^2 + f(w_n) - f(z_n),
\end{aligned}$$

for all $n \in \mathbb{N}$. Then, we obtain

$$\frac{1}{\gamma_n}\langle z_n - w_n, w_n - x\rangle \geq (f+g)(w_n) - (f+g)(x) - \frac{4\delta}{\gamma_n}\|w_n - z_n\|^2, \quad \text{for all } n \in \mathbb{N}.$$

It follows from Lemma 2.3 (iv) that $\langle z_n - w_n, w_n - x\rangle = \frac{1}{2}(\|z_n - x\|^2 - \|w_n - z_n\|^2 - \|w_n - x\|^2)$. Then, we have

$$\frac{1}{2\gamma_n}(\|z_n - x\|^2 - \|z_n - w_n\|^2 - \|w_n - x\|^2) \geq (f+g)(w_n) - (f+g)(x) - \frac{4\delta}{\gamma_n}\|z_n - w_n\|^2,$$

for all $n \in \mathbb{N}$. Hence, for any $x \in domg$, we have

$$\|z_n - x\|^2 - \|w_n - x\|^2 \geq 2\gamma_n[(f+g)(w_n) - (f+g)(x)] + (1-8\delta)\|w_n - z_n\|^2,$$

and (I) is proved. Next we show that (II) also hold. To prove our claim, we know that

$$\frac{z_n - w_n}{\gamma_n} - \nabla f(z_n) \in \partial g(w_n), \text{ and}$$

$$\frac{w_n - v_n}{\gamma_n} - \nabla f(w_n) \in \partial g(v_n).$$

Then,

$$g(x) - g(w_n) \geq \langle \frac{z_n - w_n}{\gamma_n} - \nabla f(z_n), x - w_n\rangle, \text{ and}$$

$$g(x) - g(v_n) \geq \langle \frac{w_n - v_n}{\gamma_n} - \nabla f(w_n), x - v_n\rangle, \quad \text{for all } n \in \mathbb{N}.$$

Moreover,

$$f(x) - f(z_n) \geq \langle \nabla f(z_n), x - z_n \rangle,$$
$$f(x) - f(w_n) \geq \langle \nabla f(w_n), x - w_n \rangle,$$
$$f(z_n) - f(w_n) \geq \langle \nabla f(w_n), z_n - w_n \rangle, \text{ and}$$
$$f(w_n) - f(v_n) \geq \langle \nabla f(v_n), w_n - v_n \rangle, \quad \text{for all } n \in \mathbb{N}.$$

These inequalities imply that

$$\begin{aligned}
&f(x) - f(z_n) + f(x) - f(w_n) + g(x) - g(w_n) + g(x) - g(v_n) \\
&\geq \frac{1}{\gamma_n}\langle z_n - w_n, x - w_n \rangle + \langle \nabla f(z_n), w_n - z_n \rangle + \frac{1}{\gamma_n}\langle w_n - v_n, x - v_n \rangle + \langle \nabla f(w_n), v_n - w_n \rangle, \\
&= \frac{1}{\gamma_n}\langle z_n - w_n, x - w_n \rangle + \langle \nabla f(z_n) - \nabla f(w_n), w_n - z_n \rangle + \langle \nabla f(w_n), w_n - z_n \rangle \\
&\quad + \frac{1}{\gamma_n}\langle w_n - v_n, x - v_n \rangle + \langle \nabla f(w_n) - \nabla f(v_n), v_n - w_n \rangle + \langle \nabla f(v_n), v_n - w_n \rangle, \\
&\geq \frac{1}{\gamma_n}\langle z_n - w_n, x - w_n \rangle + \frac{1}{\gamma_n}\langle w_n - v_n, x - v_n \rangle - \|\nabla f(w_n) - \nabla f(z_n)\|\|w_n - z_n\| \\
&\quad + \langle \nabla f(w_n), w_n - z_n \rangle - \|\nabla f(v_n) - \nabla f(w_n)\|\|v_n - w_n\| + \langle \nabla f(v_n), v_n - w_n \rangle, \\
&\geq \frac{1}{\gamma_n}\langle z_n - w_n, x - w_n \rangle + \frac{1}{\gamma_n}\langle w_n - v_n, x - v_n \rangle - \|\nabla f(w_n) - \nabla f(z_n)\|(\|w_n - z_n\| + \|v_n - w_n\|) \\
&\quad + \langle \nabla f(w_n), w_n - z_n \rangle - \|\nabla f(v_n) - \nabla f(w_n)\|(\|w_n - z_n\| + \|v_n - w_n\|) + \langle \nabla f(v_n), v_n - w_n \rangle, \\
&= \frac{1}{\gamma_n}\langle z_n - w_n, x - w_n \rangle + \frac{1}{\gamma_n}\langle w_n - v_n, x - v_n \rangle + \langle \nabla f(w_n), w_n - z_n \rangle + \langle \nabla f(v_n), v_n - w_n \rangle \\
&\quad - (\|\nabla f(w_n) - \nabla f(z_n)\| + \|\nabla f(v_n) - \nabla f(w_n)\|)(\|w_n - z_n\| + \|v_n - w_n\|), \\
&\geq \frac{1}{\gamma_n}\langle z_n - w_n, x - w_n \rangle + \frac{1}{\gamma_n}\langle w_n - v_n, x - v_n \rangle + \langle \nabla f(w_n), w_n - z_n \rangle + \langle \nabla f(v_n), v_n - w_n \rangle \\
&\quad - \frac{2\delta}{\gamma_n}(\|w_n - z_n\| + \|v_n - w_n\|)^2, \\
&\geq \frac{1}{\gamma_n}\langle z_n - w_n, x - w_n \rangle + \frac{1}{\gamma_n}\langle w_n - v_n, x - v_n \rangle + f(v_n) - f(z_n) - \frac{4\delta}{\gamma_n}(\|w_n - z_n\|^2 + \|v_n - w_n\|^2),
\end{aligned}$$

for all $x \in dom\ g$, and $n \in \mathbb{N}$.

Hence,
$$\begin{aligned}
&\frac{1}{\gamma_n}\langle z_n - w_n, w_n - x \rangle + \frac{1}{\gamma_n}\langle w_n - v_n, v_n - x \rangle \\
&\geq (f+g)(w_n) + (f+g)(v_n) - 2(f+g)(x) - \frac{4\delta}{\gamma_n}\|w_n - z_n\|^2 - \frac{4\delta}{\gamma_n}\|v_n - w_n\|^2.
\end{aligned}$$

Moreover, by Lemma 2.3 (iv), we also obtain the following, for all $n \in \mathbb{N}$,

$$\langle z_n - w_n, w_n - x \rangle = \frac{1}{2}(\|z_n - x\|^2 - \|z_n - w_n\|^2 - \|w_n - x\|^2), \text{ and}$$
$$\langle w_n - v_n, v_n - x \rangle = \frac{1}{2}(\|w_n - x\|^2 - \|w_n - v_n\|^2 - \|v_n - x\|^2).$$

As a result, we obtain

$$\begin{aligned}
&\frac{1}{2\gamma_n}(\|z_n - x\|^2 - \|z_n - w_n\|^2) - \frac{1}{2\gamma_n}(\|w_n - v_n\|^2 + \|v_n - x\|^2) \\
&\geq (f+g)(w_n) + (f+g)(v_n) - 2(f+g)(x) - \frac{4\delta}{\gamma_n}\|w_n - z_n\|^2 - \frac{4\delta}{\gamma_n}\|v_n - w_n\|^2,
\end{aligned}$$

for all $x \in dom\ g$, and $n \in \mathbb{N}$. Therefore,

$$\|z_n - x\|^2 - \|v_n - x\|^2 \geq 2\gamma_n[(f+g)(w_n) + (f+g)(v_n) - 2(f+g)(x)] + (1-8\delta)(\|w_n - z_n\|^2 + \|v_n - w_n\|^2),$$

for all $x \in dom\ g$, and $n \in \mathbb{N}$, and (II) is proved. □

Theorem 3.1. *Let x_n be a sequence generated by Algorithm 3. Suppose that the following hold:*

C1. $\lim_{n\to+\infty} \zeta_n = 0$ *and* $\sum_{n=1}^{+\infty} \zeta_n = +\infty$,
C2. there exists $\gamma > 0$ *such that* $\gamma_n \geq \gamma$, *for all* $n \in \mathbb{N}$,
C3. $\lim_{n\to+\infty} \frac{\tau_n}{\zeta_n} = 0$.

Then $\{x_n\}$ converges strongly to $x^* = P_{S_*}F(x^*)$.

Proof. We know that S_* is closed and convex, so mapping $P_{S_*}F$ has a fixed point. It also follows from C3 that $\lim_{n\to+\infty} \frac{\beta_n}{\zeta_n}\|x_n - x_{n-1}\| = 0$. Now, let $x^* = P_{S_*}F(x^*)$, we know that $x_n \in dom\ g$. Consequently, the following hold, for all $n \in \mathbb{N}$,

$$\|y_n - \hat{x}_n\| \leq \|x_n - \hat{x}_n\| = \beta_n\|x_n - x_{n-1}\| = \frac{\beta_n}{\zeta_n}\|x_n - x_{n-1}\|\zeta_n \to 0, \text{ as } n \to \infty, \tag{3.8}$$

$$\|z_n - y_n\| = \zeta_n\|Fy_n - y_n\|. \tag{3.9}$$

Since $x^* \in domg$ and P_{domg} is nonexpansive, we also have

$$\|y_n - x^*\| \leq \|\hat{x}_n - x^*\| \leq \|x_n - x^*\| + \frac{\beta_n}{\zeta_n}\|x_n - x_{n-1}\|\zeta_n \leq \|x_n - x^*\| + \zeta_n M_1, \tag{3.10}$$

for some $M_1 \geq \frac{\beta_n}{\zeta_n}\|x_n - x_{n-1}\|$. Moreover, we obtain the following , for all $n \in \mathbb{N}$,

$$\begin{aligned} \|y_n - x^*\|^2 &\leq \|x_n - x^*\|^2 + 2\beta_n\|x_n - x^*\|\|x_n - x_{n-1}\| + \beta_n^2\|x_n - x_{n-1}\|^2 \\ &= \|x_n - x^*\|^2 + \frac{\beta_n}{\zeta_n}\|x_n - x_{n-1}\|(2\zeta_n\|x_n - x^*\| + \zeta_n\beta_n\|x_n - x_{n-1}\|). \end{aligned} \tag{3.11}$$

It follows directly from Leamma 3.7 that

$$\|z_n - x^*\|^2 - \|w_n - x^*\|^2 \geq (1-8\delta)\|w_n - z_n\|^2, \tag{3.12}$$

and

$$\|z_n - x^*\|^2 - \|v_n - x^*\|^2 \geq (1-8\delta)(\|w_n - z_n\|^2 + \|v_n - w_n\|^2), \tag{3.13}$$

where $v_n = prox_{\gamma_n g}(w_n - \gamma_n \nabla f(w_n))$, for all $n \in \mathbb{N}$. So,

$$\begin{aligned} \|x_{n+1} - x^*\|^2 &\leq (1-\alpha_n)\|w_n - x^*\|^2 + \alpha_n\|v_n - x^*\|^2 \\ &\leq \|z_n - x^*\|^2, \quad \text{for all } n \in \mathbb{N}. \end{aligned} \tag{3.14}$$

Next, we prove that $\{x_n\}$ is bounded. Indeed, from (3.10) and (3.14), we have

$$\begin{aligned}\|x_{n+1}-x^*\| \leq \|z_n-x^*\| &= \|\zeta_n F y_n + (1-\zeta_n)y_n - x^*\| \\ &\leq \zeta_n\|Fy_n - Fx^*\| + \zeta_n\|Fx^* - x^*\| + (1-\zeta_n)\|y_n - x^*\| \\ &\leq (1-(1-c)\zeta_n)\|y_n - x^*\| + \zeta_n\|Fx^* - x^*\| \\ &\leq (1-(1-c)\zeta_n)\|x_n - x^*\| + (1-(1-c)\zeta_n)\zeta_n M_1 + \zeta_n\|Fx^* - x^*\| \\ &\leq (1-(1-c)\zeta_n)\|x_n - x^*\| + \zeta_n(1-c)(\frac{M_1 + \|Fx^* - x^*\|}{1-c}) \\ &\leq \max\{\|x_n - x^*\|, \frac{M_1 + \|Fx^* - x^*\|}{1-c}\}.\end{aligned}$$

So, $\|x_{n+1} - x^*\| \leq \max\{\|x_0 - x^*\|, \frac{M_1 + \|Fx^* - x^*\|}{1-c}\}$, and hence $\{x_n\}$ is bounded. Furthermore, from (3.9) and (3.10), we also obtain that $\{y_n\}$ and $\{z_n\}$ are bounded.
Next, we prove that $\{x_n\}$ converges strongly to x^*. The proof is divided into two cases.

Case 1: There exists $N_0 \in \mathbb{N}$ such that $\|x_{n+1} - x^*\| \leq \|x_n - x^*\|$, for all $n \geq N_0$. So $\lim_{n\to\infty} \|x_n - x^*\| = a$, for some $a \in \mathbb{R}$. By using (3.8), (3.9) and (3.14) we obtain following inequality:

$$\|x_{n+1} - x^*\| \leq \|z_n - x^*\| \leq \|z_n - y_n\| + \|y_n - x_n\| + \|x_n - x^*\|,$$

we conclude that $\lim_{n\to\infty} \|z_n - x^*\| = a$. A boundedness of $\{z_n\}$ implies that there exists a subsequence $\{z_{n_k}\}$ of $\{z_n\}$ such that $z_{n_k} \rightharpoonup w$, for some $w \in H$, and

$$\limsup_{n\to\infty}\langle Fx^* - x^*, z_n - x^*\rangle = \lim_{k\to\infty}\langle Fx^* - x^*, z_{n_k} - x^*\rangle = \langle Fx^* - x^*, w - x^*\rangle.$$

If $w \in S_*$, then $\limsup_{n\to\infty}\langle Fx^* - x^*, z_n - x^*\rangle \leq 0$. So, in the next step, we show that $w \in S_*$. In order to prove this, we first show that $\lim_{n\to+\infty} \|z_n - w_n\| = 0$.

If $\limsup_{n\to+\infty} \alpha_n = r < 1$, then from (3.14), we obtain

$$\limsup_{n\to+\infty} \|w_n - x^*\|^2 = \limsup_{n\to+\infty} \|x_n - x^*\|^2 = \limsup_{n\to+\infty} \|z_n - x^*\|^2 = a^2.$$

So, it follows from (3.12) that $\lim_{n\to+\infty} \|w_n - z_n\| = 0$.

If $\limsup_{n\to+\infty} \alpha_n = 1$, then by (3.14), we get

$$\limsup_{n\to+\infty} \|v_n - x^*\|^2 = \limsup_{n\to+\infty} \|x_n - x^*\|^2 = \limsup_{n\to+\infty} \|z_n - x^*\|^2 = a^2.$$

Using (3.13), we have $\lim_{n\to+\infty} \|w_n - z_n\| = 0$. Thus, we conclude that $w_{n_k} \rightharpoonup w$. Since ∇f is uniformly continuous, we have $\lim_{k\to+\infty} \|\nabla f(z_{n_k}) - \nabla f(w_{n_k})\| = 0$. From (2.5), we get

$$\frac{z_{n_k} - \gamma_{n_k}\nabla f(z_{n_k}) - w_{n_k}}{\gamma_{n_k}} \in \partial g(w_{n_k}), \quad \text{for all } k \in \mathbb{N}.$$

Hence,

$$\frac{z_{n_k} - w_{n_k}}{\gamma_{n_k}} - \nabla f(z_{n_k}) + \nabla f(w_{n_k}) \in \partial g(w_{n_k}) + \nabla f(w_{n_k}) = \partial(f+g)(w_{n_k}), \quad \text{for all } k \in \mathbb{N}.$$

By letting $k \to +\infty$ in the above inequality and the fact that $Gph(\partial(f+g))$ is demiclosed, we obtain that $0 \in \partial(f+g)(w)$. Hence $w \in S_*$.

From (3.11) and Lemma 2.3, we have

$$\begin{aligned}\|z_n - x^*\|^2 &= \|\zeta_n F y_n + (1-\zeta_n) y_n - x^*\|^2,\\ &= \|\zeta_n(Fy_n - Fx^*) + \zeta_n(Fx^* - x^*) + (1-\zeta_n)(y_n - x^*)\|^2,\\ &\le \|\zeta_n(Fy_n - Fx^*) + (1-\zeta_n)(y_n - x^*)\|^2 + 2\zeta_n\langle Fx^* - x^*, z_n - x^*\rangle,\\ &\le (1-(1-c)\zeta_n)\|y_n - x^*\|^2 + 2\zeta_n\langle Fx^* - x^*, z_n - x^*\rangle,\\ &\le (1-(1-c)\zeta_n)\|x_n - x^*\|^2 + \frac{\beta_n}{\zeta_n}\|x_n - x_{n-1}\|(2\zeta_n\|x_n - x^*\|\\ &\quad + \zeta_n\beta_n\|x_n - x_{n-1}\|) + 2\zeta_n\langle Fx^* - x^*, z_n - x^*\rangle,\\ &\le (1-(1-c)\zeta_n)\|x_n - x^*\|^2 + (1-c)\zeta_n(\frac{\beta_n}{\zeta_n}\|x_n - x_{n-1}\|\frac{M_2}{1-c}\\ &\quad + \frac{2}{1-c}\langle Fx^* - x^*, z_n - x^*\rangle),\end{aligned}$$

for some $M_2 \ge 2\|x_n - x^*\| + \beta_n\|x_n - x_{n-1}\|$. Hence, from (3.14), we obtain

(3.15)
$$\|x_{n+1} - x^*\|^2 \le (1-(1-c)\zeta_n)\|x_n - x^*\|^2 + (1-c)\zeta_n(\frac{\beta_n}{\zeta_n}\|x_n - x_{n-1}\|\frac{M_2}{1-c} + \frac{2}{1-c}\langle Fx^* - x^*, z_n - x^*\rangle).$$

Since $\frac{\beta_n}{\zeta_n}\|x_n - x_{n-1}\|\frac{M_2}{1-c} \to 0$ and $\limsup_{n\to\infty}\langle Fx^* - x^*, z_n - x^*\rangle \le 0$, We can invoke Lemma 2.5 to conclude that $\|x_n - x^*\|^2 \to 0$. That is $\{x_n\}$ converges strongly to x^*.

Case 2: Suppose that there exists a subsequence $\{x_{m_j}\}$ of $\{x_n\}$ such that $\|x_{m_j} - x^*\| < \|x_{m_j+1} - x^*\|$, for all $j \in \mathbb{N}$. As a consequence of Lemma 2.4, there exists a nondecreasing sequence $\{n_k\}$ of $\mathbb{N}$ such that $\lim_{k\to+\infty} n_k = +\infty$ and the following is satisfied, for any sufficiently large $k \in \mathbb{N}$,

$$\|x_{n_k} - x^*\| \le \|x_{n_k+1} - x^*\| \text{ and } \|x_k - x^*\| \le \|x_{n_k+1} - x^*\|.$$

From the definition of z_{n_k} and (3.11), we have, for all $k \in \mathbb{N}$,

$$\begin{aligned}\|z_{n_k} - x^*\|^2 &\le \zeta_{n_k}\|Fy_{n_k} - x^*\|^2 + (1-\zeta_{n_k})\|y_{n_k} - x^*\|^2\\ &\le \zeta_{n_k}\|Fy_{n_k} - x^*\|^2 + \|y_{n_k} - x^*\|^2\\ &\le \|x_{n_k} - x^*\|^2 + \zeta_{n_k}\|Fy_{n_k} - x^*\|^2\\ &\quad + \frac{\beta_{n_k}}{\zeta_{n_k}}\|x_{n_k} - x_{n_k-1}\|(2\zeta_{n_k}\|x_{n_k} - x^*\| + \zeta_{n_k}\beta_{n_k}\|x_{n_k} - x_{n_k-1}\|)\\ &\le \|x_{n_k} - x^*\|^2 + \zeta_{n_k}\|Fy_{n_k} - x^*\|^2 + \zeta_{n_k}(\frac{\beta_{n_k}}{\zeta_{n_k}}\|x_{n_k} - x_{n_k-1}\|M_3)\\ &\le \|x_{n_k+1} - x^*\|^2 + \zeta_{n_k}\|Fy_{n_k} - x^*\|^2 + \zeta_{n_k}(\frac{\beta_{n_k}}{\zeta_{n_k}}\|x_{n_k} - x_{n_k-1}\|M_3),\end{aligned} \tag{3.16}$$

for some $M_3 \ge 2\|x_{n_k} - x^*\| + \beta_{n_k}\|x_{n_k} - x_{n_k-1}\|$. So, it follows from (3.12) and (3.13) that

$$\begin{aligned}\|z_{n_k} - x^*\|^2 - \|x_{n_k+1} - x^*\|^2 &\ge \|z_{n_k} - x^*\|^2 - (1-\zeta_{n_k})\|w_{n_k} - x^*\|^2 - \zeta_{n_k}\|v_{n_k} - x^*\|^2\\ &\ge (1-\zeta_{n_k})(1-8\delta)\|w_{n_k} - z_{n_k}\|^2 + \zeta_{n_k}(1-8\delta)\|w_{n_k} - z_{n_k}\|^2\\ &= (1-8\delta)\|w_{n_k} - z_{n_k}\|^2\end{aligned} \tag{3.17}$$

By combining (3.16) and (3.17), we get

$$\zeta_{n_k}\|Fy_{n_k} - x^*\|^2 + \zeta_{n_k}(\frac{\beta_{n_k}}{\zeta_{n_k}}\|x_{n_k} - x_{n_k-1}\|M_3) \ge (1-8\delta)\|w_{n_k} - z_{n_k}\|, \quad \text{for all } k \in \mathbb{N}.$$

So, $\lim_{k\to+\infty}\|w_{n_k}-z_{n_k}\|=0$. Again, the boundedness of $\{z_{n_k}\}$ implies that there exists a subsequence $\{z_{n_{k_j}}\}$ such that $z_{n_{k_j}}\rightharpoonup w$, for some $w\in H$, and

$$\limsup_{k\to\infty}\langle Fx^*-x^*,z_{n_k}-x^*\rangle=\lim_{j\to\infty}\langle Fx^*-x^*,z_{n_{k_j}}-x^*\rangle=\langle Fx^*-x^*,w-x^*\rangle.$$

By the same argument as in case 1, we conclude that $w\in S_*$ and

$$\limsup_{k\to\infty}\langle Fx^*-x^*,z_{n_k}-x^*\rangle=\langle Fx^*-x^*,w-x^*\rangle\leq 0.$$

Moreover, it follows from (3.15) that

$$\begin{aligned}\|x_{n_k+1}-x^*\|^2&\leq(1-(1-c)\zeta_{n_k})\|x_{n_k}-x^*\|^2\\&\quad+(1-c)\zeta_{n_k}(\frac{\beta_{n_k}}{\zeta_{n_k}}\|x_{n_k}-x_{n_k-1}\|\frac{M_2}{1-c}+\frac{2}{1-c}\langle Fx^*-x^*,z_{n_k}-x^*\rangle),\\&\leq(1-(1-c)\zeta_{n_k})\|x_{n_k+1}-x^*\|^2\\&\quad+(1-c)\zeta_{n_k}(\frac{\beta_{n_k}}{\zeta_{n_k}}\|x_{n_k}-x_{n_k-1}\|\frac{M_2}{1-c}+\frac{2}{1-c}\langle Fx^*-x^*,z_{n_k}-x^*\rangle).\end{aligned}$$

Consequently, $\|x_{n_k+1}-x^*\|^2\leq\frac{\beta_{n_k}}{\zeta_{n_k}}\|x_{n_k}-x_{n_k-1}\|\frac{M_2}{1-c}+\frac{2}{1-c}\langle Fx^*-x^*,z_{n_k}-x^*\rangle$. Hence,

$$0\leq\limsup_{k\to\infty}\|x_k-x^*\|^2\leq\limsup_{k\to\infty}\|x_{n_k+1}-x^*\|^2\leq 0.$$

Thus, we conclude that $\{x_n\}$ converges strongly to x^*, and the proof is complete. $\square$

We note that, the twice computations of proximal map per iterations in Algorithm 3 might be computationally expensive. However, in real world applications, the performance of Algorithm 3 is better than the linesearch algorithms which use only one computation of proximal map per iteration in term of accuracy as evidence by the numerical experiments in Section 4. Moreover, the difference in the computational time is very small.

4. Applications to Data Classification Problems

In this section, we apply Algorithm 1, 2 and 3 to solving some classification problems based on a learning technique called *Extreme Learning Machine(ELM)* introduced by Huang et. al. [9]. It is defined as follows:

Let $S:=\{(x_k,t_k):x_k\in\mathbb{R}^n,t_k\in\mathbb{R}^m,k=1,2,...,N\}$ be a training set of N samples, x_k is an *input* and t_k is a *target*. The output of ELM with M hidden nodes and activation function G is defined by

$$o_j=\sum_{i=1}^{M}\eta_i G(\langle w_i,x_j\rangle+b_i),$$

where w_i is the weight vector connecting the i-th hidden node and the input node, η_i is the weight vector connecting the i-th hidden node and the output node and b_i is bias. The hidden layer output matrix $\mathbf{H}$ is defined as

$$\mathbf{H}=\begin{bmatrix}G(\langle w_1,x_1\rangle+b_1)&\cdots&G(\langle w_M,x_1\rangle+b_M)\\\vdots&\ddots&\vdots\\G(\langle w_1,x_N\rangle+b_1)&\cdots&G(\langle w_M,x_N\rangle+b_M)\end{bmatrix}.$$

The main objective of ELM is to calculate an optimal weight $\eta=[\eta_1^T,...,\eta_M^T]^T$ such that $\mathbf{H}\eta=\mathbf{T}$, where $\mathbf{T}=[t_1^T,...,t_N^T]^T$ is the training set. If *Moore-Penrose generalized inverse* $\mathbf{H}^\dagger$ of $\mathbf{H}$ exists, then $\eta=\mathbf{H}^\dagger\mathbf{T}$ is the solution. However, in general cases, $\mathbf{H}^\dagger$ may not exist or difficult to obtain. Thus, in order to avoid such difficulties, we utilize the convex minimization to find η without relying on $\mathbf{H}^{\dagger}$.

In machine learning, a model can be overfit in the sense that it is very accurate on a training set but inaccurate on testing sets. In other words, it can not be used to predict unknown data. In order to prevent overfitting, the *Least absolute shrinkage and selection operator (LASSO)* [21] is used. It can be formulated as follows:

$$\min_{\eta}\{\|\mathbf{H}\eta - \mathbf{T}\|_2^2 + \lambda\|\eta\|_1\}, \tag{4.18}$$

where λ is a regularization parameter. If we set $f(x) = \|\mathbf{H}x - \mathbf{T}\|_2^2$ and $g(x) = \lambda\|x\|_1$, then it is problem (1.1).

In our experiments, we classify four data sets from https://archive.ics.uci.edu namely Iris [7], Heart disease [6], Breast cancer and Wine [8]. The detail of these data sets can be seen in Table 1.

TABLE 1. Detail of data sets

	Instances	Attributes	Classes
Iris	150	4	3
Heart disease	303	13	2
Breast cancer	569	30	2
Wine	178	13	3

We use sigmoid as an activation function with number of hidden nodes $M = 30$, We use 10-fold cross validation and *Average accuracy* to evaluate the performance. It is define as follows:

$$\text{Average ACC} = \sum_{i=1}^{N} \frac{x_i}{y_i} \times 100\%/N,$$

where N is a number of sets considered during cross validation ($N = 10$), x_i is a number of correctly predicted data at fold i and y_i is a number of all data at fold i. All parameters of Algorithm 1, 2 and 3 are chosen as seen in Table 2.

TABLE 2. Chosen parameters of each algorithm

	Algorithm 1	Algorithm 2	Algorithm 3
σ	0.49	0.124	0.124
δ	0.1	0.1	0.1
θ	0.1	0.1	0.1
ζ_n	-	-	$\frac{1}{100n}$
α_n	-	-	$\frac{1}{2}$

The inertial parameters β_n of Algorithm 3 is chosen as

$$\beta_n = \begin{cases} \min\{0.95, \frac{10^{30}}{n^2\|x_n - x_{n-1}\|}\}, & \text{if } x_n \neq x_{n-1}, \\ 0.95, & \text{otherwise}, \end{cases}$$

and the contractive mapping F for Algorithm 3 is defined by $F(x) = 0.99x$. Regularization parameters λ for each algorithm and data set are chosen using 10-fold cv. The parameters λ can be seen in Table 3.

TABLE 3. Chosen λ of each algorithm

	Iris	Heart disease	Breast cancer	Wine
Algorithm 1	0.001	0.003	0.2	0.02
Algorithm 2	0.01	0.03	0.9	0.006
Algorithm 3	0.01	0.03	0.07	0.0001

TABLE 4. Average accuracy of each algorithm at 300th iteration with 10-fold cv.

	Algorithm 1			Algorithm 2			Algorithm 3		
	train	test	time	train	test	time	train	test	time
Iris	92.37	90.67	0.095	94.37	94.00	0.105	98.59	98.67	0.102
Heart disease	81.85	80.52	0.119	83.53	81.84	0.148	84.31	82.85	0.139
Breast cancer	95.82	95.77	0.149	95.90	95.77	0.193	96.93	96.31	0.183
Wine	97.57	97.16	0.095	98.00	97.19	0.112	99.63	99.44	0.106

We measure the performance of each algorithm at 300th iteration with average accuracy. We also compare their computational time (in seconds) on training sets, and the results can be seen in Table 4.

From Table 4, with our choice of regularization parameters, we see that all models are not overfitting because the accuracy differences between training and testing set are small. Moreover, Algorithm 3 obtains higher accuracy than other algorithms in both training and testing sets for each data set. We also observe that the difference between the computational time of each algorithm is very small. Based on these experiment, Algorithm 3 performs better than other linesearch algorithms.

5. CONCLUSIONS

In conclusion, in this work we introduced a new accelerated forward-backward algorithm with a modified linesearch. A strong convergence to a solution of (1.1) of the proposed algorithm is obtained without assuming ∇f to be $L-$Lipschitz continuity. Furthermore, we apply our new algorithm to classification problems and evaluate its performance and compare with other linesearch algorithm, namely Algorithm 1 and 2. We observed from these experiments that the proposed algorithm performs better than Algorithm 1 and 2. Our model and algorithm can also be applied to prediction and classification of other data sets which will be useful for other real world applications in various fields such as medical science, economics, engineering, and business.

Acknowledgments. This research has received funding support from the NSRF via the Program Management Unit for Human Resources & Institutional Development, Research and Innovation [grant number B05F640183]. D. Chumpungam was also supported by Post-Doctoral Fellowship of Chiang Mai University, Thailand.

REFERENCES

[1] Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2** (2009), 183–202.

[2] Bello Cruz, J. Y.; Nghia, T. T. On the convergence of the forward-backward splitting method with linesearches. *Optim. Methods Softw.* **31** (2016), 1209–1238.

[3] Burachik, R. S.; Iusem, A. N. *Set-Valued Mappings and Enlargements of Monotone Operators*. Springer, Berlin, 2008.

[4] Chen, M.; Zhang, H.; Lin, G.; Han, Q. A new local and nonlocal total variation regularization model for image denoising. *Cluster Comput.* **22** (2019), 7611–7627.

[5] Combettes, P. L.; Wajs, V. Signal recovery by proximal forward–backward splitting. *Multiscale Model. Simul.* **4** (2005), 1168–1200.

[6] Detrano, R.; Janosi, A.; Steinbrunn, W.; Pfisterer, M.,; Schmid, J. J.; Sandhu, S.; Guppy, K. H.; Lee, S., Froelicher, V. International application of a new probability algorithm for the diagnosis of coronary artery disease. *Am. J. Cardiol.*, **64** (1989), no. 5, 304–10.

[7] Fisher, R. A. The use of multiple measurements in taxonomic problems. *Ann. Eugen*., **7** (1936), no. 2, 179–188

[8] Forina, M.; Leardi, R.; Armanino C.; Lanteri, S. *PARVUS: An Extendable Package of Programs for Data Exploration*. Elsevier, Amsterdam, 1988.

[9] Huang, G. B.; Zhu, Q. Y.; Siew, C. K. Extreme learning machine: theory and applications. *Neurocomputing* **70** (2006), 489–501.

[10] Huang, Y.; Dong, Y. New properties of forward-backward splitting and a practical proximal-descent algorithm. *Appl. Math. Comput.* **237** (2014), 60–68.

[11] Kankam, K.; Pholasa, N.; Cholamjiak, C. On convergence and complexity of the modified forward-backward method involving new linesearches for convex minimization. *Math Meth Appl Sci.* **42** (2019), 1352–1362.

[12] Kowalski, M.; Meynard, A.; Wu, H. Convex optimization approach to signals with fast varying instantaneous frequency. *Appl. Comput. Harmon. Anal. 44* (2018), 89–122.

[13] Lions, P. L.; Mercier, B. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **16** (1979), 964–979.

[14] Mainge, P. E. Strong convergence of projected subgradient methods for nonsmooth and nonstrictly convex minimization. *Set-Valued Anal.* **16** (2008), 899–912.

[15] Moudafi, A.; Oliny, M. Convergence of a splitting inertial proximal method for monotone operators. *J. Comput. Appl. Math.* **155** (2003), 447–454.

[16] Polyak, B. T. Some methods of speeding up the convergence of iteration methods. *USSR Comput. Math. Math. Phys.* **4** (1964), 1–17.

[17] Shehu, Y.; Iyiola, O. S.; Ogbuisi, F. U. Iterative method with inertial terms for nonexpansive mappings: applications to compressed sensing. *Numer. Algorithms* **83** (2020), 1321–1347.

[18] Shehu, Y.; Cai, G.; Iyiola, O. S. Iterative approximation of solutions for proximal split feasibility problems. *Fixed Point Theory Appl.* **2015** (2015), no. 123, 18 pp.

[19] Shehu, Y.; Ogbuisi, F. U. Convergence analysis for proximal split feasibility problems and fixed point problems. *J. Appl. Math. Comput.* **48** (2015), 221–239.

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

[21] Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B Methodol.* **58** (1996), 267–288.

[22] Verma, M.; Shukla, K. K. A new accelerated proximal gradient technique for regularized multitask learning framework. *Pattern Recogn. Lett.* **95** (2017), 98–103.

[23] Wu, Z.; Li, C.; Li, M.; Lim A. Inertial proximal gradient methods with Bregman regularization for a class of nonconvex optimization problems. *J. Global Optim.* **79** (2021), no. 3, 617–644.

[24] Xu, H.K. Another control condition in an iterative method for nonexpansive mappings. *Bull. Austral. Math. Soc.* **65** (2002), 109–113.

[25] Zhang, Y.; Li, X.; Zhao, G.; Cavalcante, C. C. Signal reconstruction of compressed sensing based on alternating direction method of multipliers. *Circuits Syst. Signal Process* **39** (2020), 307–323.

[1]DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE, CHIANG MAI UNIVERSITY
CHIANG MAI 50200, THAILAND
Email address: dawan_c@cmu.ac.th
Email address: suthep.s@cmu.ac.th

[2]KOSEN-KMITL, BANGKOK 10520, THAILAND
Email address: panitarn.sa@kmitl.ac.th