Research article

An accelerated forward-backward algorithm with a new linesearch for convex minimization problems and its applications

Adisak Hanjing, Pachara Jailoka and Suthep Suantai

Research Center in Mathematics and Applied Mathematics, Department of Mathematics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

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

Abstract: We study and investigate a convex minimization problem of the sum of two convex functions in the setting of a Hilbert space. By assuming the Lipschitz continuity of the gradient of the function, many optimization methods have been invented, where the stepsizes of those algorithms depend on the Lipschitz constant. However, finding such a Lipschitz constant is not an easy task in general practice. In this work, by using a new modification of the linesearches of Cruz and Nghia [7] and Kankam et al. [14] and an inertial technique, we introduce an accelerated algorithm without any Lipschitz continuity assumption on the gradient. Subsequently, a weak convergence result of the proposed method is established. As applications, we apply and analyze our method for solving an image restoration problem and a regression problem. Numerical experiments show that our method has a higher efficiency than the well-known methods in the literature.

Keywords: convex minimization problems; forward-backward methods; linesearches; inertial techniques; image restoration problems; regression problems

Mathematics Subject Classification: 65K05, 90C25, 90C30

1. Introduction

Throughout this article, we suppose that $\mathcal{H}$ is a real Hilbert space with an inner product $\langle \cdot, \cdot \rangle$ and the induced norm $\|\cdot\|$. We are interested in studying the following unconstrained convex minimization problem:

$$
\begin{align*}
\text{minimize} \quad & f_1(x) + f_2(x), \\
\text{subject to} \quad & x \in \mathcal{H}
\end{align*}
$$

where $f_1, f_2 : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ are two proper, lower semi-continuous and convex functions such that $f_1$ is differentiable on an open set containing the domain of $f_2$. Problem (1.1) has been widely studied
due to its applications which can be used in various real-world applications such as in signal and image processing, in regression problems, and in classification problems, etc., see [3, 5, 8, 10, 11, 13] and the references therein. One of the important topics of studying Problem (1.1) is to invent some efficient procedures for approximating minimizers of $f_1 + f_2$. Various optimization methods were introduced and developed by many researchers, see [3, 5–8, 12, 14–16, 25–27, 32], for instance.

If a minimizer $x^*$ of $f_1 + f_2$ exists, it is known that $x^*$ is characterized by the fixed point equation of the forward-backward operator

$$x^* = FB_\alpha(x^*) := \text{prox}_{\alpha f_2} (x^* - \alpha \nabla f_1(x^*)), \quad (1.2)$$

where $\alpha > 0$, $\text{prox}_{f_2}$ is the proximity operator of $f_2$ and $\nabla f_1$ stands for the gradient of $f_1$. The above equation leads to the following iterative method:

**Method 1.** Let $x_1 \in \text{dom} \ f_2$. For $k \geq 1$, let

$$x_{k+1} = \text{prox}_{\alpha_k f_2} (x_k - \alpha_k \nabla f_1(x_k)),$$

where $0 < \alpha_k < \frac{2}{L}$ and $L$ is a Lipschitz constant of $\nabla f_1$.

This method is well known as the *forward-backward splitting* algorithm [8, 15], which includes the proximal point algorithm [17,24], the gradient method [4,9] and the $CQ$ algorithm [1] as special cases.

It is observed from Method 1 that we need to assume the Lipschitz continuity condition on the gradient of $f_1$ and the stepsize $\alpha_k$ depends on the Lipschitz constant $L$, which is not an easy task to find in general practice (see [3, 5, 8, 12] for other relevant methods).

In the sequel, we set the standing hypotheses on Problem (1.1) as follows:

**H1** $f_1, f_2 : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ are two proper, lower semi-continuous and convex functions with $\text{dom} \ f_2 \subseteq \text{dom} \ f_1$, and $\text{dom} \ f_2$ is nonempty, closed and convex;

**H2** $f_1$ is differentiable on an open set containing $\text{dom} \ f_2$. The gradient $\nabla f_1$ is uniformly continuous on any bounded subset of $\text{dom} \ f_2$ and maps any bounded subset of $\text{dom} \ f_2$ to a bounded set in $\mathcal{H}$.

We note that the second part of **H2** is a weaker assumption than the Lipschitz continuity assumption on $\nabla f_1$.

Cruz and Nghia [7] proposed a technique for selecting the stepsize $\alpha_k$ which is independent of the Lipschitz constant $L$ by using the following linesearch process.

**Linesearch 1.** Given $x \in \text{dom} \ f_2$, $\sigma > 0$, $\theta \in (0, 1)$ and $\delta > 0$.

**Input** $\alpha = \sigma$.

**While** $\alpha \left\| \nabla f_1(FB_\alpha(x)) - \nabla f_1(x) \right\| > \delta \left\| FB_\alpha(x) - x \right\|$, **do**

\[ \alpha = \theta \alpha. \]

**End**

**Output** $\alpha$. 
Linesearch 1 is a particular case of the linesearch proposed in [29] for inclusion problems and it was shown that this linesearch is well defined, that is, it stops after finitely many steps, see [7, Lemma 3.1] and [29, Theorem 3.4(a)]. Cruz and Nghia [7] employed the forward-backward iteration where the stepsize $\alpha_k$ is generated by Linesearch 1.

**Method 2.** Let $x_1 \in \text{dom } f_2$, $\sigma > 0$, $\theta \in (0, 1)$ and $\delta \in (0, \frac{1}{2})$. For $k \geq 1$, let $$x_{k+1} = \text{prox}_{\alpha_k f_2}(x_k - \alpha_k \nabla f_1(x_k)),$$
where $\alpha_k := \text{Linesearch 1}(x_k, \sigma, \theta, \delta)$.

In optimization theory, to speed up the convergence of iterative methods, many mathematicians often use the inertial-type extrapolation [20, 22] by adding the technical term $\beta_k(x_k - x_{k-1})$. The control parameter $\beta_k$ is called an inertial parameter, which controls the momentum $x_k - x_{k-1}$. Using Linesearch 1, Cruz and Nghia [7] also introduced an accelerated algorithm with the inertial technical term as follows.

**Method 3.** Let $x_0 = x_1 \in \text{dom } f_2$, $\alpha_0 = \sigma > 0$, $\theta \in (0, 1)$, $\delta \in (0, \frac{1}{2})$, and $t_1 = 1$. For $k \geq 1$, let

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \quad \beta_k = \frac{t_k - 1}{t_{k+1}},$$

$$y_k = P_{\text{dom } f_2}(x_k + \beta_k(x_k - x_{k-1})),$$

$$x_{k+1} = \text{prox}_{\alpha_k f_2}(y_k - \alpha_k \nabla f_1(y_k)),$$

where $\alpha_k := \text{Linesearch 1}(y_k, \alpha_{k-1}, \theta, \delta)$.

The technique of choosing $\beta_k$ in Method 3 was first mentioned in the fast iterative shrinkage-thresholding algorithm (FISTA) by Beck and Teboulle [5]. Weak convergence results of Methods 2 and 3 were obtained for solving Problem (1.1) with (H1) and (H2).

Recently, Kankam et al. [14] proposed a modification of Linesearch 1 as follows.

**Linesearch 2.** Given $x \in \text{dom } f_2$, $\sigma > 0$, $\theta \in (0, 1)$ and $\delta > 0$.

**Input** $\alpha = \sigma$.

**While** $\alpha \max \{\|\nabla f_1(FB_\sigma^2(x)) - \nabla f_1(FB_\sigma(x))\|, \|\nabla f_1(FB_\sigma(x)) - \nabla f_1(x)\|\}$

$> \delta \left(\|FB_\sigma^2(x) - FB_\sigma(x)\| + \|FB_\sigma(x) - x\|\right)$

**do**

$\alpha = \theta \alpha$.

**End**

**Output** $\alpha$,

where $FB_\sigma^2(x) := FB_\sigma(FB_\sigma(x))$.

They showed the well-definedness of Linesearch 2 and introduced the following double forward-backward algorithm whose stepsize is generated by Linesearch 2.
Method 4. Let \( x_1 \in \text{dom} f_2, \sigma > 0, \theta \in (0, 1) \) and \( \delta \in (0, \frac{1}{8}) \). For \( k \geq 1 \), let

\[
y_k = \text{prox}_{\alpha_k f_2}(x_k - \alpha_k \nabla f_1(x_k)),
\]

\[
x_{k+1} = \text{prox}_{\alpha_k f_2}(y_k - \alpha_k \nabla f_1(y_k)),
\]

where \( \alpha_k := \text{Linesearch 2}(x_k, \sigma, \theta, \delta) \).

A weak convergence theorem of Method 4 was proved and an application in signal recovery was illustrated, see [14].

In this paper, inspired and motivated by the results of Cruz and Nghia [7] and Kankam et al. [14], and other related researches, we aim to improve Linesearches 1 and 2 and introduce a new accelerated algorithm using our proposed linesearch for the convex minimization problem of the sum of two convex functions. The paper is organized as follows. Basic definitions, notations and some useful tools for proving our convergence results are given in Section 2. Our main result is in Section 3. In this section, we introduce a new modification of Linesearches 1 and 2 and present a double forward-backward algorithm by using an inertial technique for solving Problem (1.1) with Hypotheses (H1) and (H2). After that, a weak convergence theorem of the proposed method is proved. The complexity of our reduced algorithm is also discussed. In Section 4, we apply the convex minimization problem to an image restoration problem and a regression problem. We analyze and illustrate the convergence behavior of our method, and also compare its efficiency with the well-known methods in the literature.

2. Notations and tools

The mathematical symbols used throughout this paper are as follows. \( \mathbb{R}, \mathbb{R}_+ \) and \( \mathbb{N} \) are the set of real numbers, the set of nonnegative real numbers, and the set of positive integers, respectively. \( I_d \) stands for the identity operator on \( \mathcal{H} \). Denote weak and strong convergence of a sequence \( \{x_k\} \subset \mathcal{H} \) to \( x \in \mathcal{H} \) by \( x_k \rightharpoonup x \) and \( x_k \to x \), respectively. The set of all weak-cluster points of \( \{x_k\} \) is denoted by \( \omega_w(x_k) \). If \( C \) is a nonempty closed convex subset of \( \mathcal{H} \), then \( P_C \) stands for the metric projection from \( \mathcal{H} \) onto \( C \), i.e., for each \( x \in \mathcal{H} \), \( P_C x \) is the unique element in \( C \) such that \( \|x - P_C x\| = \text{dist}(x, C) := \inf_{y \in C} \|x - y\| \).

Let us recall the concept of the proximity operator which extends the notion of the metric projection. Let \( f : \mathcal{H} \to \mathbb{R} \cup \{\infty\} \) be a proper, lower semi-continuous and convex function. The proximity (or proximal) operator [2, 18] of \( f \), denoted by \( \text{prox}_f \), is defined for each \( x \in \mathcal{H} \), \( \text{prox}_f x \) is the unique solution of the minimization problem:

\[
\min_{y \in \mathcal{H}} f(y) + \frac{1}{2}\|x - y\|^2.
\]  

(2.1)

If \( f := i_C \) is an indicator function on \( C \) (defined by \( i_C(x) = 0 \) if \( x \in C \); otherwise \( i_C(x) = \infty \)), then \( \text{prox}_f = P_C \).

The proximity operator can be formulated in the equivalent form

\[
\text{prox}_f = (I_d + \partial f)^{-1} : \mathcal{H} \to \text{dom} f,
\]  

(2.2)

where \( \partial f \) is the subdifferential of \( f \) defined by

\[
\partial f(x) := \{u \in \mathcal{H} : f(x) + \langle u, y - x \rangle \leq f(y), \forall y \in \mathcal{H}\}, \quad \forall x \in \mathcal{H}.
\]
Moreover, we have the following useful fact:

$$
\frac{x - \text{prox}_{\alpha f}(x)}{\alpha} \in \partial f(\text{prox}_{\alpha f}(x)), \quad \forall x \in \mathcal{H}, \alpha > 0.
$$

(2.3)

The following is a property of the subdifferential operator.

**Lemma 2.1 ([23]).** If $f : \mathcal{H} \to \mathbb{R} \cup \{\infty\}$ is a proper, lower semi-continuous and convex function, then the graph of $\partial f$ defined by $\text{Gph}(\partial f) := \{(x, y) \in \mathcal{H} \times \mathcal{H} : y \in \partial f(x)\}$ is demiclosed, i.e., if the sequence $\{(x_k, y_k)\} \subset \text{Gph}(\partial f)$ satisfies that $x_k \rightharpoonup x$ and $y_k \to y$, then $(x, y) \in \text{Gph}(\partial f)$.

We end this section by providing useful tools for proving our main results.

**Fact.** Let $x, y \in \mathcal{H}$. The following inequalities hold on $\mathcal{H}$:

$$
\|x \pm y\|^2 = \|x\|^2 \pm 2 \langle x, y \rangle + \|y\|^2.
$$

(2.4)

**Lemma 2.2 ([12]).** Let $\{a_k\}$ and $\{t_k\}$ be two sequences of nonnegative real numbers such that

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

Then the following holds

$$
a_{k+1} \leq K \prod_{j=1}^{k} (1 + 2t_j), \quad \text{where } K = \max\{a_1, a_2\}.
$$

Moreover, if $\sum_{k=1}^{\infty} t_k < \infty$, then $\{a_k\}$ is bounded.

**Lemma 2.3 ([31]).** Let $\{a_k\}$ and $\{b_k\}$ be two sequences of nonnegative real numbers such that $a_{k+1} \leq a_k + b_k$ for all $k \in \mathbb{N}$. If $\sum_{k=1}^{\infty} b_k < \infty$, then $\lim_{k \to \infty} a_k$ exists.

**Lemma 2.4 (Opial [19]).** Let $\{x_k\}$ be a sequence in $\mathcal{H}$ such that there exists a nonempty set $\Omega \subset \mathcal{H}$ satisfying:

(i) For every $p \in \Omega$, $\lim_{k \to \infty} \|x_k - p\|$ exists;

(ii) $\omega_p(x_k) \subset \Omega$.

Then, $\{x_k\}$ converges weakly to a point in $\Omega$.

3. Methods and convergence analysis

In this section, using the idea of Linesearches 1 and 2, we introduce a new linesearch and present an inertial double forward-backward algorithm with the proposed linesearch for solving the convex minimization problem of the sum of two convex functions without any Lipschitz continuity assumption on the gradient. A weak convergence result of our proposed method is analyzed and established.

We now focus on Problem (1.1) and assume that $f_1$ and $f_2$ satisfy Hypotheses (H1) and (H2). The solution set of (1.1) is denoted by $\Omega$. Also, suppose that $\Omega \neq \emptyset$. For simplicity, we let $\mathcal{F} := f_1 + f_2$ and denote by $\mathcal{F}_{\alpha}$ the forward-backward operator of $f_1$ and $f_2$ with respect to $\alpha$, that is, $\mathcal{F}_{\alpha} := \text{prox}_{\alpha f_1}(I_d - \alpha \nabla f_1)$. Here, our linesearch is designed as follows.
The loop termination condition of Linesearch 3 is weaker than that of Linesearch 2. Indeed, since
\[
\alpha \left(1 - \mu \right) \left\| \nabla f_1(FB_\sigma^2(x)) - \nabla f_1(FB_\sigma(x)) \right\| + \mu \left\| \nabla f_1(FB_\sigma(x)) - \nabla f_1(x) \right\|
\]
\[
\leq \alpha \max \left\{ \left\| \nabla f_1(FB_\sigma^2(x)) - \nabla f_1(FB_\sigma(x)) \right\|, \left\| \nabla f_1(FB_\sigma(x)) - \nabla f_1(x) \right\| \right\},
\]
it follows from the well-definedness of Linesearch 2 that our linesearch also stops after finitely many steps, see [14, Lemma 3.2] for more details.

Using Linesearch 3, we propose the following iterative method with the inertial technical term.

**Method 5.**

**Initialization:** Take \( x_1 = y_0 \in \text{dom } f_2, \sigma > 0, \theta \in (0, 1), \mu \in (0, 1/2) \) and \( \delta \in (0, \frac{\alpha}{3}) \). Let \( \beta_k \subset \mathbb{R}_+ \).

**Iterative steps:** For \( k \geq 1 \), calculate \( x_{k+1} \) as follows:

**Step 1.** Compute
\[
z_k = FB_\sigma^2(x_k) = \text{prox}_{\alpha_k f_1}(x_k - \alpha_k \nabla f_1(x_k)), \tag{3.1}
\]
\[
y_k = FB_\sigma(x_k) = \text{prox}_{\alpha_k f_2}(z_k - \alpha_k \nabla f_1(z_k)), \tag{3.2}
\]
where \( \alpha_k := \text{Linesearch } 3(x_k, \sigma, \theta, \mu, \delta) \).

**Step 2.** Compute
\[
x_{k+1} = P_{\text{dom } f_1}(y_k + \beta_k (y_k - y_{k-1})). \tag{3.3}
\]

Set \( k := k + 1 \) and return to Step 1.

To verify the convergence of Method 5, the following result is needed.

**Lemma 3.2.** Let \( \{x_k\} \) be the sequence generated by Method 5. For each \( k \geq 1 \) and \( p \in \text{dom } f_2 \), we have
\[
\| x_k - p \|^2 - \| y_k - p \|^2 \geq 2\alpha_k \left[ \mathcal{F}(y_k) + \mathcal{F}(z_k) - 2\mathcal{F}(p) \right] + \left(1 - \frac{4\delta}{\mu}\right) \left(\| x_k - z_k \|^2 + \| z_k - y_k \|^2 \right).
\]

**Proof.** Let \( k \geq 1 \) and \( p \in \text{dom } f_2 \). By (2.3), (3.1) and (3.2), we get
\[
\frac{x_k - z_k}{\alpha_k} - \nabla f_1(x_k) \in \partial f_2(z_k) \quad \text{and} \quad \frac{z_k - y_k}{\alpha_k} - \nabla f_1(z_k) \in \partial f_2(y_k).
\]
Then, by the definition of the subdifferential of \( f_2 \), we have
\[
f_2(p) - f_2(z_k) \geq \frac{x_k - z_k}{\alpha_k} - \nabla f_1(x_k), \quad p - z_k
\]

(3.4)

and
\[
f_2(p) - f_2(y_k) \geq \frac{z_k - y_k}{\alpha_k} - \nabla f_1(z_k), \quad p - y_k
\]

(3.5)

By the assumptions on \( f_1 \) and \( f_2 \), we have the following fact:
\[
f_1(x) - f_1(y) \geq \langle \nabla f_1(y), x - y \rangle, \quad \forall x \in \text{dom} \ f_1, \ \forall y \in \text{dom} \ f_2.
\]

(3.6)

From (3.6), we get
\[
f_1(p) - f_1(x_k) \geq \langle \nabla f_1(x_k), p - x_k \rangle,
\]

(3.7)

and
\[
f_1(p) - f_1(z_k) \geq \langle \nabla f_1(z_k), p - z_k \rangle.
\]

(3.8)

Summing (3.4), (3.5), (3.7) and (3.8) yields
\[
2\mathcal{F}(p) - \mathcal{F}(z_k) - f_2(y_k) - f_1(x_k)
\]
\[
\geq \langle \nabla f_1(x_k), z_k - p \rangle + \langle \nabla f_1(z_k), y_k - p \rangle + \langle \nabla f_1(x_k), p - x_k \rangle + \langle \nabla f_1(z_k), p - z_k \rangle
\]
\[
+ \frac{1}{\alpha_k} \left[ \langle x_k - z_k, p - z_k \rangle + \langle z_k - y_k, p - y_k \rangle \right]
\]
\[
= \langle \nabla f_1(x_k), z_k - x_k \rangle + \langle \nabla f_1(z_k), z_k - x_k \rangle + \langle \nabla f_1(y_k), y_k - z_k \rangle
\]
\[
+ \langle \nabla f_1(z_k) - \nabla f_1(y_k), y_k - z_k \rangle + \frac{1}{\alpha_k} \left[ \langle x_k - z_k, p - z_k \rangle + \langle z_k - y_k, p - y_k \rangle \right]
\]
\[
\geq \langle \nabla f_1(z_k), z_k - x_k \rangle + \langle \nabla f_1(y_k), y_k - z_k \rangle - \| \nabla f_1(x_k) - \nabla f_1(z_k) \| \| z_k - x_k \|
\]
\[-\| \nabla f_1(z_k) - \nabla f_1(y_k) \| \| y_k - z_k \| + \frac{1}{\alpha_k} \left[ \langle x_k - z_k, p - z_k \rangle + \langle z_k - y_k, p - y_k \rangle \right].
\]

Using (3.6) again, the above inequality becomes
\[
2\mathcal{F}(p) - \mathcal{F}(z_k) - f_2(y_k) - f_1(x_k)
\]
\[
\geq f_1(y_k) - f_1(x_k) - \| \nabla f_1(x_k) - \nabla f_1(z_k) \| \| z_k - x_k \| - \| \nabla f_1(z_k) - \nabla f_1(y_k) \| \| y_k - z_k \|
\]
\[
+ \frac{1}{\alpha_k} \left[ \langle x_k - z_k, p - z_k \rangle + \langle z_k - y_k, p - y_k \rangle \right]
\]
\[
\geq f_1(y_k) - f_1(x_k) - \| \nabla f_1(x_k) - \nabla f_1(z_k) \| (\| y_k - z_k \| + \| z_k - x_k \|)
Therefore, we conclude from (3.11)–(3.13) that

\[
\alpha_k (x_k - z_k, z_k - p) + \langle z_k - y_k, y_k - p \rangle
\]

\[
= f_1(y_k) - f_1(x_k) + \frac{1}{\alpha_k} \left[ \langle x_k - z_k, p - z_k \rangle + \langle z_k - y_k, p - y_k \rangle \right]
\]

\[
- (\|\nabla f_1(x_k) - \nabla f_1(z_k)\| + \|\nabla f_1(z_k) - \nabla f_1(y_k)\|) (\|y_k - z_k\| + \|z_k - x_k\|).
\]

(3.9)

Since \( \alpha_k := \text{Linesearch } 3(x_k, \sigma, \theta, \mu, \delta) \), we have

\[
\alpha_k \{(1 - \mu) \|\nabla f_1(y_k) - \nabla f_1(z_k)\| + \mu \|\nabla f_1(z_k) - \nabla f_1(x_k)\| \} \leq \delta (\|y_k - z_k\| + \|z_k - x_k\|).
\]

(3.10)

It follows from (3.9) and (3.10) that

\[
\frac{1}{\alpha_k} \left[ \langle x_k - z_k, z_k - p \rangle + \langle z_k - y_k, y_k - p \rangle \right]
\]

\[
\geq \mathcal{F}(y_k) + \mathcal{F}(z_k) - 2\mathcal{F}(p)
\]

\[
- (\|\nabla f_1(x_k) - \nabla f_1(z_k)\| + \|\nabla f_1(z_k) - \nabla f_1(y_k)\|) (\|y_k - z_k\| + \|z_k - x_k\|)
\]

\[
\geq \mathcal{F}(y_k) + \mathcal{F}(z_k) - 2\mathcal{F}(p)
\]

\[
- \frac{1}{\mu} \{(1 - \mu) \|\nabla f_1(y_k) - \nabla f_1(z_k)\| + \mu \|\nabla f_1(z_k) - \nabla f_1(x_k)\| \} (\|y_k - z_k\| + \|z_k - x_k\|)
\]

\[
\geq \mathcal{F}(y_k) + \mathcal{F}(z_k) - 2\mathcal{F}(p) - \frac{\delta}{\alpha_k \mu} (\|y_k - z_k\|^2 + \|z_k - x_k\|^2)
\]

\[
\geq \mathcal{F}(y_k) + \mathcal{F}(z_k) - 2\mathcal{F}(p) - \frac{2\delta}{\alpha_k \mu} (\|y_k - z_k\|^2 + \|z_k - x_k\|^2).
\]

(3.11)

From (2.4), we get

\[
\langle x_k - z_k, z_k - p \rangle = \frac{1}{2} \left( \|x_k - p\|^2 - \|x_k - z_k\|^2 - \|z_k - p\|^2 \right),
\]

(3.12)

and

\[
\langle z_k - y_k, y_k - p \rangle = \frac{1}{2} \left( \|z_k - p\|^2 - \|z_k - y_k\|^2 - \|y_k - p\|^2 \right).
\]

(3.13)

Therefore, we conclude from (3.11)–(3.13) that

\[
\|x_k - p\|^2 - \|y_k - p\|^2 \geq 2\alpha_k \left[ \mathcal{F}(y_k) + \mathcal{F}(z_k) - 2\mathcal{F}(p) \right] + \left( 1 - \frac{4\delta}{\mu} \right) (\|x_k - z_k\|^2 + \|z_k - y_k\|^2). \]

Now we are ready to prove a convergence result of Method 5, which is the main theorem of this paper.

**Theorem 3.3.** Let \( \{x_k\} \) be the sequence generated by Method 5. If \( \alpha_k \geq \alpha > 0, \beta_k \geq 0 \) for all \( k \in \mathbb{N} \) and \( \sum_{k=1}^{\infty} \beta_k < \infty \), then \( \{x_k\} \) converges weakly to a point in \( \Omega \).

**Proof.** Let \( p^* \in \Omega, \alpha_k \geq \alpha > 0, \beta_k \geq 0 \) for all \( k \in \mathbb{N} \) and \( \sum_{k=1}^{\infty} \beta_k < \infty \). By applying Lemma 3.2, we have

\[
\|x_k - p^*\|^2 - \|y_k - p^*\|^2 \geq 2\alpha_k \left[ \mathcal{F}(y_k) - \mathcal{F}(p^*) + \mathcal{F}(z_k) - \mathcal{F}(p^*) \right] + \left( 1 - \frac{4\delta}{\mu} \right) (\|x_k - z_k\|^2 + \|z_k - y_k\|^2)
\]
we get \( \omega \) that \( \lim_{k \to \infty} \| x_k - z_k \|^2 \geq 0 \), \hspace{1cm} (3.14)

which implies \( \| y_k - p^* \| \leq \| x_k - p^* \| \). \hspace{1cm} (3.15)

Let \( x_0 = x_1 \). By (3.3) and (3.15), we have

\[
\| x_{k+1} - p^* \| = \| P_{\Omega}(y_k + \beta_k(y_k - y_{k-1})) - p^* \|
\leq \| y_k + \beta_k(y_k - y_{k-1}) - p^* \|
\leq \| y_k - p^* \| + \beta_k \| y_k - y_{k-1} \|
\leq (1 + \beta_k) \| y_k - p^* \| + \beta_k \| y_{k-1} - p^* \|,
\forall k \in \mathbb{N}. \hspace{1cm} (3.16)
\]

By Lemma 2.2, we have \( \| x_{k+1} - p^* \| \leq K \cdot \prod_{j=1}^{k} (1 + 2\beta_j) \), where \( K := \max(\| x_1 - p^* \|, \| x_2 - p^* \|) \), and \( \{ x_k \} \) is bounded. Thus, \( \sum_{k=1}^{\infty} \beta_k (\| x_k - p^* \| + \| x_{k-1} - p^* \|) < \infty \). In view of (3.16), it follows from Lemma 2.3 that \( \lim_{k \to \infty} \| x_k - p^* \| \) exists. Also, we have \( \lim_{k \to \infty} \| x_k - p^* \| = \lim_{k \to \infty} \| y_k - p^* \| \). Then, by (3.14), we get

\[
\lim_{k \to \infty} \| x_k - z_k \| = 0. \hspace{1cm} (3.17)
\]

Next, we show that \( \omega_n(x_k) \subset \Omega \). Pick \( \bar{p} \in \omega_n(x_k) \), then there is a subsequence \( \{ x_k \} \) of \( \{ x_k \} \) such that \( x_k \rightharpoonup \bar{p} \). Thus, \( z_k \rightharpoonup \bar{p} \). By (H2), we have

\[
\lim_{k \to \infty} \| \nabla f_1(x_k) - \nabla f_1(z_k) \| = 0. \hspace{1cm} (3.18)
\]

From (2.3), we get

\[
\frac{x_k - z_k}{\alpha_k} + \nabla f_1(x_k) - \nabla f_1(z_k) \in \partial f_2(z_k) + \nabla f_1(z_k) = \partial F(z_k). \hspace{1cm} (3.19)
\]

Here, by \( z_k \rightharpoonup \bar{p} \) and (3.17)–(3.19), it follows from Lemma 2.1 that \( 0 \in \partial F(\bar{p}) \), that is, \( \bar{p} \in \Omega \). So, \( \omega_n(x_k) \subset \Omega \). Using Lemma 2.4, we obtain that \( x_k \rightharpoonup p \) for some \( p \in \Omega \). This completes the proof. \( \square \)

The following method is obtained by reducing the inertial step in Method 5.

\begin{center}
\textbf{Method 6.}
\end{center}

\textbf{Initialization:} Take \( x_1 \in \text{dom} f_2, \sigma > 0, \theta \in (0, 1), \mu \in (0, \frac{1}{2}] \) and \( \delta \in (0, \frac{4}{3}) \).

\textbf{Iterative steps:} For \( k \geq 1 \), compute

\[
z_k = FB_{\alpha_k}(x_k) = \text{prox}_{\alpha_k f_2}(x_k - \alpha_k \nabla f_1(x_k)), \hspace{1cm} (3.20)
\]

\[
x_{k+1} = FB_{\alpha_k}(z_k) = \text{prox}_{\alpha_k f_2}(z_k - \alpha_k \nabla f_1(z_k)), \hspace{1cm} (3.21)
\]

where \( \alpha_k := \text{Linesearch} 3(x_k, \sigma, \theta, \mu, \delta) \).

We also analyze the convergence and the complexity of Method 6.
Theorem 3.4. Let \( \{x_k\} \) be the sequence generated by Method 6. If \( \alpha_k \geq \alpha > 0 \) for all \( k \in \mathbb{N} \), then the following statements hold:

(i) \( \{x_k\} \) converges weakly to a point in \( \Omega \);

(ii) If \( \delta \in (0, \frac{\beta}{8}) \), then \( \{F(x_k)\} \) is decreasing and

\[
F(x_k) - \min_{x \in H} F(x) \leq \frac{[\text{dist}(x_1, \Omega)]^2}{2\alpha k}, \quad \forall k \in \mathbb{N}. \tag{3.22}
\]

Proof. Setting \( \beta_k := 0 \) in Method 5, (i) is obtained directly from Theorem 3.3. To prove (ii), let \( \delta \in (0, \frac{\beta}{8}) \). We first show that \( \{F(x_k)\} \) is decreasing. By Lemma 3.2, we get

\[
\|x_k - p\|^2 - \|x_{k+1} - p\|^2 \geq 2\alpha_k [F(x_{k+1}) - F(z_k)] + \left(1 - \frac{4\delta}{\mu}\right) (\|x_k - z_k\|^2 + \|z_k - x_{k+1}\|^2), \quad \forall p \in \text{dom } f_2. \tag{3.23}
\]

Taking \( p := z_k \) in (3.23), we have

\[
\|x_k - z_k\|^2 - \|x_{k+1} - z_k\|^2 \geq 2\alpha_k [F(x_{k+1}) - F(z_k)] + \left(1 - \frac{4\delta}{\mu}\right) (\|x_k - z_k\|^2 + \|z_k - x_{k+1}\|^2). \tag{3.24}
\]

Also, taking \( p := x_k \) in (3.23), we have

\[
-\|x_{k+1} - x_k\|^2 \geq 2\alpha_k [F(x_{k+1}) + F(z_k) - 2F(x_k)] + \left(1 - \frac{4\delta}{\mu}\right) (\|x_k - z_k\|^2 + \|z_k - x_{k+1}\|^2). \tag{3.25}
\]

Combining (3.24) and (3.25), we get

\[
\|x_k - z_k\|^2 - \|x_{k+1} - z_k\|^2 - \|x_{k+1} - x_k\|^2 \geq 4\alpha_k [F(x_{k+1}) - F(x_k)] + 2 \left(1 - \frac{4\delta}{\mu}\right) (\|x_k - z_k\|^2 + \|z_k - x_{k+1}\|^2),
\]

which implies

\[
F(x_{k+1}) - F(x_k) \leq \frac{1}{4\alpha_k} \left(\left(1 - \frac{8\delta}{\mu}\right) \|x_k - z_k\|^2 + \left(3 - \frac{8\delta}{\mu}\right) \|z_k - x_{k+1}\|^2 + \|x_{k+1} - x_k\|^2\right).
\]

This together with \( \delta \in (0, \frac{\beta}{8}) \) yields \( F(x_{k+1}) \leq F(x_k) \), that is, \( \{F(x_k)\} \) is decreasing. Next, we will show that (3.22) is true. Pick any \( p^* \in \Omega \). From (3.23), we have

\[
\|x_k - p^*\|^2 - \|x_{k+1} - p^*\|^2 \geq 2\alpha_k [F(x_{k+1}) - F(p^*) + F(z_k) - F(p^*)] + \left(1 - \frac{4\delta}{\mu}\right) (\|x_k - z_k\|^2 + \|z_k - x_{k+1}\|^2)
\]

\[
\geq 2\alpha [F(x_{k+1}) - F(p^*)].
\]

Hence

\[
F(x_{i+1}) - F(p^*) \leq \frac{1}{2\alpha} (\|x_i - p^*\|^2 - \|x_{i+1} - p^*\|^2), \quad \forall i \in \mathbb{N}.
\]
It follows that
\[
\sum_{i=1}^{k} \mathcal{F}(x_{i+1}) - k\mathcal{F}(p^*) \leq \frac{1}{2\alpha} \sum_{i=1}^{k} \left( ||x_i - p^*||^2 - ||x_{i+1} - p||^2 \right)
= \frac{1}{2\alpha} \left( ||x_1 - p^*||^2 - ||x_{k+1} - p||^2 \right)
\leq \frac{||x_1 - p^*||^2}{2\alpha}.
\]

By the decreasing property of \{\mathcal{F}(x_k)\}, we get
\[
k[\mathcal{F}(x_k) - \mathcal{F}(p^*)] \leq \sum_{i=1}^{k} \mathcal{F}(x_{i+1}) - k\mathcal{F}(p^*) \leq \frac{||x_1 - p^*||^2}{2\alpha}, \quad \forall p^* \in \Omega. \quad (3.26)
\]

It follows from (3.26) that
\[
\mathcal{F}(x_k) - \min_{x \in H} \mathcal{F}(x) \leq \inf_{p \in \Omega} \frac{||x_1 - p^*||^2}{2ak} = \frac{[\text{dist}(x_1, \Omega)]^2}{2ak}.
\]

We note that the stepsize condition on \{\alpha_k\} in Theorems 3.3 and 3.4 needs the boundedness from below by a positive real number. Next, we show that this condition can be guaranteed by the Lipschitz continuity assumption on the gradient.

Let \( C \subset H \) be a nonempty closed convex set. Recall that an operator \( T : C \to H \) is said to be Lipschitz continuous if there exists \( L > 0 \) such that \( ||Tx - Ty|| \leq L||x - y|| \) for all \( x, y \in C \).

Proposition 3.5. Let \( \{\alpha_k\} \) be the sequence generated by Linesearch 3 of Method 5 (or Method 6). If \( \nabla f_1 \) is Lipschitz continuous on \( dom f_2 \) with a coefficient \( L > 0 \), then \( \alpha_k \geq \min \{\sigma, \frac{\sigma}{L}\} \) for all \( k \in \mathbb{N} \).

Proof. Let \( \nabla f_1 \) be \( L \)-Lipschitz continuous on \( dom f_2 \). From Linesearch 3, we know that \( \alpha_k \leq \sigma \) for all \( k \in \mathbb{N} \). If \( \alpha_k < \sigma \), then \( \alpha_k = \sigma\theta^m \) where \( m_k \) is the smallest positive integer such that
\[
\alpha_k \left( 1 - \mu \right) \left( \nabla f_1 (FB_{\alpha_k}(x_k)) - \nabla f_1 (FB_{\alpha_k}(x_k)) \right) + \mu \left( \nabla f_1 (FB_{\alpha_k}(x_k)) - \nabla f_1 (x_k) \right)
\leq \delta \left( \|FB_{\alpha_k}(x_k) - FB_{\alpha_k}(x_k)\| + \|FB_{\alpha_k}(x_k) - x_k\| \right).
\]

Let \( \tilde{\alpha}_k := \frac{\alpha_k}{\sigma} > 0 \). By the Lipschitz continuity of \( \nabla f_1 \) and the above expression, we have
\[
\tilde{\alpha}_k L \left( \|FB_{\tilde{\alpha}_k}(x_k) - FB_{\tilde{\alpha}_k}(x_k)\| + \|FB_{\tilde{\alpha}_k}(x_k) - x_k\| \right)
\geq \tilde{\alpha}_k \left( \|\nabla f_1 (FB_{\tilde{\alpha}_k}(x_k)) - \nabla f_1 (FB_{\tilde{\alpha}_k}(x_k))\| + \|\nabla f_1 (FB_{\tilde{\alpha}_k}(x_k)) - \nabla f_1 (x_k)\| \right)
\geq \tilde{\alpha}_k \left( 1 - \mu \right) \left( \|\nabla f_1 (FB_{\tilde{\alpha}_k}(x_k)) - \nabla f_1 (FB_{\tilde{\alpha}_k}(x_k))\| + \mu \|\nabla f_1 (FB_{\tilde{\alpha}_k}(x_k)) - \nabla f_1 (x_k)\| \right)
\geq \delta \left( \|FB_{\tilde{\alpha}_k}(x_k) - FB_{\tilde{\alpha}_k}(x_k)\| + \|FB_{\tilde{\alpha}_k}(x_k) - x_k\| \right),
\]

which implies that \( \alpha_k > \frac{\sigma}{L} \). Consequently, \( \alpha_k \geq \min \{\sigma, \frac{\sigma}{L}\} \) for all \( k \in \mathbb{N} \). \( \Box \)

Remark 3.6. It is worth mentioning again that the Lipschitz continuity assumption on the gradient of \( f_1 \) is sufficient for Hypothesis (H2). However, if we assume this assumption further, the calculation of the stepsize \( \alpha_k \) generated by Linesearch 3 is still independent of the Lipschitz constant.
4. Applications

In this section, we apply the convex minimization problem, Problem (1.1), to an image restoration problem and a regression problem. To solve these problems, we analyze and illustrate the convergence behavior of our method (Method 5) and compare its efficiency with Methods 1–4 and 6. We also consider the following method with another linesearch strategy [21]:

**Linesearch 4.** Given \( x \in \text{dom} f_2, \sigma > 0 \), and \( \theta \in (0, 1) \).

**Input** \( \alpha = \sigma \).

While \( F(FB_\alpha(x)) > f_1(x) + <FB_\alpha(x) - x, \nabla f_1(x)> + \frac{1}{2\alpha}||FB_\alpha(x) - x||^2 + f_2(FB_\alpha(x)), \) do

\[ \alpha = \theta \alpha. \]

End

**Output** \( \alpha \).

**Method 7.** Let \( x_1 = y_0 \in \text{dom} f_2, \sigma > 0, \theta \in (0, 1), \rho > 0, \) and \( t_1 = 1 \). For \( k \geq 1 \), let

\[
y_k = \text{prox}_{\alpha_k f_2}(x_k - \alpha_k \nabla f_1(x_k)),
\]

\[
t_{k+1} = \frac{1 + \sqrt{1 + 4\rho_k t_k^2}}{2}, \quad \beta_k = \frac{t_k - 1}{t_{k+1}},
\]

\[
x_{k+1} = y_k + \beta_k(y_k - y_{k-1}),
\]

where \( \alpha_k := \text{Linesearch 4}(x_k, \sigma, \theta) \).

Note that Method 7 is well known as the FISTA with backtracking, see [5, 25].

All experiments and visualizations are done with MATLAB and are performed on a laptop computer (Intel Core-i5/4.00 GB RAM/Windows 8/64-bit).

4.1. Image restoration problems

Many problems rising in image and signal processing, especially the image restoration problem, can be formulated as the following equation:

\[
y = Ax + \varepsilon,
\]

(4.1)

where \( x \in \mathbb{R}^N \) is an original image, \( y \in \mathbb{R}^M \) is the observed image, \( \varepsilon \) is an additive noise and \( A \in \mathbb{R}^{M \times N} \) is the blurring operation. To approximate the original image, we need to minimize the value of \( \varepsilon \) by using the LASSO model [28]:

\[
\min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} ||y - Ax||_2^2 + \lambda ||x||_1 \right\},
\]

(4.2)

where \( \lambda > 0 \) is a regularization parameter, \( || \cdot ||_1 \) is the \( l_1 \)-norm, and \( || \cdot ||_2 \) is the Euclidean norm. In this situation, we apply Problem (1.1) to the LASSO model (4.2) by setting

\[
f_1(x) = \frac{1}{2} ||y - Ax||_2^2 \quad \text{and} \quad f_2(x) = \lambda ||x||_1.
\]
where \( A = RW \), \( R \) is the matrix representing the blur operator and \( W \) is the inverse of a three stage Haar wavelet transform. We use the peak signal-to-noise ratio (PSNR) in decibel (dB) [30] and the structural similarity index metric (SSIM) [33] as the image quality measures, which are formulated as follows:

\[
PSNR(x_k) = 10 \log_{10} \left( \frac{255^2}{MS\,E(k)} \right),
\]

where \( MS\,E(k) = \frac{1}{M} \| x_k - x \|^2 \), \( M \) is the number of image samples, and \( x \) is the original image and

\[
SSIM(x_k, x) = \frac{(2\nu_{x_k}\nu_x + C_1)(2\xi_{x_k} + C_2)}{(\nu_x^2 + \nu_{x_k}^2 + C_1)(\xi_x^2 + \xi_{x_k}^2 + C_2)},
\]

where \( \{\nu_{x_k}, \xi_{x_k}\} \) and \( \{\nu_x, \xi_x\} \) denote the mean intensity and standard deviation set of the deblurring image \( x_k \) and the original image \( x \), respectively, \( \xi_{x_{k\times}} \) denotes its cross correlation, and \( C_1, C_2 \) are small constants value to avoid instability problem when the denominator is too close to zero.

**Example 4.1.** Consider two grayscale images, Cameraman (see Figure 1 (a)) and Boy (see Figure 2 (a)), with size of \( 256 \times 256 \), which are contaminated by Gaussian blur of filter size \( 9 \times 9 \) with standard deviation \( \hat{\sigma} = 4 \), noise \( 10^{-5} \) (see Figure 1 (b) and Figure 2 (b)).

Firstly, we test the efficiency of Method 5 for recovering the Cameraman image at the 300th iteration by choosing various linesearch parameters and different starting points as shown in Table 1.

**Table 1.** The test for recovering the Cameraman image of Method 5.

<table>
<thead>
<tr>
<th>Initial points ( x_1 )</th>
<th>Linesearch parameters</th>
<th>PSNR</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \sigma ) ( \theta ) ( \mu ) ( \delta )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blurred image</td>
<td>5 0.1 0.1 0.02</td>
<td>31.554</td>
<td>0.6193</td>
</tr>
<tr>
<td></td>
<td>1 0.5 0.1 0.02</td>
<td>31.096</td>
<td>0.6081</td>
</tr>
<tr>
<td></td>
<td>0.1 0.9 0.5 0.12</td>
<td>28.721</td>
<td>0.5717</td>
</tr>
<tr>
<td></td>
<td>10 0.9 0.5 0.12</td>
<td>33.051</td>
<td>0.6708</td>
</tr>
<tr>
<td>Random selection</td>
<td>5 0.1 0.1 0.02</td>
<td>30.554</td>
<td>0.6043</td>
</tr>
<tr>
<td></td>
<td>1 0.5 0.1 0.02</td>
<td>29.660</td>
<td>0.5934</td>
</tr>
<tr>
<td></td>
<td>0.1 0.9 0.5 0.12</td>
<td>28.694</td>
<td>0.5640</td>
</tr>
<tr>
<td></td>
<td>10 0.9 0.5 0.12</td>
<td>32.798</td>
<td>0.6554</td>
</tr>
</tbody>
</table>

It is observed from Table 1 that the linesearch parameter \( \sigma \) influences the image recovery performance of Method 5, while choosing the starting point \( x_1 \) has no significant impact on the convergence behavior of our method.

Next, we test and analyze the image recovery performance of Methods 1–7 for recovering the images (Cameraman and Boy) by setting the parameters as in Table 2 and by choosing the blurred images as the starting points. The maximum iteration number for all methods is fixed at 500. The comparative experiments for recovering the Cameraman and Boy images are as follows. Original images, blurry image contaminated by Gaussian blur and restored images by Methods 1–7 are shown in Figures 1 and 2. The PSNR and SSIM results are shown in Figures 3 and 4. It can be seen that
Method 5 gives the higher values of PSNR and SSIM than the others, meanwhile Method 6 gives the better results than the other methods without the inertial step. Therefore, our method has the highest image recovery efficiency comparing with the studied methods.

### Table 2. The parameters for the deblurring methods.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = 10^{-5}$</td>
<td>$\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{}}}}}$</td>
</tr>
<tr>
<td>$\alpha_k = \frac{k}{(k+1)L}, L = \lambda_{\text{max}}(A^\top A)$</td>
<td>$\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{}}}}}$</td>
</tr>
<tr>
<td>$\sigma = 10, \theta = 0.9, \delta = 0.1$</td>
<td>$\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{}}}}}$</td>
</tr>
<tr>
<td>$\mu = 0.5$</td>
<td>$\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{}}}}}$</td>
</tr>
<tr>
<td>$\beta_k = \begin{cases} \frac{k}{k+1} &amp; \text{if } 1 \leq k \leq 500 \ \frac{1}{2} &amp; \text{otherwise} \end{cases}$</td>
<td>$\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{}}}}}$</td>
</tr>
<tr>
<td>$\bar{\sigma} = 10, \bar{\theta} = 0.9, \rho_k = 1$</td>
<td>$\sqrt{\sqrt{\sqrt{\sqrt{\sqrt{}}}}}$</td>
</tr>
</tbody>
</table>

**Figure 1.** Restoration for Cameraman image. (c)–(g): Restored images by the deblurring methods.
Figure 2. Restoration for Boy image. (c)–(i): Restored images by the deblurring methods.

Figure 3. Plot of PSNR and SSIM results by the deblurring methods.
The main goal of ELM is to find $\xi = [\xi_1^\top, \ldots, \xi_M^\top]^\top$ such that

$$H \xi = T, \quad (35)$$

where $H$ is defined by

$$H = \begin{bmatrix}
G(a_1, b_1, x_1) & \cdots & G(a_M, b_M, x_1) \\
\vdots & \ddots & \vdots \\
G(a_1, b_1, x_N) & \cdots & G(a_M, b_M, x_N)
\end{bmatrix}_{N \times M}.$$

The main goal of ELM is to find $\xi = [\xi_1^\top, \ldots, \xi_M^\top]^\top$ such that

$$H \xi = T, \quad (4.3)$$

where $T = [t_1^\top, \ldots, t_N^\top]^\top$. From (4.3), we can estimate the weight $\xi$ by $\xi = H^\dagger T$ where $H^\dagger = (H^\top H)^{-1}H^\top$ is the pseudo-inverse matrix of $H$. 

Figure 4. The comparison of PSNR and SSIM values for blurred images and restored images by the deblurring methods.

4.2. Regression problems

We first introduce a brief concept of extreme learning machine (ELM) for regression problems. Let $S = \{(x_k, t_k) : x_k \in \mathbb{R}^n, t_k \in \mathbb{R}^m, k = 1, 2, \ldots, N\}$ be a training set of $N$ distinct samples where $x_k$ is an input data and $t_k$ is a target. For any single hidden layer of ELM, the output of the $i$-th hidden node is

$$h_i(x) = G(a_i, b_i, x),$$

where $G$ is an activation function, $a_i$ and $b_i$ are parameters of the $i$-th hidden node. The output function of ELM for a single-hidden layer feedforward neural networks (SLFNs) with $M$ hidden nodes is

$$O_k = \sum_{i=1}^{M} \xi_i h_i(x_k),$$

where $\xi_i$ is the output weight of the $i$-th hidden node. The hidden layer output matrix $H$ is defined by

$$H = \begin{bmatrix}
G(a_1, b_1, x_1) & \cdots & G(a_M, b_M, x_1) \\
\vdots & \ddots & \vdots \\
G(a_1, b_1, x_N) & \cdots & G(a_M, b_M, x_N)
\end{bmatrix}_{N \times M}.$$
Example 4.2. This example aims to predict the sine function by using the ELM method defined as follows:

- Create randomly 10 distinct points \( y_1, y_2, ..., y_{10} \in [-4, 4] \);
- Create a training matrix \( R = [y_1 \ y_2 \ ... \ y_{10}]^T \) and a training target matrix \( S = [\sin(y_1) \ \sin(y_2) \ ... \ \sin(y_{10})]^T \);
- Create a testing matrix \( V = [-4 - 3.99 - 3.98 ... 4] \) and a target matrix \( T = [\sin(-4) \ \sin(-3.99) \ ... \ \sin(y_{10})]^T \);
- Using sigmoid as an activation function generates the hidden layer output matrix \( H_1 \) of the training matrix \( R \) with the hidden node \( M = 100 \);
- Choose a regularization parameter \( \lambda = 10^{-5} \). Formulate a convex minimization problem via the LASSO model that
  \[
  f_1(\xi) = \frac{1}{2} \|H_1 \xi - S\|_2^2 \quad \text{and} \quad f_2(\xi) = \lambda \|\xi\|_1,
  \]
  and find the optimal weight \( \xi_k \) of this problem by employing Methods 1–5 with certain number of iterations;
- Using sigmoid as an activation function generates the hidden layer output matrix \( H_2 \) of the testing matrix \( V \) with the hidden node \( M = 100 \);
- Calculate the output \( O_k = H_2 \xi_k \) and calculate the mean squared error (MSE) at \( k \)-th iteration by
  \[
  MSE(k) = \frac{1}{N} \|O_k - T\|_2^2,
  \]
  where \( N \) is the number of distinct samples.

The parameters for each prediction method are set as in Table 3.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Parameters</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha_k = \frac{k}{(k+1)^2} ), ( L = \lambda_{\max}(H^T H) )</td>
<td>\checkmark</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \sigma = 0.1, \ \theta = 0.49, \ \delta = 0.1 )</td>
<td>-</td>
<td>\checkmark</td>
<td>\checkmark</td>
<td>\checkmark</td>
<td>\checkmark</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>( \mu = 0.5 )</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>\checkmark</td>
<td>\checkmark</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
| \( \beta_k = \begin{cases} 
\frac{k}{k+1} & \text{if } 1 \leq k \leq 10000 \\
\frac{1}{2} & \text{otherwise} 
\end{cases} \) | - | - | - | \checkmark | \checkmark | - | - | - |
| \( \bar{\sigma} = 10^{-5}, \ \bar{\theta} = 0.49, \ \rho_k = 1 \) | - | - | - | - | - | - | \checkmark | - |

Now a graph of predicting the sine function by Methods 1–7 at the 500th iteration is shown in Figure 5. We show a comparative result of iteration numbers, mean squared error values and times in Table 4 when the stopping criterion is set as: \( MSE \leq 10^{-3} \) or at the 10000th iteration. Also, a graph of the mean squared error is shown in Figure 6. It can be observed that Method 5 gives a better performance for predicting the sine function than the other tested methods.
Figure 5. The regression for the sine function by the prediction methods at the 500th iteration.

Table 4. The comparative result by the prediction methods.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Iteration No.</th>
<th>Time (s)</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10000</td>
<td>0.7986</td>
<td>0.0408</td>
</tr>
<tr>
<td>2</td>
<td>10000</td>
<td>12.1314</td>
<td>0.0032</td>
</tr>
<tr>
<td>3</td>
<td>3999</td>
<td>1.3463</td>
<td>$9.9995 \times 10^{-4}$</td>
</tr>
<tr>
<td>4</td>
<td>10000</td>
<td>17.7834</td>
<td>0.0014</td>
</tr>
<tr>
<td>5</td>
<td>338</td>
<td>0.6531</td>
<td>$9.9239 \times 10^{-4}$</td>
</tr>
<tr>
<td>6</td>
<td>10000</td>
<td>18.3197</td>
<td>0.0014</td>
</tr>
<tr>
<td>7</td>
<td>10000</td>
<td>3.0340</td>
<td>0.0067</td>
</tr>
</tbody>
</table>

Figure 6. Plot of the MSE value by the prediction methods.
5. Conclusions

In this paper, we discuss the convex minimization problem of the sum of two convex functions in a Hilbert space. The challenge of removing the Lipschitz continuity assumption on the gradient of the function attracts us to study the concept of the linesearch process. We introduce a new linesearch and propose an inertial forward-backward algorithm whose stepsize does not depend on any Lipschitz constant for solving the considered problem without any Lipschitz continuity condition on the gradient. It is shown that the sequence generated by our proposed method converges weakly to a minimizer of the sum of those two convex functions under some mild control conditions. As applications, we employ our method to recover blurry images in an image restoration problem and predict the sine function in a regression problem. The results of the experiments show that our method has a higher efficiency than the well-known methods in [5, 7, 14, 15, 25].

Acknowledgments

This research was supported by Thailand Science Research and Innovation under the project IRN62W0007 and Chiang Mai University.

Conflict of interest

The authors declare no conflict of interest.

References


© 2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)