

**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:

$$\begin{aligned} & \text{minimize} && f_1(x) + f_2(x), \\ & \text{subject to} && x \in \mathcal{H} \end{aligned} \tag{1.1}$$

where $f_1, f_2 : \mathcal{H} \rightarrow \mathbb{R} \cup \{\infty\}$ 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

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

where $\alpha > 0$, prox_{f_2} is the proximity operator of f_2 and ∇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 ∇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 α_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) $f_1, f_2 : \mathcal{H} \rightarrow \mathbb{R} \cup \{\infty\}$ 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 ∇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 ∇f_1 .

Cruz and Nghia [7] proposed a technique for selecting the stepsize α_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 $\alpha \|\nabla f_1(FB_\alpha(x)) - \nabla f_1(x)\| > \delta \|FB_\alpha(x) - x\|$, **do**

$\alpha = \theta \alpha$.

End

Output α .

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 α_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 β_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

$$\begin{aligned} t_{k+1} &= \frac{1 + \sqrt{1 + 4t_k^2}}{2}, \quad \beta_k = \frac{t_k - 1}{t_{k+1}}, \\ y_k &= P_{\text{dom } f_2}(x_k + \beta_k(x_k - x_{k-1})), \\ x_{k+1} &= \text{prox}_{\alpha_k f_2}(y_k - \alpha_k \nabla f_1(y_k)), \end{aligned}$$

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

The technique of choosing β_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 $\alpha \max \left\{ \|\nabla f_1(FB_\alpha^2(x)) - \nabla f_1(FB_\alpha(x))\|, \|\nabla f_1(FB_\alpha(x)) - \nabla f_1(x)\| \right\}$
 $> \delta \left(\|FB_\alpha^2(x) - FB_\alpha(x)\| + \|FB_\alpha(x) - x\| \right)$, **do**
 $\alpha = \theta\alpha$.

End

Output α ,

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

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

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

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 $x_k \rightarrow x$, 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 $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{\infty\}$ be a proper, lower semi-continuous and convex function. The *proximity* (or *proximal*) operator [2, 18] of f , denoted by prox_f is defined for each $x \in \mathcal{H}$, $\text{prox}_f x$ is the unique solution of the minimization problem:

$$\underset{y \in \mathcal{H}}{\text{minimize}} \quad f(y) + \frac{1}{2} \|x - y\|^2. \quad (2.1)$$

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

$$\text{prox}_f = (I_d + \partial f)^{-1} : \mathcal{H} \rightarrow \text{dom } f, \quad (2.2)$$

where ∂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:

$$\frac{x - \text{prox}_{\alpha f}(x)}{\alpha} \in \partial f(\text{prox}_{\alpha f}(x)), \quad \forall x \in \mathcal{H}, \alpha > 0. \quad (2.3)$$

The following is a property of the subdifferential operator.

Lemma 2.1 ([23]). *If $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{\infty\}$ is a proper, lower semi-continuous and convex function, then the graph of ∂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 $y_k \rightarrow y$, 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} :

$$\|x \pm y\|^2 = \|x\|^2 \pm 2\langle x, y \rangle + \|y\|^2. \quad (2.4)$$

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

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

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 $\lim_{k \rightarrow \infty} a_k$ 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 $p \in \Omega$, $\lim_{k \rightarrow \infty} \|x_k - p\|$ exists;*
- (ii) *$\omega_w(x_k) \subset \Omega$.*

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

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 Ω . Also, suppose that $\Omega \neq \emptyset$. For simplicity, we let $\mathcal{F} := f_1 + f_2$ and denote by FB_{α} the forward-backward operator of f_1 and f_2 with respect to α , that is, $FB_{\alpha} := \text{prox}_{\alpha f_2}(I_d - \alpha \nabla f_1)$. Here, our linesearch is designed as follows.

Linesearch 3. Given $x \in \text{dom } f_2$, $\sigma > 0$, $\theta \in (0, 1)$, $\mu \in (0, 1)$ and $\delta > 0$.

Input $\alpha = \sigma$.

While $\alpha \left\{ (1 - \mu) \|\nabla f_1(FB_\alpha^2(x)) - \nabla f_1(FB_\alpha(x))\| + \mu \|\nabla f_1(FB_\alpha(x)) - \nabla f_1(x)\| \right\}$
 $> \delta \left(\|FB_\alpha^2(x) - FB_\alpha(x)\| + \|FB_\alpha(x) - x\| \right)$, **do**
 $\alpha = \theta\alpha$.

End

Output α .

Remark 3.1. The loop termination condition of Linesearch 3 is weaker than that of Linesearch 2. Indeed, since

$$\begin{aligned} & \alpha \left\{ (1 - \mu) \|\nabla f_1(FB_\alpha^2(x)) - \nabla f_1(FB_\alpha(x))\| + \mu \|\nabla f_1(FB_\alpha(x)) - \nabla f_1(x)\| \right\} \\ & \leq \alpha \max \left\{ \|\nabla f_1(FB_\alpha^2(x)) - \nabla f_1(FB_\alpha(x))\|, \|\nabla f_1(FB_\alpha(x)) - \nabla f_1(x)\| \right\}, \end{aligned}$$

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 $x_1 = y_0 \in \text{dom } f_2$, $\sigma > 0$, $\theta \in (0, 1)$, $\mu \in (0, \frac{1}{2}]$ and $\delta \in (0, \frac{\mu}{4})$. Let $\{\beta_k\} \subset \mathbb{R}_+$.

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

Step 1. Compute

$$z_k = FB_{\alpha_k}(x_k) = \text{prox}_{\alpha_k f_2}(x_k - \alpha_k \nabla f_1(x_k)), \quad (3.1)$$

$$y_k = FB_{\alpha_k}(z_k) = \text{prox}_{\alpha_k f_2}(z_k - \alpha_k \nabla f_1(z_k)), \quad (3.2)$$

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

Step 2. Compute

$$x_{k+1} = P_{\text{dom } f_2}(y_k + \beta_k(y_k - y_{k-1})). \quad (3.3)$$

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

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

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

$$\begin{aligned} f_2(p) - f_2(z_k) &\geq \left\langle \frac{x_k - z_k}{\alpha_k} - \nabla f_1(x_k), p - z_k \right\rangle \\ &= \frac{1}{\alpha_k} \langle x_k - z_k, p - z_k \rangle + \langle \nabla f_1(x_k), z_k - p \rangle \end{aligned} \quad (3.4)$$

and

$$\begin{aligned} f_2(p) - f_2(y_k) &\geq \left\langle \frac{z_k - y_k}{\alpha_k} - \nabla f_1(z_k), p - y_k \right\rangle \\ &= \frac{1}{\alpha_k} \langle z_k - y_k, p - y_k \rangle + \langle \nabla f_1(z_k), y_k - p \rangle. \end{aligned} \quad (3.5)$$

By the assumptions on f_1 and f_2 , we have the following fact:

$$f_1(x) - f_1(y) \geq \langle \nabla f_1(y), x - y \rangle, \quad \forall x \in \text{dom } f_1, \forall y \in \text{dom } f_2. \quad (3.6)$$

From (3.6), we get

$$f_1(p) - f_1(x_k) \geq \langle \nabla f_1(x_k), p - x_k \rangle, \quad (3.7)$$

and

$$f_1(p) - f_1(z_k) \geq \langle \nabla f_1(z_k), p - z_k \rangle. \quad (3.8)$$

Summing (3.4), (3.5), (3.7) and (3.8) yields

$$\begin{aligned} &2\mathcal{F}(p) - \mathcal{F}(z_k) - f_2(y_k) - f_1(x_k) \\ &\geq \langle \nabla f_1(x_k), z_k - p \rangle + \langle \nabla f_1(z_k), y_k - p \rangle + \langle \nabla f_1(x_k), p - x_k \rangle + \langle \nabla f_1(z_k), p - z_k \rangle \\ &\quad + \frac{1}{\alpha_k} [\langle x_k - z_k, p - z_k \rangle + \langle z_k - y_k, p - y_k \rangle] \\ &= \langle \nabla f_1(x_k), z_k - x_k \rangle + \langle \nabla f_1(z_k), y_k - z_k \rangle + \frac{1}{\alpha_k} [\langle x_k - z_k, p - z_k \rangle + \langle z_k - y_k, p - y_k \rangle] \\ &= \langle \nabla f_1(x_k) - \nabla f_1(z_k), z_k - x_k \rangle + \langle \nabla f_1(z_k), z_k - x_k \rangle + \langle \nabla f_1(y_k), y_k - z_k \rangle \\ &\quad + \langle \nabla f_1(z_k) - \nabla f_1(y_k), y_k - z_k \rangle + \frac{1}{\alpha_k} [\langle x_k - z_k, p - z_k \rangle + \langle z_k - y_k, p - y_k \rangle] \\ &\geq \langle \nabla f_1(z_k), z_k - x_k \rangle + \langle \nabla f_1(y_k), y_k - z_k \rangle - \|\nabla f_1(x_k) - \nabla f_1(z_k)\| \|z_k - x_k\| \\ &\quad - \|\nabla f_1(z_k) - \nabla f_1(y_k)\| \|y_k - z_k\| + \frac{1}{\alpha_k} [\langle x_k - z_k, p - z_k \rangle + \langle z_k - y_k, p - y_k \rangle]. \end{aligned}$$

Using (3.6) again, the above inequality becomes

$$\begin{aligned} &2\mathcal{F}(p) - \mathcal{F}(z_k) - f_2(y_k) - f_1(x_k) \\ &\geq f_1(y_k) - f_1(x_k) - \|\nabla f_1(x_k) - \nabla f_1(z_k)\| \|z_k - x_k\| - \|\nabla f_1(z_k) - \nabla f_1(y_k)\| \|y_k - z_k\| \\ &\quad + \frac{1}{\alpha_k} [\langle x_k - z_k, p - z_k \rangle + \langle z_k - y_k, p - y_k \rangle] \\ &\geq f_1(y_k) - f_1(x_k) - \|\nabla f_1(x_k) - \nabla f_1(z_k)\| (\|y_k - z_k\| + \|z_k - x_k\|) \end{aligned}$$

$$\begin{aligned}
& - \|\nabla f_1(z_k) - \nabla f_1(y_k)\| (\|y_k - z_k\| + \|z_k - x_k\|) + \frac{1}{\alpha_k} [\langle x_k - z_k, p - z_k \rangle + \langle z_k - y_k, p - y_k \rangle] \\
& = f_1(y_k) - f_1(x_k) + \frac{1}{\alpha_k} [\langle x_k - z_k, p - z_k \rangle + \langle z_k - y_k, p - y_k \rangle] \\
& \quad - (\|\nabla f_1(x_k) - \nabla f_1(z_k)\| + \|\nabla f_1(z_k) - \nabla f_1(y_k)\|) (\|y_k - z_k\| + \|z_k - x_k\|). \tag{3.9}
\end{aligned}$$

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

$$\alpha_k \{(1 - \mu) \|\nabla f_1(y_k) - \nabla f_1(z_k)\| + \mu \|\nabla f_1(z_k) - \nabla f_1(x_k)\|\} \leq \delta (\|y_k - z_k\| + \|z_k - x_k\|). \tag{3.10}$$

It follows from (3.9) and (3.10) that

$$\begin{aligned}
& \frac{1}{\alpha_k} [\langle x_k - z_k, z_k - p \rangle + \langle z_k - y_k, y_k - p \rangle] \\
& \geq \mathcal{F}(y_k) + \mathcal{F}(z_k) - 2\mathcal{F}(p) \\
& \quad - (\|\nabla f_1(x_k) - \nabla f_1(z_k)\| + \|\nabla f_1(z_k) - \nabla f_1(y_k)\|) (\|y_k - z_k\| + \|z_k - x_k\|) \\
& \geq \mathcal{F}(y_k) + \mathcal{F}(z_k) - 2\mathcal{F}(p) \\
& \quad - \frac{1}{\mu} \{(1 - \mu) \|\nabla f_1(y_k) - \nabla f_1(z_k)\| + \mu \|\nabla f_1(z_k) - \nabla f_1(x_k)\|\} (\|y_k - z_k\| + \|z_k - x_k\|) \\
& \geq \mathcal{F}(y_k) + \mathcal{F}(z_k) - 2\mathcal{F}(p) - \frac{\delta}{\alpha_k \mu} (\|y_k - z_k\| + \|z_k - x_k\|)^2 \\
& \geq \mathcal{F}(y_k) + \mathcal{F}(z_k) - 2\mathcal{F}(p) - \frac{2\delta}{\alpha_k \mu} (\|y_k - z_k\|^2 + \|z_k - x_k\|^2). \tag{3.11}
\end{aligned}$$

From (2.4), we get

$$\langle x_k - z_k, z_k - p \rangle = \frac{1}{2} (\|x_k - p\|^2 - \|x_k - z_k\|^2 - \|z_k - p\|^2), \tag{3.12}$$

and

$$\langle z_k - y_k, y_k - p \rangle = \frac{1}{2} (\|z_k - p\|^2 - \|z_k - y_k\|^2 - \|y_k - p\|^2). \tag{3.13}$$

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

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

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 Ω .*

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

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

$$\geq \left(1 - \frac{4\delta}{\mu}\right) \|x_k - z_k\|^2 \geq 0, \quad (3.14)$$

which implies

$$\|y_k - p^*\| \leq \|x_k - p^*\|. \quad (3.15)$$

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

$$\begin{aligned} \|x_{k+1} - p^*\| &= \|P_{\text{dom } f_2}(y_k + \beta_k(y_k - y_{k-1})) - p^*\| \\ &\leq \|y_k + \beta_k(y_k - y_{k-1}) - p^*\| \\ &\leq \|y_k - p^*\| + \beta_k \|y_k - y_{k-1}\| \\ &\leq (1 + \beta_k) \|y_k - p^*\| + \beta_k \|y_{k-1} - p^*\| \\ &\leq (1 + \beta_k) \|x_k - p^*\| + \beta_k \|x_{k-1} - p^*\|, \quad \forall k \in \mathbb{N}. \end{aligned} \quad (3.16)$$

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 $\lim_{k \rightarrow \infty} \|x_k - p^*\|$ exists. Also, we have $\lim_{k \rightarrow \infty} \|x_k - p^*\| = \lim_{k \rightarrow \infty} \|y_k - p^*\|$. Then, by (3.14), we get

$$\lim_{k \rightarrow \infty} \|x_k - z_k\| = 0. \quad (3.17)$$

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

$$\lim_{k \rightarrow \infty} \|\nabla f_1(x_k) - \nabla f_1(z_k)\| = 0. \quad (3.18)$$

From (2.3), we get

$$\frac{x_{k_i} - z_{k_i}}{\alpha_{k_i}} + \nabla f_1(z_{k_i}) - \nabla f_1(x_{k_i}) \in \partial f_2(z_{k_i}) + \nabla f_1(z_{k_i}) = \partial \mathcal{F}(z_{k_i}). \quad (3.19)$$

Here, by $z_{k_i} \rightharpoonup \bar{p}$ and (3.17)–(3.19), it follows from Lemma 2.1 that $0 \in \partial \mathcal{F}(\bar{p})$, that is, $\bar{p} \in \Omega$. So, $\omega_w(x_k) \subset \Omega$. 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.

Method 6.

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

Iterative steps: For $k \geq 1$, compute

$$z_k = FB_{\alpha_k}(x_k) = \text{prox}_{\alpha_k f_2}(x_k - \alpha_k \nabla f_1(x_k)), \quad (3.20)$$

$$x_{k+1} = FB_{\alpha_k}(z_k) = \text{prox}_{\alpha_k f_2}(z_k - \alpha_k \nabla f_1(z_k)), \quad (3.21)$$

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 Ω ;
- (ii) If $\delta \in (0, \frac{\mu}{8})$, then $\{\mathcal{F}(x_k)\}$ is decreasing and

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

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

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

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

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

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

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

Combining (3.24) and (3.25), we get

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

which implies

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

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

$$\begin{aligned} \|x_k - p^*\|^2 - \|x_{k+1} - p^*\|^2 &\geq 2\alpha_k [\mathcal{F}(x_{k+1}) - \mathcal{F}(p^*) + \mathcal{F}(z_k) - \mathcal{F}(p^*)] \\ &\quad + \left(1 - \frac{4\delta}{\mu}\right) (\|x_k - z_k\|^2 + \|z_k - x_{k+1}\|^2) \\ &\geq 2\alpha [\mathcal{F}(x_{k+1}) - \mathcal{F}(p^*)]. \end{aligned}$$

Hence

$$\mathcal{F}(x_{i+1}) - \mathcal{F}(p^*) \leq \frac{1}{2\alpha} (\|x_i - p^*\|^2 - \|x_{i+1} - p^*\|^2), \quad \forall i \in \mathbb{N}.$$

It follows that

$$\begin{aligned} \sum_{i=1}^k \mathcal{F}(x_{i+1}) - k\mathcal{F}(p^*) &\leq \frac{1}{2\alpha} \sum_{i=1}^k (\|x_i - p^*\|^2 - \|x_{i+1} - p^*\|^2) \\ &= \frac{1}{2\alpha} (\|x_1 - p^*\|^2 - \|x_{k+1} - p^*\|^2) \\ &\leq \frac{\|x_1 - p^*\|^2}{2\alpha}. \end{aligned}$$

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

$$\mathcal{F}(x_k) - \min_{x \in \mathcal{H}} \mathcal{F}(x) \leq \inf_{p^* \in \Omega} \frac{\|x_1 - p^*\|^2}{2\alpha k} = \frac{[\text{dist}(x_1, \Omega)]^2}{2\alpha k}.$$

□

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 $C \subset \mathcal{H}$ be a nonempty closed convex set. Recall that an operator $T : C \rightarrow \mathcal{H}$ 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 ∇f_1 is Lipschitz continuous on $\text{dom } f_2$ with a coefficient $L > 0$, then $\alpha_k \geq \min\{\sigma, \frac{\delta\theta}{L}\}$ for all $k \in \mathbb{N}$.*

Proof. Let ∇f_1 be L -Lipschitz continuous on $\text{dom } f_2$. From Linesearch 3, we know that $\alpha_k \leq \sigma$ for all $k \in \mathbb{N}$. If $\alpha_k < \sigma$, then $\alpha_k = \sigma\theta^{m_k}$ where m_k is the smallest positive integer such that

$$\begin{aligned} \alpha_k \{(1 - \mu) \|\nabla f_1(FB_{\alpha_k}^2(x_k)) - \nabla f_1(FB_{\alpha_k}(x_k))\| + \mu \|\nabla f_1(FB_{\alpha_k}(x_k)) - \nabla f_1(x_k)\| \} \\ \leq \delta (\|FB_{\alpha_k}^2(x_k) - FB_{\alpha_k}(x_k)\| + \|FB_{\alpha_k}(x_k) - x_k\|). \end{aligned}$$

Let $\bar{\alpha}_k := \frac{\alpha_k}{\theta} > 0$. By the Lipschitz continuity of ∇f_1 and the above expression, we have

$$\begin{aligned} \bar{\alpha}_k L (\|FB_{\bar{\alpha}_k}^2(x_k) - FB_{\bar{\alpha}_k}(x_k)\| + \|FB_{\bar{\alpha}_k}(x_k) - x_k\|) \\ \geq \bar{\alpha}_k (\|\nabla f_1(FB_{\bar{\alpha}_k}^2(x_k)) - \nabla f_1(FB_{\bar{\alpha}_k}(x_k))\| + \|\nabla f_1(FB_{\bar{\alpha}_k}(x_k)) - \nabla f_1(x_k)\|) \\ \geq \bar{\alpha}_k \{(1 - \mu) \|\nabla f_1(FB_{\bar{\alpha}_k}^2(x_k)) - \nabla f_1(FB_{\bar{\alpha}_k}(x_k))\| + \mu \|\nabla f_1(FB_{\bar{\alpha}_k}(x_k)) - \nabla f_1(x_k)\| \} \\ > \delta (\|FB_{\bar{\alpha}_k}^2(x_k) - FB_{\bar{\alpha}_k}(x_k)\| + \|FB_{\bar{\alpha}_k}(x_k) - x_k\|), \end{aligned}$$

which implies that $\alpha_k > \frac{\delta\theta}{L}$. Consequently, $\alpha_k \geq \min\{\sigma, \frac{\delta\theta}{L}\}$ for all $k \in \mathbb{N}$. □

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 α_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 $x \in \text{dom } f_2$, $\bar{\sigma} > 0$, and $\bar{\theta} \in (0, 1)$.

Input $\alpha = \bar{\sigma}$.

While $\mathcal{F}(FB_\alpha(x)) > f_1(x) + \langle FB_\alpha(x) - x, \nabla f_1(x) \rangle + \frac{1}{2\alpha} \|FB_\alpha(x) - x\|^2 + f_2(FB_\alpha(x))$, **do**
 $\alpha = \bar{\theta}\alpha$.

End

Output α .

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

$$\begin{aligned} y_k &= \text{prox}_{\alpha_k f_2}(x_k - \alpha_k \nabla f_1(x_k)), \\ t_{k+1} &= \frac{1 + \sqrt{1 + 4\rho_k t_k^2}}{2}, \quad \beta_k = \frac{t_k - 1}{t_{k+1}}, \\ x_{k+1} &= y_k + \beta_k(y_k - y_{k-1}), \end{aligned}$$

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

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:

$$y = Ax + \varepsilon, \quad (4.1)$$

where $x \in \mathbb{R}^N$ is an original image, $y \in \mathbb{R}^M$ is the observed image, ε 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 ε by using the LASSO model [28]:

$$\min_{x \in \mathbb{R}^N} \left\{ \frac{1}{2} \|y - Ax\|_2^2 + \lambda \|x\|_1 \right\}, \quad (4.2)$$

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:

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

where $MSE(k) = \frac{1}{M} \|x_k - x\|_2^2$, M is the number of image samples, and x is the original image and

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

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, ξ_{x_kx} 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 256×256 , which are contaminated by Gaussian blur of filter size 9×9 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.

Initial points x_1	Linesearch parameters				PSNR	SSIM
	σ	θ	μ	δ		
Blurred image	5	0.1	0.1	0.02	31.554	0.6193
	1	0.5	0.1	0.02	31.096	0.6081
	0.1	0.9	0.5	0.12	28.721	0.5717
	10	0.9	0.5	0.12	33.051	0.6708
Random selection	5	0.1	0.1	0.02	30.554	0.6043
	1	0.5	0.1	0.02	29.660	0.5934
	0.1	0.9	0.5	0.12	28.694	0.5640
	10	0.9	0.5	0.12	32.798	0.6554

It is observed from Table 1 that the linesearch parameter σ 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.

Parameters	Methods						
	1	2	3	4	5	6	7
$\lambda = 10^{-5}$	✓	✓	✓	✓	✓	✓	✓
$\alpha_k = \frac{k}{(k+1)L}, L = \lambda_{\max}(A^T A)$	✓	—	—	—	—	—	—
$\sigma = 10, \theta = 0.9, \delta = 0.1$	—	✓	✓	✓	✓	—	—
$\mu = 0.5$	—	—	—	—	✓	✓	—
$\beta_k = \begin{cases} \frac{k}{k+1} & \text{if } 1 \leq k \leq 500 \\ \frac{1}{2^k} & \text{otherwise} \end{cases}$	—	—	—	—	✓	—	—
$\bar{\sigma} = 10, \bar{\theta} = 0.9, \rho_k = 1$	—	—	—	—	—	—	✓

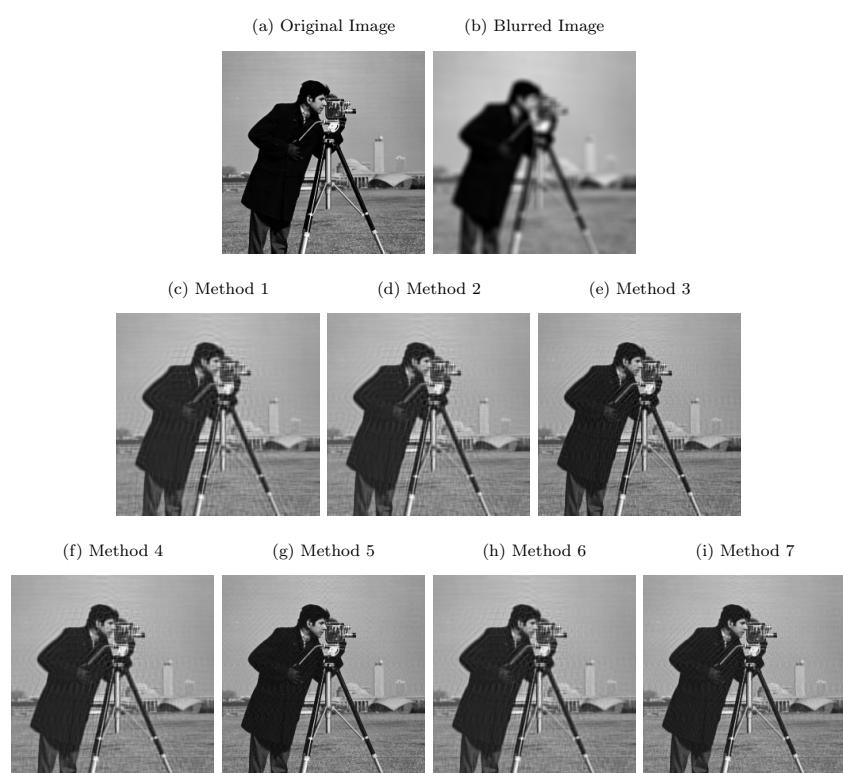


Figure 1. Restoration for Cameraman image. (c)–(i): Restored images by the deblurring methods.

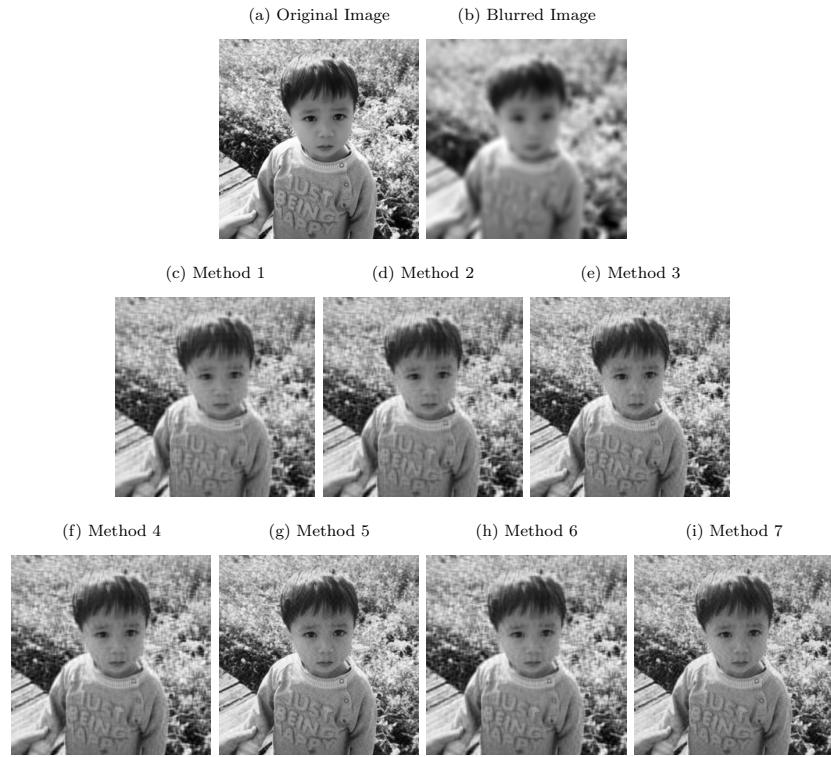


Figure 2. Restoration for Boy image. (c)–(i): Restored images by the deblurring methods.

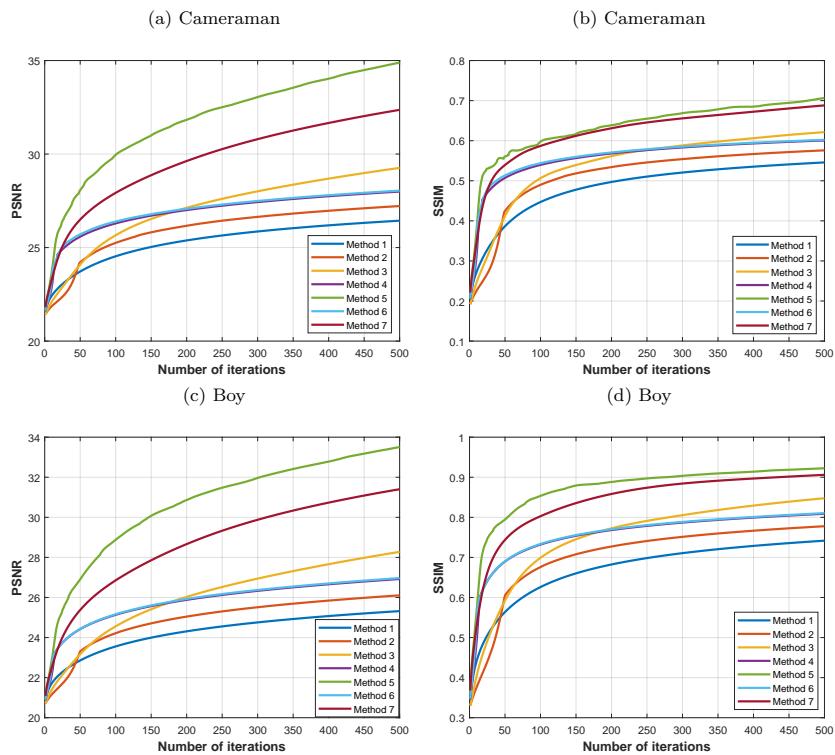


Figure 3. Plot of PSNR and SSIM results by the deblurring methods.

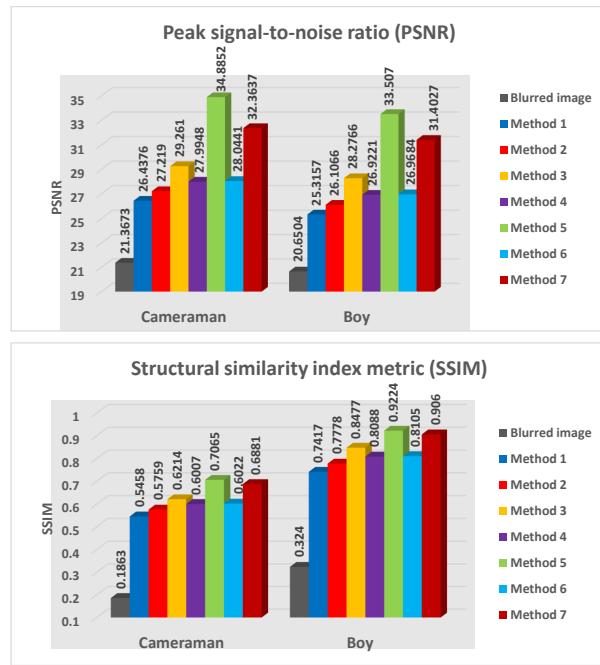


Figure 4. The comparison of PSNR and SSIM values for blurred images and restored images by the deblurring methods.

4.2. Regression problems

We first introduce a brief concept of *extreme learning machine* (ELM) for regression problems. Let $S = \{(x_k, t_k) : x_k \in \mathbb{R}^n, t_k \in \mathbb{R}^m, k = 1, 2, \dots, N\}$ 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 ξ_i is the *output weight* of the i -th hidden node. The hidden layer output matrix H is defined by

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

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

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

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

Example 4.2. This example aims to predict the sine function by using the ELM method defined as follows:

- Create randomly 10 distinct points $y_1, y_2, \dots, y_{10} \in [-4, 4]$;
- Create a training matrix $R = [y_1 \ y_2 \ \dots \ y_{10}]^\top$ and a training target matrix $S = [\sin(y_1) \ \sin(y_2) \ \dots \ \sin(y_{10})]^\top$;
- Create a testing matrix $V = [-4 \ -3.99 \ -3.98 \ \dots \ 4]$ and a target matrix $T = [\sin(-4) \ \sin(-3.99) \ \dots \ \sin(y_{10})]^\top$;
- 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

$$f_1(\xi) = \frac{1}{2} \|H_1\xi - S\|_2^2 \text{ and } f_2(\xi) = \lambda\|\xi\|_1,$$

and find the optimal weight ξ_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 3. The parameters for the prediction methods.

Parameters	Methods						
	1	2	3	4	5	6	7
$\alpha_k = \frac{k}{(k+1)L}$, $L = \lambda_{\max}(H^\top H)$	✓	–	–	–	–	–	–
$\sigma = 0.1$, $\theta = 0.49$, $\delta = 0.1$	–	✓	✓	✓	✓	✓	–
$\mu = 0.5$	–	–	–	–	✓	✓	–
$\beta_k = \begin{cases} \frac{k}{k+1} & \text{if } 1 \leq k \leq 10000 \\ \frac{1}{2^k} & \text{otherwise} \end{cases}$	–	–	–	–	✓	–	–
$\bar{\sigma} = 10^{-5}$, $\bar{\theta} = 0.49$, $\rho_k = 1$	–	–	–	–	–	–	✓

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.

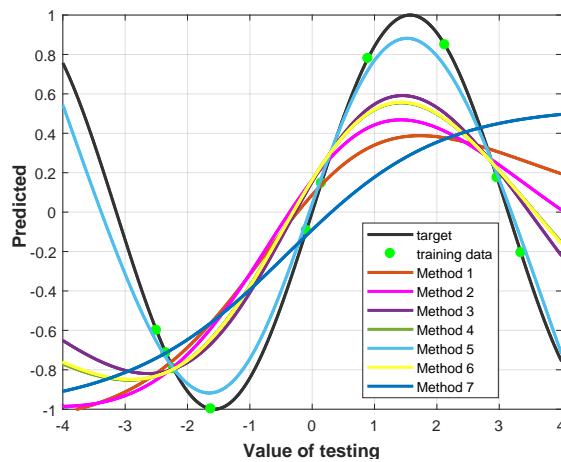


Figure 5. The regression for the sine function by the prediction methods at the 500th iteration.

Table 4. The comparative result by the prediction methods.

Methods	Iteration No.	Time (s)	MSE
1	10000	0.7986	0.0408
2	10000	12.1314	0.0032
3	3999	1.3463	9.9995×10^{-4}
4	10000	17.7834	0.0014
5	338	0.6531	9.9239×10^{-4}
6	10000	18.3197	0.0014
7	10000	3.0340	0.0067

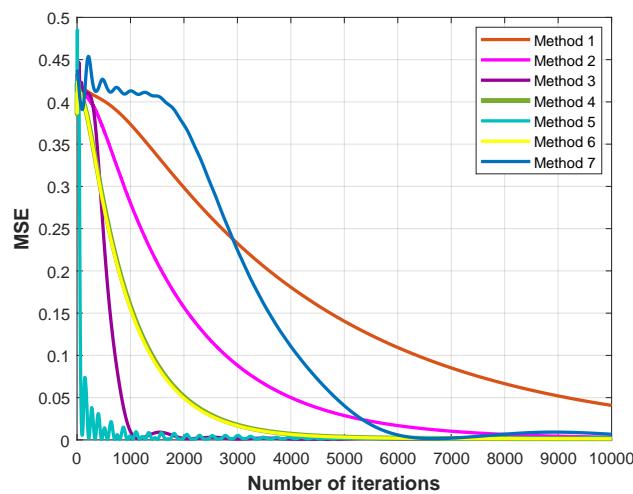


Figure 6. Plot of the MSE value by the prediction methods.

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

1. C. Byrne, Iterative oblique projection onto convex sets and the split feasibility problem, *Inverse Probl.*, **18** (2002), 441–453.
2. H. H. Bauschke, P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, New York: Springer, 2011.
3. L. Bussaban, S. Suantai, A. Kaewkhao, A parallel inertial S-iteration forward-backward algorithm for regression and classification problems, *Carpathian J. Math.*, **36** (2020), 35–44.
4. D. P. Bertsekas, J. N. Tsitsiklis, *Parallel and distributed computation numerical methods*, Belmont: Athena Scientific, 1997.
5. A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.*, **2** (2009), 183–202.
6. P. L. Combettes, J. C. Pesquet, A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery, *IEEE J. STSP*, **1** (2007), 564–574.
7. J. Y. B. Cruz, T. T. A. Nghia, On the convergence of the forward-backward splitting method with linesearchs, *Optim. Method. Softw.*, **31** (2016), 1209–1238.
8. P. L. Combettes, V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.*, **4** (2005), 1168–1200.
9. J. C. Dunn, Convexity, monotonicity, and gradient processes in Hilbert space, *J. Math. Anal. Appl.*, **53** (1976), 145–158.

10. I. Daubechies, M. Defrise, C. D. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Commun. Pure Appl. Math.*, **57** (2004), 1413–1457.
11. M. Figueiredo, R. Nowak, An EM algorithm for wavelet-based image restoration, *IEEE T. Image Process.*, **12** (2003), 906–916.
12. A. Hanjing, S. Suantai, A fast image restoration algorithm based on a fixed point and optimization, *Mathematics*, **8**, (2020), 378.
13. E. Hale, W. Yin, Y. Zhang, *A fixed-point continuation method for l_1 -regularized minimization with applications to compressed sensing*, Rice University: Department of Computational and Applied Mathematics, 2007.
14. K. Kankam, N. Pholasa, P. Cholamjiak, On convergence and complexity of the modified forward-backward method involving new linesearches for convex minimization, *Math. Meth. Appl. Sci.*, **42** (2019), 1352–1362.
15. P. L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.*, **16** (1979), 964–979.
16. L. J. Lin, W. Takahashi, A general iterative method for hierarchical variational inequality problems in Hilbert spaces and applications, *Positivity*, **16** (2012), 429–453.
17. B. Martinet, Régularisation d'inéquations variationnelles par approximations successives, *Rev. Française Informat. Rech. Opér.*, **4** (1970), 154–158.
18. J. J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris Sér. A Math.*, **255** (1962), 2897–2899.
19. A. Moudafi, E. Al-Shemas, Simultaneous iterative methods for split equality problem, *Trans. Math. Program. Appl.*, **1** (2013), 1–11.
20. Y. Nesterov, A method for solving the convex programming problem with convergence rate $O(1/k^2)$, *Dokl. Akad. Nauk SSSR*, **269** (1983), 543–547.
21. Y. Nesterov, *Gradient Methods for Minimizing Composite Objective Function*, CORE Report, 2007. Available from: http://www.ecore.be/DPs/dp_1191313936.pdf.
22. B. T. Polyak, Some methods of speeding up the convergence of iteration methods, *USSR Comput. Math. Math. Phys.*, **4** (1964), 1–17.
23. R. T. Rockafellar, On the maximal monotonicity of subdifferential mappings, *Pac. J. Math.*, **33** (1970), 209–216.
24. R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.*, **17** (1976), 877–898.
25. K. Scheinberg, D. Goldfarb, X. Bai, Fast first order methods for composite convex optimization with backtracking, *Found. Comput. Math.*, **14** (2014), 389–417.
26. S. Suantai, K. Kankam, P. Cholamjiak, A novel forward-backward algorithm for solving convex minimization problem in Hilbert spaces, *Mathematics*, **8** (2020), 42.
27. S. Suantai, P. Jailoka, A. Hanjing, An accelerated viscosity forward-backward splitting algorithm with the linesearch process for convex minimization problems, *J. Inequal. Appl.*, **2021** (2021), 42.

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

29. P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, *SIAM J. Control Optim.*, **38** (2000), 431–446.

30. K. Thung, P. Raveendran, A survey of image quality measures, In: *Proceedings of the International Conference for Technical Postgraduates (TECHPOS)*, Kuala Lumpur, Malaysia, 14–15 December 2009, 1–4.

31. K. Tan, H. K. Xu, Approximating fixed points of nonexpansive mappings by the ishikawa iteration process, *J. Math. Anal. Appl.*, **178** (1993), 301–308.

32. K. Toh, S. Yun. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems, *Pac. J. Optim.*, **6** (2010), 615–640.

33. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE T. Image Process.*, **13** (2004), 600–612.



© 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>)