Results in Nonlinear Analysis **5** (2022) No. 4, 412–422.
https://doi.org/10.53006/rna.1143531
Available online at www.nonlinear-analysis.com

Results in Nonlinear Analysis
ISSN 2636-7556

Peer Reviewed Scientific Journal

A double proximal gradient method with new linesearch for solving convex minimization problem with application to data classification

Suparat Kesornprom[a], Prasit Cholamjiak[a]

[a] *School of Science, University of Phayao, Phayao 56000, Thailand.*

Abstract

In this paper, we propose an inertial double proximal forward-backward method (IDFB) for convex minimization problem in real Hilbert spaces. We suggest a new linesearch that does not require the condition of Lipschitz constant and improve conditions of inertial term to speed up performance of convergence. Moreover, we prove the weak convergence of the proposed method under some suitable conditions. The numerical implementations in data classification from cervical cancer behaviour risk data set are reported to show its efficiency.

Keywords: Minimization problem proximal method linesearch rule inertial method data classification.
2010 MSC: 65K05, 90C25, 90C30.

1. Introduction

Convex minimization theory is critical in the fields of pure and applied mathematics as well as in many other branches of science; see [14, 15, 20, 27]. The convex minimization problem is formulated as follows:

$$\min_{a\in\mathcal{H}}(p(a) + q(a)), \tag{1.1}$$

where $\mathcal{H}$ is a real Hilbert space, $q : \mathcal{H} \to (-\infty, +\infty]$ is proper, lower semicontinuous and convex and $p : \mathcal{H} \to \mathbb{R}$ is convex and differentiable with the Lipschitz continuous gradient. If a^* is a minimizer of (1.1), then it is the solution of (1.1), i.e.,

$$0 \in (\nabla p + \partial q)(a^*),$$

Email addresses: suparat.ke@gmail.com (Suparat Kesornprom), prasitch2008@yahoo.com, prasit.ch@up.ac.th (Prasit Cholamjiak)

Received :July 13, 2022; Accepted: September 4, 2022; Online: September 17, 2022

where ∇p is the gradient of p and ∂q is the subdifferential of q. It is well-known that the minimization problem is related to image processing [1, 6, 7], signal processing [11, 21, 28], machine learning and others, see [8, 9, 29, 30].

Nowadays, methods have been proposed for solving convex minimization problem, see [4, 12, 17, 22, 25]. The method classically used in the field for solving this problem is the forward-backward method (FB) which is defined by:

$$a_n = \text{prox}_{\lambda_n q}(a_n - \lambda \nabla p(a_n)),\ n \geq 1,$$

where the stepsize $\lambda \in (0, 2/\mathcal{L})$, $\mathcal{L}$ is the Lipschitz constant of ∇p and prox_q is the proximal operator of q. Subsequently, the forward-backward method (FB) was modified as well as improved the stepsize to avoid the Lipschitz constant such as linesearch rules and self-adaptive techniques [10, 13, 16, 21, 28, 29, 30]. In addition, the acceleration of FB was studied by adding inertial terms which have the convergence speed at order of time complexity $O(1/n^2)$ with respect to the size of the input as follows:

Method 1.1. A fast iterative shrinkage-thresholding algorithm (FISTA)
Let $s_0 = 1$ and $a_0 = a_1 \in \mathcal{H}$. Compute

$$\begin{aligned} c_n &= a_n + \theta_n(a_n - a_{n-1}) \\ a_{n+1} &= \text{prox}_{\frac{1}{\mathcal{L}}q}(c_n - \frac{1}{\mathcal{L}}\nabla p(c_n)),\ n \geq 1, \end{aligned}$$

where $\theta_n = \frac{s_{n-1}-1}{s_n}$, $s_{n+1} = \frac{1+\sqrt{1+4s_{n-1}^2}}{2}$ and $\mathcal{L}$ is the Lipschitz constant of ∇p. This method was proposed by Beck and Teboulle [3]. In 2016, Cruz and Nghia [4] proposed the method that does not require the condition of Lipschitz constant as follows:

Method 1.2. A fast multistep forward-backward method with linesearch (FMFB)
Let $a_0, a_1 \in \mathcal{H}$, $s_0 = 1$, $\gamma > 0$, $\ell \in (0, 1)$, $\delta \in (0, \frac{1}{2})$ and define

$$\begin{aligned} c_n &= a_n + \theta_n(a_n - a_{n-1}) \\ a_{n+1} &= \text{prox}_{\lambda_n q}(c_n - \lambda_n \nabla p(c_n)),\ n \geq 1, \end{aligned}$$

where $\theta_n = \frac{s_{n-1}-1}{s_n}$, $s_{n+1} = \frac{1+\sqrt{1+4s_{n-1}^2}}{2}$ and $\lambda_n = \gamma\ell^{m_n}$ is the smallest nonnegative integer such that

$$\lambda_n \|\nabla p(\text{prox}_{\lambda_n q}(c_n - \lambda_n \nabla p(c_n))) - \nabla p(c_n)\| \leq \delta \|\text{prox}_{\lambda_n q}(c_n - \lambda_n p(c_n)) - c_n\|.$$

Many effective methods have been proposed to solve the minimization problem. For instance, Kankam et al. [19] proposed two proximal gradient method using linesearch and proved that a convergence rate better than the others. Motivated by this idea, we propose a new forward-backward method with a new linesearch for solving the convex minimization problem. Moreover, we introduce weak convergence theorem under some mild assumptions. Finally, we apply our methods to data classification problem from cervical cancer behaviour risk data set [23].

2. Preliminaries

Let $q : \mathcal{H} \to (-\infty, +\infty]$ be a proper, lower semicontinuous and convex function. We denote the domain of q by $\text{dom}q = \{a \in \mathcal{H} | q(a) < +\infty\}$. For any $a \in \text{dom}q$, the subdifferential of q at a is defined by

$$\partial q(a) = \{v \in \mathcal{H} | \langle v, c - a\rangle \leq q(c) - q(a),\ c \in \mathcal{H}\}.$$

- The proximal operator $\text{prox}_q : \text{dom}(q) \to \mathcal{H}$ is defined by $\text{prox}_q(a) = (I + \partial q)^{-1}(b),\ b \in \mathcal{H}$.
- The proximal operator is single-valued and we have

$$\frac{b - \text{prox}_{\lambda q}(b)}{\lambda} \in \partial q(\text{prox}_{\lambda q}(b)) \qquad \text{for all } b \in \mathcal{H},\ \lambda > 0. \tag{2.1}$$

- A differentiable function p is convex if and only if there holds the inequality

$$p(d) \geq p(a) + \langle \nabla p(a), d - a\rangle,\ \forall d \in \mathcal{H}. \tag{2.2}$$

Definition 2.1. Let Ω be a nonempty subset of $\mathcal{H}$. A sequence $\{a_n\}$ in $\mathcal{H}$ is said to be quasi-Fejér convergent to Ω if and only if for all $a \in \Omega$ there exists a positive sequence $\{\alpha\}$ such that $\sum_{n=0}^{\infty} \alpha_n < +\infty$ and $\|a_{n+1} - a\|^2 \leq \|a_n - a\|^2 + \alpha_n$ for all $n \geq 1$. When $\{\alpha_n\}$ is a null sequence, we say that $\{a_n\}$ is Fejér convergent to Ω.

Lemma 2.1. *[5] The graph of ∂q, $\mathrm{Graph}(\partial q) = \{(a, v) \in \mathcal{H} \times \mathcal{H} : v \in \partial q(a)\}$ is demiclosed, i.e., if the sequence $\{(a_n, v_n)\} \subset \mathrm{Graph}(\partial q)$ satisfies that $\{a_n\}$ converges weakly to a and $\{v_n\}$ converges strongly to v, then $(a, v) \in \mathrm{Graph}(\partial q)$.*

Lemma 2.2. *[24] Let $\{a_n\}$, $\{c_n\}$ and $\{d_n\}$ be real positive sequences such that*

$$a_{n+1} \leq (1 + d_n)a_n + c_n, \ n \geq 1.$$

If $\Sigma_{n=1}^{\infty} d_n < +\infty$ and $\Sigma_{n=1}^{\infty} c_n < +\infty$, then $\lim\limits_{n\to+\infty} a_n$ exists.

Lemma 2.3. *[16] Let $\{a_n\}$ and $\{\theta_n\}$ be real positive sequences such that*

$$a_{n+1} \leq (1 + \theta_n)a_n + \theta_n a_{n-1}, \ n \geq 1.$$

Then, $a_{n+1} \leq K \cdot \prod_{i=1}^{n}(1 + 2\theta_i)$ where $K = \max\{a_1, a_2\}$. Moreover, if $\sum_{n=1}^{\infty} \theta_n < +\infty$, then $\{a_n\}$ is bounded.

Lemma 2.4. *[2, 18] If $\{a_n\}$ is quasi-Fejér convergent to Ω, then we have:*
(i) $\{a_n\}$ is bounded.
(ii) If all weak accumulation points of $\{a_n\}$ is in S, then $\{a_n\}$ weakly converges to a point in Ω.

3. Main results

3.1. An inertial double proximal forward-backward method (IDFB)

In this section, we introduce an inertial double forward-backward method for solving (1.1) with new stepsize as follows:

Method 3.1. An inertial double proximal forward-backward method (IDFB)
Initialization: *Let* $a_0 = a_1 \in \mathcal{H}$, $\theta_1 > 0$, $\gamma > 0$, $\ell \in (0, 1)$ *and* $0 < \mu < 1$.
Iterative step: *For* $n \geq 1$, *calculate* a_{n+1} *as follows:*
Step 1. *Compute the inertial step:*

$$b_n = a_n + \theta_n(a_n - a_{n-1}). \tag{3.1}$$

Step 2. *Compute the forward-backward step:*

$$c_n = prox_{\lambda_n q}(b_n - \lambda_n \nabla p(b_n)).$$

Step 3. *Compute the* a_{n+1} *step:*

$$a_{n+1} = prox_{\lambda_n q}(c_n - \lambda_n \nabla p(c_n))$$

where the linesearch $\lambda_n = \gamma \ell^{m_n}$ *is the smallest nonnegative integer such that*

$$\begin{aligned} & \lambda_n(\langle \nabla p(a_{n+1}) - \nabla p(c_n), a_{n+1} - c_n\rangle + \langle \nabla p(c_n) - \nabla p(b_n), c_n - b_n\rangle) \\ \leq \ & \frac{\mu^2 + 1}{4}\|a_{n+1} - c_n\|^2 + \frac{\mu}{\mu + 1}\|c_n - b_n\|^2. \end{aligned} \tag{3.2}$$

We instance $n = n + 1$ *and go to* **Step 1**.

3.2. Convergence theorems

From Method 3.1, we assume that the following conditions are satisfied for our convergence analysis:
(A1) Define $\Omega = \operatorname{argmin}(p+q) \neq \emptyset$ is the solution set (1.1).
(A2) $p, q : \mathcal{H} \to (-\infty, +\infty]$ are convex functions, two proper and lower semicontinuous.
(A3) The gradient ∇p is uniformly continuous on bounded subset of $\mathcal{H}$.

Lemma 3.1. *Let $a \in \mathcal{H}$, $\gamma > 0$, $\ell \in (0,1)$ and $0 < \mu < 1$. For $i = 1, 2, 3, ...,$ set*

$$\begin{aligned} U(a,i) &= \operatorname{prox}_{\gamma\ell^i q}(a - \gamma\ell^i \nabla p(a)) \\ W(a,i) &= \operatorname{prox}_{\gamma\ell^i q}(U(a,i) - \gamma\ell^i \nabla p(W(a,i))). \end{aligned}$$

If

$$\begin{aligned} &\gamma\ell^i(\langle \nabla p(W(a,i)) - \nabla p(U(a,i)), W(a,i) - U(a,i)\rangle + \langle \nabla p(U(a,i)) - \nabla p(a), U(a,i) - a\rangle) \\ \leq\ &\frac{\mu^2+1}{4}\|W(a,i) - U(a,i)\|^2 + \frac{\mu}{\mu+1}\|U(a,i) - a\|^2, \end{aligned}$$

then $\lambda = \gamma\ell^i$.
Else $i = i + 1$. The linesearch (3.2) stops after finitely many steps.

Proof. If $a \in \Omega$, then $a = \operatorname{prox}_{\gamma q}(a - \gamma\nabla p(a)) = U(a,0)$. It follows that $U(a,0) = a$ and the linesearch stops with zero step, hence $\lambda = \gamma$.

If $a \notin \Omega$, then

$$\begin{aligned} &\gamma\ell^i(\langle \nabla p(W(a,i)) - \nabla p(U(a,i)), W(a,i) - U(a,i)\rangle + \langle \nabla p(U(a,i)) - \nabla p(a), U(a,i) - a\rangle) \\ >\ &\frac{\mu^2+1}{4}\|W(a,i) - U(a,i)\|^2 + \frac{\mu}{\mu+1}\|U(a,i) - a\|^2, \end{aligned}$$

it follows that

$$\begin{aligned} &\gamma\ell^i(\|\nabla p(W(a,i)) - \nabla p(U(a,i))\|\|W(a,i) - U(a,i)\| \\ &+\|\nabla p(U(a,i)) - \nabla p(a)\|\|U(a,i) - a\|) \\ >\ &\frac{\mu^2+1}{4}\|W(a,i) - U(a,i)\|^2 + \frac{\mu}{\mu+1}\|U(a,i) - a\|^2. \end{aligned} \tag{3.3}$$

So we have as $i \to \infty$, $\|W(a,i) - U(a,i)\| \to 0$ and $\|U(a,i) - a\| \to 0$. Since ∇p is uniformly continuous, we get $\|\nabla p(W(a,i)) - \nabla p(U(a,i))\| \to 0$ and $\|\nabla p(U(a,i)) - \nabla p(a)\| \to 0$ as $i \to \infty$. By equation (3.3), we have

$$\frac{\|U(a,i) - a\|}{\gamma\ell^i} \to 0 \text{ as } i \to \infty.$$

We see that

$$\frac{a - \gamma\ell^i\nabla p(a) - U(a,i)}{\gamma\ell^i} \in \partial q(U(a,i)).$$

Hence,

$$\frac{a - U(a,i)}{\gamma\ell^i} \in \partial q(U(a,i)) + \nabla p(a).$$

By Lemma 2.1, we have $0 \in \partial q(a) + \nabla p(a)$. Therefore, $a \in \Omega$ which is a contradiction. □

Next, we have the following theorem.

Theorem 3.1. *Let $\{a_n\}$ be generated by Method 3.1. If $\sum_{n=1}^{\infty} \theta_n < \infty$ and $\lambda_n \geq \lambda$ for some $\lambda > 0$, then $\{a_n\}$ weakly converges to point in Ω.*

Proof. By equation (2.1), we obtain

$$\frac{b_n - c_n}{\lambda_n} - \nabla p(b_n) = \frac{b_n - prox_{\lambda_n g}(b_n - \lambda_n \nabla p(b_n))}{\lambda_n} - \nabla p(b_n) \in \partial q(c_n).$$

By equation (2.2), we get

$$q(a) - q(c_n) \geq \langle \frac{b_n - c_n}{\lambda_n} - \nabla p(b_n), a - c_n \rangle, \ \forall a \in \mathcal{H}. \tag{3.4}$$

Also, we have

$$\frac{c_n - a_{n+1}}{\lambda_n} - \nabla p(c_n) = \frac{c_n - prox_{\lambda_n q}(c_n - \lambda_n \nabla p(c_n))}{\lambda_n} - \nabla p(c_n) \in \partial q(a_{n+1}),$$

from equation (2.2) again, we obtain

$$q(a) - q(a_{n+1}) \geq \langle \frac{c_n - a_{n+1}}{\lambda_n} - \nabla p(c_n), a - a_{n+1} \rangle, \ \forall a \in \mathcal{H}. \tag{3.5}$$

For any $a \in \mathcal{H}$, we have

$$p(a) - p(b_n) \geq \langle \nabla p(b_n), a - b_n \rangle \tag{3.6}$$

and

$$p(a) - p(c_n) \geq \langle \nabla p(c_n), a - c_n \rangle. \tag{3.7}$$

Using equation (2.2) and equation (3.2) and combining equations (3.4), (3.5), (3.6) and (3.7), we obtain

$$\begin{aligned}
& q(a) - q(a_{n+1}) + q(a) - q(c_n) + p(a) - p(b_n) + p(a) - p(c_n) \\
\geq\ & \langle \frac{c_n - a_{n+1}}{\lambda_n} - \nabla p(c_n), a - a_{n+1} \rangle + \langle \frac{b_n - c_n}{\lambda_n} - \nabla p(b_n), a - c_n \rangle \\
& + \langle \nabla p(b_n), a - b_n \rangle + \langle \nabla p(c_n), a - c_n \rangle \\
=\ & \frac{1}{\lambda_n} \langle c_n - a_{n+1}, a - a_{n+1} \rangle + \langle \nabla p(c_n), a_{n+1} - a \rangle + \frac{1}{\lambda_n} \langle b_n - c_n, a - c_n \rangle + \langle \nabla p(b_n), c_n - a \rangle \\
& + \langle \nabla p(b_n), a - b_n \rangle + \langle \nabla p(c_n), a - c_n \rangle \\
=\ & \frac{1}{\lambda_n} \langle c_n - a_{n+1}, a - a_{n+1} \rangle + \frac{1}{\lambda_n} \langle b_n - c_n, a - c_n \rangle \\
& + \langle \nabla p(c_n) - \nabla p(a_{n+1}) + \nabla p(a_{n+1}), a_{n+1} - c_n \rangle + \langle \nabla p(b_n) - \nabla p(c_n) + \nabla p(c_n), c_n - b_n \rangle \\
=\ & \frac{1}{\lambda_n} [\langle c_n - a_{n+1}, a - a_{n+1} \rangle + \langle b_n - c_n, a - c_n \rangle] \\
& - [\langle \nabla p(a_{n+1}) - \nabla p(c_n), a_{n+1} - c_n \rangle + \langle \nabla p(c_n) - \nabla p(b_n), c_n - b_n \rangle] \\
& + \langle \nabla p(a_{n+1}), a_{n+1} - c_n \rangle + \langle \nabla p(c_n), c_n - b_n \rangle \\
\geq\ & \frac{1}{\lambda_n} [\langle c_n - a_{n+1}, a - a_{n+1} \rangle + \langle b_n - c_n, a - c_n \rangle] - [\frac{\mu^2 + 1}{4\lambda_n} \|a_{n+1} - c_n\|^2 \\
& + \frac{\mu}{(\mu + 1)\lambda_n} \|c_n - b_n\|^2] + p(a_{n+1}) - p(c_n) + p(c_n) - p(b_n).
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
& 2\langle c_n - a_{n+1}, a_{n+1} - a \rangle + 2\langle b_n - c_n, c_n - a \rangle \\
\geq\ & 2\lambda_n [q(a_{n+1}) - q(a) + q(c_n) - q(a) - p(a) + p(c_n) - p(a) + p(a_{n+1})] \\
& -2\lambda_n [\frac{\mu^2 + 1}{4\lambda_n} \|a_{n+1} - c_n\|^2 + \frac{\mu}{(\mu + 1)\lambda_n} \|c_n - b_n\|^2] \\
=\ & 2\lambda_n [(p + q)(a_{n+1}) - (p + q)(a) + (p + q)(c_n) - (p + q)(a)] \\
& -[\frac{\mu^2 + 1}{2} \|a_{n+1} - c_n\|^2 + \frac{2\mu}{\mu + 1} \|c_n - b_n\|^2].
\end{aligned} \tag{3.8}$$

We have

$$2\langle c_n - a_{n+1}, a_{n+1} - a\rangle = \|c_n - a\|^2 - \|c_n - a_{n+1}\|^2 - \|a_{n+1} - a\|^2, \tag{3.9}$$

and

$$2\langle b_n - c_n, c_n - a\rangle = \|b_n - a\|^2 - \|b_n - c_n\|^2 - \|c_n - a\|^2. \tag{3.10}$$

By equations (3.8)-(3.10), we have

$$\begin{aligned}
& -\|c_n - a_{n+1}\|^2 - \|a_{n+1} - a\|^2 + \|b_n - a\|^2 - \|b_n - c_n\|^2 \\
\geq\ & 2\lambda_n[(p+q)(a_{n+1}) - (p+q)(a) + (p+q)(c_n) - (p+q)(a)] \\
& -[\frac{\mu^2+1}{2}\|a_{n+1} - c_n\|^2 + \frac{2\mu}{\mu+1}\|c_n - b_n\|^2].
\end{aligned}$$

It follows that

$$\begin{aligned}
\|a_{n+1} - a\|^2 &\leq \|b_n - a\|^2 - \|c_n - a_{n+1}\|^2 - \|b_n - c_n\|^2 \\
&\quad -2\lambda_n[(p+q)(a_{n+1}) - (p+q)(a) + (p+q)(c_n) - (p+q)(a)] \\
&\quad +\frac{\mu^2+1}{2}\|a_{n+1} - c_n\|^2 + \frac{2\mu}{\mu+1}\|c_n - b_n\|^2 \\
&= \|b_n - a\|^2 - (1 - \frac{\mu^2+1}{2})\|c_n - a_{n+1}\|^2 - (1 - \frac{2\mu}{\mu+1})\|b_n - c_n\|^2 \\
&\quad -2\lambda_n[(p+q)(a_{n+1}) - (p+q)(a) + (p+q)(c_n) - (p+q)(a)].
\end{aligned}$$

Setting $a = a^* \in \Omega$ and using $0 < \mu < 1$, we derive

$$\begin{aligned}
\|a_{n+1} - a^*\|^2 &= \|b_n - a^*\|^2 - (1 - \frac{\mu^2+1}{2})\|c_n - a_{n+1}\|^2 - (1 - \frac{2\mu}{\mu+1})\|b_n - c_n\|^2 \\
&\quad -2\lambda_n[(p+q)(a_{n+1}) - (p+q)(a^*) + (p+q)(c_n) - (p+q)(a^*)] \\
&\leq \|b_n - a^*\|^2.
\end{aligned} \tag{3.11}$$

So,

$$\begin{aligned}
\|a_{n+1} - a^*\| &\leq \|b_n - a^*\| \\
&= \|a_n + \theta_n(a_n - a_{n-1}) - a^*\| \\
&\leq \|a_n - a^*\| + \theta_n(\|a_n - a^*\| + \|a_{n-1} - a^*\|),
\end{aligned} \tag{3.12}$$

which gives $\|a_{n+1} - a^*\| \leq (1+\theta_n)\|a_n - a^*\| + \theta_n\|a_{n-1} - a^*\|$. By Lemma 2.3, we obtain

$$\|a_{n+1} - a^*\| \leq K \cdot \prod_{i=1}^{n}(1 + 2\theta_i),$$

where $K = \max\{\|a_1 - a^*\|, \|a_2 - a^*\|\}$. By Lemma 2.3 and $\sum_{n=1}^{\infty}\theta_n < +\infty$, we obtain $\{a_n\}$ is bounded. So

$$\sum_{n=1}^{\infty}\theta_n\|a_n - a_{n-1}\| < +\infty. \tag{3.13}$$

From Lemma 2.2 and equation (3.12), we obtain $\lim_{n\to\infty}\|a_n - a^*\|$ exists.

Consider,

$$\begin{aligned}
\|b_n - a^*\|^2 &= \|a_n + \theta_n(a_n - a_{n-1}) - a^*\|^2 \\
&\leq \|a_n - a^*\|^2 + 2\theta_n\|a_n - a^*\|\|a_n - a_{n-1}\| + \theta_n^2\|a_n - a_{n-1}\|^2.
\end{aligned} \tag{3.14}$$

From equation (3.11) and equation (3.14), we get

$$\begin{aligned}
\|a_{n+1} - a^*\|^2 &\leq \|a_n - a^*\|^2 + 2\theta_n\|a_n - a^*\|\|a_n - a_{n-1}\| + \theta_n^2\|a_n - a_{n-1}\|^2 \\
&\quad -(1 - \frac{\mu^2 + 1}{2})\|c_n - a_{n+1}\|^2 - (1 - \frac{2\mu}{\mu + 1})\|b_n - c_n\|^2 \\
&\quad -2\lambda_n[(p + q)(a_{n+1}) - (p + q)(a^*) + (p + q)(c_n) - (p + q)(a^*)] \\
&\leq \|a_n - a^*\|^2 + 2\theta_n\|a_n - a^*\|\|a_n - a_{n-1}\| + \theta_n^2\|a_n - a_{n-1}\|^2 \\
&\quad -(1 - \frac{\mu^2 + 1}{2})\|c_n - a_{n+1}\|^2 - (1 - \frac{2\mu}{\mu + 1})\|b_n - c_n\|^2. \qquad (3.15)
\end{aligned}$$

From equation (3.13) and $\lim_{n\to\infty} \|a_n - a^*\|$ exists, from equation (3.15), we have

$$\lim_{n\to\infty} \|c_n - a_{n+1}\| = 0 \qquad (3.16)$$

and

$$\lim_{n\to\infty} \|b_n - c_n\| = 0. \qquad (3.17)$$

From equation (3.1), we have

$$\lim_{n\to\infty} \|a_n - b_n\| = 0. \qquad (3.18)$$

From equation (3.17) and equation (3.18), we consider

$$\begin{aligned}
\lim_{n\to\infty} \|a_n - c_n\| &\leq \lim_{n\to\infty} \|a_n - b_n\| + \lim_{n\to\infty} \|b_n - c_n\| \\
&= 0. \qquad (3.19)
\end{aligned}$$

From equation (3.16) and equation (3.19), we get

$$\begin{aligned}
\lim_{n\to\infty} \|a_{n+1} - a_n\| &\leq \lim_{n\to\infty} \|a_{n+1} - c_n\| + \lim_{n\to\infty} \|c_n - a_n\| \\
&= 0.
\end{aligned}$$

Since $\{a_n\}$ is bounded, then there exists a subsequence $\{a_{n_k}\}$ of $\{a_n\}$ such that $a_{n_k} \rightharpoonup \bar{a} \in \mathcal{H}$. Moreover, we obtain $a_{n_k+1} \rightharpoonup \bar{a}$. Since $\{a_{n_k}\}$ is bounded, $\lim_{k\to\infty} \|a_{n_k+1} - c_{n_k}\| = 0$ and ∇p is uniformly continuous, we have

$$\lim_{k\to\infty} \|\nabla p(a_{n_k+1}) - \nabla p(c_{n_k})\| = 0. \qquad (3.20)$$

From

$$a_{n_k+1} = prox_{\lambda_{n_k} q}(c_{n_k} - \lambda_{n_k}\nabla p(c_{n_k})),$$

it follows that

$$\frac{c_{n_k} - \lambda_{n_k}\nabla p(c_{n_k}) - a_{n_k+1}}{\lambda_{n_k}} \in \partial q(a_{n_k+1}).$$

Hence,

$$\frac{c_{n_k} - a_{n_k+1}}{\lambda_{n_k}} + \nabla p(a_{n_k+1}) - \nabla p(c_{n_k}) \in \nabla p(a_{n_k+1}) + \partial q(a_{n_k+1}). \qquad (3.21)$$

Using equation (3.20), letting $k \to \infty$ in equation (3.21) and applying Lemma 2.1, we get

$$0 \in (\nabla p + \partial q)(\bar{a}). \qquad (3.22)$$

Hence $\bar{a} \in \Omega$. From equation (3.15) and Definition 2.1, we have $\{a_n\}$ is a quasi- Fejér sequence. Hence, the sequence $\{a_n\}$ weakly converges to a point in Ω by Lemma 2.4. □

4. Application to Data Classification

In this section, we discuss data classification problems based on a learning technique called extreme learning machine (ELM). Let $\{(a_n, c_n) : a_n \in \mathbb{R}^N, c_n \in \mathbb{R}^M, n = 1, 2, 3, ..., K\}$ be a training set of K distinct samples, a_n is an input training data and c_n is a training target. For the output of ELM with single hidden layer at the i-th hidden node is

$$\eta_i(a) = p(u_i, v_i, a),$$

where p is an activation function, u_i is the weight at the i-th hidden node and v_i is the bias at the i-th hidden node. The output function with L hidden nodes is the single-hidden layer feed forward neural networks (SLFNs)

$$O_n = \sum_{i=1}^{L} \beta_i \eta_i(a_n),$$

where β_i is the optimal output weight at the i-th hidden node. The hidden layer output matrix H is defined by

$$H = \begin{bmatrix} p(u_1, v_1, a_1) & \cdots & p(u_L, v_L, a_1) \\ \vdots & \ddots & \vdots \\ p(u_1, v_1, a_K) & \cdots & p(u_L, v_L, a_K) \end{bmatrix}$$

The principal objective of ELM is to calculate an optimal weight $\beta = [\beta_1, ..., \beta_L]^T$ such that $H\beta = B$, where $B = [t_1, ..., t_K]^T$ is the training target data. We find the solution β via convex minimization problem. Next, we introduce the least absolute shrinkage and selection operator (LASSO) [26] to find the parameter β. It can be modeled as follows:

$$\min_{\beta \in \mathbb{R}^L} \{\|H\beta - B\|_2^2 + \tau \|\beta\|_1\}, \tag{4.1}$$

where τ is a regularization parameter. We see that if $p(\beta) = \|H\beta - B\|_2^2$ and $q(\beta) = \tau\|\beta\|_1$, then the problem (4.1) is reduced to the problem (1.1).

In experiments, we use a cervical cancer behaviour risk data set from UCI Machine Learning Repository [23] for training processing. This data set contains 72 samples which has 19 attributes. We classify two classes of data. We use the sigmoid as the activation function and the hidden nodes $L = 300$. For efficiency of algorithms, we measure by the accuracy of the output data as follows:

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

For the loss of an example, it is computed by the binary cross entropy loss function:

$$\text{Loss} = -\frac{1}{\text{output size}} \sum_{i=1}^{\text{output size}} b_i \log \hat{b}_i + (1 - b_i) \log(1 - \hat{b}_i),$$

where $\hat{b}_i$ is the i-th scalar value in the model output, b_i is the corresponding target value, and output size is the number of scalar values in the model output. In Table 1, we fix parameters for each methods as follows:

Table 1: Chosen control parameters of each methods

Methods	γ	ℓ	δ	μ
FISTA		$L = 1/\|A\|$		
FMFB	2	0.5	0.1	-
IDFB	2	0.5	-	0.8

In our method (IDFB), we set $s_0 = 1$, $s_n = \frac{1+\sqrt{1+4s_{n-1}^2}}{2}$ and

$$\theta_n = \begin{cases} s_{n-1} - 1 & \text{if } n \leq 1000, \\ 0 & \text{otherwise.} \end{cases}$$

Table 2: The result of each methods with the stopping criteria

Methods	Iter	Training time	Acc(%)
FISTA	49	0.0491	90.91
FMFB	32	0.9747	90.91
IDFB	25	0.8014	90.91

The regularization parameter is $\tau = 10^{-5}$. The stopping criteria is the binary cross entropy (Loss=0.119). We report measured quantities in Table 2.

Next, we show graphs of the accuracy and loss of training data and testing data for overfitting of IDFB.

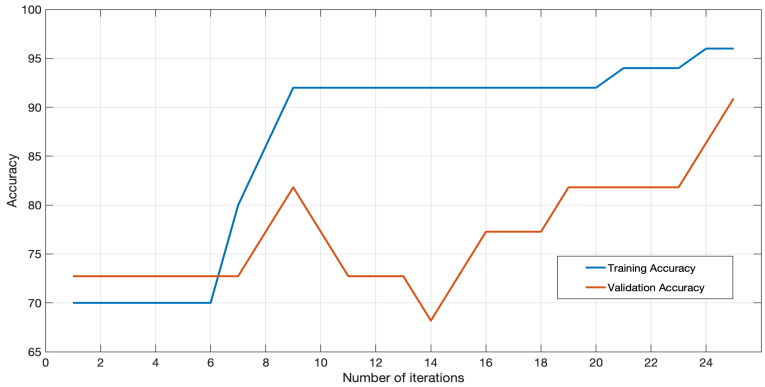

Figure 1: Plot accuracy of IDFB

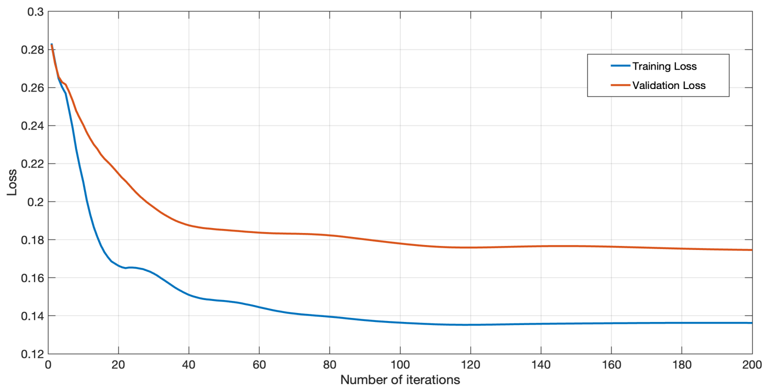

Figure 2: Plot loss of IDFB

5. Discussion

We see that in Table 2, IDFB has the number of iterations less than FISTA and FMFB at the testing accuracy 90.91. It shows that IDFB has a better efficiency than other methods. In Figure 1, we see that training accuracy and validation accuracy have a high gap. It shows that a few training data set are not good enough to train model. Also, Figure 2 has a gap between training loss and testing loss. However, graphs of accuracy and loss values tends in the same way which show that our method (IDFB) can still classify data set even if there are a few data set.

6. Conclusions

In this work, we have proposed an inertial double proximal gradient method with a new linesearch for solving convex minimization problem. We provided weak convergence theorem under some suitable conditions. It was shown that our method has a better performance than FISTA and FMFB in data classification problem. In future work, we study double proximal gradient method with a new linesearch in Banach spaces.

Acknowledgments

This project is funded by the Thailand Science Research and Innovation Fund, and University of Phayao (Grant No. FF65-UoE001) and National Research Council of Thailand (NRCT) under grant no. N41A640094.

S. Kesornprom was supported by Thailand Science Research and Innovation under the project IRN62W0007.

References

[1] Q.H. Ansari, A. Rehan, Split feasibility and fixed point problems. In: Ansari, Q.H. (ed) Nonlinear Analysis: Approximation Theory, Optimization and Applications (2014)(pp. 281-322). Birkhäuser, Springer.

[2] H.H. Bauschke, P.L. Combettes, Convex analysis and monotone operator theory in Hilbert spaces (Vol. 408)(2011). New York: Springer.

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

[4] J.Y. Bello Cruz, T.T. Nghia, On the convergence of the forward-backward splitting method with linesearches. Optimization Methods and Software, 31(6)(2016), 1209-1238.

[5] R.S. Burachik, A.N. Iusem, Enlargements of Monotone Operators. In Set-Valued Mappings and Enlargements of Monotone Operators (2008)(pp. 161-220). Springer, Boston, MA.

[6] C. Byrne, Iterative oblique projection onto convex sets and the split feasibility problem. Inverse Problems, 18(2)(2002), 441.

[7] C. Byrne, A unified treatment of some iterative algorithms in signal processing and image reconstruction. Inverse Problems, 20(1)(2003), 103.

[8] L.C. Ceng, Q.H. Ansari, J.C. Yao, An extragradient method for solving split feasibility and fixed point problems. Computers and Mathematics with Applications, 64(4)(2012), 633-642.

[9] Y. Censor, T. Elfving, A multiprojection algorithm using Bregman projections in a product space. Numerical Algorithms, 8(2)(1994), 221-239.

[10] W. Cholamjiak, P. Cholamjiak, S. Suantai, An inertial forward-backward splitting method for solving inclusion problems in Hilbert spaces. Journal of Fixed Point Theory and Applications, 20(1)(2018), 1-17.

[11] P. Cholamjiak, Y. Shehu, Inertial forward-backward splitting method in Banach spaces with application to compressed sensing. Applications of Mathematics, 64(4)(2019), 409-435.

[12] F. Cui, Y. Tang, C. Zhu, Convergence analysis of a variable metric forward–backward splitting algorithm with applications. Journal of Inequalities and Applications, 2019(1)(2019), 1-27.

[13] Y. Dang, Y. Gao, The strong convergence of a KM-CQ-like algorithm for a split feasibility problem. Inverse Problems, 27(1)(2010), 015007.

[14] O. Ege, I. Karaca, Banach fixed point theorem for digital images. J. Nonlinear Sci. Appl, 8(3)(2015), 237-245.

[15] J. Franklin, Methods of Mathematical Economics Springer-Verlag New York. Inc, January(1980).

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

[17] D. Hieu Van, P. K. Anh, L. D. Muu, Modified forward-backward splitting method for variational inclusions. 4OR, 19(1)(2021), 127-151.

[18] A.N. Iusem, B.F. Svaiter, M. Teboulle, Entropy-like proximal methods in convex programming. Mathematics of Operations Research, 19(4)(1994), 790-814.

[19] K. Kankam, N. Pholasa, P. Cholamjiak, On convergence and complexity of the modified forward-backward method involving new linesearches for convex minimization. Mathematical Methods in the Applied Sciences, 42(5)(2019),1352-1362.

[20] J. L. Lions, G. Stampacchia, Variational inequalities. Communications on pure and applied mathematics, 20(3)(1967), 493-519.

[21] G. López, V. Martín-Márquez, F. Wang, H.K. Xu, Solving the split feasibility problem without prior knowledge of matrix norms. Inverse Problems, 28(8)(2012), 085004.

[22] Y. Malitsky, M. K. Tam, A forward-backward splitting method for monotone inclusions without cocoercivity. SIAM Journal on Optimization, 30(2)(2020), 1451-1472.

[23] R. Machmud, A. Wijaya, Behavior determinant based cervical cancer early detection with machine learning algorithm. Advanced Science Letters, 22(10)(2016), 3120-3123.

[24] M.O. Osilike, S.C. Aniagbosor, B.G. Akuchu, Fixed points of asymptotically demicontractive mappings in arbitrary Banach spaces, Panamerican Mathematical Journal, 12(2)(2002), 77-88.

[25] P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, SIAM J. Control Optim., 2000, 38(2), 431-446.

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

[27] L. U. Uko, Generalized equations and the generalized Newton method. Mathematical programming, 73(3)(1996), 251-268.

[28] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4)(2004), 600-612.

[29] H.K. Xu, Iterative methods for the split feasibility problem in infinite-dimensional Hilbert spaces. Inverse Problems, 26(10)(2010), 105018.

[30] Q. Yang, The relaxed CQ algorithm solving the split feasibility problem. Inverse Problems, 20(4)(2004), 1261.