CARPATHIAN J. MATH. 35 (2019), No. 3, 349 - 363
Online version at http://carpathian.ubm.ro
Print Edition: ISSN 1584 - 2851 Online Edition: ISSN 1843 - 4401

Dedicated to Prof. Qamrul Hasan Ansari on the occasion of his 60^{th} anniversary

A new forward-backward penalty scheme and its convergence for solving monotone inclusion problems

NATTHAPHON ARTSAWANG[1] and KASAMSUK UNGCHITTRAKOOL[1,2]

ABSTRACT. The purposes of this paper are to establish an alternative forward-backward method with penalization terms called new forward-backward penalty method (**NFBP**) and to investigate the convergence behavior of the new method via numerical experiment. It was proved that the proposed method (**NFBP**) converges in norm to a zero point of the monotone inclusion problem involving the sum of a maximally monotone operator and the normal cone of the set of zeros of another maximally monotone operator. Under the observation of some appropriate choices for the available properties of the considered functions and scalars, we can generate a suitable method that weakly ergodic converges to a solution of the monotone inclusion problem. Further, we also provide a numerical example to compare the new forward-backward penalty method with the algorithm introduced by Attouch [Attouch, H., Czarnecki, M.-O. and Peypouquet, J., *Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities*, SIAM J. Optim., **21** (2011), 1251-1274].

1. INTRODUCTION

Let $\mathcal{H}$ be a real Hilbert space with inner product and the corresponding norm be, respectively, denoted by the notations $\langle \cdot, \cdot \rangle$ and $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. Let $f : \mathcal{H} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ be proper convex lower semicontinuous. The following classical convex optimization problem is as follows:

$$\min_{x \in \mathcal{C}} f(x), \tag{1.1}$$

where $\mathcal{C}$ is a nonempty closed convex subset of $\mathcal{H}$.

Many problems in the real world, such as optimal control problems, economic modelings, computational chemistry and biology, data analysis, etc. can be formulated as the problem (1.1) (see [8]).

Most well known algorithms to approximate the solution of (1.1) use the metric projection onto the constrained set $\mathcal{C}$. However, in some situations such as set $\mathcal{C}$ is not simple form, the projection cannot be easily implemented.

For instance, if we take $\mathcal{H} = \mathbb{R}^n$ and $\mathcal{C} = \{x \in \mathbb{R}^n : Ax = b\}$ where A is an $m \times n$ matrix with $m < n$ and $b \in \mathbb{R}^m$, then it is not hard to verify that $\mathrm{proj}_{\mathcal{C}}(\hat{x}) = \hat{x} - A^T(AA^T)^{-1}(A\hat{x} - b)$ where $\mathrm{proj}_{\mathcal{C}} : \mathbb{R}^n \to \mathcal{C}$ is the metric projection. However, the disadvantage of this way is the complication in computing the term of inverse of matrix. On the other hand, if $g : \mathbb{R}^n \to \mathbb{R}$ is defined by $g(x) = \frac{1}{2}\|Ax - b\|^2$ for all $x \in \mathbb{R}^n$, then it is not hard to verify that $\mathcal{C} = \arg\min g = \left\{x \in \mathbb{R}^n : 0 = \nabla g(x) = A^T(Ax - b)\right\}$, where the calculations are less complicated than the calculations using the metric projection.

As a results, Attouch et al. [5] proposed an algorithm concerning the gradient method and exterior penalization scheme for constrained minimization of convex functions

Received: 07.03.2019. In revised form: 25.08.2019. Accepted: 29.08.2019
2010 *Mathematics Subject Classification.* 47H05, 65K05, 65K10, 90C25.
Key words and phrases. *forward-backward penalty method, monotone inclusion, constraint convex minimization.*
Corresponding author: Kasamsuk Ungchittrakool; kasamsuku@nu.ac.th

instead of computing the metric projection onto constrained sets directly. However, if the set $\mathcal{C}$ is a simple form then the metric projection has a closed form expression, it may happen that the computation can be high-price.

It is worth mentioning that the consideration of convex optimization with the constrained function has been applied in several problems such as partial differential equation, signal and image processing, see [4, 5, 6, 9, 16, 19, 20, 30, 32] for more informaton details. These advantages naturally motivate us to consider the particular structure of the constrained set $\mathcal{C} = \arg\min g$, which leads us to consider the following constrained convex optimization problem:

$$\min_{x \in \arg\min g} f(x), \tag{1.2}$$

where $f : \mathcal{H} \to \overline{\mathbb{R}}$ is proper convex lower semicontinuous and $g : \mathcal{H} \to \mathbb{R}$ is (Fréchet) differentiable on the space $\mathcal{H}$. Assume that the solution set of the problem (1.2) is nonempty and some qualifications in [10, Proposition 27.8] hold. Then, problem (1.2) is equivalent to the following problem: find $x \in \mathcal{H}$ such that

$$0 \in \partial f(x) + N_{\arg\min g}(x). \tag{1.3}$$

Problem (Monotone Inclusion Problem (MIP)) Find $x \in \mathcal{H}$ such that

$$0 \in A(x) + N_{(\mathbf{zer}(B))}(x), \tag{1.4}$$

where, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator and $B : \mathcal{H} \to \mathcal{H}$ is a *cocoercive operator* with parameter $\omega > 0$.

We let the set of all zeros of the operator B be denoted by $\mathbf{zer}(B) := \{z \in \mathcal{H} : 0 = B(z)\}$.

Note that if $A = \partial f$ and $B = \nabla g$, then the problem (1.3) is a special case of MIP (1.4).

The aim of this work is to employ the forward-backward penalty method to solve (1.4) from [6] with a new inertial effect. We refer the reader to [1, 2, 3, 7, 11, 12, 13, 14, 15, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27, 28, 29, 31] for more intensive research efforts dedicated to algorithms of inertial type. Inspired by the research works mentioned above, we wish to develop the algorithm called a new forward-backward penalty algorithm (**NFBP**) for solving (1.4) as follows:

$$(\mathbf{NFBP}) \quad \begin{cases} x_1 \in \mathcal{H}; \\ y_n = J_{\lambda_n A}(x_n - \lambda_n \beta_n B(x_n)) \\ x_{n+1} = y_n + \alpha_n (y_n - x_n) \qquad \text{for all } n \geq 1, \end{cases} \tag{1.5}$$

where $\{\lambda_n\}_{n=1}^{\infty}$, $\{\beta_n\}_{n=1}^{\infty}$ and $\{\alpha_n\}_{n=1}^{\infty}$ are sequences of positive parameters.

The proposed numerical scheme can be reduced to the algorithm investigated in [5] which is called forward-backward method (**FB**) when $\alpha_n = 0,\ \forall n \geq 1$.

2. Notations and preliminalies

In this section, we recall some elements of convex analysis which are needed in the sequel. Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a set-value operator. We denote domain of A by $\mathbf{Dom}(A) := \{x \in \mathcal{H} : Ax \neq \varnothing\}$, the range of A by $\mathbf{Ran}(A) := \{u \in \mathcal{H} : \exists x \in \mathcal{H}, u \in Ax\}$, and the graph of A by $\mathbf{gra}(A) := \{(x,u) \in \mathcal{H} \times \mathcal{H} : u \in Ax\}$. The operator A is *monotone* if $\langle x - y, u - v\rangle \geq 0$ for all $(x,u), (y,v) \in \mathbf{gra}(A)$ and it is called *maximally monotone* if there exists no proper monotone extension of the graph of A. The operator A is said to be *ρ-strongly monotone* with modulus $\rho > 0$ if $\langle x - y, u - v\rangle \geq \rho\|x - y\|^2$ for all $(x,u), (y,v) \in \mathbf{gra}(A)$. Moreover, if A is maximally monotone, then $\mathbf{zer}(A)$ is a nonempty closed convex set [10]. We refer to [10] for characterization of its zeros, for a maximally monotone operator A, we have

$$x \in \mathbf{zer}(A) \Leftrightarrow \langle y - x, v\rangle \geq 0\ \forall (y,v) \in \mathbf{gra}(A).$$

Furthermore, if A is maximally monotone and strongly monotone, then $\mathbf{zer}(A)$ is a singleton.

The *resolvent* of A, $J_A : \mathcal{H} \rightrightarrows \mathcal{H}$ is defined by $J_A := (I + A)^{-1}$, where I is the *identity operator* from $\mathcal{H}$ to $\mathcal{H}$. If A is maximally monotone, the resolvent of A is a single-valued. Let $T : \mathcal{H} \to \mathcal{H}$ be operator and let $\omega > 0$. The operator T is said to be *cocoercive* (or *inverse strongly monotone*) with parameter ω if $\langle x - y, Tx - Ty\rangle \geq \omega\|Tx - Ty\|^2$ for all $x, y \in \mathcal{H}$.

Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be monotone, the *Fitzpatrick function* of A, $F_A : \mathcal{H} \times \mathcal{H} \to \overline{\mathbb{R}}$, is defined by $F_A(x, u) := \sup\limits_{(y,v)\in\mathbf{gra}(A)} \{\langle y, u\rangle + \langle x, v\rangle - \langle y, v\rangle\}$ for all $(x, u) \in \mathbf{gra}(A)$ and F_A is a convex lower semicontinuous function. Notice that , if A is maximally monotone then proper and $F_A(x, u) \geq \langle x, u\rangle$ for all $(x, u) \in \mathcal{H} \times \mathcal{H}$.

For a function $h : \mathcal{H} \to \overline{\mathbb{R}}$ we denote its *effective domain* by $\mathbf{Dom}(h) = \{x \in \mathcal{H} : h(x) < +\infty\}$ and say that h is proper, if $\mathbf{Dom}(h) \neq \emptyset$ and $h(x) \neq -\infty$ for all $x \in \mathcal{H}$. The *Fenchel conjugate* of h is $h^* : \mathcal{H} \to \overline{\mathbb{R}}$, which is defined by

$$h^*(z) = \sup_{x\in\mathcal{H}}\{\langle z, x\rangle - h(x)\} \ \text{ for all } z \in \mathcal{H}.$$

The subdifferential of h at $x \in \mathcal{H}$, with $h(x) \in \mathbb{R}$, is the set

$$\partial h(x) := \{v \in \mathcal{H} : h(y) - h(x) \geq \langle v, y - x\rangle \ \forall y \in \mathcal{H}\}.$$

Notice that $\partial h(x) := \emptyset$, if $h(x) \in \{\pm\infty\}$. We know that the subdifferential of a proper convex lower semicontinuous function is a maximally monotone operator and hence

$$F_{\partial h}(x, u) \leq h(x) + h^*(u) \ \text{ for all } \ (x, u) \in \mathcal{H} \times \mathcal{H}.$$

For $\gamma > 0$ and $x \in \mathcal{H}$, we denote the *proximal point* of parameter γ of a proper convex lower semicontinuous function f at x by $\mathbf{prox}_{\gamma f}(x)$, which is the unique optimal solution of the optimization problem

$$\min_{u\in\mathcal{H}} f(u) + \tfrac{1}{2\gamma}\|u - x\|^2.$$

Note that $\mathbf{prox}_{\gamma f} = J_{\gamma\partial f}$ and it is a single-valued operator.

We will call the convex and differentiable function $T : \mathcal{H} \to \mathbb{R}$ has a Lipschitz continuous gradient with Lipschitz constant $L_T > 0$, if $\|\nabla T(x) - \nabla T(y)\| \leq L_T\|x - y\|$ for all $x, y \in \mathcal{H}$.

Let $\mathcal{C} \subset \mathcal{H}$ be a nonempty closed convex set. The *indicator function* is defined as:

$$\delta_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C} \\ +\infty & \text{otherwise.} \end{cases}$$

The *support function* of $\mathcal{C}$ is defined as: $\sigma_{\mathcal{C}}(x) := \sup_{c\in\mathcal{C}}\langle x, c\rangle$ for all $x \in \mathcal{H}$. The *normal cone* $\mathcal{C}$ at a point x is

$$N_{\mathcal{C}}(x) := \begin{cases} \{\overline{x} \in \mathcal{H} : \langle \overline{x}, c - x\rangle \leq 0 \ \text{ for all } c \in \mathcal{C}\}, & \text{if } x \in \mathcal{C} \\ \emptyset, \ \text{otherwise.} \end{cases}$$

We denote the range of $N_{\mathcal{C}}$ by $\mathbf{Ran}(N_{\mathcal{C}})$. Notice that $\delta^*_{\mathcal{C}} = \sigma_{\mathcal{C}}$. Moreover, it holds that $\overline{x} \in N_{\mathcal{C}}(x)$ if and only if $\sigma_{\mathcal{C}}(\overline{x}) = \langle \overline{x}, x\rangle$. Let $\{x_n\}_{n=1}^{\infty}$ be a sequence generated by (**NFBP**) (1.5) and $\{z_n\}_{n=1}^{\infty}$ be the sequence of weighted averages

$$z_n = \tfrac{1}{\tau_n}\sum_{k=1}^{n} \lambda_k x_k, \qquad \text{where} \qquad \tau_n = \sum_{k=1}^{n} \lambda_k. \tag{2.6}$$

3. TECHNICAL LEMMAS

In this section, we will carry out the convergence analysis for new gradient penalty algoritm (**NFBP**) (1.5) which is settled by the following hypotheses.

Assumption 3.1. (I) The qualification condition $\mathbf{zer}(B) \cap \text{int}\, \mathbf{Dom}(A) \neq \emptyset$ holds.

(II) $\{\lambda_n\} \in l^2 \setminus l^1$, $\lim_{n\to\infty} \alpha_n = 0$ and $0 < \liminf_{n\to\infty} \lambda_n\beta_n \leq \limsup_{n\to\infty} \lambda_n\beta_n < \omega$.

(III) For each $p \in \mathbf{Ran}(N_{\mathbf{zer}(B)})$, we have

$$\sum_{n=1}^{+\infty} \lambda_n\beta_n \left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{p}{\beta_n}, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{p}{\beta_n}\right)\right] < +\infty.$$

We present some situations that satisfy the **Assumption** 3.1 as the following remark.

Remark 3.1. (i) Since A and $N_{\mathbf{zer}(B)}$ are maximal monotone and **Assumption** 3.1 (I), we obtain that $A + N_{\mathbf{zer}(B)}$ is maximal monotone operator (see [10, Example 20.26 and Corollary 25.5]).

(ii) There are some examples satisfying **Assumption** 3.1 (II) e.g. sequences $\lambda_n \sim \frac{1}{n}$, $\beta_n \sim n$ and $\alpha_n \sim \frac{1}{n}$ for all $n \in \mathbb{N}$.

(iii) **Assumption** 3.1 (III) has already been used in [11] in order to show the convergence of the proposed algorithm (see [11, Assumption (H_{fitz})]). They also pointed out that for each $p \in \mathrm{Ran}(N_{\mathbf{zer}(B)})$ and any $n \in \mathbb{N}$ one has $\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\frac{p}{\beta_n}, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\frac{p}{\beta_n}\right) \geq 0$. Some examples of the operator B satisfying **Assumption** 3.1 (III) can be found in [9, Section 5].

Lemma 3.1 ([10], Lemma 2.12 and Corollary 2.15)**.** *Let x and y be in Hilbert space $\mathcal{H}$ and $\alpha \in \mathbb{R}$. Then*

(i) $2\langle x, y\rangle = \|x\|^2 + \|y\|^2 - \|x-y\|^2$.

(ii) $\|\alpha x + (1-\alpha)y\|^2 + \alpha(1-\alpha)\|x-y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2$.

Let us denote an arbitrary sequence verifying (**NFBP**) (1.5) by $\{x_n\}_{n=1}^{\infty}$ and provide some estimations.

Lemma 3.2. *Let $x^* \in \mathbf{zer}(B) \cap \mathbf{Dom}(A)$ and $v \in A(x^*)$. Then the following inequality holds for each $n \in \mathbb{N}$ and $\varepsilon \geq 0$*

$$\begin{aligned}&\|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2 + (1+\alpha_n)\left(\tfrac{2\varepsilon}{1+\varepsilon}\right)\lambda_n\beta_n\langle B(x_n), x_n - x^*\rangle\\ &\qquad + (1+\alpha_n)\left(\tfrac{\varepsilon}{1+\varepsilon} - \alpha_n\right)\|y_n - x_n\|^2\\ &\qquad \leq (1+\alpha_n)\lambda_n\beta_n\left((1+\varepsilon)\lambda_n\beta_n - \tfrac{2\omega}{1+\varepsilon}\right)\|B(x_n)\|^2 + 2(1+\alpha_n)\lambda_n\langle v, x^* - y_n\rangle. \end{aligned} \tag{3.7}$$

Proof. It is not hard to verify from (1.5) that for each $n \in \mathbb{N}$, $\frac{x_n - y_n}{\lambda_n} - \beta_n B(x_n) \in A(y_n)$. By the monotonicity of A and $v \in A(x^*)$,

$$\left\langle \tfrac{x_n-y_n}{\lambda_n} - \beta_n B(x_n) - v, y_n - x^*\right\rangle \geq 0.$$

It follows that

$$\langle x_n - y_n, x^* - y_n\rangle \leq \lambda_n\langle \beta_n B(x_n) + v, x^* - y_n\rangle, \text{ for all } n \in \mathbb{N}.$$

From Lemma 3.1 (i), we obtain that for each $n \in \mathbb{N}$

$$\|x_n - y_n\|^2 + \|x^* - y_n\|^2 - \|x_n - x^*\|^2 \leq 2\lambda_n\langle \beta_n B(x_n) + v, x^* - y_n\rangle, \tag{3.8}$$

which mean that

$$\|y_n - x^*\|^2 - \|x_n - x^*\|^2 + \|x_n - y_n\|^2 \leq 2\lambda_n\langle v, x^* - y_n\rangle + 2\lambda_n\beta_n\langle B(x_n), x^* - x_n\rangle + 2\lambda_n\beta_n\langle B(x_n), x_n - y_n\rangle. \tag{3.9}$$

Note that B is ω-cocoercive and $B(x^*) = 0$, we have

$$2\lambda_n\beta_n\langle B(x_n), x^* - x_n\rangle \leq -2\omega\lambda_n\beta_n\|B(x_n)\|^2 \tag{3.10}$$

for all $n \in \mathbb{N}$. From (3.10), we observe that

$$\begin{aligned}2\lambda_n\beta_n\langle B(x_n), x^* - x_n\rangle &= \tfrac{1}{1+\varepsilon}2\lambda_n\beta_n\langle B(x_n), x^* - x_n\rangle + \tfrac{\varepsilon}{1+\varepsilon}2\lambda_n\beta_n\langle B(x_n), x^* - x_n\rangle\\ &\leq -\tfrac{2\omega}{1+\varepsilon}\lambda_n\beta_n\|B(x_n)\|^2 + \tfrac{2\varepsilon}{1+\varepsilon}\lambda_n\beta_n\langle B(x_n), x^* - x_n\rangle.\end{aligned} \tag{3.11}$$

For each $n \in \mathbb{N}$, let us consider

$$\begin{aligned}0 \leq \tfrac{1}{1+\varepsilon}\|y_n - x_n + (1+\varepsilon)\lambda_n\beta_n B(x_n)\|^2 &= \tfrac{1}{1+\varepsilon}\|y_n - x_n\|^2 + (1+\varepsilon)\lambda_n^2\beta_n^2\|B(x_n)\|^2\\ &\quad + 2\lambda_n\beta_n\langle B(x_n), y_n - x_n\rangle,\end{aligned}$$

which implies that

$$2\lambda_n\beta_n\langle B(x_n), x_n - y_n\rangle \leq \tfrac{1}{1+\varepsilon}\|y_n - x_n\|^2 + (1+\varepsilon)\lambda_n^2\beta_n^2\|B(x_n)\|^2. \tag{3.12}$$

Joining (3.11) and (3.12) to (3.9) together with some simple calculations, we have that

$$\begin{aligned}\|y_n - x^*\|^2 &\leq 2\lambda_n\langle v, x^* - y_n\rangle + \tfrac{2\varepsilon}{1+\varepsilon}\lambda_n\beta_n\langle B(x_n), x^* - x_n\rangle\\ &\quad + \lambda_n\beta_n\left((1+\varepsilon)\lambda_n\beta_n - \tfrac{2\omega}{1+\varepsilon}\right)\|B(x_n)\|^2 + \|x_n - x^*\|^2 - \tfrac{\varepsilon}{1+\varepsilon}\|y_n - x_n\|^2.\end{aligned} \tag{3.13}$$

On the other hand, by using Lemma 3.1 (ii), we have the following equation

$$\begin{aligned}\|x_{n+1} - x^*\|^2 &= \|y_n + \alpha_n(y_n - x_n) - x^*\|^2 = \|(1+\alpha_n)(y_n - x^*) - \alpha_n(x_n - x^*)\|^2\\ &= (1+\alpha_n)\|y_n - x^*\|^2 - \alpha_n\|x_n - x^*\|^2 + \alpha_n(1+\alpha_n)\|x_n - y_n\|^2.\end{aligned} \tag{3.14}$$

Multiplying both sides of (3.13) by $(1+\alpha_n)$ and then connecting to (3.14), it yields that

$$\begin{aligned}\|x_{n+1} - x^*\|^2 &\leq 2(1+\alpha_n)\lambda_n\langle v, x^* - y_n\rangle - (1+\alpha_n)\left(\tfrac{2\varepsilon}{1+\varepsilon}\right)\lambda_n\beta_n\langle B(x_n), x_n - x^*\rangle\\ &\quad + (1+\alpha_n)\lambda_n\beta_n\left((1+\varepsilon)\lambda_n\beta_n - \tfrac{2\omega}{1+\varepsilon}\right)\|B(x_n)\|^2 + (1+\alpha_n)\|x_n - x^*\|^2\\ &\quad - (1+\alpha_n)\left(\tfrac{\varepsilon}{1+\varepsilon}\right)\|y_n - x_n\|^2 - \alpha_n\|x_n - x^*\|^2 + \alpha_n(1+\alpha_n)\|x_n - y_n\|^2\end{aligned}$$

for all $n \in \mathbb{N}$. This completes the proof. □

Lemma 3.3. *Let $(x^*, w) \in \mathbf{gra}(A + N_{\mathbf{zer}(B)})$, $v \in A(x^*)$ and $p \in N_{\mathbf{zer}(B)}(x^*)$ be such that $w = v + p$. Suppose that $\limsup_{n\to\infty} \lambda_n\beta_n < \omega$. Then there exist $\overline{n} \in \mathbb{N}$, $\varepsilon_0 > 0$ and $K > 0$ such that for each $n \geq \overline{n}$,*

$$\begin{aligned}&\|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2 + \left(\tfrac{\varepsilon_0}{4(1+\varepsilon_0)}\right)\|y_n - x_n\|^2\\ &\quad + \left(\tfrac{\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\right)\langle B(x_n), x_n - x^*\rangle + \left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_n\beta_n\|B(x_n)\|^2\\ &\quad \leq \tfrac{(1+K)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{2p(1+\varepsilon_0)}{\varepsilon_0\beta_n}, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{2p(1+\varepsilon_0)}{\varepsilon_0\beta_n}\right)\right]\\ &\quad + 2(1+K)\lambda_n\langle w, x^* - x_n\rangle + 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+K)\lambda_n^2\|v\|^2.\end{aligned}$$

Proof. Since $\limsup_{n\to\infty} \lambda_n\beta_n < \omega$, there exists $N_0 \in \mathbb{N}$ such that $\lambda_n\beta_n < \omega$ for all $n \geq N_0$. So, we can find $\varepsilon_0 \in \left(0, \sqrt{\frac{\omega}{\limsup_{n\to\infty} \lambda_n\beta_n}} - 1\right)$ and hence, $(1+\varepsilon_0)\lambda_n\beta_n < \frac{\omega}{1+\varepsilon_0}$ for all $n \geq N_0$. Note that $\alpha_n \to 0$ as $n \to +\infty$, there exists $N_1 \in \mathbb{N}$ such that $\alpha_n < \frac{\varepsilon_0}{4(1+\varepsilon_0)}$ for all $n \geq N_1$. Choose $\overline{n} := \max\{N_0, N_1\}$. For each $n \in \mathbb{N}$, by applying Lemma 3.1 (i), the following inequality holds:

$$
\begin{aligned}
2(1+\alpha_n)\lambda_n\langle v, x^*-y_n\rangle &= 2(1+\alpha_n)\lambda_n\langle v, x^*-x_n\rangle + 2(1+\alpha_n)\langle \lambda_n v, x_n-y_n\rangle \\
&\leq 2(1+\alpha_n)\lambda_n\langle v, x^*-x_n\rangle + 2(1+\alpha_n)\langle \sqrt{\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}}\lambda_n v, \sqrt{\tfrac{\varepsilon_0}{2(1+\varepsilon_0)}}(x^*-x_n)\rangle \\
&\leq 2(1+\alpha_n)\lambda_n\langle v, x^*-x_n\rangle + (1+\alpha_n)\left(\tfrac{\varepsilon_0}{2(1+\varepsilon_0)}\right)\|y_n-x_n\|^2 \\
&\quad + (1+\alpha_n)\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)\lambda_n^2\|v\|^2.
\end{aligned} \tag{3.15}
$$

Combining (3.15) to (3.7), we obtain that for each $n \geq \overline{n}$,

$$
\begin{aligned}
&\|x_{n+1}-x^*\|^2 - \|x_n-x^*\|^2 + (1+\alpha_n)\left(\tfrac{\varepsilon_0}{2(1+\varepsilon_0)} - \alpha_n\right)\|y_n-x_n\|^2 \\
&+ (1+\alpha_n)\left(\tfrac{2\varepsilon_0}{1+\varepsilon_0}\right)\lambda_n\beta_n\langle B(x_n), x_n-x^*\rangle + (1+\alpha_n)\left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_n\beta_n\|B(x_n)\|^2 \\
&\leq (1+\alpha_n)\lambda_n\beta_n\left((1+\varepsilon_0)\lambda_n\beta_n - \tfrac{\omega}{1+\varepsilon_0}\right)\|B(x_n)\|^2 \\
&+ (1+\alpha_n)\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)\lambda_n^2\|v\|^2 + 2(1+\alpha_n)\lambda_n\langle v, x^*-x_n\rangle.
\end{aligned} \tag{3.16}
$$

From (3.16), we get that for each $n \geq \overline{n}$,

$$
\begin{aligned}
&\|x_{n+1}-x^*\|^2 - \|x_n-x^*\|^2 + (1+\alpha_n)\left(\tfrac{\varepsilon_0}{4(1+\varepsilon_0)}\right)\|y_n-x_n\|^2 \\
&+ (1+\alpha_n)\left(\tfrac{2\varepsilon_0}{1+\varepsilon_0}\right)\lambda_n\beta_n\langle B(x_n), x_n-x^*\rangle + (1+\alpha_n)\left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_n\beta_n\|B(x_n)\|^2 \\
&\leq 2(1+\alpha_n)\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)\lambda_n^2\|v\|^2 + 2(1+\alpha_n)\lambda_n\langle v, x^*-x_n\rangle.
\end{aligned} \tag{3.17}
$$

Next, for each $n \geq \overline{n}$, we focus on the following terms of (3.17)

$$
\begin{aligned}
&2(1+\alpha_n)\lambda_n\langle v, x^*-x_n\rangle - \tfrac{(1+\alpha_n)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\langle B(x_n), x_n-x^*\rangle \\
&= 2(1+\alpha_n)\lambda_n\langle w-p, x^*-x_n\rangle - \tfrac{(1+\alpha_n)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\langle B(x_n), x_n-x^*\rangle \\
&= 2(1+\alpha_n)\lambda_n\langle w, x^*-x_n\rangle + 2(1+\alpha_n)\lambda_n\langle p, x_n\rangle \\
&\quad - \tfrac{(1+\alpha_n)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\langle B(x_n), x_n-x^*\rangle - 2(1+\alpha_n)\lambda_n\langle p, x^*\rangle \\
&= \tfrac{(1+\alpha_n)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\left[\left\langle \tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p, x_n\right\rangle + \langle B(x_n), x^*\rangle\right. \\
&\quad \left. - \langle B(x_n), x_n\rangle - \left\langle \tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p, x^*\right\rangle\right] + (1+\alpha_n)2\lambda_n\langle w, x^*-x_n\rangle \\
&\leq \tfrac{(1+\alpha_n)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p, x^*\right) - \left\langle \tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p, x^*\right\rangle\right] \\
&\quad + 2(1+\alpha_n)\lambda_n\langle w, x^*-x_n\rangle.
\end{aligned} \tag{3.18}
$$

Since $p \in N_{\mathbf{zer}(B)}(x^*)$, we have $\frac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p \in N_{\mathbf{zer}(B)}(x^*)$ for all $n \in \mathbb{N}$. It is equivalent to say that $\sigma_{\mathbf{zer}(B)}\left(\frac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p\right) = \left\langle \frac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p, x^*\right\rangle$ for all $n \in \mathbb{N}$. It follows from (3.18) that

$$\begin{aligned} 2(1+\alpha_n)\lambda_n\langle v, x^*-x_n\rangle &\leq \tfrac{(1+\alpha_n)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\langle B(x_n), x_n - x^*\rangle \\ &\quad + \tfrac{(1+\alpha_n)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p, x^*\right)\right. \\ &\quad \left. -\sigma_{\mathbf{zer}(B)}\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p\right)\right] + 2(1+\alpha_n)\lambda_n\langle w, x^*-x_n\rangle. \end{aligned} \tag{3.19}$$

Combining (3.19) to (3.16), it appears the result that for each $n \geq \overline{n}$,

$$\begin{aligned} &\|x_{n+1}-x^*\|^2 - \|x_n - x^*\|^2 + (1+\alpha_n)\left(\tfrac{\varepsilon_0}{4(1+\varepsilon_0)}\right)\|y_n - x_n\|^2 \\ &\quad + \tfrac{(1+\alpha_n)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\langle B(x_n), x_n - x^*\rangle + (1+\alpha_n)\left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_n\beta_n\|B(x_n)\|^2 \\ &\quad \leq \tfrac{(1+\alpha_n)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p\right)\right] \\ &\quad + 2(1+\alpha_n)\lambda_n\langle w, x^*-x_n\rangle + 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+\alpha_n)\lambda_n^2\|v\|^2. \end{aligned}$$

Note that the positive sequence $\{\alpha_n\}_{n=1}^{\infty}$ is bounded, there exists $K > 0$ such that $\alpha_n \leq K$ for all $n \in \mathbb{N}$. Since $\langle B(x_n), x_n - x^*\rangle$ is nonnegative for all $n \in \mathbb{N}$, we obtain that

$$\begin{aligned} &\|x_{n+1}-x^*\|^2 - \|x_n - x^*\|^2 + \left(\tfrac{\varepsilon_0}{4(1+\varepsilon_0)}\right)\|y_n - x_n\|^2 \\ &\quad + \left(\tfrac{\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\right)\langle B(x_n), x_n - x^*\rangle + \left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_n\beta_n\|B(x_n)\|^2 \\ &\quad \leq \tfrac{(1+K)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p\right)\right] \\ &\quad + 2(1+K)\lambda_n\langle w, x^*-x_n\rangle + 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+K)\lambda_n^2\|v\|^2, \ \ \forall n \geq \overline{n}. \end{aligned}$$

This completes the proof. □

The next lemma plays an important role in the convergence analysis (see in [6, Lemma 2] or [25, Lemma 3.1]).

Lemma 3.4. *Let $\{\gamma_n\}_{n=1}^{\infty}$, $\{\delta_n\}_{n=1}^{\infty}$ and $\{\varepsilon_n\}_{n=1}^{\infty}$ be real sequences. Assume that $\{\gamma_n\}_{n=1}^{\infty}$ is bounded from below, $\{\delta_n\}_{n=1}^{\infty}$ is non-negative and $\sum_{n=1}^{+\infty}\varepsilon_n < +\infty$ such that*

$$\gamma_{n+1} - \gamma_n + \delta_n \leq \varepsilon_n \ \text{ for all } \ n \geq 1.$$

Then $\lim_{n\to\infty}\gamma_n$ exists and $\sum_{n=1}^{+\infty}\delta_n < +\infty$.

4. Convergence results for monotone inclusion problem

In this section, some convergence results for (**NFBP**) (1.5) are demonstrated. Before going into the main results, it is useful to know the following propositions.

Proposition 4.1 ([10], Opial Lemma)**.** *Let $\mathcal{H}$ be a real Hilbert space, $\mathcal{C} \subseteq \mathcal{H}$ be nonempty set, $\{x_n\}_{n=1}^{\infty}$ be any arbitrary sequence and $\{z_n\}_{n=1}^{\infty}$ defined as* (2.6) *such that:*

(i) *For every $z \in \mathcal{C}$, $\lim_{n\to\infty}\|x_n - z\|$ exists;*

(ii) *Every weak cluster point of the sequence* $\{x_n\}_{n=1}^{\infty}$ *(resp.,* $\{z_n\}_{n=1}^{\infty}$*) lies in* $\mathcal{C}$.

Then the sequence $\{x_n\}_{n=1}^{\infty}$ *(resp.,* $\{z_n\}_{n=1}^{\infty}$*) converges weakly to a point in* $\mathcal{C}$.

Proposition 4.2. *Let* $\{x_n\}_{n=1}^{\infty}$ *be a sequence generated by* (**NFBP**) (1.5). *If all assumptions in* **Assumption** 3.1 *hold, then the following hold:*

(i) *For each* $x^* \in \mathbf{zer}(A + N_{\mathbf{zer}(B)})$, $\lim_{n\to+\infty} \|x_n - x^*\|$ *exists.*

(ii) *The series* $\sum_{n=1}^{+\infty} \|y_n - x_n\|^2$, $\sum_{n=1}^{+\infty} \lambda_n\beta_n\|B(x_n)\|^2$, *and* $\sum_{n=1}^{+\infty} \lambda_n\beta_n\langle B(x_n), x_n - x^*\rangle$ *are convergent.*

(iii) $\lim_{n\to+\infty} \|y_n - x_n\|^2 = \lim_{n\to+\infty} \|B(x_n)\| = \lim_{n\to+\infty} \langle B(x_n), x_n - x^*\rangle = 0.$

Proof. Let $x^* \in \mathbf{zer}(A + N_{\mathbf{zer}(B)})$. Taking $w = 0$ in Lemma 3.3, we get that

$$\begin{aligned}
&\|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2 + \left(\tfrac{\varepsilon_0}{4(1+\varepsilon_0)}\right)\|y_n - x_n\|^2 \\
&\quad + \left(\tfrac{\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\right)\langle B(x_n), x_n - x^*\rangle + \left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_n\beta_n\|B(x_n)\|^2 \\
&\quad \le \tfrac{(1+K)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{2p(1+\varepsilon_0)}{\varepsilon_0\beta_n}, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{2p(1+\varepsilon_0)}{\varepsilon_0\beta_n}\right)\right] \\
&\quad + 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+K)\lambda_n^2\|v\|^2, \ \ \forall n \ge \overline{n}.
\end{aligned}$$

By **Assumption 3.1**, the conclusion in (i) and (ii) follows from Lemma 3.4.

(iii) From (ii), we have $\lim_{n\to+\infty} \|y_n - x_n\|^2 = 0$. Since $\liminf_{n\to+\infty} \lambda_n\beta_n > 0$, we obtain that $\lim_{n\to+\infty} \|B(x_n)\| = \lim_{n\to+\infty} \langle B(x_n), x_n - x^*\rangle = 0.$ □

Theorem 4.2. *Let* $\{x_n\}_{n=1}^{\infty}$ *be a sequence generated by (***NFBP***)* (1.5) *and* $\{z_n\}_{n=1}^{\infty}$ *be a sequence of weighted averages as* (2.6). *Suppose that all assumptions in* **Assumption** *3.1 hold. Then the sequence* $\{z_n\}_{n=1}^{\infty}$ *converges weakly to an element in* $\mathbf{zer}(A + N_{\mathbf{zer}(B)})$.

Proof. Let z be a weak cluster point of $\{z_n\}_{n=1}^{\infty}$. Then there exists a subsequence $\{z_{n_k}\}_{k=1}^{\infty}$ of $\{z_n\}_{n=1}^{\infty}$ such that $z_{n_k} \rightharpoonup z$ as $k \to +\infty$. We will show that $z \in \mathbf{zer}(A + N_{\mathbf{zer}(B)})$. Since $A + N_{\mathbf{zer}(B)}$ is a maximal monotone operator, it suffices to show that $\langle w, x^* - z\rangle \ge 0$ for all $(x^*, w) \in \mathbf{gra}(A + N_{\mathbf{zer}(B)})$.

Let $(x^*, w) \in \mathbf{gra}(A + N_{\mathbf{zer}(B)})$, $v \in A(x^*)$ and $p \in N_{\mathbf{zer}(B)}(x^*)$ be such that $w = v + p$. Recall from Lemma 3.3 that

$$\begin{aligned}
&\|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2 + \left(\tfrac{\varepsilon_0}{4(1+\varepsilon_0)}\right)\|y_n - x_n\|^2 \\
&\quad + \left(\tfrac{\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\right)\langle B(x_n), x_n - x^*\rangle + \left(\tfrac{\omega}{1+\varepsilon_0}\right)\lambda_n\beta_n\|B(x_n)\|^2 \\
&\quad \le \tfrac{(1+K)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p\right)\right] \\
&\quad + 2(1+K)\lambda_n\langle w, x^* - x_n\rangle + 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+K)\lambda_n^2\|v\|^2, \ \ \forall n \ge \overline{n}.
\end{aligned}$$

Discarding nonnegative terms $\langle B(x_n), x_n - x^*\rangle$, $\|B(x_n)\|^2$ and $\|y_n - x_n\|^2$, we deduce to

$$\begin{aligned}
&\|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2 \\
&\quad \le \tfrac{(1+K)\varepsilon_0\lambda_n\beta_n}{1+\varepsilon_0}\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p\right)\right]
\end{aligned}$$

$$+ 2(1+K)\lambda_n\langle w, x^* - x_n\rangle + 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+K)\lambda_n^2\|v\|^2, \;\; \forall n \geq \overline{n}.$$

Summing up for $n = \overline{n}, \overline{n}+1, ..., n_k$ in the above inequality, we have

$$\begin{aligned}\|x_{n_k+1} - x^*\|^2 - \|x_{\overline{n}} - x^*\|^2 &\leq 2(1+K)\left\langle w, \sum_{n=\overline{n}}^{n_k}\lambda_n x^* - \sum_{n=\overline{n}}^{n_k}\lambda_n x_n\right\rangle + L_1 \\ &= 2(1+K)\left\langle w, \sum_{n=1}^{n_k}\lambda_n x^* - \sum_{n=1}^{\overline{n}-1}\lambda_n x^* - \sum_{n=1}^{n_k}\lambda_n x_n + \sum_{n=1}^{\overline{n}-1}\lambda_n x_n\right\rangle + L_1\end{aligned}$$

where

$$\begin{aligned}L_1 :=& \tfrac{(1+K)}{2}\sum_{n=\overline{n}}^{n_k}\lambda_n\beta_n\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0\beta_n}p\right)\right] \\ &+ 2\left(\tfrac{2(1+\varepsilon_0)}{\varepsilon_0}\right)(1+K)\sum_{n=\overline{n}}^{n_k}\lambda_n^2\|v\|^2.\end{aligned}$$

Discarding the nonnegative term $\|x_{n_k+1} - x^*\|^2$ and dividing inequality above by $2(1+K)\tau_{n_k}$, we obtain

$$-\tfrac{\|x_{\overline{n}} - x^*\|^2}{2(1+K)\tau_{n_k}} \leq \langle w, x^* - z_{n_k}\rangle + \tfrac{L_2}{2(1+K)\tau_{n_k}}, \tag{4.20}$$

where $L_2 := L_1 + 2(1+K)\left\langle w, -\sum_{n=1}^{\overline{n}-1}\lambda_n x^* + \sum_{n=1}^{\overline{n}-1}\lambda_n x_n\right\rangle$. Note that L_2 is a finite real number. Taking $k \to +\infty$ (so that $\lim_{k\to+\infty}\tau_{n_k} = +\infty$) on both sides of (4.20), we get that

$$0 \leq \langle w, x^* - z\rangle.$$

Since $(x^*, w) \in \mathbf{gra}(A+N_{\mathbf{zer}(B)})$ is arbitrary, we have $z \in \mathbf{zer}(A+N_{\mathbf{zer}(B)})$. By Proposition 4.1 (ii), we can conclude that the sequence $\{z_n\}_{n=1}^{\infty}$ converges weakly to an element in $\mathbf{zer}(A+N_{\mathbf{zer}(B)})$. □

Next, we will prove the strong convergence of the sequence $\{x_n\}_{n=1}^{\infty}$.

Theorem 4.3. *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence generated by (**NFBP**) (1.5) and the operator A be a γ-strongly monotone with $\gamma > 0$. If all assumptions in **Assumption** 3.1 hold, then the sequence $\{x_n\}_{n=1}^{\infty}$ converges in norm to the unique $x^* \in \mathbf{zer}(A+N_{\mathbf{zer}(B)})$.*

Proof. Let x^* be the unique element in $\mathbf{zer}(A+N_{\mathbf{zer}(B)})$. Then there exists $v \in A(x^*)$ and $p \in N_{\mathbf{zer}(B)}(x^*)$ such that $0 = v + p$. Since $\frac{x_n - y_n}{\lambda_n} - \beta_n B(x_n) \in A(y_n)$ and $v \in A(x^*)$, the strong monotonicity of A implies

$$\lambda_n\gamma\|y_n - x^*\|^2 \leq \langle x_n - y_n - \lambda_n(\beta_n B(x_n) + v), y_n - x^*\rangle$$

for all $n \in \mathbb{N}$. It follows that

$$\lambda_n\gamma\|y_n - x^*\|^2 + \langle x_n - y_n, x^* - y_n\rangle \leq \lambda_n\langle\beta_n B(x_n) + v, x^* - y_n\rangle \tag{4.21}$$

for all $n \in \mathbb{N}$. By applying Lemma 3.1 (i), we have

$$2\lambda_n\gamma\|y_n - x^*\|^2 + \|y_n - x^*\|^2 - \|x_n - x^*\|^2 \leq 2\lambda_n\langle\beta_n B(x_n) + v, x^* - y_n\rangle - \|x_n - y_n\|^2 \tag{4.22}$$

for all $n \in \mathbb{N}$. Focusing on the right hand side of (4.22), we see that

$$\begin{aligned}&2\lambda_n\langle\beta_n B(x_n) + v, x^* - y_n\rangle - \|x_n - y_n\|^2 \\ &= 2\lambda_n\langle\beta_n B(x_n) + v, x^* - x_n\rangle + 2\lambda_n\langle\beta_n B(x_n) + v, x_n - y_n\rangle - \|x_n - y_n\|^2 \qquad (4.23)\end{aligned}$$

$$\leq 2\lambda_n\langle \beta_n B(x_n) + v, x^* - x_n\rangle + \lambda_n^2\|\beta_n B(x_n) + v\|^2$$
$$\leq 2\lambda_n\langle \beta_n B(x_n) + v, x^* - x_n\rangle + 2\lambda_n^2\beta_n^2\|B(x_n)\|^2 + 2\lambda_n^2\|v\|^2, \quad \forall n \in \mathbb{N}.$$

Next, we consider the first term on right hand side of (4.23),

$$\begin{aligned}
(4.24)\qquad & 2\lambda_n\langle \beta_n B(x_n) + v, x^* - x_n\rangle \\
&= 2\lambda_n\langle \beta_n B(x_n), x^* - x_n\rangle + 2\lambda_n\langle v, x^* - x_n\rangle \\
&= 2\lambda_n\langle \beta_n B(x_n), x^* - x_n\rangle + 2\lambda_n\langle p, x_n\rangle - 2\lambda_n\langle p, x^*\rangle \\
&= 2\lambda_n\beta_n\left[\left\langle \tfrac{p}{\beta_n}, x_n\right\rangle + \langle B(x_n), x^*\rangle - \langle B(x_n), x_n\rangle - \left\langle \tfrac{p}{\beta_n}, x^*\right\rangle\right] \\
&\leq 2\lambda_n\beta_n\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{p}{\beta_n}, x^*\right) - \left\langle \tfrac{p}{\beta_n}, x^*\right\rangle\right] \\
&= 2\lambda_n\beta_n\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{p}{\beta_n}, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{p}{\beta_n}, x^*\right)\right], \quad \forall n \in \mathbb{N}.
\end{aligned}$$

Combining (4.22), (4.23) and (4.24), we have

$$\begin{aligned}
(4.25)\qquad & 2\lambda_n\gamma\|y_n - x^*\|^2 + \|y_n - x^*\|^2 - \|x_n - x^*\|^2 \\
&\leq 2\lambda_n\beta_n\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{p}{\beta_n}, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{p}{\beta_n}, x^*\right)\right] \\
&\quad + 2\lambda_n^2\beta_n^2\|B(x_n)\|^2 + 2\lambda_n^2\|v\|^2, \quad \forall n \in \mathbb{N}.
\end{aligned}$$

By simple calculation using (4.25), we get the result that

$$\begin{aligned}
(4.26)\qquad \|y_n - x^*\|^2 &\leq \tfrac{2\lambda_n\beta_n}{2\lambda_n\gamma+1}\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{p}{\beta_n}, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{p}{\beta_n}, x^*\right)\right] \\
&\quad + \tfrac{1}{2\lambda_n\gamma+1}\|x_n - x^*\|^2 + \tfrac{2\lambda_n^2\beta_n^2}{2\lambda_n\gamma+1}\|B(x_n)\|^2 + \tfrac{2\lambda_n^2}{2\lambda_n\gamma+1}\|v\|^2.
\end{aligned}$$

Combining (4.26) to (3.14), we have the following inequality

$$\begin{aligned}
\|x_{n+1} - x^*\|^2 &\leq (1+\alpha_n)\tfrac{2\lambda_n\beta_n}{2\lambda_n\gamma+1}\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{p}{\beta_n}, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{p}{\beta_n}, x^*\right)\right] \\
(4.27)\qquad &\quad + (1+\alpha_n)\left[\tfrac{1}{2\lambda_n\gamma+1}\|x_n - x^*\|^2 + \tfrac{2\lambda_n^2\beta_n^2}{2\lambda_n\gamma+1}\|B(x_n)\|^2\right] \\
&\quad + \left(\tfrac{1+\alpha_n}{2\lambda_n\gamma+1}\right)2\lambda_n\|v\|^2 - \tfrac{\alpha_n}{2\lambda_n\gamma+1}\|x_n - x^*\|^2 + \alpha_n(1+\alpha_n)\|x_n - y_n\|^2.
\end{aligned}$$

It is not hard to verify from (4.27) and it yields that

$$\begin{aligned}
& 2\lambda_n\gamma\|x_{n+1} - x^*\|^2 + \|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2 \\
&\leq (1+\alpha_n)2\lambda_n\beta_n\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\tfrac{p}{\beta_n}, x^*\right) - \sigma_{\mathbf{zer}(B)}\left(\tfrac{p}{\beta_n}, x^*\right)\right] + (1+\alpha_n)2\lambda_n^2\beta_n^2\|B(x_n)\|^2 \\
&\quad + (1+\alpha_n)\,2\lambda_n^2\|v\|^2 + \alpha_n(1+\alpha_n)(2\lambda_n\gamma+1)\|x_n - y_n\|^2.
\end{aligned}$$

Since nonnegative sequences $\{\lambda_n\}_{n=1}^\infty$, $\{\lambda_n\beta_n\}_{n=1}^\infty$ and $\{\alpha_n\}_{n=1}^\infty$ are bounded, there exists positive numbers M, c and K such that $\lambda_n \leq M, \ \ \lambda_n\beta_n \leq c,$ and $\alpha_n \leq K$ for all $n \in \mathbb{N}$.

Hence,

$$2\lambda_n\gamma\|x_{n+1} - x^*\|^2 + \|x_{n+1} - x^*\|^2 - \|x_n - x^*\|^2$$

$$\leq (1+K)2\lambda_n\beta_n\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\frac{p}{\beta_n},x^*\right)-\sigma_{\mathbf{zer}(B)}\left(\frac{p}{\beta_n},x^*\right)\right]$$

$$+(1+K)2c\lambda_n\beta_n\|B(x_n)\|^2+(1+K)\,2\lambda_n^2\|v\|^2+K(1+K)(2M\gamma+1)\|x_n-y_n\|^2, \tag{4.28}$$

and then

$$\begin{aligned}2\gamma\sum_{n=1}^{+\infty}\lambda_n\|x_{n+1}-x^*\|^2 \leq \|x_1-x^*\|^2+(1+K)&\left[2c\sum_{n=1}^{+\infty}\lambda_n\beta_n\|B(x_n)\|^2+2\sum_{n=1}^{+\infty}\lambda_n^2\|v\|^2\right]\\ &+(1+K)2\sum_{n=1}^{+\infty}\lambda_n\beta_n\left[\sup_{x^*\in\mathbf{zer}(B)} F_B\left(\frac{p}{\beta_n},x^*\right)-\sigma_{\mathbf{zer}(B)}\left(\frac{p}{\beta_n},x^*\right)\right]\\ &+K(1+K)(2M\gamma+1)\sum_{n=1}^{+\infty}\|x_n-y_n\|^2.\end{aligned}$$

By all assumptions in **Assumption** 3.1 and Proposition 4.2, we have

$$2\gamma\sum_{n=1}^{+\infty}\lambda_n\|x_{n+1}-x^*\|^2<+\infty.$$

From (4.28) and Lemma 3.4, we obtain that $\lim_{n\to+\infty}\|x_n-x^*\|$ exists.

Since $\sum_{n=1}^{+\infty}\lambda_n=+\infty$, we have $\lim_{n\to+\infty}\|x_n-x^*\|=0$. This completes the proof. □

5. Applications to constrained convex minimization problem

In this section, we will apply the results obtained in the previous section to solve constrained convex optimization problems. The constrained convex minimization problem:

$$\min_{x\in\arg\min g} f(x), \tag{5.29}$$

where $f:\mathcal{H}\to\overline{\mathbb{R}}$ is proper convex lower semicontinuous and $g:\mathcal{H}\to\mathbb{R}$ is (Fréchet) differentiable function on the space $\mathcal{H}$ and the gradient ∇g is Lipschitz continuous operator with constants L_g. We will assume that the set $\arg\min g$ is nonempty. Throughout the paper we also assume that the solution set $\mathcal{S}:=\arg\min\{f(x):x\in\arg\min g\}$ of the problem (5.29) is a nonempty set. Furthermore, without loss of generality, we may assume that $\min g=0$. Notice that f is proper convex lower semicontinuous, we have that the subdifferential ∂f is maximally monotone. Moreover, since the function g is convex differentiable, by using the Theorem of Baillon-Haddad (see [10, Corollary 18.16]), ∇g is $\frac{1}{L_g}$-cocoercive and $\arg\min g=\mathbf{zer}(\nabla g)$ can be used to solve the constrained convex monimization problem (5.29) reduces to solve the monotone inclusion

$$0\in\partial f(x)+N_{\arg\min g}(x).$$

By using this and algorithm (**NFBP**) (1.5), we will consider the following algorithm.

Algorithm 5.4. *Let $\{\lambda_n\}_{n=1}^{\infty}$, $\{\beta_n\}_{n=1}^{\infty}$ and $\{\alpha_n\}_{n=1}^{\infty}$ be sequences of positive real numbers. Algorithm for solving* (5.29) *as follows:*

$$\begin{cases} x_1\in\mathcal{H};\\ y_n=\mathbf{prox}_{\lambda_n f}(x_n-\lambda_n\beta_n\nabla g(x_n))\\ x_{n+1}=y_n+\alpha_n(y_n-x_n) \qquad \text{for all } n\geq 1.\end{cases}$$

In order to obtain the convergence of the sequence generated by **Algorithm** (5.4), we have to assume the following assumption.

Assumption 5.5. (a) The qualification condition $\arg\min g \cap \operatorname{int} \mathbf{Dom}(f) \neq \emptyset$ holds.
(b) $\{\lambda_n\} \in l^2 \setminus l^1$, $\lim\limits_{n\to\infty} \alpha_n = 0$ and $0 < \liminf\limits_{n\to\infty} \lambda_n\beta_n \leq \limsup\limits_{n\to\infty} \lambda_n\beta_n < \frac{1}{L_g}$.
(c) For each $p \in \mathbf{Ran}(N_{\arg\min g})$, we have

$$\sum_{n=1}^{+\infty} \lambda_n\beta_n \left[g^*\left(\frac{p}{\beta_n}\right) - \sigma_{\arg\min g}\left(\frac{p}{\beta_n}\right)\right] < +\infty.$$

Note that $F_{\nabla g}\left(x^*, \frac{p}{\beta_n}\right) \leq g(x^*) + g^*\left(\frac{p}{\beta_n}\right) = g^*\left(\frac{p}{\beta_n}\right)$ for all $x^* \in \arg\min g$, we have $\sup\limits_{x^*\in\arg\min g} F_{\nabla g}\left(x^*, \frac{p}{\beta_n}\right) \leq g^*\left(\frac{p}{\beta_n}\right)$. Hence, conditions (a)-(c) in **Assumption** 5.5 imply hypotheses (I)-(III) in **Assumption** 3.1.

Corollary 5.1. *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence generated by* **Algorithm** 5.4 *and $\{z_n\}_{n=1}^{\infty}$ be a sequence of weighted averages as* (2.6). *Suppose that all assumptions in* **Assumption** 5.5 *hold. Then the sequence $\{z_n\}_{n=1}^{\infty}$ converges weakly to an element in $\mathcal{S}$.*

If we assume that the function f is strongly convex, then its subdifferential ∂f is strongly monotone.

Corollary 5.2. *Let $\{x_n\}_{n=1}^{\infty}$ be a sequence generated by* **Algorithm** 5.4 *and the function f be a γ-strongly convex with $\gamma > 0$. If all assumptions in* **Assumption** 5.5 *hold, then the sequence $\{x_n\}_{n=1}^{\infty}$ converges strongly to the unique $\mathcal{S}$.*

6. NUMERICAL EXPERIMENTS

In this section, we present an example of numerical set for testing the purposed algorithm. Some comparisons of our algorithm (**NFBP**) (1.5) with the algorithm (**FB**) introduced by Attouch [5] are also reported. All the experiments are implemented in MATLAB R2015b running on a MacBook air 13-inch, Early 2017 with a 1.8 GHz Intel Core i5 processor and 8 GB 1600 MHz DDR3 memory.

We consider the problem with equality constraints:

$$\begin{aligned} &\text{minimize } \|x\|_1 \\ &\text{subject to } \mathbf{A}x = \mathbf{b}, \end{aligned} \tag{6.30}$$

where $\mathbf{A} \in \mathbb{R}^{l\times k}$, $\mathbf{b} \in \mathbb{R}^l$. In addition, we assume that $k > l$. The problem (6.30) can be written in the form of the problem (5.29) as :

$$\begin{aligned} &\text{minimize } f(x) := \|x\|_1 \\ &\text{subject to } x \in \arg\min g, \end{aligned}$$

where $g(z) := \frac{1}{2}\|\mathbf{A}z - \mathbf{b}\|^2$, for all $z \in \mathbb{R}^k$.

In this setting, we have $\nabla g(z) = \mathbf{A}^T(\mathbf{A}x - \mathbf{b})$ and notice that ∇g is $\|\mathbf{A}\|^2$-Lipschitz continuous. We also get that
$\mathbf{prox}_{\lambda_n f}(x) = \left(\max\left(0, 1 - \frac{\lambda_n}{|x_1|}\right)x_1, \max\left(0, 1 - \frac{\lambda_n}{|x_2|}\right)x_2, ..., \max\left(0, 1 - \frac{\lambda_n}{|x_k|}\right)x_k\right)$.

We begin with the problem by random vectors $x_1 \in \mathbb{R}^k$, $\mathbf{b} \in \mathbb{R}^l$ and matrix $\mathbf{A} \in \mathbb{R}^{l\times k}$. Next, we compare the performance of our algorithm (**NFBP**) (1.5) with the algorithm (**FB**). The used of parameters in two algorithms are chosen as follows:

$\beta_n = \frac{n}{(\|\mathbf{A}\|^2)+1}$, $\lambda_n = \frac{1}{n}$, $\forall n \geq 1$. We obtain the CPU times (seconds) and the number of iterations by using the stopping criteria : $\|x_n - x_{n-1}\| \leq 10^{-6}$.

Algorithm	CPU times (s)	Iterations
(**FB**) ($\alpha_n = 0$)	180.44	38352
(**NFBP**) ($\alpha_n = 1/\sqrt{n+1}$)	140.36	35649
(**NFBP**) ($\alpha_n = 1/(n+1)$)	155.79	35589
(**NFBP**) ($\alpha_n = 1/(n+1)^2$)	136.01	33841
(**NFBP**) ($\alpha_n = 1/(n+1)^4$)	150.45	37164
(**NFBP**) ($\alpha_n = 1/(n+1)^{10}$)	154.40	38344

TABLE 1. Comparison of number of iterations and CPU computation time between (**NFBP**) (1.5) and (**FB**) with difference of parameter sequences $\{\alpha_n\}_{n=1}^{\infty}$.

We compare the performance of our algorithm (**NFBP**) (1.5) and (**FB**) algorithm for case $k = 4000$, $l = 1000$ with difference of parameter sequences $\{\alpha_n\}_{n=1}^{\infty}$. The results are reported in table 1. We observe that (**FB**) spends more CPU computation time than our algorithm (**NFBP**) (1.5). We can see that when $\alpha_n = \frac{1}{(n+1)^2}$, it leads to the lowest CPU computation time and number of iterations for (**NFBP**) (1.5) of 136.01 seconds and 33841 times, respectively. We also observe that our algorithm (**NFBP**) (1.5) requires less iterations than (**FB**) for all choice of parameter sequences $\{\alpha_n\}_{n=1}^{\infty}$.

(l,k)	(**NFBP**) (1.5)		(**FB**)	
	CPU time (s)	Iterations	CPU time (s)	Iterations
(20,1000)	1.99	34160	5.13	83860
(50,1000)	2.32	32986	5.72	77435
(100,1000)	2.92	30352	7.38	79054
(200,1000)	3.94	30546	7.35	56337
(300,1000)	5.17	26191	6.70	33513
(20,2000)	4.14	37505	11.14	98780
(50,2000)	6.14	45289	10.55	78691
(100,2000)	5.00	27642	10.42	58504
(200,2000)	8.33	24207	23.45	67317
(300,2000)	15.71	27088	28.22	48109
(20,5000)	10.17	40463	25.04	96251
(50,5000)	7.09	22812	21.79	68287
(100,5000)	18.70	29416	42.56	66194
(200,5000)	40.66	33008	92.56	77144
(300,5000)	51.59	27193	123.60	58909

TABLE 2. The comparison of two algorithms with different sizes of matrix **A**.

In table 2, we present a comparison between the numerical results of (**NFBP**) (1.5) and (**FB**) cases for $\alpha_n = \frac{1}{\sqrt{n+1}}, \forall n \geq 1$ and different sizes of matrix **A**. We can see that the number of iterations of (**NFBP**) (1.5) are smaller than of (**FB**) for all different sizes of matrix **A**. Furthermore, (**NFBP**) (1.5) requires less CPU computation time to reach the optimality tolerance than (**FB**) for all cases.

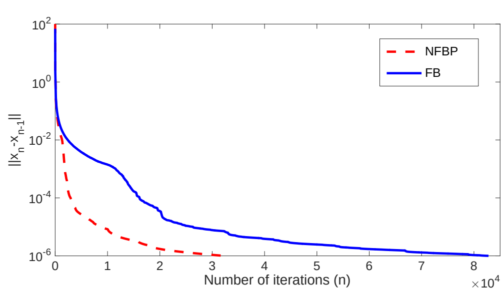

FIGURE 1. Illustration of the behavior of $\|x_n - x_{n-1}\|$ for (**NFBP**) (1.5) and (**FB**) methods when $\alpha_n = \frac{1}{\sqrt{n+1}}$ and $(l,k) = (100, 3000)$.

Figure 1 shows the behavior of $\|x_n - x_{n-1}\|$ for (**NFBP**) (1.5) and (**FB**) methods when $\alpha_n = \frac{1}{\sqrt{n+1}}$ and $(l, k) = (100, 3000)$. We can observe that by using our algorithm (**NFBP**) (1.5) the behavior of the red line (**NFBP**) (1.5) performs better than the blue line (**FB**).

7. Conclusions

In this paper, we purposed a new forward-backward method with penalization term (**NFBP**) (1.5) for solving monotone inclusion problems (1.4). We provide the sufficient conditions to guarantee the convegences of **(NFBP)** (1.5) for the considered problems. We also provide a numerical example to compare between our algorithm (**NFBP**) (1.5) and the (**FB**) algorithm. As a result, we observe that our algorithm (**NFBP**) (1.5) performs better behavior when comparing with algorithm (**FB**) for all different cases.

Acknowledgements. The second author would like to thank Naresuan University and The Thailand Research Fund for financial support. Moreover, N. Artsawang is also supported by The Royal Golden Jubilee Program under Grant PHD/0158/2557, Thailand.
Disclosure statement The authors declare that there is no conflict of interests regarding the publication of this paper.
Funding N. Artsawang was supported by the Thailand Research Fund through the Royal Golden Jubilee PhD Program under Grant PHD/0158/2557, Thailand.

References

[1] Alvarez, F., *On the minimizing property of a second order dissipative system in Hilbert spaces*, SIAM J. Control. Optim., **38** (2000), 1102–1119
[2] Alvarez, F., *Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space*, SIAM J. Optim., **14** (2004), 773–782
[3] Alvarez, F. and Attouch, H., *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping*, Set-Valued Anal., **9** (2001), 3–11
[4] Attouch, H. and Czarnecki, M.-O., *Asymptotic behavior of coupled dynamical systems with multiscale aspects* , J. Differ. Equat., **248** (2010), 1315–1344
[5] Attouch, H., Czarnecki, M.-O. and Peypouquet, J., *Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities*, SIAM J. Optim., **21** (2011), 1251–1274
[6] Attouch, H., Czarnecki, M.-O. and Peypouquet, J., *Prox-penalization and splitting methods for constrained variational problems*, SIAM J. Optim., **21** (2011), 149–173
[7] Attouch, H., Peypouquet, J. and Redont, P., *A dynamical approach to an inertial forward-backward algorithm for convex minimization*, SIAM J. Optim., **24** (2014), 232–256
[8] Bagirov, A., Karmitsa, N. and Mäkelä, M. M., *Introduction to Nonsmooth Optimization*, Theory, Practice and Software, Springer, New York, 2014
[9] Banert, S. and Boţ, R. I., *Backward penalty schemes for monotone inclusion problems*, J. Optim. Theory Appl., **166** (2015), 930–948
[10] Bauschke, H. H. and Combettes, P. L., *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics. New York:Springer, 2011
[11] Boţ, R. I. and Csetnek, E. R., *Forward-backward and Tseng's type penalty schemes for monotone inclusion problems*, Set-Valued Var. Anal., **22** (2014), 313–331
[12] Boţ, R. I. and Csetnek, E. R., *An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems*, Numer. Algorithms, **71** (2016), 519–540
[13] Boţ, R. I. and Csetnek, E. R., *An inertial alternating direction method of multipliers*, Minimax Theory Appl., **1** (2016), 29–49
[14] Boţ, R. I. and Csetnek, E. R., *A hybrid proximal-extragradient algorithm with inertial effects*, Numer. Funct. Anal. Optim., **36** (2015), 951–963
[15] Boţ, R. I. and Csetnek, E. R., *An inertial Tsengs type proximal algorithm for nonsmooth and nonconvex optimization problems*, J. Optim. Theory Appl., **171** (2016), 600–616
[16] Boţ, R. I. and Csetnek, E. R., *Approaching the solving of constrained variational inequalities via penalty term-based dynamical systems*, J. Math. Anal. Appl., **435** (2016), 1688–1700
[17] Boţ, R. I., Csetnek, E. R. and Hendrich, C., *Inertial DouglasRachford splitting for monotone inclusion problems*, Appl. Math. Comput., **256** (2015), 472–487

[18] Boţ, R. I., Csetnek, E. R. and László, S., *An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions*, EURO J. Comput. Optim., **4** (2016), 3–25

[19] Boţ, R. I., Csetnek, E. R. and Nimana, N., *Gradient-type penalty method with inertial effects for solving constrained convex optimization problems with smooth data*, Optim. Lett., https://doi.org/10.1007/s11590-017-1158-1, (2017)

[20] Boţ, R. I., Csetnek, E. R. and Nimana, N., *An Inertial Proximal-Gradient Penalization Scheme for Constrained Convex Optimization Problems*, Vietnam J. Math., 2017

[21] Boţ, R. I. and Nguyen, D.-K., *A forward-backward penalty scheme with inertial effects for montone inclusions*, Applications to convex bilevel programming, http://www. optimization-online.org/DB_HTML/2018/01/6416.html (2018)

[22] Cabot, A. and Frankel, P., *Asymptotics for some proximal-like method involving inertia and memory aspects*, Set-valued Var. Anal., **19** (2011), 59–74

[23] Chen, C., Chan, R. H., Ma, S. and Yang, J., *Inertial proximal ADMM for linearly constrained separable convex optimization*, SIAM J. Imag. Sci., **8** (2015), 2239–2267

[24] Chen, C., Ma, S. and Yang, J., *A general inertial proximal point algorithm for mixed variational inequality problem*, SIAM J. Optim., **25** (2015), 2120–2142

[25] Combettes, P. L., *Quasi-Fejérian analysis of some optimization algorithms*, in: Butnariu, D., Censor, Y., Reich, S. (eds.), Inherently Parallel Algorithms for Feasibility and Optimization, pp. 115–152, Elsevier, Amsterdam, (2001)

[26] Maingé, P.-E., *Convergence theorems for inertial KM-type algorithms*, J. Comput. Appl. Math., **219** (2008), 223–236

[27] Maingé, P.-E. and Moudafi, A., *Convergence of new inertial proximal methods for DC programming*, SIAM J. Optim., **19** (2008), 397–413

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

[29] Nimana, N. and Petrot, N., *Subgradient algorithm for split hierarchical optimization problems*, Carpathian J. Math., **34** (2018), No. 3, 391–399

[30] Noun, N. and Peypouquet, J., *Forward-backward penalty scheme for constrained convex minimization without inf-compactness*, J. Optim. Theory Appl., **158** (2013), 787–795

[31] Ochs, P., Chen, Y., Brox, T. and Pock, T., *iPiano: Inertial proximal algorithm for nonconvex optimization*, SIAM J. Imag. Sci., **7** (2014), 1388–1419

[32] Peypouquet, J., *Coupling the gradient method with a general exterior penalization scheme for convex minimization*, J. Optim. Theory Appl., **153** (2012), 123–138

[1]DEPARTMENT OF MATHEMATICS
FACULTY OF SCIENCE
NARESUAN UNIVERSITY
PHITSANULOK 65000, THAILAND
E-mail address: na_tta_pon@hotmail.com
E-mail address: kasamsuku@nu.ac.th

[2]RESEARCH CENTER FOR ACADEMIC EXCELLENCE IN NONLINEAR ANALYSIS AND OPTIMIZATION
FACULTY OF SCIENCE, PHITSANULOK 65000, THAILAND
E-mail address: kasamsuku@nu.ac.th