4OR (2021) 19:127–151
https://doi.org/10.1007/s10288-020-00440-3

RESEARCH PAPER

Modified forward–backward splitting method for variational inclusions

Dang Van Hieu[1] · **Pham Ky Anh**[2] · **Le Dung Muu**[3]

Received: 2 January 2019 / Revised: 28 March 2020 / Published online: 21 April 2020

Abstract
In this paper we propose an explicit algorithm for solving a variational inclusion problem of the sum of two operators, the one is maximally monotone and the other is monotone and Lipschitz continuous. The algorithm uses the variable stepsizes which are updated over each iteration by some cheap comptutations. These stepsizes are found without the prior knowledge of the Lipschitz constant of operator as well as without using lineseach procedure. The algorithm thus can be implemented easily. The convergence and the convergence rate of the algorithm are established under mild conditions. Several preliminary numerical results are provided to demonstrate the theoretical results and also to compare the new algorithm with some existing ones.

Keywords Forward–backward method · Tseng's method · Operator splitting method

Mathematics Subject Classification 65J15 · 47H05 · 47J25 · 47J20 · 91B50

1 Introduction

Let H be a real Hilbert space. In this paper, we consider the variational inclusion problem (shortly, VI) of finding a zero of the sum of two operators. More precisely, this problem is stated as follows:

✉ Dang Van Hieu
dangvanhieu@tdtu.edu.vn

Pham Ky Anh
anhpk@vnu.edu.vn

Le Dung Muu
ldmuu@math.ac.vn

1 Applied Analysis Research Group, Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, Vietnam

2 Department of Mathematics, Vietnam National University, Hanoi, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam

3 TIMAS, Thang Long University, Hanoi, Vietnam

$$\text{Find } x^* \in H \text{ such that } 0 \in (A+B)(x^*), \tag{VI}$$

where $A : H \rightrightarrows H$ is a (multi-valued) maximally monotone operator and $B : H \to H$ is a monotone and Lipschitz continuous operator. It was known that the problem (VI) plays a central role in nonlinear analysis around known mathematical models as optimization problems, split feasibility problems and convex programming with applications in signal and image processing, machine learning and others, see, e.g., Combettes and Wajs (2005), Daubechies et al. (2004), Duchi and Singer (2009), Raguet et al. (2013). Problem (VI) also contains properly the class of classical variational inequalities (Facchinei and Pang 2002; Gibali and Hieu 2019; Hieu et al. 2017, 2019a, b, c, 2020a, b; Hieu and Quy 2019; Kinderlehrer and Stampacchia 1980). We restrict here in our interest to the following two simple models:

Firstly, let f and g be two proper, lower semicontinuous and convex functions from H to the set of extended real numbers $\bar{\Re} := \Re \cup \{+\infty\}$ such that f is differentiable with Lipschitz continuous gradient ∇f and g is subdifferentiable with its computable proximal mapping. Consider the optimization problem of sum of two functions,

$$\min_{x \in H} f(x) + g(x). \tag{OP}$$

Then, the problem (OP) is equivalent to the problem (VI) with $A = \partial g$ (the subdiffential of g) and $B = \nabla f$. One particular example of this case is the so-called basis pursuit denoising problem in Chen et al. (1998),

$$\min_{x \in H_1} \frac{1}{2}||Tx - b||^2 + \lambda ||x||_1,$$

where $T : H_1 \to H_2$ is a bounded linear operator, $b \in H_2$ and $\lambda > 0$. This problem with the l_1—penalty is relative to a technique, called the LASSO, for least absolute shrinkage and selection operator in Tibshirami (1996).

Secondly, let H_1, H_2 be two Hilbert spaces and $T : H_1 \to H_2$ be a bounded linear operator with its adjoint T^*. Let $C \subset H_1$ and $Q \subset H_2$ be two nonempty closed covex subsets. The split feasibility problem (Censor and Elfving 1994) is stated as follows:

$$\text{Find } x^* \in C \text{ such that } Tx^* \in Q. \tag{SFP}$$

The problem (SFP) can be reformulated equivalently under the form of the problem (VI) by setting $B(x) = \nabla \left(\frac{1}{2}||Tx - P_Q Tx||^2\right) = T^*(I - P_Q)T(x)$ and $A(x) = N_C(x)$, the normal cone mapping of C. Several other models, which are formulated under the variational inclusion form (VI), can be found in Malitsky and Tam (2018), Ryu and Boyd (2016).

The problem (VI) has received a lot of attention by many authors who devoted their works to theoretical results as well as iterative algorithms, see, e.g., Attouch et al. (2018), Bruck (1977), Censor and Elfving (1994), Combettes and Wajs (2005), Davis and Yin (2017), Dong and Fischer (2010), Goldstein (1964), Huang and Dong (2014), Lions and Mercier (1979), Passty (1979), Raguet et al. (2013), Rockafellar (1976), Ryu and Boyd (2016), Zong et al. (2018). A commonly used method for solving problem

(VI) is the forward–backward method (FBM) (Lions and Mercier 1979; Passty 1979). Precisely, the FBM generates a sequence $\{x_n\}$, from a starting point $x_0 \in H$, defined by

$$x_{n+1} = J_{\lambda A}(x_n - \lambda B(x_n)), \tag{FBM}$$

where $J_{\lambda A} := (I + \lambda A)^{-1}$ is the resolvent of λA. Under the hypothesis of the β -cocoercivity of the operator B, i.e.,

$$\langle B(x) - B(y), x - y\rangle \geq \beta||B(x) - B(y)||^2,$$

then the sequence $\{x_n\}$ generated by Algorithm 3 converges weakly to some solution of problem (VI) . Note that the cocoercivity of an operator implies the Lipschitz continuity of operator. Without this assumption, the modified forward–backward splitting method (MFBSM) (Tseng 2000) can be more useful. More precisely, from a starting poin $x_0 \in H$, the MFBSM generates two sequences $\{x_n\}$ and $\{y_n\}$, defined by

$$y_n = J_{\lambda A}(x_n - \lambda B(x_n)), \quad x_{n+1} = y_n - \lambda(B(y_n) - B(x_n)). \tag{MFBSM}$$

The weak convergence of the sequences $\{x_n\}$ and $\{y_n\}$ is established under the assumptions of monotonicity and Lipschitzness of B. As be seen, the MFBSM requires an additional evaluation of value of operator B. Very recently, the authors in Malitsky and Tam (2018) have introduced an algorithm, named the forward-reflected-backward splitting method (FRBSM) without the cocoercivity, for solving problem (VI) . This algorithm is of the form,

$$x_{n+1} = J_{\lambda_n A}(x_n - \lambda_n B(x_n) - \lambda_{n-1}(B(x_n) - B(x_{n-1}))),\ x_0 = x_{-1} \in H, \tag{FRBSM}$$

where $\{\lambda_n\}$ is the sequence of suitable stepsizes. The convergence of the sequences $\{x_n\}$ and $\{y_n\}$ are established under the hypotheses of maximal monotonicity of A, of monotonicity and (local) Lipschitz continuity of B. The advantage of the (FRBSM) is that it only requires to compute one resolvent $J_{\lambda_n A}$ of A and one value of B at the current approximation. Then, the complexity of this algorithm is almost equivalent to the (FBM), while its convergence is formulated without the strong hypothesis of the cocoercivity of B. At this stage, it is worth mentioning that the stepsizes used in Malitsky and Tam (2018) either depend on the Lipschitz constant of B or are found by a linesearch procedure. In general, the Lipschitz constant is often unknown in practice, or even in nonlinear problems it can be difficult to approximate. In those cases, an algorithm with a linesearch is often used. However, a linesearch algorithm needs an inner loop with some stopping criterion over iteration, and this task may be time-consuming.

In this paper, we revisit the (FRBSM) and present an explicit algorithm for solving problem (VI). The resulting algorithm generates the variable stepsizes which do not depend on the Lipschitz constant of operator and without any linesearch procedure, and it thus can be implemented more easily. The convergence as well as the convergence rate of the algorithm are established under mild conditions. Several numerical experiments are also given. The organization of this paper is as follows: Sect. 2 recalls some definitions and results used further. Section 3 deals with the description of the new algorithm and the analyses of its convergence. Finally, in Sect. 4 we present a preliminary numerical experiment to illustrate the convergence of our algorithm in comparison with others.

2 Preliminaries

Let H be a real Hilbert space and $A : H \rightrightarrows H$ be a multi-valued mapping. Let us begin with some concepts of monotonicity of an operator. The multi-valued operator A is called:(i) *monotone* if $\langle u - v, x - y\rangle \geq 0$ for all $x,\ y \in H$ and $u \in A(x),\ v \in A(y)$; (ii) *maximally monotone* if it is monotone and its graph is not properly contained in the graph of any other monotone operator; (iii) *strongly monotone* if there exists a number $\gamma > 0$ such that $\langle u - v, x - y\rangle \geq \gamma||x - y||^2$ for all $x,\ y \in H$ and $u \in A(x),\ v \in A(y)$; (iv) *Lipschitz continuous* if there exists a number $L > 0$ such that $||u - v|| \leq L||x - y||$ for all $x,\ y \in H$ and $u \in A(x),\ v \in A(y)$.

It follows from the definition of maximal monotonicity that if for any $(x, u) \in H \times H$, $\langle u - v, x - y\rangle \geq 0$ for all $(y, v) \in$ Graph(A) (the *graph* of operator A) then $u \in A(x)$.

Let $A : H \rightrightarrows H$ be a multi-valued maximally monotone mapping. Recall the resolvent of A is a single-valued mapping $J_A : H \to H$, defined by

$$J_A(x) = (I + A)^{-1}(x),\ x \in H,$$

where I stands for the identity operator on H. We have the following lemmas.

Lemma 2.1 (Takahashi 2000, Sect. 4) *Let $A : H \rightrightarrows H$ be a maximally monotone mapping and $B : H \to H$ is a mapping. For each $\lambda > 0$, define $T_\lambda(x) := J_{\lambda A}(x - \lambda B(x))$ for all $x \in H$. Then, $x \in (A + B)^{-1}(0)$ if and only if $x \in Fix(T_\lambda)$, where $Fix(T_\lambda)$ is the fixed point set of T_λ.*

Lemma 2.2 (Brézis and Chapitre 1973, Lemma 2.4) *Let $A : H \rightrightarrows H$ be a maximally monotone mapping and $B : H \to H$ be a Lipschitz continuous and monotone mapping. Then the mapping $A + B$ is a maximally monotone mapping.*

3 Algorithm

In this section, we present an explicit algorithm with a new stepsize rule for approximating a solution of problem (VI). We assume that A is maximally monotone and B is monotone and Lipschitz continuous with some constant L. However, the prior knowledge or an estimate of this constant is not necessary to be required. Moreover, the solution set $(A + B)^{-1}(0)$ of problem (VI) is assumed to be nonempty. For the sake of simplicity, we adopt the convention $\frac{0}{0} = +\infty$. The algorithm is described as follows:

Algorithm 3.1 (Modified FRBSM with New Stepsizes).

Initialization: *Choose* $x_{-1},\ x_0 \in H, \lambda_{-1},\ \lambda_0 > 0$ *and* $\mu \in \left(0, \frac{1}{2}\right)$.

Iterative Step: *Assume that* $x_{n-1},\ x_n \in H$ *are known, calculate* x_{n+1} *as follows*:

$$\begin{cases} x_{n+1} = J_{\lambda_n A}(x_n - \lambda_n B(x_n) - \lambda_{n-1}(B(x_n) - B(x_{n-1}))), \\ \lambda_{n+1} = \min\left\{\lambda_n, \frac{\mu||x_{n+1}-x_n||}{||B(x_{n+1})-B(x_n)||}\right\}. \end{cases}$$

Stoppig Criterion: *If* $x_{n+1} = x_n = x_{n-1}$ *then* x_n *is a solution of problem* (VI).

Unlike the forward backward splitting methods with fixed stepsize, the Lipschitz constant must not be the input parameter of Algorithm 3. It is particularly interesting because this constant is not often known, or even for nonlinear problems it is difficult to approximate. Comparing with algorithms with linesearch, the stepsizes $\{\lambda_n\}$ generated by Algorithm 3 are found more easily by some cheap computations.

Lemma 3.1 *If* $x_{n+1} = x_n = x_{n-1}$ *then* x_n *is a solution of problem* (VI), *i.e.,* $x_n \in (A + B)^{-1}(0)$.

The proof of Lemma 3.1 follows directly from Lemma 2.1 and the definition of x_{n+1}. Thus, it is omitted. Lemma 3.1 ensures that if Algorithm 3 terminates at some iteration n then a solution of problem (VI) can be found. In what follows, we assume that Algorithm 3 does not stop, and then we focus on studying the asymptotic behavior of the sequence $\{x_n\}$ generated by Algorithm 3.

3.1 Convergence of Algorithm 3

In this part, we study the convergence of the sequence $\{x_n\}$ generated by Algorithm 3. Let us define $\Xi_n(x^*)$ for each $n \geq 1$ and $x^* \in (A + B)^{-1}(0)$ by

$$\Xi_n(x^*) = ||x_n - x^*||^2 + 2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_n - x^*\right\rangle + \frac{\mu\lambda_{n-1}}{\lambda_n}||x_{n-1} - x_n||^2. \quad (1)$$

The following lemma plays an important role in formulating the convergence of $\{x_n\}$.

Lemma 3.2 *The following estimate holds for all* $n \geq 1$ *and* $x^* \in (A + B)^{-1}(0)$

$$\Xi_{n+1}(x^*) \leq \Xi_n(x^*) - \left(1 - \frac{\mu\lambda_{n-1}}{\lambda_n} - \frac{\mu\lambda_n}{\lambda_{n+1}}\right)||x_{n+1} - x_n||^2.$$

Proof It follows from the definitions of x_{n+1} and $J_{\lambda_n A}$ that $x_n - \lambda_n B(x_n) - \lambda_{n-1}(B(x_n) - B(x_{n-1})) \in (I + \lambda_n A)(x_{n+1})$ or $x_n - x_{n+1} - \lambda_n B(x_n) - \lambda_{n-1}(B(x_n) - B(x_{n-1})) \in \lambda_n A(x_{n+1})$. Moreover, since x^* is the solution of problem (VI), we have that $-\lambda_n B(x^*) \in \lambda_n A(x^*)$. Thus, from the monotonicity of A, we obtain

$$\left\langle x_n - x_{n+1} - \lambda_n B(x_n) - \lambda_{n-1}(B(x_n) - B(x_{n-1})) + \lambda_n B(x^*), x_{n+1} - x^* \right\rangle \geq 0.$$

Thus, multiplying both sides of this inequality by 2 and using the monotonicity of B, we obtain

$$\begin{aligned}
0 &\leq 2\left\langle x_n - x_{n+1}, x_{n+1} - x^* \right\rangle - 2\lambda_n \left\langle B(x_n) - B(x^*), x_{n+1} - x^* \right\rangle \\
&\quad +2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_{n+1} - x^* \right\rangle \\
&= 2\left\langle x_n - x_{n+1}, x_{n+1} - x^* \right\rangle - 2\lambda_n \left\langle B(x_n) - B(x_{n+1}), x_{n+1} - x^* \right\rangle \\
&\quad -2\lambda_n \left\langle B(x_{n+1}) - B(x^*), x_{n+1} - x^* \right\rangle + 2\lambda_{n-1}\langle B(x_{n-1}) \\
&\qquad -B(x_n), x_{n+1} - x_n \rangle \\
&\quad +2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_n - x^* \right\rangle \\
&\leq 2\left\langle x_n - x_{n+1}, x_{n+1} - x^* \right\rangle - 2\lambda_n \left\langle B(x_n) - B(x_{n+1}), x_{n+1} - x^* \right\rangle \\
&\quad +2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_n - x^* \right\rangle \\
&\quad +2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_{n+1} - x_n \right\rangle.
\end{aligned} \tag{2}$$

Thus, by using the equality $2\langle a, b \rangle = ||a+b||^2 - ||a||^2 - ||b||^2$, we obtain

$$\begin{aligned}
0 &\leq ||x_n - x^*||^2 - ||x_n - x_{n+1}||^2 - ||x_{n+1} - x^*||^2 \\
&\quad -2\lambda_n \left\langle B(x_n) - B(x_{n+1}), x_{n+1} - x^* \right\rangle \\
&\quad +2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_n - x^* \right\rangle \\
&\quad +2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_{n+1} - x_n \right\rangle.
\end{aligned} \tag{3}$$

Now, we estimate the last term of inequality (3). From the facts $\langle a, b \rangle \leq ||a||||b||$, $2xy \leq x^2 + y^2$ and the definition of λ_n, we obtain

$$\begin{aligned}
2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_{n+1} - x_n \right\rangle &\leq 2\lambda_{n-1}||B(x_{n-1}) - B(x_n)||||x_{n+1} - x_n|| \\
&\leq 2\lambda_{n-1}\frac{\mu||x_{n-1} - x_n||}{\lambda_n}||x_{n+1} - x_n|| \\
&\leq \frac{\mu\lambda_{n-1}}{\lambda_n}||x_{n-1} - x_n||^2 \\
&\qquad + \frac{\mu\lambda_{n-1}}{\lambda_n}||x_{n+1} - x_n||^2.
\end{aligned} \tag{4}$$

Combining the relations (3) and (4), we obtain

$$
\begin{aligned}
&||x_{n+1}-x^*||^2 + 2\lambda_n \left\langle B(x_n) - B(x_{n+1}), x_{n+1} - x^* \right\rangle \\
&\quad + \left(1 - \frac{\mu\lambda_{n-1}}{\lambda_n}\right) ||x_{n+1} - x_n||^2 \\
&\le ||x_n - x^*||^2 + 2\lambda_{n-1} \left\langle B(x_{n-1}) - B(x_n),\ \ x_n - x^* \right\rangle \\
&\quad + \frac{\mu\lambda_{n-1}}{\lambda_n} ||x_{n-1} - x_n||^2. \qquad (5)
\end{aligned}
$$

Using the definition of $\Xi_n(x^*)$, we can rewrite the relation (5) shortly as

$$
\Xi_{n+1}(x^*) \le \Xi_n(x^*) - \left(1 - \frac{\mu\lambda_{n-1}}{\lambda_n} - \frac{\mu\lambda_n}{\lambda_{n+1}}\right) ||x_{n+1} - x_n||^2.
$$

This finishes the proof. ◻

Before establishing the convergence of Algorithm 3, we have the following remark.

Remark 3.1 From the definition, we see that the sequence $\{\lambda_n\}$ is non-increasing. Moreover, since B is Lipschitz continuous, we obtain that $||B(x_{n+1}) - B(x_n)|| \le L||x_{n+1} - x_n||$ where L is some positive number. Thus, by induction, we can prove that the sequence $\{\lambda_n\}$ is bounded from below by $\min\left\{\lambda_0, \frac{\mu}{L}\right\}$, hence $\lim_{n\to\infty} \lambda_n = \lambda > 0$. This means that the sequence of stepsizes $\{\lambda_n\}$ generated by Algorithm 3 is separated from 0. Then, the stepsize rule here in general is much better than the requirement of non-summable and diminishing stepsizes in Hieu et al. (2019d), because it is well known that an algorithm which uses a sequence of diminishing stepsizes often provides slow convergence.

Theorem 3.1 *Let $A : H \rightrightarrows H$ be maximally monotone and $B : H \to H$ be monotone and Lipschitz continuous, and assume that $(A+B)^{-1}(0) \ne \emptyset$. Then, the sequence $\{x_n\}$ generated by Algorithm 3 converges weakly to some solution of problem (VI).*

Proof Let $\bar{\delta} \in (0, 1-2\mu)$ be fixed. Since $\lambda_n \to \lambda > 0$, we derive

$$
\lim_{n\to\infty} \left(1 - \frac{\mu\lambda_{n-1}}{\lambda_n} - \frac{\mu\lambda_n}{\lambda_{n+1}}\right) = 1 - 2\mu > \bar{\delta}.
$$

Thus, there exists a number $n_1 \ge 1$ such that

$$
1 - \frac{\mu\lambda_{n-1}}{\lambda_n} - \frac{\mu\lambda_n}{\lambda_{n+1}} > \bar{\delta},\ \forall n \ge n_1. \qquad (6)
$$

It follows from Lemma 3.2 and the relation (6) that

$$
\Xi_{n+1} \le \Xi_n - \bar{\delta}||x_{n+1} - x_n||^2,\ \forall n \ge n_1. \qquad (7)
$$

Thus, the sequence $\{\Xi_n\}_{n\geq n_1}$ is non-increasing. From the definitions of Ξ_n and λ_n, for each $n \geq n_1$, we see that

$$\begin{aligned}
\Xi_n &= ||x_n - x^*||^2 + 2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_n - x^*\right\rangle \\
&\quad + \frac{\mu\lambda_{n-1}}{\lambda_n}||x_{n-1} - x_n||^2 \\
&\geq ||x_n - x^*||^2 - 2\lambda_{n-1}||B(x_{n-1}) - B(x_n)||||x_n - x^*|| \\
&\quad + \frac{\mu\lambda_{n-1}}{\lambda_n}||x_{n-1} - x_n||^2 \\
&\geq ||x_n - x^*||^2 - 2\frac{\mu\lambda_{n-1}}{\lambda_n}||x_{n-1} - x_n||||x_n - x^*|| \\
&\quad + \frac{\mu\lambda_{n-1}}{\lambda_n}||x_{n-1} - x_n||^2 \\
&\geq ||x_n - x^*||^2 - \frac{\mu\lambda_{n-1}}{\lambda_n}\left(||x_{n-1} - x_n||^2 + ||x_n - x^*||^2\right) \\
&\quad + \frac{\mu\lambda_{n-1}}{\lambda_n}||x_{n-1} - x_n||^2 \\
&= \left(1 - \frac{\mu\lambda_{n-1}}{\lambda_n}\right)||x_n - x^*||^2 \\
&\geq \left(1 - \frac{\mu\lambda_{n-1}}{\lambda_n} - \frac{\mu\lambda_n}{\lambda_{n+1}}\right)||x_n - x^*||^2 \\
&\geq \bar{\delta}||x_n - x^*||^2. \qquad (8)
\end{aligned}$$

Thus, the sequence $\{\Xi_n\}_{n\geq n_1}$ is non-negative, hence the limit of $\{\Xi_n\}$ exists, which from the relation (8), follows the boundedness of $\{x_n\}$. Morever, from the relation (7), we obtain

$$\bar{\delta}\sum_{n=n_1}^{\infty}||x_{n+1} - x_n||^2 \leq \Xi_{n_1} - \lim_{n\to\infty}\Xi_{n+1} < +\infty.$$

Thus

$$\lim_{n\to\infty}||x_{n+1} - x_n||^2 = 0. \qquad (9)$$

Since B is Lipschitz continuous, the boundedness of $\{x_n\}$ and the fact $\lambda_n \to \lambda > 0$, we obtain from the relation (9) that

$$\lim_{n\to\infty}\left[2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_n - x^*\right\rangle + \frac{\mu\lambda_{n-1}}{\lambda_n}||x_{n-1} - x_n||^2\right] = 0.$$

Thus, from the definition of Ξ_n, we find that $\lim_{n\to\infty}\Xi_n = \lim_{n\to\infty}||x_n - x^*||^2$. This means that the limit of $\left\{||x_n - x^*||^2\right\}$ exists for each x^* which solves problem (VI).

Now, we prove that every weak cluster point of $\{x_n\}$ belongs to $(A + B)^{-1}(0)$. Indeed, let p be a weakly cluster point of $\{x_n\}$, i.e., there exists a subsequence $\{x_m\}$ of the sequence $\{x_n\}$ converging weakly to p. It follows from the definition of x_{n+1} that $x_n - \lambda_n B(x_n) - \lambda_{n-1}(B(x_n) - B(x_{n-1})) \in (I + \lambda_n A)(x_{n+1})$. Thus

$$\begin{aligned}&\frac{x_n - x_{n+1}}{\lambda_n} + (B(x_{n+1}) - B(x_n))\\&\quad - \frac{\lambda_{n-1}}{\lambda_n}(B(x_n) - B(x_{n-1})) \in (A + B)(x_{n+1}).\end{aligned} \tag{10}$$

Since B is Lipschitz continuous, we obtain from the relation (9) that

$$\lim_{n\to\infty}(B(x_{n+1}) - B(x_n)) = \lim_{n\to\infty}(B(x_n) - B(x_{n-1})) = 0. \tag{11}$$

From Lemma 2.2, we see that $A + B$ is also maximally monotone. Thus, its graph is demiclosed. Now, passing to the limit in (10) as $n = m \to \infty$ and using relations (9), (11) and the fact $\lambda_n \to \lambda > 0$, we obtain that $0 \in (A+B)(p)$, i.e., $p \in (A+B)^{-1}(0)$.

To complete the proof, we show the whole sequence $\{x_n\}$ converges weakly to p. Indeed, assume that $p' \neq p$ is another weakly cluster point of $\{x_n\}$, i.e., there exists a subsequence $\{x_k\}$ of $\{x_n\}$ converging weakly to p' which solves problem (VI). Note that $p' \in (A + B)^{-1}(0)$, as mentioned above, the limits $\lim\limits_{n\to\infty} ||x_n - p||^2$ and $\lim\limits_{n\to\infty} ||x_n - p'||^2$ exist. We have the following equality,

$$2\left\langle p - p', x_n\right\rangle = ||x_n - p'||^2 - ||x_n - p||^2 + ||p||^2 - ||p'||^2.$$

Thus, the limit of $\left\{\left\langle p - p', x_n\right\rangle\right\}$ denoted by l exists, i.e.,

$$\lim_{n\to\infty}\left\langle p - p', x_n\right\rangle = l. \tag{12}$$

Passing to the limit in (12) firstly as $n = m \to \infty$ and after $n = k \to \infty$, we obtain

$$\left\langle p - p', p\right\rangle = \lim_{m\to\infty}\left\langle p - p', x_m\right\rangle = l = \lim_{k\to\infty}\left\langle p - p', x_k\right\rangle = \left\langle p - p', p'\right\rangle$$

which follows $||p - p'||^2 = 0$ or $p = p'$. This completes the proof. □

Now, we consider the following variational inequality problem (VIP),

$$\text{Find } x^* \in C \text{ such that } \left\langle F(x^*), y - x^*\right\rangle \geq 0,\ \forall y \in C, \tag{VIP}$$

where C is a nonempty closed convex subset in H and $F : H \to H$ is a monotone and Lipschitz continuous operator. Let δ_C be the indicator mapping of C and N_C is the normal cone of C. Let $A(x) = \partial\delta_C(x) = N_C(x)$ and $B(x) = F(x)$. Note that N_C is maximally monotone. In this case, the problem (VIP) is equivalent to our problem (VI). Moreover, we have that $J_{\lambda A}(x) = P_C(x)$ for all $x \in H$ and $\lambda > 0$. The following corollary follows directly from Theorem 3.1.

Corollary 3.1 *Let C be a nonempty closed convex subset of H and $B : H \to H$ be a monotone and Lipschitz continuous operator. Choose $x_{-1},\ x_0 \in H$, $\lambda_{-1},\ \lambda_0 > 0$ and $\mu \in \left(0, \frac{1}{2}\right)$. Let $\{x_n\}$ be the sequence, defined by*

$$\begin{cases} x_{n+1} = P_C(x_n - \lambda_n B(x_n) - \lambda_{n-1}(B(x_n) - B(x_{n-1}))), \\ \lambda_{n+1} = \min\left\{\lambda_n, \frac{\mu||x_{n+1}-x_n||}{||B(x_{n+1})-B(x_n)||}\right\}. \end{cases}$$

Then, the sequence $\{x_n\}$ converges weakly to some solution of problem (VIP).

3.2 Convergence rate of Algorithm 3

In this subsection, we study the convergence rate of Algorithm 3. For that purpose, we suppose that the conditions of Theorem 3.1 are satisfied. Moreover, the operator $A : H \rightrightarrows H$ is assumed to be strongly monotone (SM), i.e, there exists some number $\gamma > 0$ such that

$$\langle u - v, x - y\rangle \geq \gamma||x - y||^2,\ \forall x,\ y \in H,\ u \in A(x),\ v \in A(y). \tag{SM}$$

This is a common assumption to establish the convergece rate of algorithms. Denote by $x^\dagger$ the unique solution of problem (VI). We prove the following theorem.

Theorem 3.2 *Under the conditions in Theorem 3.1 and condition (SM), the sequence $\{x_n\}$ generated by Algorithm 3 converges R-linearly to the unique solution $x^\dagger$ of problem (VI).*

Proof We first take two numbers p and k such that

$$\mu < p < 1, \qquad \frac{\mu}{p} < k < 1. \tag{13}$$

Since $0 < \mu < \frac{1}{2}$, we obtain that $\frac{1-\mu}{\mu} > 1$. Hence $\frac{1-\mu}{\mu} - p > 1 - p$ or

$$\frac{\frac{1-\mu}{\mu} - p}{1 - p} > 1.$$

Thus, there exists a number r such that

$$1 < r < \min\left\{\frac{\frac{1-\mu}{\mu} - p}{1 - p}, 1 + \frac{2\gamma\lambda}{1 - k}\right\}, \tag{14}$$

where $0 < \lambda = \lim_{n\to\infty} \lambda_n$. For each $n \geq 0$, we define Γ_n and Υ_n by

$$\Gamma_n = (1-k)||x_n - x^\dagger||^2 + \frac{(1-p)\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2. \tag{15}$$

$$\Upsilon_n = k||x_n - x^\dagger||^2 + 2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_n - x^\dagger\right\rangle + \frac{p\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2. \tag{16}$$

We divide the proof of Theorem 3.2 into four steps.

Claim 1: There exists a number $n_2 \geq 1$ such that $\Gamma_n \geq 0$ and $\Upsilon_n \geq 0$ for all $n \geq n_2$.

The proof of Claim 1. It is obvious that $\Gamma_n \geq 0$ for all $n \geq 0$. Using the definitions of Υ_n, λ_n, and the facts $\langle x, y\rangle \leq ||x||||y||$ and $2ab \leq \frac{1}{p}a^2 + pb^2$, we obtain

$$\begin{aligned}
\Upsilon_n &= k||x_n - x^\dagger||^2 + 2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_n - x^\dagger\right\rangle \\
&\quad + \frac{p\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2 \\
&\geq k||x_n - x^\dagger||^2 - 2\lambda_{n-1}||B(x_{n-1}) - B(x_n)||||x_n - x^\dagger|| \\
&\quad + \frac{p\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2 \\
&\geq k||x_n - x^\dagger||^2 - 2\lambda_{n-1}\frac{\mu||x_{n-1} - x_n||}{\lambda_n}||x_n - x^\dagger|| \\
&\quad + \frac{p\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2 \\
&\geq k||x_n - x^\dagger||^2 - \frac{\mu\lambda_{n-1}}{\lambda_n}\left(\frac{1}{p}||x_n - x^\dagger||^2 + p||x_{n-1} - x_n||^2\right) \\
&\quad + \frac{p\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2 \\
&= \left(k - \frac{\mu\lambda_{n-1}}{p\lambda_n}\right)||x_n - x^\dagger||^2.
\end{aligned} \tag{17}$$

From the relation (13) and the fact $\lambda_n \to \lambda > 0$, we obtain

$$\lim_{n\to\infty}\left(k - \frac{\mu\lambda_{n-1}}{p\lambda_n}\right) = k - \frac{\mu}{p} > 0.$$

Thus, there exists $n_2 \geq 1$ such that $k - \frac{\mu\lambda_{n-1}}{p\lambda_n} > 0$ for all $n \geq n_2$. Hence $\Upsilon_n \geq 0$ for all $n \geq n_2$.

Claim 2: There exist two numbers $\rho > 0$ and $n_3 \geq n_2$ such that $\rho\Upsilon_n \leq \Gamma_n$ for all $n \geq n_3$.

The proof of Claim 2. As the relation (17), we have

$$\begin{aligned}
\Upsilon_n &= k||x_n - x^\dagger||^2 + 2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_n - x^\dagger\right\rangle \\
&\quad + \frac{p\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2 \\
&\le k||x_n - x^\dagger||^2 + 2\lambda_{n-1}||B(x_{n-1}) - B(x_n)||||x_n - x^\dagger|| \\
&\quad + \frac{p\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2 \\
&\le k||x_n - x^\dagger||^2 + 2\lambda_{n-1}\frac{\mu||x_{n-1} - x_n||}{\lambda_n}||x_n - x^\dagger|| \\
&\quad + \frac{p\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2 \\
&\le k||x_n - x^\dagger||^2 + \frac{\mu\lambda_{n-1}}{\lambda_n}\left(\frac{1}{p}||x_n - x^\dagger||^2 + p||x_{n-1} - x_n||^2\right) \\
&\quad + \frac{p\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2 \\
&= \left(k + \frac{\mu\lambda_{n-1}}{p\lambda_n}\right)||x_n - x^\dagger||^2 + \frac{(1+p)\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2. \qquad (18)
\end{aligned}$$

Recall that

$$\Gamma_n = (1-k)||x_n - x^\dagger||^2 + \frac{(1-p)\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2. \qquad (19)$$

Now, for each n, we set

$$\rho_n = \min\left\{\frac{1-k}{k + \frac{\mu\lambda_{n-1}}{p\lambda_n}}, \frac{1-p}{1+p}\right\}. \qquad (20)$$

It follows from the relations (18)–(20) that

$$\rho_n \Upsilon_n \le \Gamma_n \qquad (21)$$

Take ρ such that

$$0 < \rho < \min\left\{\frac{1-k}{k + \frac{\mu}{p}}, \frac{1-p}{1+p}\right\}. \qquad (22)$$

Thus, since $\frac{\lambda_{n-1}}{\lambda_n} \to 1$, from the relation (20), we derive

$$\lim_{n\to\infty}\rho_n = \min\left\{\frac{1-k}{k + \frac{\mu}{p}}, \frac{1-p}{1+p}\right\} > \rho > 0$$

Hence, there exists $n_3 \geq n_2$ such that $\rho_n > \rho$. This together with the relation (21) implies that

$$\rho \Upsilon_n \leq \Gamma_n, \ \forall n \geq n_3.$$

Claim 3: There exists a number $\bar{n}_0 \geq n_3$ such that

$$r\Gamma_{n+1} + \Upsilon_{n+1} \leq \Gamma_n + \Upsilon_n$$

for all $n \geq \bar{n}_0$ where r is defined in (14).

The proof of Claim 3. Using the γ-strong monotonicity of A and arguing similarly to the relation (5), we obtain

$$\begin{aligned}
&(1+2\gamma\lambda_n)||x_{n+1}-x^{\dagger}||^2 + 2\lambda_n \left\langle B(x_n) - B(x_{n+1}), x_{n+1} - x^{\dagger} \right\rangle \\
&\quad + \left(1 - \frac{\mu\lambda_{n-1}}{\lambda_n}\right)||x_{n+1} - x_n||^2 \\
&\leq ||x_n - x^{\dagger}||^2 + 2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_n - x^{\dagger} \right\rangle \\
&\quad + \frac{\mu\lambda_{n-1}}{\lambda_n}||x_{n-1} - x_n||^2.
\end{aligned} \tag{23}$$

Note that $\Gamma_n + \Upsilon_n = ||x_n - x^{\dagger}||^2 + 2\lambda_{n-1}\left\langle B(x_{n-1}) - B(x_n), x_n - x^{\dagger}\right\rangle + \frac{\mu\lambda_{n-1}}{\lambda_n}||x_n - x_{n-1}||^2$. Thus, the relation (23) can be rewritten as

$$\begin{aligned}
&(1+2\gamma\lambda_n)||x_{n+1}-x^{\dagger}||^2 + 2\lambda_n \left\langle B(x_n) - B(x_{n+1}), x_{n+1} - x^{\dagger} \right\rangle \\
&\quad + \left(1 - \frac{\mu\lambda_{n-1}}{\lambda_n}\right)||x_{n+1} - x_n||^2 \leq \Gamma_n + \Upsilon_n.
\end{aligned} \tag{24}$$

It follows from the defintions of Γ_n and Υ_n that

$$\begin{aligned}
r\Gamma_{n+1} + \Upsilon_{n+1} = &(r(1-k)+k)||x_{n+1} - x^{\dagger}||^2 \\
&+ 2\lambda_n \left\langle B(x_n) - B(x_{n+1}), x_{n+1} - x^{\dagger} \right\rangle \\
&+ \frac{\mu\lambda_n}{\lambda_{n+1}}(r(1-p)+p)||x_n - x_{n+1}||^2.
\end{aligned} \tag{25}$$

From the relation (14), we see that

$$r < 1 + \frac{2\gamma\lambda}{1-k}.$$

Thus, from the fact $\lambda \leq \lambda_n$, we derive

$$r(1-k) + k < 1 + 2\gamma\lambda \leq 1 + 2\gamma\lambda_n. \tag{26}$$

Moreover, since $\lambda_n \to \lambda > 0$, we find

$$\begin{aligned}
&\lim_{n\to\infty}\left[\frac{\mu\lambda_n}{\lambda_{n+1}}(r(1-p)+p)-\left(1-\frac{\mu\lambda_{n-1}}{\lambda_n}\right)\right] \\
&\quad=\mu(r(1-p)+p)-(1-\mu).
\end{aligned} \tag{27}$$

Also, from the relation (14), we see that

$$r < \frac{\frac{1-\mu}{\mu}-p}{1-p}.$$

Thus, $\mu(r(1-p)+p)-(1-\mu) < 0$ which together with the relation (27) implies that there exists $\bar{n}_0 \geq n_3$ such that

$$\frac{\mu\lambda_n}{\lambda_{n+1}}(r(1-p)+p)-\left(1-\frac{\mu\lambda_{n-1}}{\lambda_n}\right) < 0, \ \forall n \geq \bar{n}_0.$$

Thus

$$\frac{\mu\lambda_n}{\lambda_{n+1}}(r(1-p)+p) < \left(1-\frac{\mu\lambda_{n-1}}{\lambda_n}\right), \ \forall n \geq \bar{n}_0. \tag{28}$$

Combining the relations (25), (26) and (28), we obtain

$$\begin{aligned}
r\Gamma_{n+1}+\Upsilon_{n+1} \leq (1+2\gamma\lambda_n)||x_{n+1}-x^\dagger||^2 \\
+2\lambda_n\left\langle B(x_n)-B(x_{n+1}), x_{n+1}-x^\dagger\right\rangle \\
+\left(1-\frac{\mu\lambda_{n-1}}{\lambda_n}\right)||x_n-x_{n+1}||^2,
\end{aligned} \tag{29}$$

which together with the relation (24), implies that $r\Gamma_{n+1}+\Upsilon_{n+1} \leq \Gamma_n+\Upsilon_n, \ \forall n \geq \bar{n}_0$.

Claim 4: The sequence $\{x_n\}$ converges R-linearly to the unique solution $x^\dagger$ of problem (VI).

The proof of Claim 4. Take $\theta \in (1, r)$ and set $\xi := \min\{\theta, 1+\rho(r-\theta)\} > 1$. From **Claim 2** we see that $\rho\Upsilon_{n+1} \leq \Gamma_{n+1}$ for $n \geq \bar{n}_0$. Thus

$$\begin{aligned}
r\Gamma_{n+1} &= \theta\Gamma_{n+1}+(r-\theta)\Gamma_{n+1} \geq \theta\Gamma_{n+1}+\rho(r-\theta)\Upsilon_{n+1} \\
&= \left[\theta\Gamma_{n+1}+(1+\rho(r-\theta))\Upsilon_{n+1}\right]-\Upsilon_{n+1} \\
&\geq \xi(\Gamma_{n+1}+\Upsilon_{n+1})-\Upsilon_{n+1}.
\end{aligned}$$

This together with **Claim 3** implies that

$$\xi(\Gamma_{n+1}+\Upsilon_{n+1}) \leq r\Gamma_{n+1}+\Upsilon_{n+1} \leq \Gamma_n+\Upsilon_n, \ \forall n \geq \bar{n}_0.$$

Therefore, we obtain from the induction that

$$\Gamma_{n+1} + \Upsilon_{n+1} \leq \left(\frac{1}{\xi}\right)^{n-\bar{n}_0+1} \left(\Gamma_{\bar{n}_0} + \Upsilon_{\bar{n}_0}\right), \ \forall n \geq \bar{n}_0,$$

which, from the definition of Γ_{n+1} and the fact $\Upsilon_{n+1} \geq 0$ for all $n \geq \bar{n}_0$, we obtain that

$$||x_{n+1} - x^{\dagger}||^2 \leq \frac{\Gamma_{n+1}}{1-k} \leq \frac{\Gamma_{n+1} + \Upsilon_{n+1}}{1-k} \leq \left(\frac{1}{\xi}\right)^{n-\bar{n}_0+1} \frac{\Gamma_{\bar{n}_0} + \Upsilon_{\bar{n}_0}}{1-k} = \frac{M}{\xi^n}, \ \forall n \geq \bar{n}_0, \tag{30}$$

where $M = \frac{\Gamma_{\bar{n}_0}+\Upsilon_{\bar{n}_0}}{(1-k)\xi^{\bar{n}_0-1}}$. This finishes the proof. □

Remark 3.2 In view of the results in Chen and Rockafellar (1997), Tseng (2000), when the Lipschitz constant and the modulus of strong monotonicity of operators are known, the convergence ratio is explicitly expressed. This can allow us to estimate the bound of the convergence ratio. Also, in the recent work (Monteiro and Svaiter 2011), R.D.C. Monteiro and B.F. Svaiter investigated complexity of variants of Tseng's modified F-B splitting and Korpelevich's methods, and obtained some notable results where pointwise complexity bounds are established. A major departure from some of these results that holds possibly at every iteration. While from Theorem 3.2, an open question here is how to give a bound of the convergence ratio of Algorithm 3. Remark that from the proof of Claim 4, the convergence ratio is given by

$$\alpha^{rat} := \frac{1}{\xi} = \frac{1}{\min\{\theta, 1 + \rho(r-\theta)\}} \in (0, 1),$$

where $\theta \in (1, r)$, r is defined in (14), and ρ is given in (22). In general, the ratio α^{rat} depends implicitly on the following quantities: the parameter μ, the modulus of strong monotonicity γ, and the number $\lambda = \lim_{n\to\infty} \lambda_n$, i.e., $\alpha^{rat} = \alpha^{rat}(\mu, \gamma, \lambda)$. Here, the sequence $\{\lambda_n\}$, from its definition, depends on not only the input data, but also on the sequence $\{x_n\}$ genarated by Algorithm 3. Besides, from relation (30), the convergence rate of the sequence $\{x_n\}$ also depends on an implicit number $M = \frac{\Gamma_{\bar{n}_0}+\Upsilon_{\bar{n}_0}}{(1-k)\xi^{\bar{n}_0-1}}$, where $\bar{n}_0$ is unknown at **Claim 3** in the proof of Theorem 3.2.

Remark 3.3 As the suggestion of a reviewer, we can define r in the relation (14) by

$$1 < r < \min\left\{\frac{\frac{1-\mu}{\mu} - p}{1-p}, 1 + \frac{2\gamma \min\left\{\lambda_0, \frac{\mu}{L}\right\}}{1-k}\right\}, \tag{31}$$

where we have replaced λ by $\min\left\{\lambda_0, \frac{\mu}{L}\right\}$. In this case, the proof of Theorem 3.2 is still true. Indeed, from the relation (31), we have

$$r < 1 + \frac{2\gamma \min\left\{\lambda_0, \frac{\mu}{L}\right\}}{1-k}.$$

Thus, $r(1-k) < 1 - k + 2\gamma \min\left\{\lambda_0, \frac{\mu}{L}\right\}$ or, equivalently

$$r(1-k) + k < 1 + 2\gamma \min\left\{\lambda_0, \frac{\mu}{L}\right\}. \tag{32}$$

From Remark 3.1, the sequence $\{\lambda_n\}$ is bounded from below by $\min\left\{\lambda_0, \frac{\mu}{L}\right\} > 0$. This implies that $\lambda_n \geq \min\left\{\lambda_0, \frac{\mu}{L}\right\}$ for all $n \geq 1$. Thus, by (32), we obtain

$$r(1-k) + k < 1 + 2\gamma \min\left\{\lambda_0, \frac{\mu}{L}\right\} \leq 1 + 2\gamma\lambda_n,$$

which follows the relation (26). Other arguments are repeated similarly. It is worth mentioning here that the choice of r as in the relation (31) allows us to remove the dependency of r on λ, and thus on the sequence $\{x_n\}$ generated by Algorithm 3. Thus, unlike Remark 3.2, the convergence ratio α^{rat} depends only on problem data.

4 Numerical illustrations

In this section, we perform some experiments to show the numerical behavior of Algorithm 3 (shortly, MFRBSM) and compare it with others. All the programs are written in Matlab 7.0 with the Optimization Toolbox and computed on a PC Desktop Intel(R) Core(TM) i5-3210M CPU @ 2.50GHz, RAM 2.00 GB.

Example 1. Consider the LASSO problem (Tibshirami 1996) (strongly related to the Basis Pursuit denosing problem Chen et al. 1998) which can be formulated as a convex constrained optimization problem

$$\begin{aligned} &\min_{x\in\Re^m} J(x) = \frac{1}{2}||Tx - b||^2, \\ &\text{s.t. } ||x||_1 \leq t, \end{aligned} \tag{33}$$

where $T \in \Re^{k\times m}$, $b \in \Re^k$ $(k < m)$ and $t > 0$ is a given constant. It was known that the solution of the Lasso problem (33) for appropriate choices $t > 0$ is a minimizer of the following unconstrained minimization problem (Anh et al. 2018; Gibali and Thong 2018):

$$\min_{x\in\Re^m} \frac{1}{2}||Tx - b||^2 + \lambda||x||_1. \tag{34}$$

Now we set $C = \{x \in \Re^m : ||x||_1 \leq t\}$. The problem is equivalent to our problem (VI) with $A(x) = \partial i_C(x)$ and $B(x) = \nabla\left(\frac{1}{2}||Tx - b||^2\right) = T^*(Tx - b)$. The projection P_C is found by using the Lagrangian method and Newton iteration. We perform a

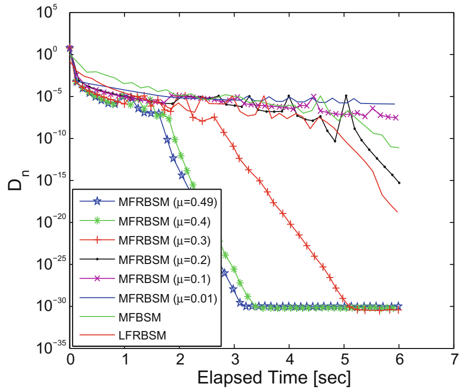


Fig. 1 The behavior of D_n for Example 1 with $T \in \Re^{128\times 1024}$ and $K = 50$. The number of iterations is 983, 918, 942, 940, 989, 931, 190, 372, respectively

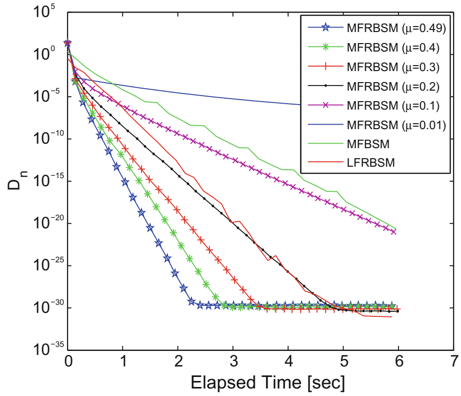


Fig. 2 The behavior of D_n for Example 1 with $T \in \Re^{128\times 1024}$ and $K = 120$. The number of iterations is 795, 776, 784, 782, 791, 775, 151, 313, respectively

sparse signal recovery $x \in \Re^m$ which contains K randomly replaced ± 1 spikes. The matrix $T \in \Re^{k\times m}$ is randomly generated (from normal distribution) with mean zero and one variance. The value of t is $K + 1$.

We compare our Algorithm 3 with two other algorithms, namely the modified forward–backward splitting method (MFBSM) in Tseng (2000) and the forward-reflected backward splitting method with linesearch (LFRBSM) in Malitsky and Tam (2018, Algorithm 1). These algorithms are done without the Lipschitz constant of the operator B.

The parameters are $\lambda_{-1} = \lambda_0 = 0.1$ and $\mu \in \{0.01, 0.1, 0.2, 0.3, 0.4, 0.49\}$ (for MFRBSM); $\lambda_0 = 0.1$, $\delta = 0.5$, $\sigma = 0.5$, $\rho = 1/\sigma$ (for LFRBSM); $\sigma = 0.1$, $\beta = 0.5$, $\theta = 0.5$ (for MFBSM). The starting points are randomly generated in the interval $(-1, 1)$. We use the sequence $D_n = \max\left\{||x_{n+1} - x_n||^2, ||x_n - x_{n-1}||^2\right\}$ for MFRBSM and LFRBSM (respectively, $D_n = ||\bar{x}_n - x_n||^2$ for MFBSM) to illustrate the convergence of the algorithms. Note that $D_n = 0$ iff x_n is the solution of the problem.

We first describe the behavior of D_n for some given data T and K. The results are shown in Figs. 1, 2, 3 and 4. In these figures, the y-axis represents for the value of D_n while the x-axis is for the execution time elapsed in the second.

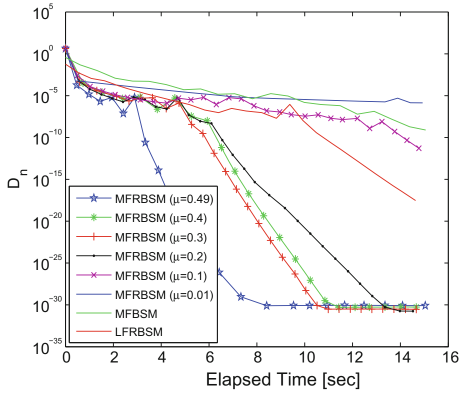


Fig. 3 The behavior of D_n for Example 1 with $T \in \Re^{256\times 2048}$ and $K = 50$. The number of iterations is 540, 551, 595, 599, 569, 563, 105, 231, respectively

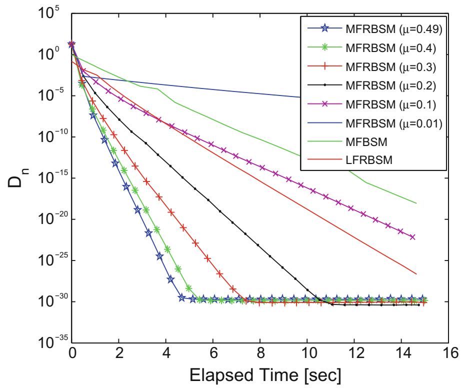


Fig. 4 The behavior of D_n for Example 1 with $T \in \Re^{256\times 2048}$ and $K = 120$. The number of iterations is 545, 561, 583, 570, 557, 560, 103, 229, respectively

Next, we fix a tolerance TOL with the stopping criterion $D_n \leq$ TOL. We show for each algorithm the number of iteration (Iter.), the number of resolvents (nJ_A) and the number of values of operator B (nB) needed to be computed. In addition, the objective function value $J(x_n)$ is also involved when a termination is reached. Note that $Iter. = nJ_A = nB$ for our MFRBSM while $nJ_A = nB$ for the LFRBSM. Tables 1 and 2 show the results for the chosen data as well as different given TOL.

In order to finish this example, we study the behavior of stepsize $\{\lambda_n\}$ generated by each aforementioned algorithm. The results are shown in Figs. 5, 6, 7 and 8.
From the reported results, we make the following observations:

(i) The convergence of Algorithm 3 (MFRBSM) depends strictly on the value of μ. The MFRBSM works well when μ is large. This is not surprising because, from Figs. 7 and 8, the larger the value of μ is, the larger the stepsize λ_n generated by the MFRBSM is.
(ii) Although the stepsize λ_n generated by the MFRBSM in general is smaller than the one of others (see, Figs. 7, 8), but it is found by cheaper computations. In view of Figs. 1, 2, 3 and 4, with the same execution time, the number of iterations of the MFRBSM is so larger than those of other algorithms. This makes the MFRBSM more effective.

Table 1 A detailed comparison between MFRBSM ($\mu = 0.49$) with MFBSM and LFRBSM for Example 1 with $T \in \Re^{128\times1024}$

K	TOL	MFRBSM ($\mu = 0.49$)			MFBSM				LFRBSM			
		CPU(s)	Iter. = $nJ_A = nB$	$J(x_n)$	CPU(s)	Iter.	nJ_A/nB	$J(x_n)$	CPU(s)	Iter.	$nJ_A = nB$	$J(x_n)$
K = 50	10^{-10}	1.31	219	2.65e–08	3.65	131	726/857	5.86e–08	3.41	221	1105	3.77e–08
	10^{-15}	1.80	260	3.42e–12	5.15	163	906/1069	6.49e–12	5.87	265	1325	5.81e–12
	10^{-20}	2.14	305	1.74e–15	5.95	198	1087/1285	7.28e–15	4.48	309	1545	4.22e–15
K = 80	10^{-10}	0.43	81	4.54e–08	2.14	66	375/441	9.40e–08	1.13	77	385	8.98e–08
	10^{-15}	0.85	136	5.13e–12	3.21	108	601/709	6.10e–12	1.81	119	595	5.98e–12
	10^{-20}	1.31	190	6.21e–15	5.41	146	831/977	7.31e–15	2.91	165	825	6.78e–15
K = 120	10^{-10}	0.49	77	5.74e–07	2.19	62	362/424	9.36e–06	1.61	83	415	7.25e–06
	10^{-15}	0.91	123	6.97e–11	3.98	101	571/672	6.90e–10	2.13	115	575	5.89e–10
	10^{-20}	1.10	175	8.23e–14	4.44	136	784/920	2.11e–13	2.21	151	755	7.32e–13

Table 2 A detailed comparison between FRBSM ($\mu = 0.49$) with MFBSM and LFRBSM for Example 1 with $T \in \Re^{256\times2048}$

K	TOL	MFRBSM ($\mu = 0.49$)			MFBSM				LFRBSM			
		CPU(s)	Iter. = $nJ_A = nB$	$J(x_n)$	CPU(s)	Iter.	nJ_A/nB	$J(x_n)$	CPU(s)	Iter.	$nJ_A = nB$	$J(x_n)$
K = 50	10^{-10}	3.65	155	5.12e–06	15.7	124	797/921	7.94e–06	10.6	181	1086	6.01e–06
	10^{-15}	4.79	203	7.84e–11	22.7	150	966/1116	8.39e–11	17.8	220	1320	6.75e–11
	10^{-20}	8.30	235	2.13e–14	35.5	177	1154/1331	5.78e–14	22.2	257	1542	4.72e–14
K = 80	10^{-10}	1.71	57	1.21e–06	7.13	54	346/400	3.98e–06	4.19	75	450	2.39e–06
	10^{-15}	2.08	86	9.91e–11	11.31	80	532/612	9.98e–11	7.10	115	690	8.91e–11
	10^{-20}	2.90	122	1.79e–14	13.81	106	717/823	4.33e–14	9.15	155	930	3.69e–14
K = 120	10^{-10}	1.36	58	8.35e–05	6.77	54	352/406	9.97e–05	5.11	75	450	9.03e–05
	10^{-15}	2.41	86	6.97e–10	12.29	77	529/606	7.68e–10	6.71	115	690	6.65e–10
	10^{-20}	2.79	121	4.38e–14	13.57	106	721/827	5.11e–14	9.63	153	918	4.97e–14

Fig. 5 The behavior of stepsizes $\{\lambda_n\}$ generated by the algorithms for Example 1 with $T \in \Re^{128\times 1024}$ and $K = 50$

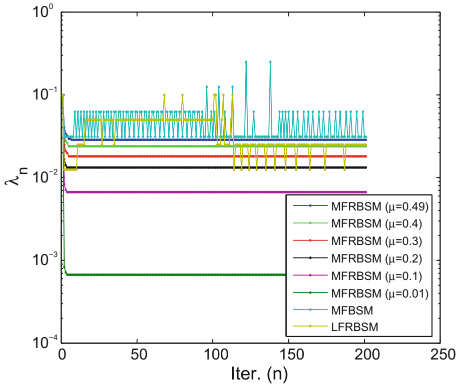


Fig. 6 The behavior of stepsizes $\{\lambda_n\}$ generated by the algorithms for Example 1 with $T \in \Re^{128\times 1024}$ and $K = 120$

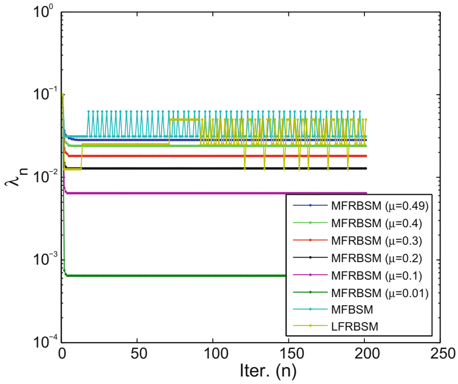
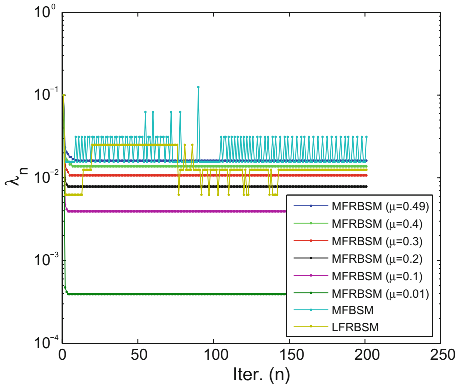


Fig. 7 The behavior of stepsizes $\{\lambda_n\}$ generated by the algorithms for Example 1 with $T \in \Re^{256\times 2048}$ and $K = 50$



(iii) With the results in Tables 1 and 2, in order to get a given error TOL, the MFRBSM is cheaper. Moreover, the objective function value which is obtained by the MFRBSM when a termination is reached, is also better than others. It is seen that at each (outer) iteration, the MFBSM and LFRBSM require many extra - computations on the resolvent $J_{\lambda A}$ and the values of operator B. This of course is more time-consuming.

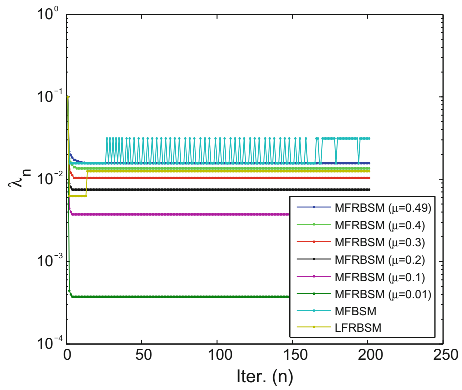


Fig. 8 The behavior of stepsizes $\{\lambda_n\}$ generated by the algorithms for Example 1 with $T \in \Re^{256\times 2048}$ and $K = 120$

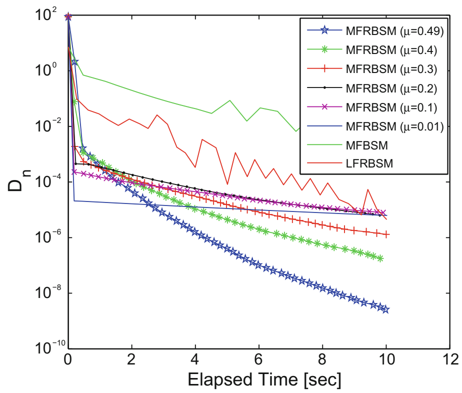


Fig. 9 The behavior of D_n for Example 2 with $m = 50$ and $k = 30$. The number of iterations is 323, 286, 288, 251, 276, 273, 21, 66, respectively

Example 2. In this example, we consider the split feasibility problem (SFP) Censor and Elfving (1994),

$$\text{Find } x^* \in C := \cap_{i=1}^{M} C_i \text{ such that } Tx^* \in Q := \cap_{j=1}^{N} Q_j. \tag{SFP}$$

where $T : \Re^m \to \Re^k$ is given by a real matrix of size $m \times k$, the set $C_i = B[c_i, r_i]$ is a closed ball in $\Re^m$ with a center $c_i \in \Re^m$ and a radius $r_i > 0$, and $Q_j = \left\{x \in \Re^k : \langle a_j, x\rangle \le b_j\right\}$ is a half-space in $\Re^k$. It is well-known that the problem (SFP) is equivalent to our problem (VI) with $B(x) = \nabla\left(\frac{1}{2}||Tx - P_Q Tx||^2\right) = T^*(I - P_Q)T(x)$ and $A(x) = N_C(x)$, the normal cone mapping of C. The resolvent $J_{\lambda A}$ in this case is the metric projection P_C for all $\lambda > 0$. The projection on the intersection C of balls C_i $(i = 1, \ldots, M)$ can be effectively found by the sequential projection method. The projection on the intersection Q of half-spaces Q_j $(j = 1, \ldots, N)$ is computed by the function *quadprog* in Matlab 7.0. For experiment, we take ten balls C_i and fifty half-spaces Q_j with all the entries of c_i, a_j being generated randomly in $(-2, 2)$. We set $r_i = ||c_i|| + 1$, and choose randomly $b_j > 0$, which ensure that both C and Q contain the original points in $\Re^m$ and $\Re^k$, respectively. The results are shown in Figs. 9, 10, 11 and 12. The convergence result is similar to the one of Example 1.

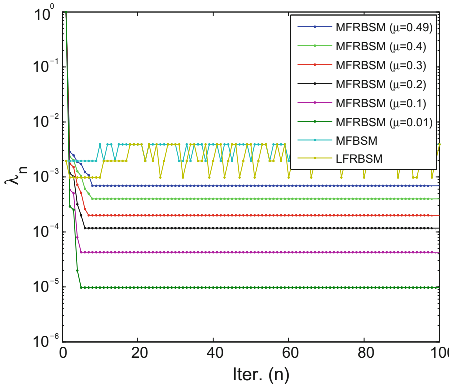


Fig. 10 The behavior of stepsizes $\{\lambda_n\}$ generated by the mentioned algorithms for Example 2 with $m = 50$ and $k = 30$

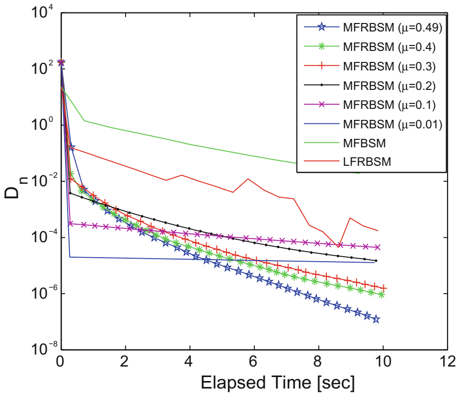


Fig. 11 The behavior of D_n for Example 2 with $m = 100$ and $k = 50$. The number of iterations is 216, 219, 217, 194, 193, 180, 13, 43, respectively

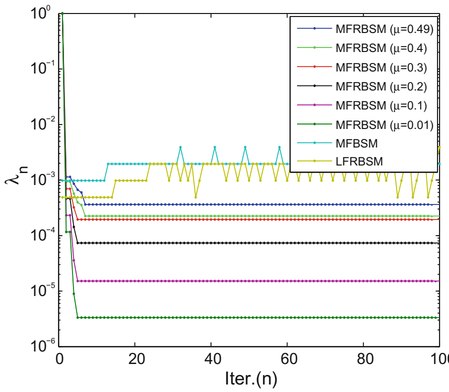


Fig. 12 The behavior of stepsizes $\{\lambda_n\}$ generated by the mentioned algorithms for Example 2 with $m = 100$ and $k = 50$

5 Conclusions

In this paper, we have presented a new algorithm for solving a variational inclusion problem involving the sum of two operators in a Hilbert space. The algorithm is delveloped from the forward-reflected-backward splitting method with a simple stepsize rule. This rule allows the algorithm to be performed easily without the prior knowledge of Lipschitz constant of operators involved. The convergence and the linear rate of convergence of the algorithm are proved. The numerical results are reported to support the obtained theoretical results.

Acknowledgements The authors would like to thank the Associate Editor and the anonymous referees for their valuable comments and suggestions which helped us very much in improving the original version of this paper. The first-named and last-named authors are supported by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under the project: 101.01-2020.06.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

References

Anh PK, Vinh NT, Dung VT (2018) A new self-adaptive CQ algorithm with an application to the LASSO problem. J Fixed Point Theory Appl 142:1–19

Attouch H, Peypouquet J, Redont P (2018) Backward–forward algorithms for structured monotone inclusions in Hilbert spaces. J Math Anal Appl 457:1095–1117

Brézis H, Chapitre II (1973) Operateurs maximaux monotones. North-Holland Math Stud 5:19–51

Bruck R (1977) On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space. J Math Anal Appl 61:159–164

Censor Y, Elfving T (1994) A multi-projection algorithm using Bregman projections in a product space. Numer Algorithms 8:221–239

Chen GH-G, Rockafellar T (1997) Convergence rates in forward–backward splitting. SIAM J Optim 7:421–444

Chen S, Donoho DL, Saunders M (1998) Atomic decomposition by basis pursuit. SIAM J Sci Comput. 20:33–61

Combettes PL, Wajs V (2005) Signal recovery by proximal forward–backward splitting. SIAM Multisc Model Simul 4:1168–1200

Daubechies I, Defrise M, De Mol C (2004) An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Commun Pure Appl Math 57:1413–1457

Davis D, Yin WT (2017) A three-operator splitting scheme and its optimization applications. Set-Valued Var Anal 25:829–858

Dong YD, Fischer A (2010) A family of operator splitting methods revisited. Nonlinear Anal 72:4307–4315

Duchi J, Singer Y (2009) Efficient online and batch learning using forward–backward splitting. J Mach Learn Res 10:2899–2934

Facchinei F, Pang JS (2002) Finite-dimensional variational inequalities and complementarity problems. Springer, Berlin

Gibali A, Hieu DV (2019) A new inertial double-projection method for solving variational inequalities. J Fixed Point Theory Appl. https://doi.org/10.1007/s11784-019-0726-7

Gibali A, Thong DV (2018) Tseng type methods for solving inclusion problems and its applications. Calcolo. https://doi.org/10.1007/s10092-018-0292-1

Goldstein AA (1964) Convex programming in Hilbert spaces. Bull Am Math Soc 70:709–710

Hieu DV, Quy PK (2019) An inertial modified algorithm for solving variational inequalities. RAIRO Oper Res. https://doi.org/10.1051/ro/2018115

Hieu DV, Anh PK, Muu LD (2017) Modified hybrid projection methods for finding common solutions to variational inequality problems. Comput Optim Appl 66:75–96

Hieu DV, Anh PK, Muu LD (2019a) Modified extragradient-like algorithms with new stepsizes for variational inequalities. Comput Optim Appl 73:913–932

Hieu DV, Cho YJ, Xiao Y-B (2019b) Golden ratio algorithms with new stepsize rules for variational inequalities. Math Methods Appl Sci. https://doi.org/10.1002/mma.5703

Hieu DV, Cho Xiao YB, Kumam P (2019c) Relaxed extragradient algorithm for solving pseudomonotone variational inequalities in Hilbert spaces. Optimization. https://doi.org/10.1080/02331934.2019.1683554

Hieu DV, Vy LV, Quy PK (2019) Three-operator splitting algorithm for a class of variational inclusion problems. Bull Iran Math Soc. https://doi.org/10.1007/s41980-019-00312-5

Hieu DV, Strodiot JJ, Muu LD (2020a) An explicit extragradient algorithm for solving variational inequalities. J Optim Theory Appl. https://doi.org/10.1007/s10957-020-01661-6

Hieu DV, Strodiot JJ, Muu LD (2020b) Strongly convergent algorithms by using new adaptive regularization parameter for equilibrium problems. J Comput Appl Math. https://doi.org/10.1016/j.cam.2020.112844

Huang YY, Dong YD (2014) New properties of forward–backward splitting and a practical proximal descent algorithm. Appl Math Comput 237:60–68

Kinderlehrer D, Stampacchia G (1980) An introduction to variational inequalities and their applications. Academic Press, New York

Lions PL, Mercier B (1979) Splitting algorithms for the sum of two nonlinear operators. SIAM J Numer Anal 16:964–979

Malitsky Y, Tam MK (2018) A Forward-Backward splitting method for monotone inclusions without cocoercivity. arXiv:1808.04162

Monteiro RDC, Svaiter BF (2011) Complexity of variants of Tseng's modified F–B splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. SIAM J Optim 21:1688–1720

Passty GB (1979) Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. J Math Anal Appl 72:383–390

Raguet H, Fadili J, Peyré G (2013) A generalized forward–backward splitting. SIAM J Imaging Sci 6:1199–1226

Rockafellar RT (1976) Monotone operators and the proximal point algorithm. SIAM J Control Optim 14:877–898

Ryu EK, Boyd S (2016) A primer on monotone operator methods. Appl Comput Math 15:3–43

Takahashi W (2000) Nonlinear functional analysis-fixed point theory and its applications. Yokohama Publishers, Yokohama

Tibshirami R (1996) Regression shrinkage and selection via lasso. J R Stat Soc Ser B 58:267–288

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

Zong C, Tang Y, Cho YJ (2018) Convergence analysis of an inexact three-operator splitting algorithm. Symmetry. https://doi.org/10.3390/sym10110563