J. Nonlinear Var. Anal. 11 (2027), No. 2, pp. 365-392
Available online at http://jnva.biemdas.com
https://doi.org/10.23952/jnva.11.2027.2.07

NONLINEAR THREE-OPERATOR SPLITTING ALGORITHMS WITH MOMENTUM FOR MONOTONE INCLUSIONS

LIQIAN QIN[1], AVIV GIBALI[2], CUIJIE ZHANG[3], YUCHAO TANG[1,*]

[1]School of Mathematics and Information Science, Guangzhou University, Guangzhou 510006, China
[2]Department of Applied Mathematics, Holon Institute of Technology, Holon 5810201, Israel
[3]College of Science, Civil Aviation University of China, Tianjin 300300, China

Abstract. In this paper, we introduce three novel splitting algorithms for solving structured monotone inclusion problems involving the sum of a maximally monotone operator, a monotone and Lipschitz continuous operator and a cocoercive operator. Each proposed method extends one of the classical schemes: the semi-forward-reflected-backward splitting algorithm, the semi-reflected-forward-backward splitting algorithm, and the outer reflected forward-backward splitting algorithm by incorporating a nonlinear momentum term. Under appropriate step-size conditions, we establish the weak convergence of all three algorithms, and further prove their R-linear convergence rates under strong monotonicity assumptions. Preliminary numerical experiments on both synthetic datasets and real-world quadratic programming problems in portfolio optimization demonstrate the effectiveness and superiority of the proposed algorithms.

Keywords. Momentum; Monotone inclusion problem; Semi-forward-reflected-backward splitting; Semi-reflected-forward-backward splitting; Outer reflected forward-backward splitting.
2020 MSC. 47H05, 90C30.

1. INTRODUCTION

Monotone inclusion problems form a cornerstone of modern optimization theory and arise in a wide range of applications, including signal and image processing, machine learning, image reconstruction, and game theory [2, 3, 4, 5, 13, 14, 15]. A key challenge in this area is the resolution of composite inclusions that involve the sum of several monotone operators, where the specific structure of each operator plays a crucial role in the development of efficient numerical methods. In this work, we focus on the following monotone inclusion problem:

Problem. Consider a real Hilbert space $\mathscr{H}$. Let $A:\mathscr{H}\to 2^{\mathscr{H}}$ be a maximally monotone operator, let $B:\mathscr{H}\to\mathscr{H}$ be a monotone and μ-Lipschitz operator, and let $C:\mathscr{H}\to\mathscr{H}$ be β^{-1}-cocoercive operator for some $\beta\in(0,+\infty)$. The problem is to find $x\in\mathscr{H}$ such that

$$0\in Ax+Bx+Cx, \tag{1.1}$$

where the solution set, denoted by $\operatorname{zer}(A+B+C)$, is assumed to be nonempty.

*Corresponding author.
E-mail address: hhaaoo1331@163.com (Y. Tang).
Received 30 November 2025; Accepted 26 March 2026; Published online 20 November 2026.
©2027 Journal of Nonlinear and Variational Analysis

Solving a monotone inclusion problem involving three operators can be viewed as a natural extension of splitting algorithms for the two-operator case. In the classical two-operator setting, Tseng's forward-backward-forward splitting method [31] finds a zero of $A+B$ by performing, at each iteration, one backward (implicit) step for A and two forward evaluations of the single-valued, monotone, and Lipschitz continuous operator B. A naive attempt to handle three operators (say A, B, and C) is to combine $B+C$ into one Lipschitz monotone operator and then apply Tseng's method [31]. In theory, this works because the sum of a cocoercive operator and a Lipschitz monotone operator is still monotone and Lipschitz. In practice, however, this direct approach is inefficient, as each iteration would require evaluating both B and C twice, failing to exploit the special cocoercivity property of C. Clearly, a more tailored approach is needed to handle three-operator problems efficiently. Briceño-Arias and Davis [6] addressed the above challenge by leveraging the cocoercivity of operator C to design the forward-backward-half forward splitting (FBHFS) algorithm as follows:

$$\begin{cases} y_k = J_{\gamma_k A}(x_k - \gamma_k(B+C)x_k), \\ x_{k+1} = y_k + \gamma_k(Bx_k - By_k), \end{cases} \tag{1.2}$$

where γ_k is the step-size, $J_{\gamma_k A} = (\mathrm{Id}+\gamma_k A)^{-1}$ is the resolvent of A. They proved the weak convergence of the FBHFS algorithm (1.2) with a variable step-size, $\gamma_k \in [\eta, \chi-\eta]$ with $\eta \in (0, \chi/2)$, $\chi = 4/(\beta+\sqrt{\beta^2+16\mu^2})$, as well as with a line search procedure. FBHFS algorithm modifies Tseng's scheme [31] to avoid redundant computations on the cocoercive part. Each iteration of the FBHFS algorithm (1.2) still performs two forward evaluations of B (the Lipschitz monotone operator), as in Tseng's method [31], but requires only one forward evaluation of C, the cocoercive operator. By halving the forward steps on C, the algorithm significantly reduces per-iteration computational cost compared to the naive approach. In particular, the FBHFS algorithm (1.2) recovers several existing splitting algorithms as special cases: if $C=0$ (no cocoercive part), it reduces to Tseng's forward-backward-forward algorithm [31], and if $B=0$ (no Lipschitz part), it becomes the classical forward-backward splitting method [20, 23]. Furthermore, the authors also considered the FBHFS algorithm (1.2) employing non-self-adjoint linear operators, which is defined by

$$\begin{cases} y_k = (M_k+A)^{-1}(M_k x_k - (B+C)x_k), \\ x_{k+1} = x_k + \lambda_k(M_k(y_k-x_k) + Bx_k - By_k), \end{cases} \tag{1.3}$$

where $M_k : \mathscr{H} \to \mathscr{H}$ are linear and bounded operators, $\forall k \in \mathbb{N}$. Under suitable conditions on M_k and λ_k, the weak convergence of the iterative scheme (1.3) was established.

Subsequently, Malitsky and Tam [21] introduced the semi-forward-reflected-back-ward splitting (SFRBS) algorithm, defined by

$$x_{k+1} = J_{\gamma A}(x_k - 2\gamma Bx_k + \gamma Bx_{k-1} - \gamma Cx_k), \tag{1.4}$$

and the convergence of $\{x_k\}$ was guaranteed when the step size satisfies $\gamma \in (0, 2/(4\mu+\beta))$. In contrast, to address problem (1.1), Cevher and Vu [16] proposed the semi-reflected-forward-backward splitting (SRFBS) algorithm, which is defined as

$$x_{k+1} = J_{\gamma A}(x_k - \gamma B(2x_k - x_{k-1}) - \gamma Cx_k). \tag{1.5}$$

The convergence of the SRFBS algorithm (1.5) was analyzed in [16] under suitable assumptions on the step size γ. More recently, Yu et al. [35] introduced the outer reflected forward-backward

splitting (ORFBS) algorithm, which iterates as follows:

$$x_{k+1} = J_{\gamma A}(x_k - \gamma B x_k - \gamma C x_k) - \gamma(B x_k - B x_{k-1}). \tag{1.6}$$

Under the step size condition $\gamma < \min\{(2/\beta - \varepsilon_2)\varepsilon_1,\ (3-\varepsilon_3)\varepsilon_2,\ (1/2 - \varepsilon_1 - 1/\varepsilon_3)/\mu\}$ with $\varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$, they proved the weak convergence of the sequence $\{x_k\}$ generated by the ORFBS algorithm (1.6). Motivated by the success of inertial and accelerated methods, several authors proposed accelerated variants of the algorithms FBHFS (1.2), SFRBS (1.4), SRFBS (1.5) and ORFBS (1.6); see, e.g., [24, 32, 33, 34, 37, 38, 39] and the references therein.

The resolvent operator $J_{\gamma A} = (\mathrm{Id} + \gamma A)^{-1}$ serves as the fundamental backward step in a broad class of operator splitting algorithms, whose analysis and design heavily rely on this operator. To enhance the applicability of these methods, this study explores an innovative approach centered around a nonlinear resolvent of the form $(M + A)^{-1}M$. Here, $M : \mathscr{H} \to \mathscr{H}$ acts as a nonlinear kernel, also referred to as a warped resolvent [7, 18]. This generalization provides a powerful modeling tool for a wide spectrum of monotone inclusion algorithms, offering a unified framework that encompasses both existing and novel schemes. The central idea is to exploit a problem adapted kernel M that captures the intrinsic structure of the underlying problem. In this context, numerous studies have proposed preconditioned variants of classic algorithms, which enable flexible and efficient problem dependent preconditioning [8, 9, 26]. In particular, the work of [10] introduced the concept of an admissible preconditioner. Specifically, for an operator $A : \mathscr{H} \to 2^{\mathscr{H}}$, a bounded linear operator M is said to be an admissible preconditioner if it is self-adjoint, positive semidefinite, and satisfies that $(M + A)^{-1}M$ is single-valued with full domain. Building upon this notion, the authors proposed the degenerate preconditioned proximal point (degenerate PPP) method [10] and established its convergence. As stated in [8], the Douglas–Rachford splitting algorithm [17] can be viewed as a particular instance of the degenerate PPP framework. Furthermore, Sun et al. [29] developed an accelerated preconditioned alternating direction method of multiplier (pADMM) by combining the degenerate PPP method [10] with the fast Krasnosel'skiĭ-Mann iteration [11]. Based on the Halpern iteration [19], Zhang et al. subsequently proposed the Halpern-type preconditioned proximal point algorithm [36].

While the convergence-guaranteeing methods in [7, 18] rely on corrective projection steps, Morin et al. [22] proposed an alternative strategy that incorporates a nonlinear momentum term, which typically results in a lower per-iteration computational cost. Specifically, they introduced a nonlinear forward-backward splitting algorithm with momentum correction to solve the two-operator monotone inclusion problem involving a maximally monotone operator and a cocoercive operator. The algorithm is stated as follows:

$$\begin{aligned} x_{k+1} &= (M_k + A)^{-1}(M_k x_k - C x_k + \gamma_k^{-1} u_k), \\ u_{k+1} &= (\gamma_k M_k - S)x_{k+1} - (\gamma_k M_k - S)x_k. \end{aligned} \tag{1.7}$$

Here, $\gamma_k > 0$ for all $k \in \mathbb{N}$, and $M_k : \mathscr{H} \to \mathscr{H}$ are general nonlinear operators. Moreover, S is a bounded linear, self-adjoint, strongly positive operator. For the convergence analysis, it is assumed that $\gamma_k M_k - S$ is L_k-Lipschitz continuous with respect to the norm induced by S for some $L_k \geq 0$, for all $k \in \mathbb{N}$. The generality of algorithm (1.7) has been demonstrated by its ability to recover several known methods, such as the forward-backward splitting algorithm [20, 23] and the forward-reflected-backward method [21], etc. Very recently, Roldán and Vega

[27] proposed an inertial and relaxation extension of (1.7). In contrast, in [25], we introduced a FBHFS algorithm with momentum.

This paper aims to further extend the momentum technique to the splitting algorithms, namely the SFRBS algorithm (1.4), the SRFBS algorithm (1.5), and the ORFBS algorithm (1.6), respectively. To this end, we propose three novel algorithms that integrate warped resolvents with momentum for solving the monotone inclusion problem (1.1). The main contributions of this paper are summarized as follows:

(i) We propose three nonlinear splitting algorithms with momentum for solving the monotone inclusion problem (1.1): the nonlinear semi-forward-reflected-backward splitting algorithm with momentum, the nonlinear semi-reflected-forward-backward splitting algorithm with momentum, and the nonlinear outer reflected forward-backward splitting algorithm with momentum. Weak convergence of all three algorithms is established under suitable parameter conditions. Furthermore, under strong monotonicity assumptions, *R*–linear convergence rates are proved for Algorithm 3.1, Algorithm 3.2, and Algorithm 3.3, respectively.

(ii) The proposed framework demonstrates strong generality and flexibility: several existing algorithms can be recovered as special cases. Beyond unifying known methods, the introduction of momentum terms enables the design of new splitting schemes for monotone inclusions involving sums of four operators.

(iii) The performance of the proposed algorithm is comprehensively validated through applications to two representative problems: constrained nonlinear optimization and portfolio optimization. Numerical experiments confirm their efficiency and robustness compared with existing approaches.

This paper is organized as follows. Section 2 reviews essential fundamental definitions and lemmas required for the subsequent analysis. In Section 3, we introduce three novel splitting algorithms with momentum for solving the problem (1.1): the nonlinear semi-forward-reflected-backward splitting algorithm with momentum, the nonlinear semi-reflected-forward-backward splitting algorithm with momentum, and the nonlinear outer reflected forward-backward splitting algorithm with momentum. We establish their weak convergence under standard assumptions and prove linear convergence under additional conditions of strong monotonicity. Section 4 presents numerical experiments on synthetic datasets and real-world quadratic programming problems from portfolio optimization. Finally, we give some conclusions in Section 5.

2. PRELIMINARIES

In this paper, $\mathscr{H}$ and $\mathscr{G}$ denote real Hilbert spaces endowed with the inner product $\langle\cdot,\cdot\rangle$ and let associated norm $\|\cdot\|$. Let $\mathbb{N}$ be the set of nonnegative integers and $\mathbb{N}_0$ the set of positive integers. The symbols "$\to$" and "$\rightharpoonup$" represent strong convergence and weak convergence, respectively. The identity operator is denoted by Id. Let $L:\mathscr{H}\to\mathscr{G}$ be a bounded linear operator with adjoint $L^*:\mathscr{G}\to\mathscr{H}$ and operator norm $\|L\|$. An operator S is said to be self-adjoint if $S^*=S$, and strongly monotone if there exists a constant $c\in(0,+\infty)$ such that $\langle Sx|x\rangle\geq c\|x\|^2$ for all x in $\mathscr{H}$. Define the set of strongly monotone self-adjoint operators on $\mathscr{H}$ as

$$\mathscr{P}(\mathscr{H})=\{S:\mathscr{H}\to\mathscr{H}|S \text{ is linear, self-adjoint and strongly monotone}\}.$$

If $S \in \mathscr{P}(\mathscr{H})$, then S is invertible and its inverse S^{-1} also belongs to $\mathscr{P}(\mathscr{H})$. The inner product induced by S is defined as $\langle x,y\rangle_S := \langle Sx,y\rangle$, and the corresponding norm is denoted by $\|\cdot\|_S$. Note that $S = S^{\frac{1}{2}} \circ S^{\frac{1}{2}}$, where $S^{\frac{1}{2}}$ is a self-adjoint, strongly monotone linear operator. Consequently, the Cauchy–Schwarz inequality with respect to the norms induced by S and S^{-1} takes the following form:

$$|\langle x,u\rangle| = |\langle S^{-\frac{1}{2}}x, S^{\frac{1}{2}}u\rangle| \leq \|S^{-\frac{1}{2}}x\|\|S^{\frac{1}{2}}u\| = \|x\|_{S^{-1}}\|u\|_S \qquad \forall (x,u) \in \mathscr{H}\times\mathscr{H}.$$

Moreover, the following identity holds:

$$2\langle x-y|y-z\rangle_S = \|x-z\|_S^2 - \|x-y\|_S^2 - \|y-z\|_S^2. \tag{2.1}$$

Definition 2.1. Let $A : \mathscr{H} \to 2^{\mathscr{H}}$ be a set-valued operator.

(i) the graph of A is defined by $\mathrm{gra}(A) = \{(x,y) \in \mathscr{H}^2 \mid y \in Ax\}$;
(ii) the set of zeros of A is defined by $\mathrm{zer}(A) = \{x \in \mathscr{H} \mid 0 \in Ax\} = A^{-1}(0)$;
(iii) the domain of A is defined by $\mathrm{dom}(A) = \{x \in \mathscr{H} \mid Ax \neq \emptyset\}$;
(iv) the resolvent of A with $\gamma > 0$ is defined by $J_{\gamma A} = (\mathrm{Id} + \gamma A)^{-1}$. Let $x,y \in \mathscr{H}$. Then

$$y \in J_{\gamma A}x \Longleftrightarrow x - y \in \gamma Ay.$$

Definition 2.2. ([1]) Let $A : \mathscr{H} \to 2^{\mathscr{H}}$ be a set-valued mapping. $A : \mathscr{H} \to 2^{\mathscr{H}}$ is said to be

(i) monotone if $\langle u-v, x-y\rangle \geq 0$ for all $(x,u),(y,v) \in \mathrm{gra}(A)$.
(ii) maximally monotone if there exists no monotone operator $B : \mathscr{H} \to 2^{\mathscr{H}}$ such that $\mathrm{gra}(B)$ properly contains $\mathrm{gra}(A)$, i.e., for every $(x,u) \in \mathscr{H}\times\mathscr{H}$,

$$(x,u) \in \mathrm{gra}(A) \;\Leftrightarrow\; \langle u-v, x-y\rangle \geq 0 \qquad \forall (y,v) \in \mathrm{gra}(A).$$

Definition 2.3. A single-valued operator $T : \mathscr{H} \to \mathscr{H}$ is said to be

(i) L-Lipschitz continuous w.r.t. S if there exists a constant $L > 0$, such that

$$\|Tx - Ty\|_{S^{-1}} \leq L\|x-y\|_S \qquad \forall x,y \in \mathscr{H}.$$

(ii) β^{-1}-cocoercive operator w.r.t. S if there exists a constant $\beta > 0$, such that

$$\langle Tx - Ty, x-y\rangle \geq \beta^{-1}\|Tx-Ty\|_{S^{-1}}^2 \qquad \forall x,y \in \mathscr{H}.$$

Lemma 2.1. ([22]) *Let C be the β^{-1}-cocoercive operator w.r.t. S, for some $\beta > 0$. Then $\langle Cx - Cy, z-y\rangle \geq -\frac{\beta}{4}\|z-x\|_S^2$ for all $x,y,z \in \mathscr{H}$.*

Lemma 2.2. ([1]) *Let $A : \mathscr{H} \to 2^{\mathscr{H}}$ be maximally monotone. Then $\mathrm{gra}(A)$ is sequentially closed in $\mathscr{H}^{weak} \times \mathscr{H}^{strong}$, i.e., for every sequence $\{(x_k,y_k)\}$ in $\mathrm{gra}(A)$ and $(x,y) \in \mathscr{H}\times\mathscr{H}$, if $x_k \rightharpoonup x$ and $y_k \to y$, then $(x,y) \in \mathrm{gra}(A)$.*

Lemma 2.3. ([1]) *Let D be a nonempty set of $\mathscr{H}$, and $\{x_k\}$ be a sequence in $\mathscr{H}$. If the following conditions hold:*

(i) *For every $x \in D$, $\lim_{k\to\infty}\|x_k - x\|$ exists;*
(ii) *Every weak sequential cluster point of $\{x_k\}$ belongs to D.*

Then the sequence $\{x_k\}$ converges weakly to a point in D.

3. Main Algorithms and Convergence Analysis

In this section, we propose three novel splitting algorithms for solving the monotone inclusion problem (1.1): the nonlinear semi-forward-reflected-backward splitting (SFRBS) algorithm, the nonlinear semi-reflected-forward-backward splitting (SRFBS) algorithm, and the nonlinear outer reflected forward-backward splitting (ORFBS) algorithm. The key assumptions required for their convergence analyses are presented below.

Assumption 3.1. Assume that

(i) $A : \mathscr{H} \to 2^{\mathscr{H}}$ is maximally monotone;
(ii) $B : \mathscr{H} \to \mathscr{H}$ is single-valued monotone and μ-Lipschitz w.r.t. S, where $S \in \mathscr{P}(\mathscr{H})$;
(iii) $C : \mathscr{H} \to \mathscr{H}$ is β^{-1}-cocoercive w.r.t. S, where $S \in \mathscr{P}(\mathscr{H})$, for some $\beta > 0$;
(iv) The solution set of (1.1), denoted by, $\operatorname{zer}(A+B+C) := \{x \in \mathscr{H} : 0 \in Ax+Bx+Cx\}$ is nonempty;
(v) For a fixed $\gamma > 0$, the nonlinear kernels $M_k : \mathscr{H} \to \mathscr{H}$ satisfy that $\gamma M_k - S$ is L_k-Lipschitz continuous w.r.t. S, where $L_k \in [0,1)$, for all $k \in \mathbb{N}$.

Proposition 3.1. ([22]) *Let Assumption 3.1(v) hold. Then, for all $k \in \mathbb{N}$, the nonlinear operators M_k are $2\gamma^{-1}$-Lipschitz continuous w.r.t. S, maximally monotone, and strongly monotone w.r.t. S.*

3.1. Nonlinear semi-forward-reflected-backward splitting algorithm with momentum.

In this subsection, we propose a nonlinear semi-forward-reflected-backward splitting (SFRBS) algorithm with momentum, denoted as Algorithm 3.1, and establish its convergence results.

Algorithm 3.1. ***Nonlinear SFRBS Algorithm with Momentum***

Let $S \in \mathscr{P}(\mathscr{H})$, $M_k : \mathscr{H} \to \mathscr{H}$ for all $k \in \mathbb{N}_0$ and $\gamma > 0$. Choose initial points $x_0, x_1 \in \mathscr{H}$, $u_1 = 0$. For $k = 1, 2, \cdots$, do

$$x_{k+1} = (M_k + A)^{-1}(M_k x_k - 2Bx_k + Bx_{k-1} - Cx_k + \gamma^{-1} u_k),$$
$$u_{k+1} = (\gamma M_k - S)x_{k+1} - (\gamma M_k - S)x_k.$$

For the iterates $\{x_k\}$ and $\{u_k\}$ generated by Algorithm 3.1 and $\forall x \in \operatorname{zer}(A+B+C)$, we define the following function,

$$\begin{aligned}\Psi_k(x) :=& \|x_k - x\|_S^2 + 2\langle u_k, x_k - x\rangle + (\gamma\mu + L_{k-1})\|x_k - x_{k-1}\|_S^2 \\ &+ 2\gamma\langle Bx_k - Bx_{k-1}, x - x_k\rangle,\end{aligned} \tag{3.1}$$

which helps to establish the weak convergence of the Algorithm 3.1.

Lemma 3.1. *Suppose that Assumption 3.1 holds. Then, for the sequences $\{x_k\}$ and $\{u_k\}$ generated by Algorithm 3.1 and $\forall x^* \in \operatorname{zer}(A+B+C)$, the following inequality holds:*

$$\Psi_{k+1}(x^*) \le \Psi_k(x^*) - \left(1 - L_{k-1} - L_k - 2\gamma\mu - \frac{\gamma\beta}{2}\right)\|x_{k+1} - x_k\|_S^2, \quad \forall k \in \mathbb{N}_0. \tag{3.2}$$

Proof. From Algorithm 3.1 we have $x_{k+1} = (M_k + A)^{-1}(M_k x_k - 2Bx_k + Bx_{k-1} - Cx_k + \gamma^{-1} u_k)$, which is equivalent to the inclusion

$$M_k x_k - M_k x_{k+1} - 2Bx_k + Bx_{k-1} - Cx_k + \gamma^{-1} u_k \in Ax_{k+1}. \tag{3.3}$$

Since $x^* \in \operatorname{zer}(A+B+C)$, we have $-(B+C)x^* \in Ax^*$, which together with (3.3) and the monotonicity of A obtains

$$0 \leq \langle M_k x_k - M_k x_{k+1} - 2Bx_k + Bx_{k-1} - Cx_k + \gamma^{-1}u_k + (B+C)x^*, x_{k+1} - x^* \rangle. \tag{3.4}$$

Multiplying both sides by 2γ and substituting the definition of u_{k+1} yields

$$\begin{aligned} 0 &\leq 2\langle Sx_k - Sx_{k+1} + u_k - u_{k+1} - 2\gamma Bx_k + \gamma Bx_{k-1} - \gamma Cx_k + \gamma(B+C)x^*, x_{k+1} - x^* \rangle \\ &= 2\langle Sx_k - Sx_{k+1}, x_{k+1} - x^* \rangle + 2\langle u_k - u_{k+1}, x_{k+1} - x^* \rangle - 2\gamma\langle Bx_k - Bx^*, x_{k+1} - x^* \rangle \\ &\quad + 2\gamma\langle Bx_k - Bx_{k-1}, x^* - x_k \rangle + 2\gamma\langle Bx_k - Bx_{k-1}, x_k - x_{k+1} \rangle \\ &\quad - 2\gamma\langle Cx_k - Cx^*, x_{k+1} - x^* \rangle. \end{aligned} \tag{3.5}$$

In the following, we analyze each term on the right-hand side of the inequality one by one. First, using the identity $2\langle x-y|y-z\rangle_S = \|x-z\|_S^2 - \|x-y\|_S^2 - \|y-z\|_S^2$, we have that

$$\begin{aligned} 2\langle Sx_k - Sx_{k+1}, x_{k+1} - x^* \rangle &= 2\langle x_k - x_{k+1}, x_{k+1} - x^* \rangle_S \\ &= \|x_k - x^*\|_S^2 - \|x_{k+1} - x^*\|_S^2 - \|x_{k+1} - x_k\|_S^2. \end{aligned} \tag{3.6}$$

Second, by Assumption 3.1(v) and the definition of u_k, for all $k \in \mathbb{N}_0$,

$$\begin{aligned} 2\langle u_k - u_{k+1}, x_{k+1} - x^* \rangle &= 2\langle u_k, x_k - x^* \rangle + 2\langle u_k, x_{k+1} - x_k \rangle - 2\langle u_{k+1}, x_{k+1} - x^* \rangle \\ &\leq 2\langle u_k, x_k - x^* \rangle - 2\langle u_{k+1}, x_{k+1} - x^* \rangle + L_{k-1}\|x_k - x_{k-1}\|_S^2 \\ &\quad + L_{k-1}\|x_{k+1} - x_k\|_S^2. \end{aligned} \tag{3.7}$$

Third, since the operator B is monotone and μ-Lipschitz w.r.t. S that

$$\begin{aligned} &- 2\gamma\langle Bx_k - Bx^*, x_{k+1} - x^* \rangle + 2\gamma\langle Bx_k - Bx_{k-1}, x^* - x_k \rangle + 2\gamma\langle Bx_k - Bx_{k-1}, x_k - x_{k+1} \rangle \\ &= 2\gamma\langle Bx_{k+1} - Bx^*, x^* - x_{k+1} \rangle + 2\gamma\langle Bx_k - Bx_{k+1}, x^* - x_{k+1} \rangle \\ &\quad + 2\gamma\langle Bx_k - Bx_{k-1}, x^* - x_k \rangle + 2\gamma\langle Bx_k - Bx_{k-1}, x_k - x_{k+1} \rangle \\ &\leq 2\gamma\langle Bx_k - Bx_{k+1}, x^* - x_{k+1} \rangle + 2\gamma\langle Bx_k - Bx_{k-1}, x^* - x_k \rangle \\ &\quad + \gamma\mu\|x_k - x_{k-1}\|_S^2 + \gamma\mu\|x_{k+1} - x_k\|_S^2. \end{aligned} \tag{3.8}$$

Since C is β^{-1}-cocoercive w.r.t. S, one sees from Lemma 2.1 that

$$-2\gamma\langle Cx_k - Cx^*, x_{k+1} - x^* \rangle \leq \frac{\gamma\beta}{2}\|x_{k+1} - x_k\|_S^2. \tag{3.9}$$

By substituting (3.6)–(3.9) into (3.5), we can obtain that

$$\begin{aligned} &\|x_{k+1} - x^*\|_S^2 + 2\langle u_{k+1}, x_{k+1} - x^* \rangle + 2\gamma\langle Bx_{k+1} - Bx_k, x^* - x_{k+1} \rangle \\ &\leq \|x_k - x^*\|_S^2 + 2\langle u_k, x_k - x^* \rangle + (\gamma\mu + L_{k-1})\|x_k - x_{k-1}\|_S^2 \\ &\quad + 2\gamma\langle Bx_k - Bx_{k-1}, x^* - x_k \rangle - \left(1 - L_{k-1} - \gamma\mu - \frac{\gamma\beta}{2}\right)\|x_{k+1} - x_k\|_S^2. \end{aligned}$$

Then, (3.2) follows from the definition of $\Psi_k(x)$ in (3.1). □

Theorem 3.1. *Suppose that Assumption 3.1 holds, and there exists an $\varepsilon > 0$ such that, for all $k \in \mathbb{N}_0$,*

$$1 - L_{k-1} - L_k - 2\gamma\mu - \frac{\gamma\beta}{2} \geq \varepsilon. \tag{3.10}$$

Then, the sequence $\{x_k\}$ generated by Algorithm 3.1 converges weakly to a point x^* in $\operatorname{zer}(A+B+C)$.

Proof. Using the definition of $\Psi_k(x)$ in (3.1), Assumption 3.1(v) and (3.10), we derive that

$$\begin{aligned}\Psi_k(x^*) &= \|x_k-x^*\|_S^2+2\langle u_k,x_k-x^*\rangle+(\gamma\mu+L_{k-1})\|x_k-x_{k-1}\|_S^2\\ &\quad+2\gamma\langle Bx_k-Bx_{k-1},x^*-x_k\rangle\\ &\geq \|x_k-x^*\|_S^2-L_{k-1}(\|x_k-x_{k-1}\|_S^2+\|x_k-x^*\|_S^2)+(\gamma\mu+L_{k-1})\|x_k-x_{k-1}\|_S^2\\ &\quad-\gamma\mu\|x_k-x_{k-1}\|_S^2-\gamma\mu\|x_k-x^*\|_S^2\\ &= (1-L_{k-1}-\gamma\mu)\|x_k-x^*\|_S^2\geq 0,\end{aligned}\tag{3.11}$$

where the first inequality is due to Cauchy-Schwarz inequality and Assumption 3.1(v). Hence, the sequence $\{\Psi_k(x^*)\}$ is nonnegative for all $k\in\mathbb{N}_0$. Since it is also nonincreasing, it follows that it converges. Let $N\in\mathbb{N}_0$. Summing both sides of (3.2) over $k=1,2,\cdots,N$ yields

$$\begin{aligned}\sum_{k=1}^{N}\Big(1-L_{k-1}-L_k-2\gamma\mu-\frac{\gamma\beta}{2}\Big)\|x_{k+1}-x_k\|_S^2 &\leq \Psi_1(x^*)-\Psi_{N+1}(x^*)\\ &\leq \Psi_1(x^*)<+\infty.\end{aligned}$$

It follows directly from (3.10) that $\lim_{k\to\infty}\|x_{k+1}-x_k\|_S=0$. Moreover, by (3.10), (3.11), and the monotonicity of $\{\Psi_k(x^*)\}$, we have

$$\varepsilon\|x_k-x^*\|_S^2\leq\Psi_k(x^*)\leq\Psi_1(x^*),\qquad k\geq 1.$$

Hence, $\{x_k\}$ is bounded. Combining this boundedness with $\|x_{k+1}-x_k\|_S\to 0$ and Assumption 3.1(ii) and (v), we obtain

$$\Psi_k(x^*)-\|x_k-x^*\|_S^2\to 0.$$

Since $\{\Psi_k(x^*)\}$ converges, the limit $\lim_{k\to\infty}\|x_k-x^*\|_S^2$ exists. The sequence $\{x_k\}$ admits at least one weakly convergent subsequence $\{x_{k_n}\}$. Say, without loss of generality, $x_{k_n}\rightharpoonup\bar{x}\in\mathscr{H}$ as $n\to\infty$. Define the sequence $\{\Delta_k\}$ by

$$\Delta_k:=M_kx_k-M_kx_{k+1}+(Bx_{k+1}-Bx_k)-(Bx_k-Bx_{k-1})+(Cx_{k+1}-Cx_k)+\gamma^{-1}u_k.$$

It follows from (3.3) that $(x_{k+1},\Delta_k)\in\operatorname{gra}(A+B+C)$ for all $k\in\mathbb{N}_0$. By Assumption 3.1(ii), (iii) and (v), we obtain that

$$\begin{aligned}\|\Delta_k\|_{S^{-1}}&\leq\|M_kx_k-M_kx_{k+1}\|_{S^{-1}}+\|(B+C)x_k-(B+C)x_{k+1})\|_{S^{-1}}+\|Bx_k-Bx_{k-1}\|_{S^{-1}}+\gamma^{-1}\|u_k\|_{S^{-1}}\\ &\leq\Big(\frac{2}{\gamma}+\mu+\beta\Big)\|x_k-x_{k+1}\|_S+\Big(\mu+\frac{L_{k-1}}{\gamma}\Big)\|x_k-x_{k-1}\|_S.\end{aligned}$$

Combining this with $\lim\limits_{k\to\infty}\|x_{k+1}-x_k\|_S=0$, we obtain $\Delta_{k_n}\to 0$. Under the assumption that the operator B has full domain, Corollary 25.5 (i) in [1] yields that $A+B$ is maximally monotone. Moreover, Lemma 2.1 in [30] and the cocoercivity of C imply that $A+B+C$ is maximally monotone. By Lemma 2.2, $(\bar{x},0)\in\operatorname{gra}(A+B+C)$, i.e., $\bar{x}\in\operatorname{zer}(A+B+C)$, Lemma 2.3 shows that the sequence $\{x_k\}$ converges weakly to a point in $\operatorname{zer}(A+B+C)$. □

Remark 3.1. We now consider three special cases of Algorithm 3.1.

(i) When $B = 0$ (hence $\mu = 0$). Then Algorithm 3.1 reduces to the nonlinear forward-backward splitting with momentum proposed in [22], and the condition on γ becomes $1 - L_{k-1} - L_k - \gamma\beta/2 \geq \varepsilon$.

(ii) When $S = \mathrm{Id}$, and $M_k \equiv \mathrm{Id}/\gamma$ for all $k \in \mathbb{N}$, we have $L_k = 0$ and $u_k = 0$ for all $k \in \mathbb{N}$. In this case, Algorithm 3.1 simplifies to the SFRBS algorithm (1.4). Furthermore, the upper bound on γ coincides with that in [21], namely, $\gamma < 2/(4\mu + \beta)$.

(iii) Suppose that $A = A_1 + A_2$, where $A_1 : \mathscr{H} \to 2^{\mathscr{H}}$ is maximally monotone, $A_2 : \mathscr{H} \to \mathscr{H}$ is monotone and L-Lipschitz continuous, and $A_1 + A_2$ is maximally monotone. Furthermore, let $S = \mathrm{Id}$, and $M_k \equiv M = \mathrm{Id}/\gamma - A_2$ for all $k \in \mathbb{N}$. In this case, with the initialization $x_0 = x_1$, Algorithm 3.1 takes the form

$$x_{k+1} = J_{\gamma A_1}(x_k - 2\gamma A_2 x_k - 2\gamma B x_k + \gamma B x_{k-1} + \gamma A_2 x_{k-1} - \gamma C x_k). \tag{3.12}$$

Since $\gamma M - \mathrm{Id}$ is γL-Lipschitz continuous, condition (3.10) reduces to $1 - 2\gamma L - 2\gamma\mu - \gamma\beta/2 \geq \varepsilon$. In fact, Algorithm (3.12) is equivalent to the SFRBS algorithm (1.4) for problem (1.1). To see this, define $D := A_2 + B$. Under these assumptions, D is monotone and Lipschitz continuous. Substituting D into (3.12) yields exactly the SFRBS iteration (1.4), under the same step-size conditions.

3.1.1. *Linear convergence.* This subsection establishes the R–linear convergence of the sequence $\{x_k\}$ generated by Algorithm 3.1, under the assumption that the operator A is strongly monotone. Specifically, A is said to be ρ-strongly monotone w.r.t. S ($\rho > 0$) if

$$\langle u - v, x - y\rangle \geq \rho\|x - y\|_S^2, \qquad \forall (x,u),(y,v) \in \mathrm{gra}(A).$$

The same convergence result holds when B is ρ-strongly monotone with respect to S, as explained in the proof below.

Theorem 3.2. *Let Assumption 3.1 hold, and assume that A is ρ-strongly monotone with respect to S. Suppose that condition (3.10) holds. Fix $\varepsilon_1 > 0$ and define*

$$t_k = \min\left\{\frac{2\gamma\rho}{1 + L_k/\varepsilon_1 + \gamma\mu}, \frac{\kappa_k}{\varepsilon_1 L_k + 2\gamma\mu + L_k}\right\}, \tag{3.13}$$

where $\kappa_k = (1 - L_{k-1} - L_k - 2\gamma\mu - \gamma\beta/2)$. The second fraction is interpreted as $+\infty$ when its denominator is zero. Set $t := \inf_{k\geq 1} t_k$. Since $L_k < 1$ and $\kappa_k \geq \varepsilon$ by (3.10), we have $t > 0$. Then the sequence $\{x_k\}$ generated by Algorithm 3.1 converges R–linearly to the unique point x^* in $\mathrm{zer}(A + B + C)$.

Proof. Using the strong monotonicity of A, inequality (3.4) can be rewritten as

$$\rho\|x_{k+1} - x^*\|_S^2 \leq \langle M_k x_k - M_k x_{k+1} - 2Bx_k + Bx_{k-1} - Cx_k + \gamma^{-1}u_k + (B + C)x^*, x_{k+1} - x^*\rangle.$$

Following the same arguments as in Lemma 3.1, we obtain

$$\begin{aligned}
&(1 + 2\gamma\rho)\|x_{k+1} - x^*\|_S^2 + 2\langle u_{k+1}, x_{k+1} - x^*\rangle + 2\gamma\langle Bx_{k+1} - Bx_k, x^* - x_{k+1}\rangle \\
\leq &\|x_k - x^*\|_S^2 + 2\langle u_k, x_k - x^*\rangle + (\gamma\mu + L_{k-1})\|x_k - x_{k-1}\|_S^2 \\
&+ 2\gamma\langle Bx_k - Bx_{k-1}, x^* - x_k\rangle - \left(1 - L_{k-1} - \gamma\mu - \frac{\gamma\beta}{2}\right)\|x_{k+1} - x_k\|_S^2.
\end{aligned} \tag{3.14}$$

If A is merely monotone and B is ρ-strongly monotone with respect to S, the same inequality follows by using

$$2\gamma\langle Bx_{k+1}-Bx^*,x^*-x_{k+1}\rangle \le -2\gamma\rho\|x_{k+1}-x^*\|_S^2$$

in the proof of Lemma 3.1. Set

$$\begin{aligned} a_k(x^*) &= 2\langle u_k,x_k-x^*\rangle+(\gamma\mu+L_{k-1})\|x_k-x_{k-1}\|_S^2+2\gamma\langle Bx_k-Bx_{k-1},x^*-x_k\rangle,\\ b_k &= \kappa_k\|x_{k+1}-x_k\|_S^2. \end{aligned}$$

Then inequality (3.14) can be written as

$$(1+2\gamma\rho)\|x_{k+1}-x^*\|_S^2+a_{k+1}(x^*)+b_k \le \|x_k-x^*\|_S^2+a_k(x^*). \tag{3.15}$$

Using Assumption 3.1(v) and Young's inequality, we can obtain the upper bound for $a_{k+1}(x^*)$,

$$\begin{aligned} a_{k+1}(x^*) =& 2\langle u_{k+1},x_{k+1}-x^*\rangle+(\gamma\mu+L_k)\|x_{k+1}-x_k\|_S^2\\ &+2\gamma\langle Bx_{k+1}-Bx_k,x^*-x_{k+1}\rangle\\ \le&(\varepsilon_1L_k+2\gamma\mu+L_k)\|x_{k+1}-x_k\|_S^2+\left(\frac{L_k}{\varepsilon_1}+\gamma\mu\right)\|x_{k+1}-x^*\|_S^2, \end{aligned} \tag{3.16}$$

where $\varepsilon_1>0$. Since $t\le t_k$, (3.13) and (3.16) give

$$ta_{k+1}(x^*) \le (2\gamma\rho-t)\|x_{k+1}-x^*\|_S^2+b_k. \tag{3.17}$$

Combining (3.15) and (3.17), we obtain $(1+t)(\|x_{k+1}-x^*\|_S^2+a_{k+1}(x^*)) \le \|x_k-x^*\|_S^2+a_k(x^*)$. On the other hand, from (3.11) and (3.10), we know that there exists a constant $k_1>0$ such that

$$\|x_{k+1}-x^*\|_S^2+a_{k+1}(x^*) \ge (1-L_k-\gamma\mu)\|x_{k+1}-x^*\|_S^2 \ge k_1\|x_{k+1}-x^*\|_S^2.$$

Therefore,

$$\begin{aligned} k_1\|x_{k+1}-x^*\|_S^2 &\le \|x_{k+1}-x^*\|_S^2+a_{k+1}(x^*)\\ &\le \frac{\|x_k-x^*\|_S^2+a_k(x^*)}{1+t} \le \cdots \le \frac{\|x_1-x^*\|_S^2+a_1(x^*)}{(1+t)^k}. \end{aligned}$$

Hence, the sequence $\{x_k\}$ generated by Algorithm 3.1 converges R–linearly to a point x^* in $\mathrm{zer}(A+B+C)$. □

3.2. **Nonlinear semi-reflected-forward-backward splitting algorithm with momentum.** In this subsection, we propose a nonlinear semi-reflected-forward-backward splitting (SRFBS) algorithm with momentum (Algorithm 3.2) and establish its convergence analysis.

Algorithm 3.2. ***Nonlinear SRFBS algorithm with momentum***

Let $S\in\mathscr{P}(\mathscr{H})$, $M_k:\mathscr{H}\to\mathscr{H}$ for all $k\in\mathbb{N}_0$ and $\gamma>0$. Choose initial points $x_0,x_1\in\mathscr{H}$, $u_1=0$. For $k=1,2,\cdots$, do

$$\begin{aligned} y_k &= 2x_k-x_{k-1},\\ x_{k+1} &= (M_k+A)^{-1}(M_kx_k-By_k-Cx_k+\gamma^{-1}u_k),\\ u_{k+1} &= (\gamma M_k-S)x_{k+1}-(\gamma M_k-S)x_k. \end{aligned}$$

For the iterates $\{x_k\}$, $\{y_k\}$, and $\{u_k\}$ generated by Algorithm 3.2, and $\forall x \in \operatorname{zer}(A+B+C)$, for $k \geq 2$, we define

$$\Gamma_k(x) := \|x_k - x\|_S^2 + 2\langle u_k, x_k - x\rangle + 2\gamma\langle By_{k-1} - Bx, x_k - x_{k-1}\rangle + 2\langle u_k - u_{k-1}, x_k - x_{k-1}\rangle, \tag{3.18}$$

which will be used to establish the weak convergence of Algorithm 3.2.

Lemma 3.2. *Suppose that Assumption 3.1 holds, and let $\varepsilon_2 > 0$. Then, for the sequences $\{x_k\}$, $\{y_k\}$ and $\{u_k\}$ generated by Algorithm 3.2, and $\forall x^* \in \operatorname{zer}(A+B+C)$, the following holds,*

$$\begin{aligned}
&\Gamma_{k+1}(x^*) + \left(2 - \frac{(1+2\varepsilon_2)\gamma\beta}{2} - L_{k-1}\right)\|x_{k+1} - x_k\|_S^2 + \|x_{k+1} - y_k\|_S^2 \\
\leq &\Gamma_k(x^*) + \left(1 + 2L_{k-1} + \frac{\gamma\beta}{\varepsilon_2} + \gamma\mu(\sqrt{2}+1)\right)\|x_k - x_{k-1}\|_S^2 + L_{k-2}\|x_{k-1} - x_{k-2}\|_S^2 \\
&+ \gamma\mu\|x_k - y_{k-1}\|_S^2 + (\sqrt{2}\gamma\mu + L_{k-1} + L_{k-2})\|x_{k+1} - y_k\|_S^2 \qquad \forall k \geq 2.
\end{aligned} \tag{3.19}$$

Proof. From Algorithm 3.2, we have $x_{k+1} = (M_k + A)^{-1}(M_k x_k - By_k - Cx_k + \gamma^{-1}u_k)$, which is equivalent to the inclusion

$$M_k x_k - M_k x_{k+1} - By_k - Cx_k + \gamma^{-1}u_k \in Ax_{k+1}. \tag{3.20}$$

Since $x^* \in \operatorname{zer}(A+B+C)$, we have $-(B+C)x^* \in Ax^*$. Using this, (3.20) and the monotonicity of A, we obtain $0 \leq \langle M_k x_k - M_k x_{k+1} - By_k - Cx_k + \gamma^{-1}u_k + (B+C)x^*, x_{k+1} - x^*\rangle$. Multiplying both sides by 2γ and applying the definition of u_{k+1} yield

$$\begin{aligned}
0 &\leq 2\langle Sx_k - Sx_{k+1} + u_k - u_{k+1} - \gamma By_k - \gamma Cx_k + \gamma(B+C)x^*, x_{k+1} - x^*\rangle \\
&= 2\langle Sx_k - Sx_{k+1}, x_{k+1} - x^*\rangle + 2\langle u_k - u_{k+1}, x_{k+1} - x^*\rangle - 2\gamma\langle By_k - Bx^*, x_{k+1} - x^*\rangle \\
&\quad - 2\gamma\langle Cx_k - Cx^*, x_{k+1} - x^*\rangle.
\end{aligned} \tag{3.21}$$

In the following, we analyze each term on the right-hand side of the inequality (3.21) separately. Similar to (3.6) and (3.7), we have that

$$2\langle Sx_k - Sx_{k+1}, x_{k+1} - x^*\rangle = \|x_k - x^*\|_S^2 - \|x_{k+1} - x^*\|_S^2 - \|x_{k+1} - x_k\|_S^2, \tag{3.22}$$

and

$$\begin{aligned}
2\langle u_k - u_{k+1}, x_{k+1} - x^*\rangle &= 2\langle u_k, x_k - x^*\rangle + 2\langle u_k, x_{k+1} - x_k\rangle - 2\langle u_{k+1}, x_{k+1} - x^*\rangle \\
&\leq 2\langle u_k, x_k - x^*\rangle - 2\langle u_{k+1}, x_{k+1} - x^*\rangle + L_{k-1}\|x_k - x_{k-1}\|_S^2 \\
&\quad + L_{k-1}\|x_{k+1} - x_k\|_S^2.
\end{aligned} \tag{3.23}$$

By using the monotonicity of B and the definition of y_k, one sees that

$$\begin{aligned}
-2\gamma\langle By_k - Bx^*, x_{k+1} - x^*\rangle &\leq -2\gamma\langle By_k - Bx^*, x_{k+1} - y_k\rangle \\
&= -2\gamma\langle By_k - Bx^*, x_{k+1} - x_k\rangle + 2\gamma\langle By_{k-1} - Bx^*, x_k - x_{k-1}\rangle \\
&\quad + 2\gamma\langle By_k - By_{k-1}, x_k - x_{k-1}\rangle \\
&= -2\gamma\langle By_k - Bx^*, x_{k+1} - x_k\rangle + 2\gamma\langle By_{k-1} - Bx^*, x_k - x_{k-1}\rangle \\
&\quad + 2\gamma\langle By_k - By_{k-1}, x_{k+1} - x_k\rangle + 2\gamma\langle By_k - By_{k-1}, y_k - x_{k+1}\rangle.
\end{aligned} \tag{3.24}$$

In view of Lemma 2.1 and the cocoercivity of C, we obtain that

$$-2\gamma\langle Cx_k - Cx^*, x_{k+1} - x^*\rangle \leq \frac{\gamma\beta}{2}\|x_{k+1} - x_k\|_S^2. \tag{3.25}$$

Substituting (3.22)–(3.25) into (3.21) and rearranging the terms yields

$$\begin{aligned}
&\|x_{k+1} - x^*\|_S^2 + 2\langle u_{k+1}, x_{k+1} - x^*\rangle + 2\gamma\langle By_k - Bx^*, x_{k+1} - x_k\rangle \\
&+ \big(1 - \frac{\gamma\beta}{2} - L_{k-1}\big)\|x_{k+1} - x_k\|_S^2 \\
\leq &\|x_k - x^*\|_S^2 + 2\langle u_k, x_k - x^*\rangle + L_{k-1}\|x_k - x_{k-1}\|_S^2 + 2\gamma\langle By_k - By_{k-1}, y_k - x_{k+1}\rangle \\
&+ 2\gamma\langle By_{k-1} - Bx^*, x_k - x_{k-1}\rangle + 2\gamma\langle By_k - By_{k-1}, x_{k+1} - x_k\rangle.
\end{aligned} \tag{3.26}$$

From Algorithm 3.2 we have $M_k x_k - M_k x_{k+1} - By_k - Cx_k + \gamma^{-1}u_k \in Ax_{k+1}$ and $M_{k-1}x_{k-1} - M_{k-1}x_k - By_{k-1} - Cx_{k-1} + \gamma^{-1}u_{k-1} \in Ax_k$. Then, it follows from the monotonicity of A that

$$\begin{aligned}
&\langle M_k x_k - M_k x_{k+1} - By_k - Cx_k + \gamma^{-1}u_k, x_{k+1} - x_k\rangle \\
&- \langle M_{k-1}x_{k-1} - M_{k-1}x_k - By_{k-1} - Cx_{k-1} + \gamma^{-1}u_{k-1}, x_{k+1} - x_k\rangle \geq 0.
\end{aligned}$$

Multiplying both sides by 2γ and rearranging the terms yields

$$\begin{aligned}
&2\gamma\langle By_k - By_{k-1}, x_{k+1} - x_k\rangle \\
\leq &2\langle Sx_k - Sx_{k+1} - u_{k+1} + u_k - Sx_{k-1} + Sx_k + u_k - u_{k-1} - \gamma Cx_k + \gamma Cx_{k-1}, x_{k+1} - x_k\rangle \\
= &-2\|x_{k+1} - x_k\|_S^2 + 2\langle y_k - x_k, x_{k+1} - x_k\rangle_S + 2\langle -u_{k+1} + 2u_k - u_{k-1}, x_{k+1} - x_k\rangle \\
&- 2\gamma\langle Cx_k - Cx_{k-1}, x_{k+1} - x_k\rangle \\
\leq &-\|x_{k+1} - x_k\|_S^2 - \|x_{k+1} - y_k\|_S^2 + \|x_k - x_{k-1}\|_S^2 - 2\langle u_{k+1} - u_k, x_{k+1} - x_k\rangle \\
&+ 2\langle u_k - u_{k-1}, x_k - x_{k-1}\rangle + 2\langle u_k - u_{k-1}, x_{k+1} - y_k\rangle + \frac{\gamma\beta}{\varepsilon_2}\|x_k - x_{k-1}\|_S^2 \\
&+ \gamma\beta\varepsilon_2\|x_{k+1} - x_k\|_S^2,
\end{aligned} \tag{3.27}$$

where the second equality is due to the definition of y_k, and the last inequality follows from (2.1) and Young's inequality. Using the Lipschitz property of B and Cauchy-Schwarz inequality, one has

$$\begin{aligned}
2\gamma\langle By_k - By_{k-1}, y_k - x_{k+1}\rangle \leq &2\gamma\mu\|y_k - y_{k-1}\|_S\|x_{k+1} - y_k\|_S \\
\leq &2\gamma\mu(\|x_k - y_k\|_S + \|x_k - y_{k-1}\|_S)\|x_{k+1} - y_k\|_S \\
\leq &\gamma\mu((\sqrt{2}+1)\|x_k - x_{k-1}\|_S^2 + \sqrt{2}\|x_{k+1} - y_k\|_S^2) + \gamma\mu\|x_k - y_{k-1}\|_S^2.
\end{aligned} \tag{3.28}$$

Substituting (3.27) and (3.28) into (3.26), we obtain that

$$\begin{aligned}
&\|x_{k+1}-x^*\|_S^2+2\langle u_{k+1},x_{k+1}-x^*\rangle+2\gamma\langle By_k-Bx^*,x_{k+1}-x_k\rangle\\
&+2\langle u_{k+1}-u_k,x_{k+1}-x_k\rangle+\Big(2-\frac{(1+2\varepsilon_2)\gamma\beta}{2}-L_{k-1}\Big)\|x_{k+1}-x_k\|_S^2+\|x_{k+1}-y_k\|_S^2\\
\le&\|x_k-x^*\|_S^2+2\langle u_k,x_k-x^*\rangle+2\gamma\langle By_{k-1}-Bx^*,x_k-x_{k-1}\rangle\\
&+2\langle u_k-u_{k-1},x_k-x_{k-1}\rangle+\Big(1+L_{k-1}+\frac{\gamma\beta}{\varepsilon_2}+\gamma\mu(\sqrt{2}+1)\Big)\|x_k-x_{k-1}\|_S^2\\
&+2\langle u_k-u_{k-1},x_{k+1}-y_k\rangle+\sqrt{2}\gamma\mu\|x_{k+1}-y_k\|_S^2+\gamma\mu\|x_k-y_{k-1}\|_S^2. \qquad (3.29)
\end{aligned}$$

It follows from Assumption 3.1(v) that

$$2\langle u_k-u_{k-1},x_{k+1}-y_k\rangle\le L_{k-1}\|x_k-x_{k-1}\|_S^2+L_{k-2}\|x_{k-1}-x_{k-2}\|_S^2+(L_{k-1}+L_{k-2})\|x_{k+1}-y_k\|_S^2.$$

Using this and (3.29), we obtain that

$$\begin{aligned}
&\|x_{k+1}-x^*\|_S^2+2\langle u_{k+1},x_{k+1}-x^*\rangle+2\gamma\langle By_k-Bx^*,x_{k+1}-x_k\rangle\\
&+2\langle u_{k+1}-u_k,x_{k+1}-x_k\rangle+\Big(2-\frac{(1+2\varepsilon_2)\gamma\beta}{2}-L_{k-1}\Big)\|x_{k+1}-x_k\|_S^2+\|x_{k+1}-y_k\|_S^2\\
\le&\|x_k-x^*\|_S^2+2\langle u_k,x_k-x^*\rangle+2\gamma\langle By_{k-1}-Bx^*,x_k-x_{k-1}\rangle\\
&+2\langle u_k-u_{k-1},x_k-x_{k-1}\rangle+\Big(1+2L_{k-1}+\frac{\gamma\beta}{\varepsilon_2}+\gamma\mu(\sqrt{2}+1)\Big)\|x_k-x_{k-1}\|_S^2\\
&+L_{k-2}\|x_{k-1}-x_{k-2}\|_S^2+\gamma\mu\|x_k-y_{k-1}\|_S^2+(\sqrt{2}\gamma\mu+L_{k-1}+L_{k-2})\|x_{k+1}-y_k\|_S^2.
\end{aligned}$$

Using the definition of $\Gamma_k(x)$ in (3.18), we obtain (3.19) immediately. □

Theorem 3.3. *Suppose that Assumption 3.1 holds, and let $\varepsilon_2>0$. Assume that there exists a constant $\varepsilon>0$ such that, for all $k\ge 2$,*

$$1-3L_k-\frac{(2+\varepsilon_2+2\varepsilon_2^2)\gamma\beta}{2\varepsilon_2}-L_{k-1}-\gamma\mu(\sqrt{2}+1)\ge\varepsilon,\ 1-L_{k-1}-L_{k-2}-\gamma\mu(\sqrt{2}+1)\ge\varepsilon. \qquad (3.30)$$

Then, the sequence $\{x_k\}$ generated by Algorithm 3.2 converges weakly to a point x^* in $\operatorname{zer}(A+B+C)$.

Proof. For any x^* in $\operatorname{zer}(A+B+C)$, for $k\ge 2$, we define

$$\begin{aligned}
\Xi_k(x^*):=&\Gamma_k(x^*)+\Big(1+3L_{k-1}+\frac{\gamma\beta}{\varepsilon_2}+\gamma\mu(\sqrt{2}+1)\Big)\|x_k-x_{k-1}\|_S^2\\
&+L_{k-2}\|x_{k-1}-x_{k-2}\|_S^2+\gamma\mu\|x_k-y_{k-1}\|_S^2. \qquad (3.31)
\end{aligned}$$

Using this and (3.19), we have that

$$\begin{aligned}
\Xi_{k+1}(x^*)\le&\Xi_k(x^*)-\Big(1-3L_k-\frac{(2+\varepsilon_2+2\varepsilon_2^2)\gamma\beta}{2\varepsilon_2}-L_{k-1}-\gamma\mu(\sqrt{2}+1)\Big)\|x_{k+1}-x_k\|_S^2\\
&-(1-L_{k-1}-L_{k-2}-\gamma\mu(\sqrt{2}+1))\|x_{k+1}-y_k\|_S^2. \qquad (3.32)
\end{aligned}$$

We now derive the lower bound of $\Xi_k(x^*)$ for all $k \geq 2$. We first bound $\Gamma_k(x^*)$ using Assumption 3.1(ii) and (v):

$$\begin{aligned}\Gamma_k(x^*) \geq & \|x_k - x^*\|_S^2 - 2\|u_k\|_{S^{-1}}\|x_k - x^*\|_S + 2\gamma\langle By_{k-1} - Bx_k, x_k - x_{k-1}\rangle \\ & + 2\gamma\langle Bx_k - Bx^*, x_k - x_{k-1}\rangle - 2\|u_k\|_{S^{-1}}\|x_k - x_{k-1}\|_S \\ & - 2\|u_{k-1}\|_{S^{-1}}\|x_k - x_{k-1}\|_S \\ \geq & (1 - L_{k-1} - \gamma\mu)\|x_k - x^*\|_S^2 - (3L_{k-1} + 2\gamma\mu + L_{k-2})\|x_k - x_{k-1}\|_S^2 \\ & - \gamma\mu\|x_k - y_{k-1}\|_S^2 - L_{k-2}\|x_{k-1} - x_{k-2}\|_S^2. \end{aligned} \tag{3.33}$$

Therefore, by combining the lower bound of $\Gamma_k(x^*)$ in (3.33) with the definition of $\Xi_k(x^*)$ in (3.31), we obtain that

$$\begin{aligned}\Xi_k(x^*) \geq & (1 - L_{k-1} - \gamma\mu)\|x_k - x^*\|_S^2 - (3L_{k-1} + 2\gamma\mu + L_{k-2})\|x_k - x_{k-1}\|_S^2 \\ & - \gamma\mu\|x_k - y_{k-1}\|_S^2 - L_{k-2}\|x_{k-1} - x_{k-2}\|_S^2 + L_{k-2}\|x_{k-1} - x_{k-2}\|_S^2 \\ & + \gamma\mu\|x_k - y_{k-1}\|_S^2 + \left(1 + 3L_{k-1} + \frac{\gamma\beta}{\varepsilon_2} + \gamma\mu(\sqrt{2}+1)\right)\|x_k - x_{k-1}\|_S^2 \\ = & (1 - L_{k-1} - \gamma\mu)\|x_k - x^*\|_S^2 + \left(1 + \frac{\gamma\beta}{\varepsilon_2} - L_{k-2} + (\sqrt{2}-1)\gamma\mu\right)\|x_k - x_{k-1}\|_S^2 \\ \geq & 0. \end{aligned} \tag{3.34}$$

By (3.30), (3.32), and (3.34),

$$\varepsilon\|x_k - x^*\|_S^2 \leq \Xi_k(x^*) \leq \Xi_2(x^*), \qquad k \geq 2.$$

Hence, $\{x_k\}$ is bounded, and the sequence $\{\Xi_k(x^*)\}$ is nonnegative for all $k \geq 2$. Since it is also nonincreasing by (3.32), it follows that $\{\Xi_k(x^*)\}$ is convergent. By (3.32), one has that $\lim_{k\to\infty}\|x_{k+1} - x_k\|_S = 0$, and $\lim_{k\to\infty}\|x_{k+1} - y_k\|_S = 0$. Using these limits, the boundedness of $\{x_k\}$, and Assumption 3.1(ii) and (v), we obtain

$$\Xi_k(x^*) - \|x_k - x^*\|_S^2 \to 0.$$

Consequently, $\lim_{k\to\infty}\|x_k - x^*\|_S^2$ exists. Hence, there exists at least one weakly convergent subsequence $\{x_{k_n}\}$ such that $x_{k_n} \rightharpoonup \bar{x} \in \mathscr{H}$ as $n \to \infty$. Define

$$\Delta_k := M_k x_k - M_k x_{k+1} + (Bx_{k+1} - By_k) + (Cx_{k+1} - Cx_k) + \gamma^{-1}u_k.$$

Then, from (3.20), we have $(x_{k+1}, \Delta_k) \in \operatorname{gra}(A + B + C)$ for all $k \geq 2$. Using Assumption 3.1(ii), (iii), and (v), we obtain that

$$\begin{aligned}\|\Delta_k\|_{S^{-1}} \leq & \|M_k x_k - M_k x_{k+1}\|_{S^{-1}} + \|Bx_{k+1} - By_k\|_{S^{-1}} + \|Cx_{k+1} - Cx_k\|_{S^{-1}} + \gamma^{-1}\|u_k\|_{S^{-1}} \\ \leq & \left(\frac{2}{\gamma} + \beta\right)\|x_k - x_{k+1}\|_S + \mu\|x_{k+1} - y_k\|_S + \frac{L_{k-1}}{\gamma}\|x_k - x_{k-1}\|_S. \end{aligned}$$

By the arguments above, we derive that $\Delta_{k_n} \to 0$. The remaining part of the proof follows the same reasoning as in Theorem 3.1 and is therefore omitted. □

Remark 3.2. We present two special cases of Algorithm 3.2.

(i) When $S = \mathrm{Id}$, and $M_k \equiv \mathrm{Id}/\gamma$, we have $L_k = 0$ and $u_k = 0$ for all $k \in \mathbb{N}$. In this case, Algorithm 3.2 reduces to the SRFBS algorithm (1.5). Moreover, the admissible step size satisfies

$$0 < \gamma < \left(\frac{(2+\varepsilon_2+2\varepsilon_2^2)\beta}{2\varepsilon_2} + (\sqrt{2}+1)\mu \right)^{-1}.$$

Notably, the optimized step-size bound coincides with that in [16].

(ii) Suppose that $A = A_1 + A_2$, where $A_1 : \mathscr{H} \to 2^{\mathscr{H}}$ is maximally monotone, $A_2 : \mathscr{H} \to \mathscr{H}$ is L-Lipschitz continuous, and $A_1 + A_2$ is maximally monotone. Let $S = \mathrm{Id}$, $\gamma_k \equiv \gamma$, and $M_k \equiv M = \mathrm{Id}/\gamma - A_2$, $\forall k \in \mathbb{N}$. Then, with the initialization $x_0 = x_1$, Algorithm 3.2 takes the form

$$\begin{cases} y_k = 2x_k - x_{k-1}, \\ x_{k+1} = J_{\gamma A_1}(x_k - 2\gamma A_2 x_k - \gamma B y_k - \gamma C x_k + \gamma A_2 x_{k-1}). \end{cases} \tag{3.35}$$

3.2.1. *Linear convergence.* This subsection establishes the R–linear convergence of the sequence $\{x_k\}$ generated by Algorithm 3.2 under the assumption that A is strongly monotone.

Theorem 3.4. *Suppose that Assumption 3.1 holds. Assume further that A is ρ-strongly monotone with respect to S, and that the conditions in (3.30) are satisfied. Fix $\varepsilon_3, \varepsilon_4 > 0$. For each $k \geq 2$, define*

$$t_k := \min\left\{ \frac{2\gamma\rho}{1 + L_k/\varepsilon_3 + 2\varepsilon_4\gamma\mu}, \frac{\delta_k^1}{1 + 5L_{k-1} + (\varepsilon_3+5)L_k + \gamma\beta/\varepsilon_2 + \gamma\mu(1/\varepsilon_4 + \sqrt{2}+1)}, \frac{\delta_k^2}{\gamma\mu(2\varepsilon_4+1) + 4L_{k-1}} \right\}, \tag{3.36}$$

where $\delta_k^1 = 1 - 3L_k - (2+\varepsilon_2+2\varepsilon_2^2)\gamma\beta/(2\varepsilon_2) - L_{k-1} - \gamma\mu(\sqrt{2}+1)$ and $\delta_k^2 = 1 - L_{k-1} - L_{k-2} - \gamma\mu(\sqrt{2}+1)$. The third fraction is interpreted as $+\infty$ when its denominator is zero. Set $t := \inf_{k\geq 2} t_k$. Since $L_k < 1$ and $\delta_k^1, \delta_k^2 \geq \varepsilon$ by (3.30), we have $t > 0$. Then the sequence $\{x_k\}$ generated by Algorithm 3.2 converges R–linearly to a point x^* in $\mathrm{zer}(A+B+C)$.

Proof. For any x^* in $\mathrm{zer}(A+B+C)$, for $k \geq 2$, we define

$$\begin{aligned} c_k(x^*) =& 2\langle u_k, x_k - x^*\rangle + 2\gamma\langle By_{k-1} - Bx^*, x_k - x_{k-1}\rangle + 2\langle u_k - u_{k-1}, x_k - x_{k-1}\rangle \\ &+ \left(1 + 3L_{k-1} + \frac{\gamma\beta}{\varepsilon_2} + \gamma\mu(\sqrt{2}+1)\right)\|x_k - x_{k-1}\|_S^2 + L_{k-2}\|x_{k-1} - x_{k-2}\|_S^2 \\ &+ \gamma\mu\|x_k - y_{k-1}\|_S^2. \end{aligned}$$

By applying the strong monotonicity of A in Lemma 3.2 and following (3.18), (3.31) and (3.32), we obtain

$$(1+2\gamma\rho)\|x_{k+1} - x^*\|_S^2 + c_{k+1}(x^*) + \delta_k^1\|x_{k+1} - x_k\|_S^2 + \delta_k^2\|x_{k+1} - y_k\|_S^2 \leq \|x_k - x^*\|_S^2 + c_k(x^*). \tag{3.37}$$

By applying the Lipschitz continuity of B, Assumption 3.1(v), and Young's inequality, we obtain

$$\begin{aligned}
c_{k+1}(x^*) \leq & \varepsilon_3 L_k \|x_{k+1}-x_k\|_S^2 + \frac{L_k}{\varepsilon_3}\|x_{k+1}-x^*\|_S^2 + 2\varepsilon_4\gamma\mu\|x_{k+1}-y_k\|_S^2 \\
& + 2\varepsilon_4\gamma\mu\|x_{k+1}-x^*\|_S^2 + \frac{\gamma\mu}{\varepsilon_4}\|x_{k+1}-x_k\|_S^2 + 2L_k\|x_{k+1}-x_k\|_S^2 \\
& + 2L_{k-1}\|(x_{k+1}-x_k)-(x_{k+1}-y_k)\|_S^2 + L_{k-1}\|x_{k+1}-x_k\|_S^2 \\
& + \Big(1+3L_k+\frac{\gamma\beta}{\varepsilon_2}+\gamma\mu(\sqrt{2}+1)\Big)\|x_{k+1}-x_k\|_S^2 + \gamma\mu\|x_{k+1}-y_k\|_S^2 \\
\leq & \Big(\frac{L_k}{\varepsilon_3}+2\varepsilon_4\gamma\mu\Big)\|x_{k+1}-x^*\|_S^2 + (\gamma\mu(2\varepsilon_4+1)+4L_{k-1})\|x_{k+1}-y_k\|_S^2 \\
& + \Big(1+5L_{k-1}+(\varepsilon_3+5)L_k+\frac{\gamma\beta}{\varepsilon_2}+\gamma\mu\big(\frac{1}{\varepsilon_4}+\sqrt{2}+1\big)\Big)\|x_{k+1}-x_k\|_S^2.
\end{aligned}$$

Since $t \leq t_k$, (3.36) and the preceding estimate give

$$tc_{k+1}(x^*) \leq (2\gamma\rho - t)\|x_{k+1}-x^*\|_S^2 + \delta_k^1\|x_{k+1}-x_k\|_S^2 + \delta_k^2\|x_{k+1}-y_k\|_S^2. \tag{3.38}$$

Combining (3.37) and (3.38), we have that

$$(1+t)(\|x_{k+1}-x^*\|_S^2 + c_{k+1}(x^*)) \leq \|x_k-x^*\|_S^2 + c_k(x^*).$$

By (3.34) and under the conditions in (3.30), there exists a constant $k_2 > 0$ such that

$$\begin{aligned}
& \|x_{k+1}-x^*\|_S^2 + c_{k+1}(x^*) = \Xi_{k+1}(x^*) \\
\geq & (1-L_k-\gamma\mu)\|x_{k+1}-x^*\|_S^2 + \Big(1-L_{k-1}+\frac{\gamma\beta}{\varepsilon_2}+(\sqrt{2}-1)\gamma\mu\Big)\|x_{k+1}-x_k\|_S^2 \\
\geq & k_2\|x_{k+1}-x^*\|_S^2.
\end{aligned}$$

Therefore, for all $k \geq 2$,

$$k_2\|x_{k+1}-x^*\|_S^2 \leq \|x_{k+1}-x^*\|_S^2 + c_{k+1}(x^*) \leq \frac{\|x_k-x^*\|_S^2 + c_k(x^*)}{1+t} \leq \cdots \leq \frac{\|x_2-x^*\|_S^2 + c_2(x^*)}{(1+t)^{k-1}}.$$

which implies that the sequence $\{x_k\}$ generated by Algorithm 3.2 converges R–linearly to a point x^* in $\operatorname{zer}(A+B+C)$. □

3.3. **Nonlinear outer reflected forward-backward splitting algorithm with momentum.** In this subsection, we introduce a nonlinear outer reflected forward-backward splitting (ORFBS) algorithm with momentum, denoted by Algorithm 3.3, and establish its convergence analysis.

Algorithm 3.3. ***Nonlinear ORFBS algorithm with momentum***
Let $S \in \mathscr{P}(\mathscr{H})$, $M_k : \mathscr{H} \to \mathscr{H}$ for all $k \in \mathbb{N}_0$ and $\gamma > 0$. Choose initial points $x_0, x_1 \in \mathscr{H}$, $u_1 = 0$. For $k = 1, 2, \cdots$, do

$$\begin{aligned}
y_k &= (M_k + A)^{-1}(M_k x_k - (B+C)x_k + \gamma^{-1}u_k), \\
x_{k+1} &= y_k - \gamma S^{-1}Bx_k + \gamma S^{-1}Bx_{k-1}, \\
u_{k+1} &= (\gamma M_k - S)y_k - (\gamma M_k - S)x_k.
\end{aligned}$$

Let $\varepsilon_5, \varepsilon_6, \alpha > 0$ and assume that $D := 1-(1+\varepsilon_6)\gamma^2\mu^2 > 0$. Set $\theta := (\alpha+\gamma\mu(\varepsilon_5\mu+1))/D$. For the iterates $\{x_k\}$, $\{y_k\}$, and $\{u_k\}$ generated by Algorithm 3.3, every $x \in \operatorname{zer}(A+B+C)$, and $k \geq 2$, define

$$\begin{aligned} S_k(x) :=& \|(x_k+\gamma S^{-1}Bx_{k-1})-(x+\gamma S^{-1}Bx)\|_S^2 + 2\langle u_k, x_k - x\rangle \\ &+ L_{k-1}\|y_{k-1}-x_{k-1}\|_S^2 + \theta\|x_k - x_{k-1}\|_S^2. \end{aligned} \tag{3.39}$$

This function will be used to establish the weak convergence of Algorithm 3.3. For convenience, define

$$\nu_k := 1 - L_{k-1} - L_k - \gamma L_k^2\mu - \frac{\gamma\beta}{2} - \frac{\gamma}{\varepsilon_5} - \theta\big(1+1/\varepsilon_6\big), \qquad k \geq 2.$$

Lemma 3.3. *Suppose that Assumption 3.1 holds. Let $\varepsilon_5, \varepsilon_6, \alpha, D, \theta$, and ν_k be as defined above. Then, for every $x^* \in \operatorname{zer}(A+B+C)$ and $k \geq 2$,*

$$S_{k+1}(x^*) \leq S_k(x^*) - \alpha\|x_k - x_{k-1}\|_S^2 - \nu_k\|y_k - x_k\|_S^2. \tag{3.40}$$

Proof. From Algorithm 3.3, we have $y_k = (M_k+A)^{-1}(M_kx_k - (B+C)x_k + \gamma^{-1}u_k)$, which is equivalent to the inclusion

$$M_kx_k - M_ky_k - (B+C)x_k + \gamma^{-1}u_k \in Ay_k. \tag{3.41}$$

Since $x^* \in \operatorname{zer}(A+B+C)$, it follows that $-(B+C)x^* \in Ax^*$. Using the monotonicity of A, we obtain

$$0 \leq \langle M_kx_k - M_ky_k - (B+C)x_k + \gamma^{-1}u_k + (B+C)x^*, y_k - x^*\rangle. \tag{3.42}$$

Multiplying both sides by 2γ and using the definition of u_{k+1} yield

$$\begin{aligned} 0 &\leq 2\langle Sx_k - Sy_k + u_k - u_{k+1} - \gamma(B+C)x_k + \gamma(B+C)x^*, y_k - x^*\rangle \\ &= 2\langle Sx_k - Sy_k, y_k - x^*\rangle + 2\langle u_k - u_{k+1}, y_k - x^*\rangle - 2\gamma\langle Bx_k - Bx^*, y_k - x^*\rangle \\ &\quad - 2\gamma\langle Cx_k - Cx^*, y_k - x^*\rangle. \end{aligned} \tag{3.43}$$

We now analyze each term on the right-hand side of (3.43), separately. Using the definition of x_{k+1}, and the identity $2\langle x-y|y-z\rangle_S = \|x-z\|_S^2 - \|x-y\|_S^2 - \|y-z\|_S^2$, we obtain that

$$\begin{aligned} 2\langle Sx_k - Sy_k, y_k - x^*\rangle =& \|x_k - x^*\|_S^2 - \|x_{k+1} + \gamma S^{-1}Bx_k - \gamma S^{-1}Bx_{k-1} - x^*\|_S^2 - \|y_k - x_k\|_S^2 \\ =& \|x_k - x^*\|_S^2 - \|x_{k+1} - x^*\|_S^2 - 2\gamma\langle x_{k+1} - x^*, Bx_k - Bx_{k-1}\rangle \\ &- \gamma^2\|Bx_k - Bx_{k-1}\|_{S^{-1}}^2 - \|y_k - x_k\|_S^2. \end{aligned} \tag{3.44}$$

By Assumptions 3.1(ii) and (v), one has

$$\begin{aligned} 2\langle u_k - u_{k+1}, y_k - x^*\rangle \leq& 2\langle u_k, x_k - x^*\rangle + 2\|u_k\|_{S^{-1}}\|y_k - x_k\|_S - 2\langle u_{k+1}, x_{k+1} - x^*\rangle \\ &+ 2\gamma\|u_{k+1}\|_{S^{-1}}\|Bx_k - Bx_{k-1}\|_{S^{-1}} \\ \leq& 2\langle u_k, x_k - x^*\rangle - 2\langle u_{k+1}, x_{k+1} - x^*\rangle + L_{k-1}\|y_{k-1} - x_{k-1}\|_S^2 \\ &+ L_{k-1}\|y_k - x_k\|_S^2 + \gamma L_k^2\mu\|y_k - x_k\|_S^2 + \gamma\mu\|x_k - x_{k-1}\|_S^2. \end{aligned} \tag{3.45}$$

Since C is β^{-1}-cocoercive w.r.t. S, we obtain from Lemma 2.1 that

$$-2\gamma\langle Cx_k - Cx^*, y_k - x^*\rangle \leq \frac{\gamma\beta}{2}\|y_k - x_k\|_S^2. \tag{3.46}$$

Substituting (3.44)–(3.46) into (3.43), we obtain that

$$\begin{aligned}
&\|x_{k+1}-x^*\|_S^2+2\langle u_{k+1},x_{k+1}-x^*\rangle\\
\le&\|x_k-x^*\|_S^2+2\langle u_k,x_k-x^*\rangle+L_{k-1}\|y_{k-1}-x_{k-1}\|_S^2\\
&-2\gamma\langle Bx_k-Bx^*,y_k-x^*\rangle-2\gamma\langle x_{k+1}-x^*,Bx_k-Bx_{k-1}\rangle-\gamma^2\|Bx_k-Bx_{k-1}\|_{S^{-1}}^2\\
&+\gamma\mu\|x_k-x_{k-1}\|_S^2-\Big(1-L_{k-1}-\gamma L_k^2\mu-\frac{\gamma\beta}{2}\Big)\|y_k-x_k\|_S^2. \qquad (3.47)
\end{aligned}$$

Finally, applying the monotonicity of B and Young's inequality, we derive

$$\begin{aligned}
&-2\gamma\langle Bx_k-Bx^*,y_k-x^*\rangle-2\gamma\langle x_{k+1}-x^*,Bx_k-Bx_{k-1}\rangle-\gamma^2\|Bx_k-Bx_{k-1}\|_{S^{-1}}^2\\
=&-2\gamma\langle Bx_k-Bx^*,x_{k+1}-x^*\rangle+\gamma^2(\|Bx_{k-1}-Bx^*\|_{S^{-1}}^2-\|Bx_k-Bx^*\|_{S^{-1}}^2\\
&-\|Bx_k-Bx_{k-1}\|_{S^{-1}}^2)+2\gamma\langle Bx_k-Bx_{k-1},x_k-x_{k+1}\rangle+2\gamma\langle Bx_{k-1}-Bx^*,x_k-x^*\rangle\\
&+2\gamma\langle Bx^*-Bx_k,x_k-x^*\rangle-\gamma^2\|Bx_k-Bx_{k-1}\|_{S^{-1}}^2\\
\le&-2\gamma\langle Bx_k-Bx^*,x_{k+1}-x^*\rangle+2\gamma\langle Bx_{k-1}-Bx^*,x_k-x^*\rangle+\gamma^2\|Bx_{k-1}-Bx^*\|_{S^{-1}}^2\\
&-\gamma^2\|Bx_k-Bx^*\|_{S^{-1}}^2+\gamma\varepsilon_5\mu^2\|x_k-x_{k-1}\|_S^2+\frac{\gamma}{\varepsilon_5}\|y_k-x_k\|_S^2, \qquad (3.48)
\end{aligned}$$

where the first equality is due to the definition of x_{k+1}, and the first inequality is due to the monotonicity of B. Substituting (3.48) into (3.47), we obtain

$$\begin{aligned}
&\|x_{k+1}-x^*\|_S^2+2\gamma\langle Bx_k-Bx^*,x_{k+1}-x^*\rangle+\gamma^2\|Bx_k-Bx^*\|_{S^{-1}}^2+2\langle u_{k+1},x_{k+1}-x^*\rangle\\
\le&\|x_k-x^*\|_S^2+2\gamma\langle Bx_{k-1}-Bx^*,x_k-x^*\rangle+\gamma^2\|Bx_{k-1}-Bx^*\|_{S^{-1}}^2+2\langle u_k,x_k-x^*\rangle\\
&+L_{k-1}\|y_{k-1}-x_{k-1}\|_S^2+\gamma\mu(\varepsilon_5\mu+1)\|x_k-x_{k-1}\|_S^2\\
&-\Big(1-L_{k-1}-\gamma L_k^2\mu-\frac{\gamma\beta}{2}-\frac{\gamma}{\varepsilon_5}\Big)\|y_k-x_k\|_S^2.
\end{aligned}$$

Note that $\|a-b\|_S^2+2\langle a-b,c-d\rangle+\|c-d\|_{S^{-1}}^2=\|(a-b)+S^{-1}(c-d)\|_S^2$. Rearranging the terms yields

$$\begin{aligned}
&\|(x_{k+1}+\gamma S^{-1}Bx_k)-(\gamma S^{-1}Bx^*+x^*)\|_S^2+2\langle u_{k+1},x_{k+1}-x^*\rangle\\
&\quad+\frac{\alpha+\gamma\mu(\varepsilon_5\mu+1)}{1-(1+\varepsilon_6)\gamma^2\mu^2}\|x_{k+1}-x_k\|_S^2\\
\le&\|(x_k+\gamma S^{-1}Bx_{k-1})-(\gamma S^{-1}Bx^*+x^*)\|_S^2+2\langle u_k,x_k-x^*\rangle+L_{k-1}\|y_{k-1}-x_{k-1}\|_S^2\\
&\quad+\gamma\mu(\varepsilon_5\mu+1)\|x_k-x_{k-1}\|_S^2-\Big(1-L_{k-1}-\gamma L_k^2\mu-\frac{\gamma\beta}{2}-\frac{\gamma}{\varepsilon_5}\Big)\|y_k-x_k\|_S^2\\
&\quad+\frac{\alpha+\gamma\mu(\varepsilon_5\mu+1)}{1-(1+\varepsilon_6)\gamma^2\mu^2}\|x_{k+1}-x_k\|_S^2\\
\le&\|(x_k+\gamma S^{-1}Bx_{k-1})-(\gamma S^{-1}Bx^*+x^*)\|_S^2+2\langle u_k,x_k-x^*\rangle+L_{k-1}\|y_{k-1}-x_{k-1}\|_S^2\\
&\quad+\gamma\mu(\varepsilon_5\mu+1)\|x_k-x_{k-1}\|_S^2-\Big(1-L_{k-1}-\gamma L_k^2\mu-\frac{\gamma\beta}{2}-\frac{\gamma}{\varepsilon_5}\Big)\|y_k-x_k\|_S^2\\
&\quad+\frac{\alpha+\gamma\mu(\varepsilon_5\mu+1)}{1-(1+\varepsilon_6)\gamma^2\mu^2}((1+\varepsilon_6)\gamma^2\|Bx_k-Bx_{k-1}\|_{S^{-1}}^2+\big(1+\frac{1}{\varepsilon_6}\big)\|y_k-x_k\|_S^2)\\
\le&\|(x_k+\gamma S^{-1}Bx_{k-1})-(\gamma S^{-1}Bx^*+x^*)\|_S^2+2\langle u_k,x_k-x^*\rangle+L_{k-1}\|y_{k-1}-x_{k-1}\|_S^2\\
&\quad+\frac{\alpha(1+\varepsilon_6)\gamma^2\mu^2+\gamma\mu(\varepsilon_5\mu+1)}{1-(1+\varepsilon_6)\gamma^2\mu^2}\|x_k-x_{k-1}\|_S^2\\
&\quad-\Big(1-L_{k-1}-\gamma L_k^2\mu-\frac{\gamma\beta}{2}-\frac{\gamma}{\varepsilon_5}-\frac{(\alpha+\gamma\mu(\varepsilon_5\mu+1))(1+1/\varepsilon_6)}{1-(1+\varepsilon_6)\gamma^2\mu^2}\Big)\|y_k-x_k\|_S^2\\
=&\|(x_k+\gamma S^{-1}Bx_{k-1})-(\gamma S^{-1}Bx^*+x^*)\|_S^2+2\langle u_k,x_k-x^*\rangle+L_{k-1}\|y_{k-1}-x_{k-1}\|_S^2\\
&\quad+\frac{\alpha+\gamma\mu(\varepsilon_5\mu+1)}{1-(1+\varepsilon_6)\gamma^2\mu^2}\|x_k-x_{k-1}\|_S^2-\alpha\|x_k-x_{k-1}\|_S^2\\
&\quad-\Big(1-L_{k-1}-\gamma L_k^2\mu-\frac{\gamma\beta}{2}-\frac{\gamma}{\varepsilon_5}-\frac{(\alpha+\gamma\mu(\varepsilon_5\mu+1))(1+1/\varepsilon_6)}{1-(1+\varepsilon_6)\gamma^2\mu^2}\Big)\|y_k-x_k\|_S^2. \qquad (3.49)
\end{aligned}$$

Adding $L_k\|y_k-x_k\|_S^2$ to both sides of (3.49) and using the definitions of S_k and ν_k, we obtain (3.40). □

Theorem 3.5. *Suppose that Assumption 3.1 holds. Let $\varepsilon_5,\varepsilon_6,\alpha>0$ and assume that $D=1-(1+\varepsilon_6)\gamma^2\mu^2>0$. Assume that there exists a constant $\varepsilon>0$ such that $\nu_k\ge\varepsilon$ for all $k\ge 2$. Then the sequence $\{x_k\}$ generated by Algorithm 3.3 converges weakly to a point $x^*\in\operatorname{zer}(A+B+C)$.*

Proof. Let $x^*\in\operatorname{zer}(A+B+C)$. Using the definition of $S_k(x^*)$ in (3.39), Assumption 3.1(v), and Young's inequality, we obtain that, for all $k\ge 2$,

$$\begin{aligned}
S_k(x^*)\ge&(1-L_{k-1})\|x_k-x^*\|_S^2+2\gamma\langle Bx_{k-1}-Bx^*,x_k-x^*\rangle\\
&+\gamma^2\|Bx_{k-1}-Bx^*\|_{S^{-1}}^2+\theta\|x_k-x_{k-1}\|_S^2\\
=&(1-L_{k-1})\|x_k-x^*\|_S^2+2\gamma\langle Bx_k-Bx^*,x_k-x^*\rangle\\
&+2\gamma\langle Bx_{k-1}-Bx_k,x_k-x^*\rangle+\gamma^2\|Bx_{k-1}-Bx^*\|_{S^{-1}}^2+\theta\|x_k-x_{k-1}\|_S^2\\
\ge&(1-L_{k-1}-\gamma\mu)\|x_k-x^*\|_S^2+(\theta-\gamma\mu)\|x_k-x_{k-1}\|_S^2\ge\varepsilon\|x_k-x^*\|_S^2,
\end{aligned}$$

where the second inequality follows from the monotonicity and Lipschitz continuity of B. The last inequality follows from $\theta > \gamma\mu$ and $1 - L_{k-1} - \gamma\mu \geq \varepsilon$ Hence, the sequence $\{S_k(x^*)\}$ is nonnegative for all $k \geq 2$. Since it is also nonincreasing by (3.40), it follows that it converges. Moreover,

$$\varepsilon\|x_k - x^*\|_S^2 \leq S_k(x^*) \leq S_2(x^*), \qquad k \geq 2,$$

which implies that $\{x_k\}$ is bounded. Let $N \geq 2$ be an integer. Summing both sides of (3.40) over $k = 2, \ldots, N$ yields

$$\sum_{k=2}^{N} \left(\alpha\|x_k - x_{k-1}\|_S^2 + \varepsilon\|y_k - x_k\|_S^2\right) \leq S_2(x^*) - S_{N+1}(x^*) \leq S_2(x^*) < +\infty.$$

Therefore,

$$\lim_{k\to\infty} \|x_k - x_{k-1}\|_S = 0, \qquad \lim_{k\to\infty} \|y_k - x_k\|_S = 0.$$

Using Assumption 3.1(v), we also obtain

$$\|u_k\|_{S^{-1}} \leq L_{k-1}\|y_{k-1} - x_{k-1}\|_S \to 0.$$

Let $F := \mathrm{Id} + \gamma S^{-1}B$, $a_k := x_k + \gamma S^{-1}Bx_{k-1}$, and $a := Fx^*$. From the definition of $S_k(x^*)$, we have

$$S_k(x^*) - \|a_k - a\|_S^2 = 2\langle u_k, x_k - x^*\rangle + L_{k-1}\|y_{k-1} - x_{k-1}\|_S^2 + \theta\|x_k - x_{k-1}\|_S^2 \to 0.$$

Thus, $\lim_{k\to\infty}\|a_k - Fx^*\|_S^2$ exists for every $x^* \in \mathrm{zer}(A+B+C)$. In particular, the sequence $\{a_k\}$ is bounded. Define $\Delta_k := M_k x_k - M_k y_k + (B+C)y_k - (B+C)x_k + \gamma^{-1}u_k$. It follows from (3.41) that $(y_k, \Delta_k) \in \mathrm{gra}(A+B+C)$ for all $k \geq 2$. By Assumption 3.1(ii), (iii), and (v), we obtain that

$$\begin{aligned}\|\Delta_k\|_{S^{-1}} &\leq \|M_k x_k - M_k y_k\|_{S^{-1}} + \|(B+C)y_k - (B+C)x_k\|_{S^{-1}} + \gamma^{-1}\|u_k\|_{S^{-1}} \\ &\leq \left(\frac{2}{\gamma} + \mu + \beta\right)\|x_k - y_k\|_S + \frac{L_{k-1}}{\gamma}\|y_{k-1} - x_{k-1}\|_S.\end{aligned}$$

Combining this with the limits above, we obtain $\Delta_k \to 0$.

To identify the weak cluster points, define $\mathscr{T}(x,v) := \left((A+C)x + v,\ B^{-1}v - x\right)$ on $\mathscr{H} \times \mathscr{H}$. The operators $A+C$ and B^{-1} are maximally monotone. Since $(x,v) \mapsto (v,-x)$ is a bounded skew linear operator with full domain, $\mathscr{T}$ is maximally monotone. Moreover, $(\Delta_k, 0) \in \mathscr{T}(y_k, By_k)$. Let $\bar{x}$ be any weak cluster point of $\{x_k\}$, and let $x_{k_n} \rightharpoonup \bar{x}$. Since $\{By_k\}$ is bounded, passing to a further subsequence if necessary, we may assume that $By_{k_n} \rightharpoonup b$. It follows from $y_k - x_k \to 0$ that $y_{k_n} \rightharpoonup \bar{x}$. By the weak–strong closedness of the graph of $\mathscr{T}$, we have

$$(0,0) \in \mathscr{T}(\bar{x}, b),$$

which implies that $b = B\bar{x}$ and $\bar{x} \in \mathrm{zer}(A+B+C)$.

On the other hand, the Lipschitz continuity of B yields

$$\|a_k - (x_k + \gamma S^{-1}By_k)\|_S \leq \gamma\mu\|x_{k-1} - y_k\|_S \to 0.$$

Let $\bar{a}$ be any weak cluster point of $\{a_k\}$, and let $a_{k_n} \rightharpoonup \bar{a}$. By the boundedness of $\{x_k\}$ and $\{By_k\}$, we may pass to a further subsequence such that $x_{k_n} \rightharpoonup \bar{x}$ and $By_{k_n} \rightharpoonup b$. The preceding argument gives $b = B\bar{x}$ and $\bar{x} \in \mathrm{zer}(A+B+C)$. Therefore,

$$\bar{a} = \bar{x} + \gamma S^{-1}B\bar{x} = F\bar{x}.$$

Hence, every weak cluster point of $\{a_k\}$ can be written as $x^* + \gamma S^{-1}Bx^*$ for some $x^* \in \operatorname{zer}(A+B+C)$. Since $\lim_{k\to\infty} \|a_k - (x^* + \gamma S^{-1}Bx^*)\|_S$ exists for every $x^* \in \operatorname{zer}(A+B+C)$, Opial's lemma shows that $\{a_k\}$ converges weakly to $a^* = \widehat{x} + \gamma S^{-1}B\widehat{x}$ for some $\widehat{x} \in \operatorname{zer}(A+B+C)$.

Finally, for any weak cluster point $\bar{x}$ of $\{x_k\}$, the same subsequence argument gives $F\bar{x} = a^*$. By the monotonicity of B, we have

$$\langle Fx - Fy, x-y\rangle_S = \|x-y\|_S^2 + \gamma\langle Bx - By, x-y\rangle \geq \|x-y\|_S^2,$$

so F is injective. Consequently, the bounded sequence $\{x_k\}$ has a unique weak cluster point and converges weakly to a point in $\operatorname{zer}(A+B+C)$. □

Remark 3.3. We now present two special cases of Algorithm 3.3.

(i) When $S = \mathrm{Id}$, and $M_k \equiv \mathrm{Id}/\gamma$, we have $L_k = 0$ and $u_k = 0$. Hence, Algorithm 3.3 reduces to the ORFBS algorithm (1.6). The conditions in Theorem 3.5 become $1-(1+\varepsilon_6)\gamma^2\mu^2 > 0$ and

$$1 - \frac{\gamma\beta}{2} - \frac{\gamma}{\varepsilon_5} - \frac{(\alpha + \gamma\mu(\varepsilon_5\mu+1))(1+1/\varepsilon_6)}{1-(1+\varepsilon_6)\gamma^2\mu^2} \geq \varepsilon.$$

(ii) Suppose that $A = A_1 + A_2$, where $A_1 : \mathscr{H} \to 2^{\mathscr{H}}$ is maximally monotone, $A_2 : \mathscr{H} \to \mathscr{H}$ is L-Lipschitz continuous, and $A_1 + A_2$ is maximally monotone. Let $S = \mathrm{Id}$ and $M_k \equiv M = \mathrm{Id}/\gamma - A_2$ for all $k \in \mathbb{N}$. Then, with the convention $y_0 = x_0$, Algorithm 3.3 becomes

$$\begin{cases} y_k = J_{\gamma A_1}(x_k - \gamma(A_2 + B + C)x_k - \gamma(A_2y_{k-1} - A_2x_{k-1})), \\ x_{k+1} = y_k - \gamma Bx_k + \gamma Bx_{k-1}. \end{cases} \tag{3.50}$$

Now, $\gamma M - \mathrm{Id}$ is γL-Lipschitz continuous. The conditions in Theorem 3.5 become $1-(1+\varepsilon_6)\gamma^2\mu^2 > 0$ and

$$1 - 2\gamma L - \gamma^3L^2\mu - \frac{\gamma\beta}{2} - \frac{\gamma}{\varepsilon_5} - \frac{(\alpha + \gamma\mu(\varepsilon_5\mu+1))(1+1/\varepsilon_6)}{1-(1+\varepsilon_6)\gamma^2\mu^2} \geq \varepsilon.$$

3.3.1. *Linear convergence.* In this subsection, we establish the R–linear convergence of the sequence $\{x_k\}$ generated by Algorithm 3.3, assuming that A is strongly monotone.

Theorem 3.6. *Suppose that the conditions of Theorem 3.5 hold and that A is ρ-strongly monotone with respect to S, where $\rho > 0$. Then the sequence $\{x_k\}$ generated by Algorithm 3.3 converges R–linearly to the unique point $x^* \in \operatorname{zer}(A+B+C)$.*

Proof. Let $x^* \in \operatorname{zer}(A+B+C)$. Using the strong monotonicity of A, inequality (3.42) can be strengthened to

$$\rho\|y_k - x^*\|_S^2 \leq \langle M_kx_k - M_ky_k - (B+C)x_k + \gamma^{-1}u_k + (B+C)x^*, y_k - x^*\rangle.$$

Following the same arguments as in Lemma 3.3, we obtain that, for all $k \geq 2$,

$$S_{k+1}(x^*) + 2\gamma\rho\|y_k - x^*\|_S^2 + \alpha\|x_k - x_{k-1}\|_S^2 + \nu_k\|y_k - x_k\|_S^2 \leq S_k(x^*). \tag{3.51}$$

Set $r_k := \|y_k - x^*\|_S$, $p_k := \|y_k - x_k\|_S$ and $h_k := \|x_k - x_{k-1}\|_S$. By the definition of x_{k+1} and the Lipschitz continuity of B, we have

$$\|x_{k+1} - x^*\|_S \leq \|y_k - x^*\|_S + \gamma\|Bx_k - Bx_{k-1}\|_{S^{-1}} \leq r_k + \gamma\mu h_k,$$

and

$$\|x_{k+1} - x_k\|_S \leq \|y_k - x_k\|_S + \gamma\|Bx_k - Bx_{k-1}\|_{S^{-1}} \leq p_k + \gamma\mu h_k.$$

Moreover,

$$\gamma\|Bx_k - Bx^*\|_{S^{-1}} \leq \gamma\mu\|x_k - x^*\|_S \leq \gamma\mu(p_k + r_k).$$

Using Assumption 3.1(v), we also obtain

$$\|u_{k+1}\|_{S^{-1}} \leq L_k\|y_k - x_k\|_S \leq p_k.$$

It follows from the definition of $S_{k+1}(x^*)$ that

$$\begin{aligned}
S_{k+1}(x^*) =& \|(x_{k+1} - x^*) + \gamma S^{-1}(Bx_k - Bx^*)\|_S^2 + 2\langle u_{k+1}, x_{k+1} - x^*\rangle \\
&+ L_k\|y_k - x_k\|_S^2 + \theta\|x_{k+1} - x_k\|_S^2 \\
\leq& \big((1+\gamma\mu)r_k + \gamma\mu p_k + \gamma\mu h_k\big)^2 + 2p_k(r_k + \gamma\mu h_k) + p_k^2 + \theta(p_k + \gamma\mu h_k)^2 \\
\leq& \big(3(1+\gamma\mu)^2 + 1\big)r_k^2 + \big(3\gamma^2\mu^2 + 2 + \gamma\mu + 2\theta\big)p_k^2 \\
&+ \big(3\gamma^2\mu^2 + \gamma\mu + 2\theta\gamma^2\mu^2\big)h_k^2 \leq \Lambda(r_k^2 + p_k^2 + h_k^2),
\end{aligned}$$

where the second inequality follows from Young's inequality, and

$$\Lambda := \max\left\{3(1+\gamma\mu)^2 + 1,\ 3\gamma^2\mu^2 + 2 + \gamma\mu + 2\theta,\ 3\gamma^2\mu^2 + \gamma\mu + 2\theta\gamma^2\mu^2\right\} > 0.$$

Set

$$t := \frac{\min\{2\gamma\rho, \varepsilon, \alpha\}}{\Lambda} > 0.$$

Since $\nu_k \geq \varepsilon$, the preceding estimate gives

$$tS_{k+1}(x^*) \leq 2\gamma\rho\|y_k - x^*\|_S^2 + \nu_k\|y_k - x_k\|_S^2 + \alpha\|x_k - x_{k-1}\|_S^2.$$

Combining this with (3.51), we obtain $(1+t)S_{k+1}(x^*) \leq S_k(x^*)$ for all $k \geq 2$. On the other hand, the lower bound established in the proof of Theorem 3.5 gives $S_k(x^*) \geq \varepsilon\|x_k - x^*\|_S^2$ for all $k \geq 2$. Therefore, for all $k \geq 2$,

$$\varepsilon\|x_{k+1} - x^*\|_S^2 \leq S_{k+1}(x^*) \leq \frac{S_k(x^*)}{1+t} \leq \cdots \leq \frac{S_2(x^*)}{(1+t)^{k-1}}.$$

Hence, the sequence $\{x_k\}$ generated by Algorithm 3.3 converges R–linearly to x^*. Uniqueness follows from the strong monotonicity of A and the monotonicity of B and C. □

4. Numerical Experiments

In this section, we focus on a special case of Algorithm 3.3, namely, Algorithm (3.50). We assess the performance of Algorithm (3.50) with the ORFBS algorithm (1.6). Note that, for the numerical experiments in this section, Algorithm (3.35) (derived in Remark 3.2) is equivalent to the SRFBS algorithm (1.5), and therefore omitted from the comparative analysis. All implementations were written in MATLAB and executed on a computer equipped with an Intel Core i7-11800H CPU (2.30GHz) and 32GB RAM, running Windows 11.

4.1. **Nonlinear constrained optimization problems.** In this subsection, we conduct numerical experiments on constrained nonlinear optimization problems to validate the effectiveness of the proposed algorithm. Specifically, we consider the following constrained optimization problem:

$$\min_{x\in C} f(x)+h(x), \tag{4.1}$$

where $C=\{x\in\mathscr{H}|(\forall i\in\{1,...,q\})\ g_i(x)\le 0\}$. Here, $f:\mathscr{H}\to(-\infty,+\infty]$ is a proper, convex and lower semi-continuous function, for each $i\in\{1,...,q\}$, $g_i:\mathrm{dom}(g_i)\subset\mathscr{H}\to\mathbb{R}$ and $h:\mathscr{H}\to\mathbb{R}$ are C^1 convex functions in $\mathrm{int}\,\mathrm{dom}\,g_i$ and $\mathscr{H}$, respectively, with ∇h being β-Lipschitz continuous. The solution to problem (4.1) corresponds to a saddle point of the Lagrangian function

$$L(x,u)=f(x)+h(x)+u^\top g(x)-\iota_{\mathbb{R}^q_+}(u),$$

where $\iota_{\mathbb{R}^q_+}$ denotes the indicator function of $\mathbb{R}^q_+$. Under standard regularity conditions, this solution can be obtained by solving the following monotone inclusion problem [6, 28]: find $x\in\mathscr{H}$ such that there exists $u\in\mathbb{R}^q_+$ satisfying

$$(0,0)\in A(x,u)+B(x,u)+C(x,u), \tag{4.2}$$

The operators are defined as follows: $A:(x,u)\mapsto\partial f(x)\times N_{\mathbb{R}^q_+}u$ is maximally monotone, $B:(x,u)\mapsto(\sum_{i=1}^q u_i\nabla g_i(x),-g_1(x),...,-g_q(x))$ is nonlinear, monotone and continuous, and $C:(x,u)\mapsto(\nabla h(x),0)$ is $\frac{1}{\beta}$-cocoercive.

Let $\mathscr{H}=\mathbb{R}^N$, $f=\iota_{[0,1]^N}$, $g_i(x)=d_i^\top x$ for all $i\in\{1,...,q\}$ with $d_1,\dots,d_q\in\mathbb{R}^N$, and $h(x)=\frac{1}{2}\|Gx-b\|^2$, where G is an $m\times N$ real matrix, $N=2m$, $b\in\mathbb{R}^m$. The operators in (4.2) then take the form:

$$\begin{aligned}
&A:(x,u)\mapsto\partial\iota_{[0,1]^N}(x)\times N_{\mathbb{R}^q_+}u,\\
&B:(x,u)\mapsto(D^\top u,-Dx),\\
&C:(x,u)\mapsto(G^\top(Gx-b),0),
\end{aligned}$$

where $x\in\mathbb{R}^N$, $u\in\mathbb{R}^q_+$ and $D=[d_1,\dots,d_q]^\top$. Here, A is maximally monotone, B is L-Lipschitz continuous with $L=\|D\|$, and C is $\frac{1}{\beta}$-cocoercive with $\beta=\|G\|^2$. Therefore, problem (4.2) can be solved by ORFBS algorithm (1.6). Alternatively, we can decompose B as B_1+B_2, where

$$B_1:(x,u)\mapsto(\frac{1}{2}D^\top u,-\frac{1}{2}Dx),\qquad B_2:(x,u)\mapsto(\frac{1}{2}D^\top u,-\frac{1}{2}Dx). \tag{4.3}$$

It is straightforward to verify that $A+B_1$ is maximally monotone, while B_1 and B_2 are both $L/2$–Lipschitz continuous, with $L=\|D\|$. Consequently, problem (4.2) can also be solved by Algorithm (3.50).

In the numerical experiments, the matrices G,D, the vector b, and the initial values (x_0,u_0) are all randomly generated. The following stopping criterion is employed:

$$E_k=\frac{\|(x_{k+1}-x_k,u_{k+1}-u_k)\|}{\|(x_k,u_k)\|}<10^{-6},$$

We test eight different problem sizes, with ten randomly generated instances for each size. Table 1 reports the average number of iterations (denoted as "av.iter") and the average CPU time (denoted as "av.time") over these ten instances. As shown in Table 1, Algorithm (3.50)

outperforms the ORFBS algorithm in both iteration count and computational efficiency in most cases.

TABLE 1. Computational results with ORFBS and Algorithm (3.50)

Problem size		**av.iter**		**av.time (s)**	
N	q	ORFBS	Algorithm (3.50)	ORFBS	Algorithm (3.50)
2000	100	7267.8	**6806.1**	25.669	**15.831**
	200	5950.1	**5742.5**	35.587	**17.538**
	500	4441.3	**4244.7**	77.558	**31.192**
	1000	3187.4	**3134**	169.68	**58.53**
4000	100	8724.7	**8525.4**	282.33	**212.37**
	200	6445.4	**6401.8**	278.65	**186.07**
	500	3427.7	**3713.6**	368.03	**158.26**
	1000	**811.7**	1295	157.74	**79.902**

4.2. **Portfolio optimization problem.** This subsection considers the classical mean-variance portfolio optimization problem for risk minimization. Suppose a portfolio consists of n distinct assets, where each asset i has a random rate of return with expected value m_i. The objective is to determine investment weights x_i that minimize portfolio risk while ensuring a specified minimum expected return r. Let $H \in \mathbb{R}^{225\times 225}$ denote the covariance matrix of asset returns ($\|H\| = 0.2263$). The optimization problem can be formulated as:

$$\begin{aligned}&\min_{x\in R^{225}} \frac{1}{2}x^T Hx,\\ &\quad s.t. \sum_{i=1}^{75} x_i \geq 0.3, \sum_{i=76}^{150} x_i \geq 0.3, \sum_{i=151}^{225} x_i \geq 0.3,\\ &\quad \sum_{i=1}^{225} m_i x_i \geq r, \sum_{i=1}^{225} x_i = 1, 0 \leq x_i \leq 1, i = 1,\cdots,225.\end{aligned} \tag{4.4}$$

Define the constraint mapping $g(x) = Dx + b$, where

$$D = \begin{bmatrix} -m_1 & \cdots & -m_{75} & -m_{76} & \cdots & -m_{150} & -m_{151} & \cdots & -m_{225} \\ -1 & \cdots & -1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & -1 & \cdots & -1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & -1 & \cdots & -1 \end{bmatrix}, b = \begin{bmatrix} r \\ 0.3 \\ 0.3 \\ 0.3 \end{bmatrix}$$

and the feasible set $C_2 = \{x \in \mathbb{R}^n | \sum_{i=1}^n x_i = 1,\ 0 \leq x_i \leq 1, i = 1,\cdots,n\}$. Solutions correspond to saddle points of the Lagrangian:

$$L(x,u) = \frac{1}{2}x^T Hx + \iota_{C_2}(x) + u^\top g(x) - \iota_{\mathbb{R}^4_+}(u).$$

Under standard qualification conditions, the solution to problem (4.4) can be obtained by solving the following monotone inclusion problem: find $x \in \mathbb{R}^{225}$ such that $\exists u \in \mathbb{R}^4_+$,

$$(0,0) \in A(x,u) + B(x,u) + C(x,u),$$

where the operators are defined as

$$A : (x,u) \mapsto \partial \iota_{C_2}(x) \times N_{\mathbb{R}^4_+} u,\ B : (x,u) \mapsto (D^\top u, -Dx - b),\ C : (x,u) \mapsto (Hx, 0).$$

Here, A is maximally monotone, B is L-Lipschitz continuous with $L = \|D\|$, and C is β^{-1}-cocoercive with $\beta = \|H\|$. Therefore, the ORFBS algorithm (1.6) can be employed to solve problem (4.4). Alternatively, we decompose $B = B_1 + B_2$, $B_1(x,u) = B_2(x,u) = 1/2(D^\top u, -Dx - b)$. Both B_1 and B_2 are monotone and $L/2$-Lipschitz continuous, and $A + B_1$ is maximally monotone. Consequently, Algorithm (3.50) can also be applied to solve problem (4.4).

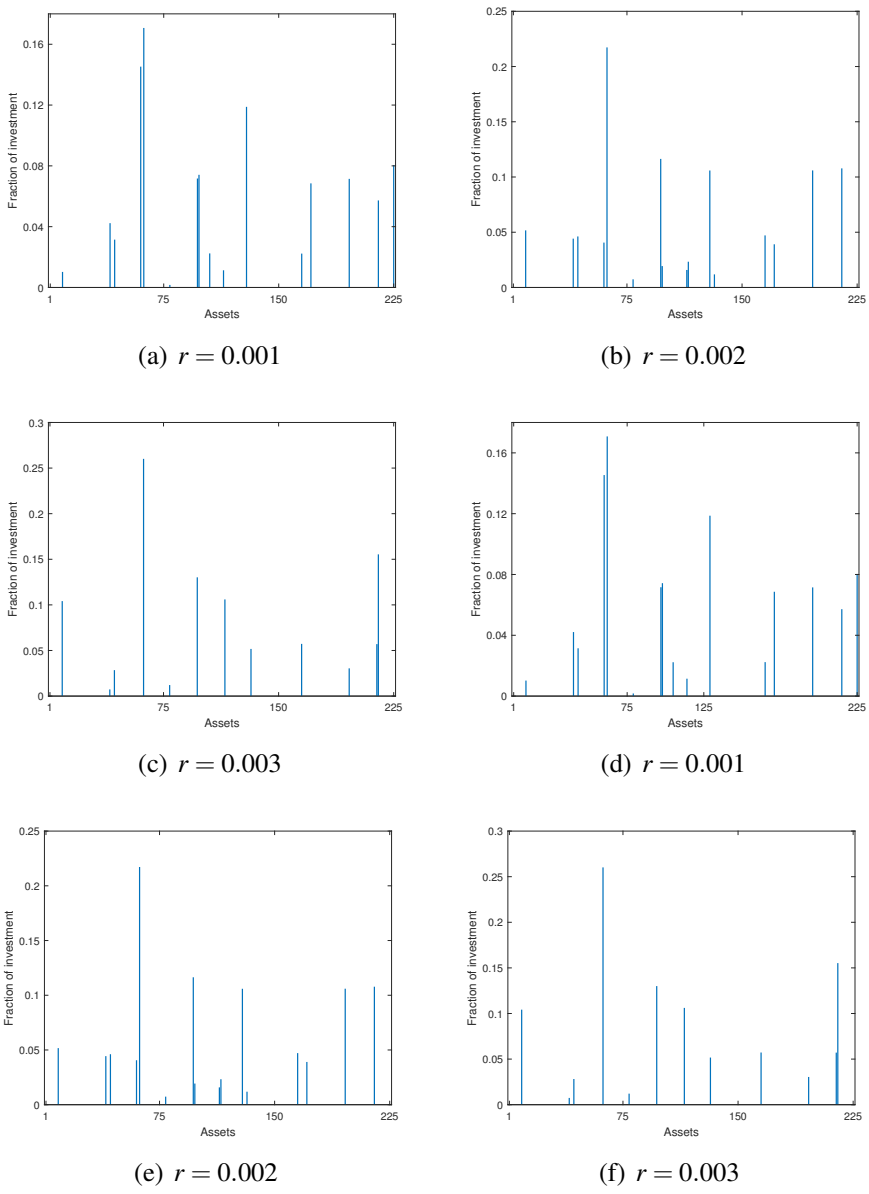
This study employs the portfolio optimization formulation (4.4), which is adopted from the MATLAB optimization case library. The benchmark dataset used in the experiment is sourced from the OR-Library [12], where the asset returns m_i range from -0.008489 to 0.003971. We evaluate the algorithms under three target return levels, $r = 0.001$, 0.002, and 0.003, in our experimental evaluations. As presented in Table 2, the solution accuracy achieved by the ORFBS algorithm is comparable to that of Algorithm (3.50). Moreover, Figure 1 illustrates that the resulting portfolio weight distributions exhibit consistent sparsity patterns across all target levels.

TABLE 2. The objective function values (Obj), number of iterations (Iter), and CPU time (in seconds) for the compared methods on the portfolio optimization problem.

r	Methods	Obj	Iter	CPU
0.001	ORFBS	$1.6386e-4$	**248954**	**6.1836**
	Algorithm (3.50)	$1.6386e-4$	249453	6.5429
0.002	ORFBS	$2.0097e-4$	**271736**	**6.6391**
	Algorithm (3.50)	$2.0097e-4$	257843	6.8260
0.003	ORFBS	$2.7686e-4$	217194	**5.7700**
	Algorithm (3.50)	$2.7686e-4$	**212883**	6.5494

5. CONCLUSIONS

This study investigated structured monotone inclusion problems that consist of the sum of three operators: a maximally monotone operator, a maximally monotone and Lipschitz continuous operator, and a cocoercive operator. To address these problems, we developed three novel splitting algorithms by enhancing the semi-forward-reflected-backward, semi-reflected-forward-backward, and outer reflected forward-backward splitting frameworks with a nonlinear momentum term. Under suitable step-size conditions, we rigorously established the weak convergence of the proposed algorithms. Furthermore, when the strong monotonicity assumption holds, we proved their linear convergence rates. The numerical results demonstrate the effectiveness of the proposed approach.



(a) $r = 0.001$ (b) $r = 0.002$

(c) $r = 0.003$ (d) $r = 0.001$

(e) $r = 0.002$ (f) $r = 0.003$

FIGURE 1. Portfolio optimization solutions for problem (4.4) obtained by the ORFBS algorithm (1.6) (panels (a)–(c)) and Algorithm (3.50) (panels (d)–(f)).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (12571491, 12571558) and the Guangdong Basic and Applied Basic Research Foundation (2026A1515010520).

REFERENCES

[1] H.H. Bauschke, PL. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, CMS Books in Mathematics, Springer, 2011.

[2] L.M. Briceño-Arias, P.L. Combettes, J.-C. Pesquet, N. Pustelnik, Proximal algorithms for multicomponent image recovery problems, Journal of Mathematical Imaging and Vision 41 (2011), 3–22.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundations and Trends in Machine learning, 3 (2011), 1–122.
[4] S. Banert, J. Rudzusika, O. Öktem, J. Adler, Accelerated forward-backward optimization using deep learning, SIAM Journal on Optimization 34 (2024), 1236–1263.
[5] L. Briceño-Arias, S.L. Rivera, A projected primal–dual method for solving constrained monotone inclusions, Journal of Optimization Theory and Applications, 180 (2019) 907–924.
[6] L.M. Briceño-Arias, D. Davis, Forward-backward-half forward algorithm for solving monotone inclusions, SIAM Journal on Optimization 28 (2018), 2839–2871.
[7] M.N. Bùi, P.L. Combettes, Warped proximal iterations for monotone inclusions, Journal of Mathematical Analysis and Applications 491 (2020), 124315.
[8] K. Bredies, H. Sun, Preconditioned douglas–rachford splitting methods for convex-concave saddle-point problems, SIAM Journal on Numerical Analysis 53 (2015), 421–444.
[9] K. Bredies, H. Sun, A proximal point analysis of the preconditioned alternating direction method of multipliers, Journal of Optimization Theory and Applications 173 (2017), 878–907.
[10] K. Bredies, E. Chenchene, D.A Lorenz, E. Naldi, Degenerate preconditioned proximal point algorithms, SIAM Journal on Optimization 32 (2022), 2376–2401.
[11] R.I. Boţ, D.K. Nguyen, Fast krasnoselkiĭ–Mann algorithm with a convergence rate of the fixed point iteration of $o(\frac{1}{k})$, SIAM Journal on Numerical Analysis 61 (2023), 2813–2843.
[12] T.J. Chang, N. Meade, J.E. Beasley, Y.M. Sharaiha, Heuristics for cardinality constrained portfolio optimisation, Computers & Operations Research 27 (2000), 1271–1302.
[13] P.L. Combettes, V.R. Wajs, Signal recovery by proximal forward-backward splitting, Multiscale Modeling & Simulation 4 (2005), 1168–1200.
[14] P.L. Combettes, J.C. Pesquet, Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators, Set-Valued and Variational Analysis 20 (2012), 307–330.
[15] L. Condat, D. Kitahara, A. Contreras, A. Hirabayashi, Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists, SIAM Review 65 (2023), 375–435.
[16] V. Cevher, B.C. Vu, A reflected forward-backward splitting method for monotone inclusions involving lipschitzian operators, arXiv preprint arXiv:1908.05912, 2019.
[17] J. Douglas, H.H. Rachford, On the numerical solution of heat conduction problems in two and three space variables, Transactions of the American mathematical Society, 82 (1956) 421–439.
[18] P. Giselsson, Nonlinear forward-backward splitting with projection correction, SIAM Journal on Optimization 31 (2021), 2199–2226.
[19] B. Halpern, Fixed points of nonexpanding maps, Bulletin of the American Mathematical Society 73 (1967), 957–961.
[20] P.L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, SIAM Journal on Numerical Analysis, 16 (1979), 964–979.
[21] Y. Malitsky, M.K. Tam, A forward-backward splitting method for monotone inclusions without cocoercivity, SIAM Journal on Optimization 30 (2020) 1451–1472.
[22] M. Morin, S. Banert, P. Giselsson, Nonlinear forward-backward splitting with momentum correction, Set-Valued and Variational Analysis 31 (2023), 37.
[23] G.B. Passty, Ergodic convergence to a zero of the sum of monotone operators in hilbert space, Journal of Mathematical Analysis and Applications, 72 (1979), 383–390.
[24] L. Qin, Q.L. Dong, Y. Zhang, D.R. Sahu, Forward-backward-half forward splitting algorithm with deviations, Optimization 74 (2023) 2137–2158.
[25] L. Qin, Y. Tang, J. Peng, Nonlinear forward-backward-half forward splitting with momentum for monotone inclusions, arXiv preprint arXiv:2510.24489, 2025.
[26] H. Raguet, L. Landrieu, Preconditioning of a generalized forward-backward splitting and application to optimization on graphs, SIAM Journal on Imaging Sciences 8 (2015), 2706–2739.
[27] F. Roldán, C. Vega, Relaxed and inertial nonlinear forward–backward with momentum, Journal of Optimization Theory and Applications 206 (2025), 23.

[28] R.T. Rockafellar, Monotone operators associated with saddle-functions and minimax problems, In: F.E. Browder (ed) Proceedings of SYmposia in Pure Math. vol I, American Math. Soc. , 1970.
[29] D. Sun, Y. Yuan, G. Zhang, X. Zhao, Accelerating preconditioned admm via degenerate proximal point mappings, SIAM Journal on Optimization 35 (2025), 1165–1193.
[30] R. Showalter, Monotone Operators in Banach Space and Nonlinear Partial Differential Equations, 1997. doi: 10.1090/surv/049
[31] P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, SIAM Journal on Control and Optimization 38 (2000), 431–446.
[32] B. Tan, S.Y. Cho, Strong convergence of inertial forward–backward methods for solving monotone inclusions, Applicable Analysis 101 (2022), 5386–5414.
[33] B. Tan, X. Qin, Two relaxed inertial forward-backward-forward algorithms for solving monotone inclusions and an application to compressed sensing, Canadian Journal of Mathematics 78 (2026), 572–593.
[34] B. Tan, Perturbed reflected forward backward splitting algorithm for monotone inclusion, Communications in Nonlinear Science and Numerical Simulation 142 (2025), 108565.
[35] H. Yu, C. Zong, Y. Tang, An outer reflected forward-backward splitting algorithm for solving monotone inclusions, arXiv preprint arXiv:2009.12493, 2020.
[36] S. Zhang, H. Zhang, H. Wang, Hppp: Halpern-type preconditioned proximal point algorithms and applications to image restoration, SIAM Journal on Imaging Sciences 18 (2025), 1493–1521.
[37] C. Zong, Y. Tang, G. Zhang, An accelerated forward-backward-half forward splitting algorithm for monotone inclusion with applications to image restoration, Optimization 73 (2024), 401–428.
[38] C. Zong, Y. Tang, G. Zhang, An inertial semi-forward-reflected-backward splitting and its application, Acta Mathematica Sinica, English Series 38 (2022), 443–464.
[39] C. Zong, G. Zhang, Y. Tang, Multi-step inertial forward-backward-half forward algorithm for solving monotone inclusion, Linear and Multilinear Algebra 71 (2023), 631–661.