Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 10(5): 10740–10763.
DOI: 10.3934/math.2025488
Received: 19 February 2025
Revised: 18 April 2025
Accepted: 24 April 2025
Published: 12 May 2025

Research article

A new operator splitting method with application to feature selection

Yunda Dong* **and Yiyi Li**

School of Mathematics and Statistics, Zhengzhou University, Zhengzhou, 450001, China

* **Correspondence:** Email: ydong@zzu.edu.cn.

Abstract: In this article, we consider the problem of finding a zero of a system of monotone inclusions in Hilbert spaces. Notably, each of these monotone inclusions comprises three operators, with two of them being linearly composed. To address this challenge, we propose a new splitting method that, at each iteration, essentially necessitates the computation of three individual resolvents, corresponding to each operator within the monotone inclusion. Under the weakest possible conditions, with the help of characteristic operator techniques, we analyze the weak convergence properties of our proposed method, which is facilitated by the introduction of a novel inequality. Numerical results demonstrate the practical usefulness of this method in solving large-scale rare feature selection in deep learning.

Keywords: monotone inclusion; splitting method; characteristic operator; weak convergence; feature selection
Mathematics Subject Classification: 65K05, 49M30, 46N10, 90C31

1. Introduction

For the Hilbert spaces $\mathcal{H}_i,\ i = 1, \dots, n,\ \mathcal{G}_1$ and $\mathcal{G}_2$, consider the following system of three-operator monotone inclusions

$$0 \in \bar{A}_i(x_i) + R_i^* A\left(\textstyle\sum_{i=1}^n R_i x_i - r\right) + Q_i^* B\left(\textstyle\sum_{i=1}^n Q_i x_i - q\right), \quad i = 1, \dots, n, \tag{1.1}$$

where each $\bar{A}_i\colon \mathcal{H}_i \rightrightarrows \mathcal{H}_i,\ A\colon \mathcal{G}_1 \rightrightarrows \mathcal{G}_1,\ B\colon \mathcal{G}_2 \rightrightarrows \mathcal{G}_2$ are all maximally monotone operators, and $R_i\colon \mathcal{H}_i \to \mathcal{G}_1,\ Q_i\colon \mathcal{H}_i \to \mathcal{G}_2$ are nonzero bounded linear operators along with their adjoint operators $R_i^*,\ Q_i^*$ respectively, and $r \in \mathcal{G}_1$ and $q \in \mathcal{G}_2$ are vectors. This problem model finds wide-ranging applications across various fields, including monotone variational inequality problems [1, 2], fused lasso [3], hyperspectral unmixing [4], image restoration [5–7], signal processing [8], and machine learning [9, 10].

A particularly notable case of the problem is given by

$$0 \in \bar{A}(x) + A(x),$$

which reminds us of the Douglas–Rachford splitting method of Lions and Mercier [11]; see [12–14] for related discussions.

In the $n = 1$ case, the problem model simplifies to

$$0 \in \bar{A}(x) + R^*A(Rx - r) + Q^*B(Qx - q), \tag{1.2}$$

which can be solved by some existing splitting methods such as those proposed in [1,9].

In the general case, this model distinguishes itself from [15] by taking into consideration the following problem:

$$0 \in \bar{A}_i(x_i) + A_i(x_i) + Q_i^*B\left(\textstyle\sum_{i=1}^n Q_i x_i - q\right), \quad i = 1, \ldots, n, \tag{1.3}$$

where each $A_i\colon \mathcal{H}_i \rightrightarrows \mathcal{H}_i$ is a maximally monotone operator. A recently proposed method [15, Algorithm 1] can be stated as follows: For $i = 1, \ldots, n + 1$, choose $\alpha_i > 0$. At the k-th iteration, for given iterates x_i^k, $a_i^k \in A_i(x_i^k)$, $i = 1, \ldots, n$, x_{n+1}^k and u^k, we update u^k in some simple way to get the intermediate $\bar{u}^k$, and compute

$$\begin{aligned}
&(\alpha_i I + \bar{A}_i)(\bar{x}_i^k) \ni \alpha_i x_i^k - a_i^k - Q_i^*\bar{u}^k,\\
&(\alpha_{n+1} I + B)(\bar{x}_{n+1}^k) \ni \alpha_{n+1} x_{n+1}^k + \bar{u}^k,
\end{aligned}$$

to get the intermediate iterates $\bar{x}_i^k$, $i = 1, \ldots, n + 1$. Calculate $\gamma_k > 0$ in some way. Finally, we obtain the new iterates

$$\begin{aligned}
&(\alpha_i I + A_i)(x_i^{k+1}) \ni \alpha_i x_i^k + a_i^k - \gamma_k(x_i^k - \bar{x}_i^k), \quad i = 1, \ldots, n\\
&\alpha_{n+1} x_{n+1}^{k+1} = \alpha_{n+1} x_{n+1}^k - \gamma_k(x_{n+1}^k - \bar{x}_{n+1}^k),\\
&u^{k+1} = u^k - \gamma_k(\bar{x}_{n+1}^k - \textstyle\sum_{i=1}^n Q_i \bar{x}_i^k + q),\\
&a_i^{k+1} = \alpha_i(x_i^k - x_i^{k+1}) + a_i^k - \gamma_k(x_i^k - \bar{x}_i^k).
\end{aligned}$$

Inspired by this work, we propose a new method to solve (1.1). Specifically speaking, for $i = 1, \ldots, n + 2$, choose $\alpha_i > 0$, $\beta > 0$, and $\hat{\beta} > 0$. At the k-th iteration, for given iterates x_i^k, $i = 1, \ldots, n+2$, u^k, and v^k, we update u^k and v^k in some simple ways to get the intermediate $\bar{u}^k$ and $\bar{v}^k$, and compute

$$\begin{aligned}
&(\alpha_i I + \bar{A}_i)(\bar{x}_i^k) \ni \alpha_i x_i^k - R_i^*\bar{u}^k - Q_i^*\bar{v}^k,\\
&(\alpha_{n+1} I + A)(\bar{x}_{n+1}^k) \ni \alpha_{n+1} x_{n+1}^k + \bar{u}^k,\\
&(\alpha_{n+2} I + B)(\bar{x}_{n+2}^k) \ni \alpha_{n+2} x_{n+2}^k + \bar{v}^k,
\end{aligned}$$

to get the intermediate iterates $\bar{x}_i^k$, $i = 1, \ldots, n + 2$. Calculate $\gamma_k > 0$ in some way. Finally, we get the new iterates

$$\begin{aligned}
&x_i^{k+1} = x_i^k - \gamma_k \alpha_i(x_i^k - \bar{x}_i^k), \quad i = 1, \ldots, n + 2,\\
&u^{k+1} = u^k - \gamma_k \beta(\bar{x}_{n+1}^k - \textstyle\sum_{i=1}^n R_i \bar{x}_i^k + r),\\
&v^{k+1} = v^k - \gamma_k \hat{\beta}(\bar{x}_{n+2}^k - \textstyle\sum_{i=1}^n Q_i \bar{x}_i^k + q).
\end{aligned}$$

This new method is not a simple extension of the method in [15], even in the $n = 1$ case, because we make use of the resolvent computations with respect to the operator A in the process of obtaining the intermediate iterates from the current ones, which is different from the method proposed in [15]. In this sense, our algorithm is new. It shares nice convergence properties with the method in [15], and their individual resolvent computations are also the same at each iteration. Moreover, the new method has a broader range of applications.

If $\alpha_1, \ldots, \alpha_n, \alpha_{n+1}, \beta, \hat{\beta}$ satisfy the inequalities (3.7) and (3.8) below, then we follow [15] to resort to characteristic operator techniques [16–18] to prove the weak convergence of our proposed method. In convergence analysis, we introduce Lemma 2, which seems new and is an extension of [19, Lemma A5].

The rest of this article is organized as follows. In Section 2, we give some basic definitions and lemmas. In Section 3, we describe our proposed splitting method in Hilbert spaces in details, specifically tailored for the monotone inclusions (1.1) mentioned earlier. In Section 4, under the weakest possible conditions, by using characteristic operator techniques [16–18] and introducing a new lemma, we prove the weak convergence of the generated primal sequence. In Section 5, we propose a variant of Algorithm 1 and analyze its weak convergence. Section 6 introduces the dual-first version of Algorithm 2. In Section 7, we performed numerical experiments to verify the practical effectiveness of our proposed method, together with its variants, in solving the large-scale rare feature selection in deep learning [9, 20]. Finally, Section 8 concludes this article with some remarks.

2. Preliminaries

In this section, we begin with some basic definitions, followed by presenting auxiliary results that will facilitate our subsequent discussions.

Let $\mathcal{H}$ be an infinite-dimensional Hilbert space, equipped with the standard inner product $\langle \cdot, \cdot \rangle$ and the corresponding norm $\|\cdot\|$. Let $Q\colon \mathcal{H} \to \mathcal{G}$ be a nonzero bounded linear operator, with its adjoint operator Q^*. The norm of Q is given by

$$\|Q\| := \max\{ \sqrt{\langle u, Q^*Qu \rangle} : \ \|u\| = 1, \ u \in \mathcal{H}\}.$$

Definition 1. *An operator $A\colon \mathcal{H} \rightrightarrows \mathcal{H}$ is said to be monotone if*

$$\langle x - x', a - a' \rangle \geq 0, \quad \forall x, \ x' \in \mathrm{dom}A, \ a \in A(x), \ a' \in A(x'),$$

where $\mathrm{dom}A := \{x \in \mathcal{H} : A(x) \neq \emptyset\}$*, and it is said to be maximally monotone if its graph* $\{(x, a) \in \mathcal{H} \times \mathcal{H} : a \in A(x)\}$ *is not properly contained in the graph of any other monotone operator in* $\mathcal{H}$.

The inverse of A defined by $A^{-1}(a) = \{x \in \mathcal{H} : a \in A(x)\}$ is maximally monotone in $\mathcal{H}$ whenever A is. And one important case of maximally monotone operators is ∂f, which is the sub-differential of a closed proper convex function $f\colon \mathcal{H} \to (-\infty, +\infty]$.

For any given maximally monotone operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and $\hat{x} \in \mathcal{H}$, the solution of $(\alpha I + A)(x) \ni \hat{x}$ or $(I + \alpha A)(x) \ni \hat{x}$ exists uniquely [21].

Lemma 1. *Denote*

$$x = (x_1^T, \ldots, x_n^T)^T, \ \bar{A} = \mathrm{diag}(\bar{A}_1, \ldots, \bar{A}_n), \ R = (R_1, \ldots, R_n), \ Q = (Q_1, \ldots, Q_n).$$

For the system of monotone inclusions (1.1), we introduce the dual variable $u \in \mathcal{G}_1$ *and the auxiliary variable* $v \in \mathcal{G}_2$*. Then*

$$T(x,u,v) = \begin{pmatrix} \bar{A} & & \\ & A^{-1} & \\ & & B^{-1} \end{pmatrix} \begin{pmatrix} x \\ u \\ v \end{pmatrix} + \begin{pmatrix} 0 & R^* & Q^* \\ -R & 0 & 0 \\ -Q & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ u \\ v \end{pmatrix} + \begin{pmatrix} 0 \\ r \\ q \end{pmatrix}$$

must be maximally monotone. And T *is termed the characteristic operator with regard to the problem (1.1).*

Proof. Its proof is similar to that of [15, Lemma 1] and thus is omitted. □

At the end of this section, we extend [22, Section 3] and [23, Lemma 5.1] to the following lemma.

Lemma 2. *Let* $R : \mathcal{H} \to \mathcal{G}_1,\ Q : \mathcal{H} \to \mathcal{G}_2$ *be nonzero, bounded, and linear operators, and let* $\alpha > 0,\ \beta > 0$*. If* $4\alpha\beta > \|(Q,\ \sqrt{\beta}R)\|^2$*, then* $\forall x \in \mathcal{H},\ \forall u \in \mathcal{G}_1,\ \forall v \in \mathcal{G}_2$*, the following holds:*

$$\alpha\|x\|^2 + \beta\|u\|^2 + \|v\|^2 - \langle x, Qu + Rv\rangle \geq \varphi(\alpha,\beta,R,Q)\left(\|x\|^2 + \|u\|^2 + \beta^{-1}\|v\|^2\right),$$

where

$$\varphi(\alpha,\beta,R,Q) = \frac{1}{2}\left(\alpha + \beta - \sqrt{(\alpha-\beta)^2 + \|(Q,\ \sqrt{\beta}R)\|^2}\right).$$

3. Method

In this section, we give a detailed description of our proposed splitting method for the system of monotone inclusions (1.1).

The underlying design of this method comes from the following considerations:

Assumption 1. *For the system of monotone inclusions* (1.1)*, we assume the existence of solutions* $x_1^* \in \mathcal{H}_1, \ldots, x_n^* \in \mathcal{H}_n,\ x_{n+1}^* \in \mathcal{G}_1,\ x_{n+2}^* \in \mathcal{G}_2,\ u^* \in \mathcal{G}_1$ *and* $v^* \in \mathcal{G}_2$ *such that these variables satisfy the system*

$$0 \in \bar{A}_i(x_i) + R_i^* u + Q_i^* v, \quad i = 1, \ldots, n, \tag{3.1}$$
$$0 \in A(x_{n+1}) - u, \tag{3.2}$$
$$0 \in B(x_{n+2}) - v, \tag{3.3}$$
$$0 = \textstyle\sum_{i=1}^n R_i x_i - r - x_{n+1}, \tag{3.4}$$
$$0 = \textstyle\sum_{i=1}^n Q_i x_i - q - x_{n+2}. \tag{3.5}$$

Furthermore, it is assumed that for $i = 1, \ldots, n,\ \emptyset \neq \mathrm{dom}\bar{A}_i \subseteq \mathrm{dom}A$ *and* $\emptyset \neq \mathrm{dom}B$*.*

Denote

$$R_{n+1} = -I,\ R_{n+2} = 0,\ Q_{n+1} = 0,\ Q_{n+2} = -I. \tag{3.6}$$

Algorithm 1. *Step 0. Choose* $x_i^0 \in \mathcal{H}_i,\ i = 1, \ldots, n,\ x_{n+1}^0 \in \mathcal{G}_1,\ u^0 \in \mathcal{G}_1,\ x_{n+2}^0 \in \mathcal{G}_2,\ v^0 \in \mathcal{G}_2$*. For* $i = 1, \ldots, n+2$*, choose* $\beta_i > 0,\ \hat{\beta}_i > 0$ *and* $0 < \theta \leq \bar{\theta} < 2$*. Set* $k := 0$*.*

Step 1. Choose α_i satisfying

$$\alpha_i > \|R_i\|^2/(4\beta_i) + \|Q_i\|^2/(4\hat{\beta}_i),\ i = 1, \dots, n, \tag{3.7}$$

$$\alpha_{n+1} > 1/(4\beta_{n+1}), \quad \alpha_{n+2} > 1/(4\hat{\beta}_{n+2}). \tag{3.8}$$

Calculate $\beta = \sum_{i=1}^{n+2}\beta_i$, $\hat{\beta} = \sum_{i=1}^{n+2}\hat{\beta}_i$. For $x_i^k \in \mathcal{H}_i$, $i = 1, \dots, n$, $x_{n+1}^k \in \mathcal{G}_1$, $u^k \in \mathcal{G}_1$, $x_{n+2}^k \in \mathcal{G}_2$, $v^k \in \mathcal{G}_2$, compute

$$\bar{u}^k = u^k - (x_{n+1}^k - \textstyle\sum_{i=1}^n R_i x_i^k + r)/\beta, \tag{3.9}$$

$$\bar{v}^k = v^k - (x_{n+2}^k - \textstyle\sum_{i=1}^n Q_i x_i^k + q)/\hat{\beta}, \tag{3.10}$$

$$(\alpha_i I + \bar{A}_i)(\bar{x}_i^k) \ni \alpha_i x_i^k - R_i^* \bar{u}^k - Q_i^* \bar{v}^k, \tag{3.11}$$

$$(\alpha_{n+1} I + A)(\bar{x}_{n+1}^k) \ni \alpha_{n+1} x_{n+1}^k + \bar{u}^k, \tag{3.12}$$

$$(\alpha_{n+2} I + B)(\bar{x}_{n+2}^k) \ni \alpha_{n+2} x_{n+2}^k + \bar{v}^k. \tag{3.13}$$

If a prescribed stopping criterion is satisfied, the algorithm terminates. Otherwise, choose $\theta_k \in [\theta, \bar{\theta}]$, compute

$$\begin{aligned}
\phi_k &:= \textstyle\sum_{i=1}^{n+2}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \langle \bar{x}_{n+1}^k - \sum_{i=1}^n R_i \bar{x}_i^k + r, u^k - \bar{u}^k\rangle + \langle \bar{x}_{n+2}^k - \sum_{i=1}^n Q_i \bar{x}_i^k + q, v^k - \bar{v}^k\rangle,\\
\psi_k &:= \textstyle\sum_{i=1}^{n+2}\|\alpha_i(x_i^k - \bar{x}_i^k)\|^2 + \|\bar{x}_{n+1}^k - \sum_{i=1}^n R_i \bar{x}_i^k + r\|^2 + \|\bar{x}_{n+2}^k - \sum_{i=1}^n Q_i \bar{x}_i^k + q\|^2,\\
\gamma_k &:= \theta_k \phi_k / \psi_k.
\end{aligned} \tag{3.14}$$

Step 2. For $i = 1, \dots, n+2$, compute in order

$$\begin{aligned}
x_i^{k+1} &= x_i^k - \gamma_k \alpha_i (x_i^k - \bar{x}_i^k),\\
u^{k+1} &= u^k - \gamma_k(\bar{x}_{n+1}^k - \textstyle\sum_{i=1}^n R_i \bar{x}_i^k + r),\\
v^{k+1} &= v^k - \gamma_k(\bar{x}_{n+2}^k - \textstyle\sum_{i=1}^n Q_i \bar{x}_i^k + q).
\end{aligned}$$

Set $k := k+1$.

Interestingly, when specialized to $0 \in \bar{A}(x) + A(x)$, in the $\beta = +\infty$ case, Algorithm 1 reduces to

$$\begin{aligned}
&(\alpha I + \bar{A})(\bar{x}^k) \ni \alpha x^k - u^k, \quad (\alpha_2 I + A)(\bar{x}_2^k) \ni \alpha_2 x_2^k + u^k,\\
&x^{k+1} = x^k - \gamma_k \alpha (x^k - \bar{x}^k), \quad x_2^{k+1} = x_2^k - \gamma_k \alpha_2 (x_2^k - \bar{x}_2^k),\\
&u^{k+1} = u^k - \gamma_k(\bar{x}_2^k - \bar{x}^k).
\end{aligned}$$

4. Weak convergence

In this section, under the weakest possible assumptions, we analyze the convergence properties exhibited by both the primal and dual sequences of iterates generated by Algorithm 1. In particular, we rigorously prove that the primal sequence of iterates converges weakly to a solution of the problem (1.1).

Theorem 1. *Let $\{x_i^k\}_{i=1}^{n+2}$, $\{u^k\}$, and $\{v^k\}$ be the sequences generated by Algorithm 1. If Assumption 1 holds, define $\beta := \sum_{i=1}^{n+2}\beta_i$, $\hat{\beta} := \sum_{i=1}^{n+2}\hat{\beta}_i$ and assume that*

$$\alpha_i > \|R_i\|^2/(4\beta_i) + \|Q_i\|^2/(4\hat{\beta}_i),\ i = 1, \dots, n, \tag{4.1}$$

$$\alpha_{n+1} > 1/(4\beta_{n+1}), \quad \alpha_{n+2} > 1/(4\hat{\beta}_{n+2}), \tag{4.2}$$

then there exists some positive number $\hat{\gamma}$ *satisfying*

$$\|w^{k+1} - w^*\|^2 \le \|w^k - w^*\|^2 - \hat{\gamma}\left(\textstyle\sum_{i=1}^{n+2}(\|x_i^k - \bar{x}_i^k\|^2 + \beta_i{}^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2) + (n+2)\|u^k - \bar{u}^k\|^2\right). \tag{4.3}$$

Proof. For $i = 1, \ldots, n$, it can be deduced from (3.11) and (3.1) that

$$\bar{A}_i(\bar{x}_i^k) \ni \alpha_i(x_i^k - \bar{x}_i^k) - R_i^*\bar{u}^k - Q_i^*\bar{v}^k, \;\; \bar{A}_i(x_i^*) \ni -R_i^*u^* - Q_i^*v^*,$$

which, together with the monotonicity of each $\bar{A}_i$, imply

$$\langle \bar{x}_i^k - x_i^*, \alpha_i(x_i^k - \bar{x}_i^k) - R_i^*(\bar{u}^k - u^*) - Q_i^*(\bar{v}^k - v^*)\rangle \ge 0.$$

By using $\bar{x}_i^k - x_i^* = x_i^k - x_i^* - (x_i^k - \bar{x}_i^k), \; i = 1, \ldots, n$, we have

$$\langle x_i^k - x_i^*, \alpha_i(x_i^k - \bar{x}_i^k)\rangle - \langle \bar{x}_i^k - x_i^*, R_i^*(\bar{u}^k - u^*) + Q_i^*(\bar{v}^k - v^*)\rangle \ge \alpha_i\|x_i^k - \bar{x}_i^k\|^2. \tag{4.4}$$

Based on (3.12) and (3.2), we derive

$$A(\bar{x}_{n+1}^k) \ni \alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k, \;\; A(x_{n+1}^*) \ni u^*,$$

and due to the monotonicity of A, it can be inferred that

$$\langle \bar{x}_{n+1}^k - x_{n+1}^*, \alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k - u^*\rangle \ge 0,$$

which indicates

$$\langle x_{n+1}^k - x_{n+1}^*, \alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k)\rangle + \langle \bar{x}_{n+1}^k - x_{n+1}^*, \bar{u}^k - u^*\rangle \ge \alpha_{n+1}\|x_{n+1}^k - \bar{x}_{n+1}^k\|^2. \tag{4.5}$$

Similarly, we can derive from (3.13) and (3.3) that

$$B(\bar{x}_{n+2}^k) \ni \alpha_{n+2}(x_{n+2}^k - \bar{x}_{n+2}^k) + \bar{v}^k, \;\; B(x_{n+2}^*) \ni v^*.$$

From the monotonicity of B, we can obtain

$$\langle \bar{x}_{n+2}^k - x_{n+2}^*, \alpha_{n+2}(x_{n+2}^k - \bar{x}_{n+2}^k) + \bar{v}^k - v^*\rangle \ge 0,$$

thereby implying that

$$\langle x_{n+2}^k - x_{n+2}^*, \alpha_{n+2}(x_{n+2}^k - \bar{x}_{n+2}^k)\rangle + \langle \bar{x}_{n+2}^k - x_{n+2}^*, \bar{v}^k - v^*\rangle \ge \alpha_{n+2}\|x_{n+2}^k - \bar{x}_{n+2}^k\|^2. \tag{4.6}$$

Combining (4.4)–(4.6) with (3.6) yields

$$\textstyle\sum_{i=1}^{n+2}\langle x_i^k - x_i^*, \alpha_i(x_i^k - \bar{x}_i^k)\rangle - \sum_{i=1}^{n+2}\langle R_i(\bar{x}_i^k - x_i^*), \bar{u}^k - u^*\rangle - \sum_{i=1}^{n+2}\langle Q_i(\bar{x}_i^k - x_i^*), \bar{v}^k - v^*\rangle \ge \sum_{i=1}^{n+2}\alpha_i\|x_i^k - \bar{x}_i^k\|^2.$$

Furthermore, according to (3.4)–(3.6), we have

$$\textstyle\sum_{i=1}^{n+2}R_ix_i^* = r, \quad \sum_{i=1}^{n+2}Q_ix_i^* = q.$$

Thus

$$\textstyle\sum_{i=1}^{n+2}\langle x_i^k - x_i^*, \alpha_i(x_i^k - \bar{x}_i^k)\rangle - \langle \sum_{i=1}^{n+2} R_i\bar{x}_i^k - r, \bar{u}^k - u^*\rangle - \langle \sum_{i=1}^{n+2} Q_i\bar{x}_i^k - q, \bar{v}^k - v^*\rangle \geq \sum_{i=1}^{n+2}\alpha_i\|x_i^k - \bar{x}_i^k\|^2.$$

In terms of

$$\bar{u}^k - u^* = u^k - u^* - (u^k - \bar{u}^k), \quad \bar{v}^k - v^* = v^k - v^* - (v^k - \bar{v}^k),$$

we can obtain

$$\begin{aligned} &\textstyle\sum_{i=1}^{n+2}\langle x_i^k - x_i^*, \alpha_i(x_i^k - \bar{x}_i^k)\rangle + \langle u^k - u^*, -\sum_{i=1}^{n+2} R_i\bar{x}_i^k + r\rangle + \langle v^k - v^*, -\sum_{i=1}^{n+2} Q_i\bar{x}_i^k + q\rangle \\ \geq &\textstyle\sum_{i=1}^{n+2}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 - \langle \sum_{i=1}^{n+2} R_i\bar{x}_i^k - r, u^k - \bar{u}^k\rangle - \langle \sum_{i=1}^{n+2} Q_i\bar{x}_i^k - q, v^k - \bar{v}^k\rangle. \end{aligned}$$

Next, denote

$$w := \begin{pmatrix} x_1 \\ \vdots \\ x_{n+2} \\ u \\ v \end{pmatrix}, \quad d := \begin{pmatrix} \alpha_1(x_1 - \bar{x}_1) \\ \vdots \\ \alpha_{n+2}(x_{n+2} - \bar{x}_{n+2}) \\ -\sum_{i=1}^{n+2} R_i\bar{x}_i + r \\ -\sum_{i=1}^{n+2} Q_i\bar{x}_i + q \end{pmatrix},$$

and

$$\phi_k = \textstyle\sum_{i=1}^{n+2}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 - \langle \sum_{i=1}^{n+2} R_i\bar{x}_i^k - r, u^k - \bar{u}^k\rangle - \langle \sum_{i=1}^{n+2} Q_i\bar{x}_i^k - q, v^k - \bar{v}^k\rangle,$$

then the inequality above can be rewritten as $\langle w^k - w^*, d^k\rangle \geq \phi_k$, where w^* is the corresponding solution. Together with (3.14) and $\psi_k = \|d^k\|^2$, we obtain

$$\begin{aligned} \|w^{k+1} - w^*\|^2 &= \|w^k - w^* - \gamma_k d^k\|^2 \\ &= \|w^k - w^*\|^2 - 2\gamma_k\langle w^k - w^*, d^k\rangle + \gamma_k^2\|d^k\|^2 \\ &\leq \|w^k - w^*\|^2 - 2\gamma_k\phi_k + \gamma_k^2\psi_k \\ &= \|w^k - w^*\|^2 - (2-\theta)\gamma_k\phi_k. \end{aligned} \tag{4.7}$$

Note that

$$\begin{aligned} -\textstyle\sum_{i=1}^{n+2} R_i\bar{x}_i^k + r &= \bar{x}_{n+1}^k - \textstyle\sum_{i=1}^{n} R_i\bar{x}_i^k + r \\ &= x_{n+1}^k - \textstyle\sum_{i=1}^{n} R_i x_i^k + r + \bar{x}_{n+1}^k - x_{n+1}^k + \sum_{i=1}^{n} R_i(x_i^k - \bar{x}_i^k) \\ &= \beta(u^k - \bar{u}^k) + \bar{x}_{n+1}^k - x_{n+1}^k + \textstyle\sum_{i=1}^{n} R_i(x_i^k - \bar{x}_i^k). \end{aligned}$$

Similarly,

$$-\textstyle\sum_{i=1}^{n+2} Q_i\bar{x}_i^k + q = \hat{\beta}(v^k - \bar{v}^k) + \bar{x}_{n+2}^k - x_{n+2}^k + \sum_{i=1}^{n} Q_i(x_i^k - \bar{x}_i^k).$$

From Lemma 2, $\beta := \sum_{i=1}^{n+2}\beta_i$ and $\hat{\beta} := \sum_{i=1}^{n+2}\hat{\beta}_i$, we have

$$\begin{aligned} \phi_k &= \textstyle\sum_{i=1}^{n+2}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 - \langle \sum_{i=1}^{n+2} R_i\bar{x}_i^k - r, u^k - \bar{u}^k\rangle - \langle \sum_{i=1}^{n+2} Q_i\bar{x}_i^k - q, v^k - \bar{v}^k\rangle \\ &= \textstyle\sum_{i=1}^{n+2}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \beta\|u^k - \bar{u}^k\|^2 + \hat{\beta}\|v^k - \bar{v}^k\|^2 + \sum_{i=1}^{n+2}\langle R_i(x_i^k - \bar{x}_i^k), u^k - \bar{u}^k\rangle + \sum_{i=1}^{n+2}\langle Q_i(x_i^k - \bar{x}_i^k), v^k - \bar{v}^k\rangle \\ &= \textstyle\sum_{i=1}^{n+2}(\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \beta_i\|u^k - \bar{u}^k\|^2 + \hat{\beta}_i\|v^k - \bar{v}^k\|^2 + \langle x_i^k - \bar{x}_i^k, R_i^*(u^k - \bar{u}^k) + Q_i^*(v^k - \bar{v}^k)\rangle) \end{aligned}$$

$$= \sum_{i=1}^{n+2}\hat{\beta}_i(\hat{\beta}_i^{-1}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \hat{\beta}_i^{-1}\beta_i\|\bar{u}^k - u^k\|^2 + \|\bar{v}^k - v^k\|^2 - \langle x_i^k - \bar{x}_i^k, \hat{\beta}_i^{-1}R_i^*(\bar{u}^k - u^k) + \hat{\beta}_i^{-1}Q_i^*(\bar{v}^k - v^k)\rangle)$$
$$\geq \sum_{i=1}^{n+2}\hat{\beta}_i\varphi(\hat{\beta}_i^{-1}\alpha_i, \hat{\beta}_i^{-1}\beta_i, \hat{\beta}_i^{-1}R_i^*, \hat{\beta}_i^{-1}Q_i^*)(\|x_i^k - \bar{x}_i^k\|^2 + \|u^k - \bar{u}^k\|^2 + \beta_i^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2)$$
$$= \sum_{i=1}^{n+2}\varphi(\alpha_i, \beta_i, R_i^*, Q_i^*)(\|x_i^k - \bar{x}_i^k\|^2 + \|u^k - \bar{u}^k\|^2 + \beta_i^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2),$$

and the conditions (4.1) and (4.2) indicate

$$4\alpha_i\beta_i > \|R_i^*\|^2 + \hat{\beta}_i^{-1}\beta_i\|Q_i^*\|^2 > \|(R_i^*, \sqrt{\hat{\beta}_i^{-1}\beta_i}Q_i^*)\|^2, \; i = 1, \ldots, n+2.$$

So each

$$\frac{1}{2}\left(\alpha_i + \beta_i - \sqrt{(\alpha_i - \beta_i)^2 + \|(R_i^*, \sqrt{\hat{\beta}_i^{-1}\beta_i}Q_i^*)\|^2}\right)$$

must be positive. Let ρ be their minimum; thus, we further obtain

$$\phi_k \geq \rho\left(\sum_{i=1}^{n+2}\|x_i^k - \bar{x}_i^k\|^2 + (n+2)\|u^k - \bar{u}^k\|^2 + \sum_{i=1}^{n+2}\beta_i^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2\right). \tag{4.8}$$

On the other hand,

$$\psi_k = \sum_{i=1}^{n+2}\|\alpha_i(x_i^k - \bar{x}_i^k)\|^2 + \| - \sum_{i=1}^{n+2}R_i\bar{x}_i^k + r\|^2 + \| - \sum_{i=1}^{n+2}Q_i\bar{x}_i^k + q\|^2$$
$$= \sum_{i=1}^{n+2}\|\alpha_i(x_i^k - \bar{x}_i^k)\|^2 + \|\beta(u^k - \bar{u}^k) + \sum_{i=1}^{n+2}R_i(x_i^k - \bar{x}_i^k)\|^2 + \|\hat{\beta}(v^k - \bar{v}^k) + \sum_{i=1}^{n+2}Q_i(x_i^k - \bar{x}_i^k)\|^2$$
$$= \sum_{i=1}^{n+2}\|\alpha_i(x_i^k - \bar{x}_i^k)\|^2 + \|\sum_{i=1}^{n+2}\beta_i(u^k - \bar{u}^k) + \sum_{i=1}^{n+2}R_i(x_i^k - \bar{x}_i^k)\|^2$$
$$+ \|\sum_{i=1}^{n+2}\sqrt{\beta_i\hat{\beta}_i}(\sqrt{\beta_i^{-1}\hat{\beta}_i}(v^k - \bar{v}^k)) + \sum_{i=1}^{n+2}Q_i(x_i^k - \bar{x}_i^k)\|^2$$
$$\leq \sum_{i=1}^{n+2}\alpha_i^2\sum_{i=1}^{n+2}\|x_i^k - \bar{x}_i^k\|^2 + (\sum_{i=1}^{n+2}\beta_i^2 + \sum_{i=1}^{n+2}\|R_i\|^2)\left((n+2)\|u^k - \bar{u}^k\|^2 + \sum_{i=1}^{n+2}\|x_i^k - \bar{x}_i^k\|^2\right)$$
$$+ (\sum_{i=1}^{n+2}\beta_i\hat{\beta}_i + \sum_{i=1}^{n+2}\|Q_i\|^2)(\sum_{i=1}^{n+2}\beta_i^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2 + \sum_{i=1}^{n+2}\|x_i^k - \bar{x}_i^k\|^2)$$
$$\leq \left(\sum_{i=1}^{n+2}(\alpha_i^2 + \beta_i^2 + \beta_i\hat{\beta}_i + \|R_i\|^2 + \|Q_i\|^2)\right)\left(\sum_{i=1}^{n+2}\|x_i^k - \bar{x}_i^k\|^2 + (n+2)\|u^k - \bar{u}^k\|^2 + \sum_{i=1}^{n+2}\beta_i^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2\right).$$

Thus, we can conclude that

$$\gamma_k = \theta_k\phi_k/\psi_k \geq (\theta\rho)/\sum_{i=1}^{n+2}(\alpha_i^2 + \beta_i^2 + \beta_i\hat{\beta}_i + \|R_i\|^2 + \|Q_i\|^2) > 0.$$

Combining this with (4.7) and (4.8) yields the desired result. □

Theorem 2. *Let $\{x_i^k\}_{i=1}^{n+2}$, $\{u^k\}$, and $\{v^k\}$ be the sequences generated by Algorithm 1. If Assumption 1 and conditions (4.1) and (4.2) hold, then the corresponding primal sequence $\{x_i^k\}_{i=1}^{n}$ weakly converges to a solution of the system of monotone inclusions (1.1) mentioned above.*

Proof. It follows from (4.3) that

$$(i)\; x_i^k - \bar{x}_i^k \to 0, \; u^k - \bar{u}^k \to 0, \; v^k - \bar{v}^k \to 0, \; i = 1, \ldots, n+2; \tag{4.9}$$
$$(ii)\; \{x_i^k\}_{i=1}^{n+2}, \; \{u^k\}, \text{ and } \{\mathrm{v}^k\} \text{ are bounded in norm.} \tag{4.10}$$

In accordance with the definition of T, we can obtain

$$T(\bar{x}^k, \alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k, \alpha_{n+2}(x_{n+2}^k - \bar{x}_{n+2}^k) + \bar{v}^k)$$

$$= \begin{pmatrix} \vdots \\ \bar{A}_i(\bar{x}_i^k) + R_i^*(\alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k) + Q_i^*(\alpha_{n+2}(x_{n+2}^k - \bar{x}_{n+2}^k) + \bar{v}^k) \\ \vdots \\ A^{-1}(\alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k) - \sum_{i=1}^n R_i \bar{x}_i^k + r \\ B^{-1}(\alpha_{n+2}(x_{n+2}^k - \bar{x}_{n+2}^k) + \bar{v}^k) - \sum_{i=1}^n Q_i \bar{x}_i^k + q \end{pmatrix}.$$

Utilizing (3.11)–(3.13), we further deduce that

$$\begin{aligned} &\bar{A}_i(\bar{x}_i^k) + R_i^*(\alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k) + Q_i^*(\alpha_{n+2}(x_{n+2}^k - \bar{x}_{n+2}^k) + \bar{v}^k) \\ \ni\ &\alpha_i(x_i^k - \bar{x}_i^k) - R_i^*\bar{u}^k - Q_i^*\bar{v}^k + \alpha_{n+1}R_i^*(x_{n+1}^k - \bar{x}_{n+1}^k) + R_i^*\bar{u}^k + \alpha_{n+2}Q_i^*(x_{n+2}^k - \bar{x}_{n+2}^k) + Q_i^*\bar{v}^k \\ =\ &\alpha_i(x_i^k - \bar{x}_i^k) + \alpha_{n+1}R_i^*(x_{n+1}^k - \bar{x}_{n+1}^k) + \alpha_{n+2}Q_i^*(x_{n+2}^k - \bar{x}_{n+2}^k), \end{aligned}$$

$$A^{-1}(\alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k) - \textstyle\sum_{i=1}^n R_i\bar{x}_i^k + r \ni \beta(u^k - \bar{u}^k) - (x_{n+1}^k - \bar{x}_{n+1}^k) + \sum_{i=1}^n R_i(x_i^k - \bar{x}_i^k),$$

$$B^{-1}(\alpha_{n+2}(x_{n+2}^k - \bar{x}_{n+2}^k) + \bar{v}^k) - \textstyle\sum_{i=1}^n Q_i\bar{x}_i^k + q \ni \hat{\beta}(v^k - \bar{v}^k) - (x_{n+2}^k - \bar{x}_{n+2}^k) + \sum_{i=1}^n Q_i(x_i^k - \bar{x}_i^k).$$

Therefore,

$$\begin{aligned} &T(\bar{x}^k, \alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k, \alpha_{n+2}(x_{n+2}^k - \bar{x}_{n+2}^k) + \bar{v}^k) \\ \ni &\begin{pmatrix} \vdots \\ \alpha_i(x_i^k - \bar{x}_i^k) + \alpha_{n+1}R_i^*(x_{n+1}^k - \bar{x}_{n+1}^k) + \alpha_{n+2}Q_i^*(x_{n+2}^k - \bar{x}_{n+2}^k) \\ \vdots \\ \beta(u^k - \bar{u}^k) - (x_{n+1}^k - \bar{x}_{n+1}^k) + \sum_{i=1}^n R_i(x_i^k - \bar{x}_i^k) \\ \hat{\beta}(v^k - \bar{v}^k) - (x_{n+2}^k - \bar{x}_{n+2}^k) + \sum_{i=1}^n Q_i(x_i^k - \bar{x}_i^k) \end{pmatrix}. \end{aligned} \tag{4.11}$$

Due to (4.9) and the boundedness of each $R_i,\ Q_i$, we can see that the right-hand side of (4.11) converges strongly to zero. Meanwhile, from (4.10), there exists at least one weak cluster point $(x^\infty,\ u^\infty,\ v^\infty)$ such that

$$x^{k_j} \rightharpoonup x^\infty,\ u^{k_j} \rightharpoonup u^\infty,\ v^{k_j} \rightharpoonup v^\infty,$$

which, together with (4.9), implies

$$\bar{x}^{k_j} \rightharpoonup x^\infty,\ \bar{u}^{k_j} \rightharpoonup u^\infty,\ \bar{v}^{k_j} \rightharpoonup v^\infty.$$

Finally, invoking [15, Lemma 3] or [24, Lemma 3.2], we can conclude that this weak cluster point is a solution point of $0 \in T(x, u, v)$, thereby also solving the problem (1.1). The proof of uniqueness of weak cluster point is standard [12, 25] and thus is omitted. □

5. The variant of Algorithm 1

In this section, we give a variant of Algorithm 1 tailored for the system of monotone inclusions (1.1). The underlying design of this method is rooted in the following:

Assumption 2. *For the system of monotone inclusions* (1.1)*, we assume the existence of solutions* $x_1^* \in \mathcal{H}_1, \ldots, x_n^* \in \mathcal{H}_n,\ x_{n+1}^* \in \mathcal{G}_1,\ u^* \in \mathcal{G}_1$ *and* $v^* \in \mathcal{G}_2$ *such that these variables collectively satisfy the system*

$$0 \in \bar{A}_i(x_i) + R_i^* u + Q_i^* v, \quad i = 1, \ldots, n, \tag{5.1}$$

$$0 \in A(x_{n+1}) - u, \tag{5.2}$$

$$0 = \textstyle\sum_{i=1}^n R_i x_i - r - x_{n+1}, \tag{5.3}$$

$$0 \in B^{-1}(v) - \textstyle\sum_{i=1}^n Q_i x_i + q. \tag{5.4}$$

Furthermore, it is assumed that for $i = 1, \ldots, n,\ \emptyset \neq \mathrm{dom}\bar{A}_i \subseteq \mathrm{dom}A$ *and* $\emptyset \neq \mathrm{dom}B$*.*

Denote

$$R_{n+1} = -I,\ Q_{n+1} = 0. \tag{5.5}$$

Algorithm 2. *Step 0. Choose* $x_i^0 \in \mathcal{H}_i$*,* $i = 1, \ldots, n$*,* $x_{n+1}^0 \in \mathcal{G}_1$*,* $u^0 \in \mathcal{G}_1$*,* $v^0 \in \mathcal{G}_2$*. For* $i = 1, \ldots, n+1$*, choose* $\beta_i > 0,\ \hat{\beta}_i > 0$ *and* $0 < \theta \leq \bar{\theta} < 2$*. Set* $k := 0$*.*

Step 1. Choose α_i *satisfying*

$$\alpha_i > \|R_i\|^2/(4\beta_i) + \|Q_i\|^2/(4\hat{\beta}_i),\ i = 1, \ldots, n+1.$$

Calculate $\beta = \sum_{i=1}^{n+1}\beta_i$*,* $\hat{\beta} = (\sum_{i=1}^{n+1}\hat{\beta}_i)^{-1}$*. For* $x_i^k \in \mathcal{H}_i$*,* $i = 1, \ldots, n$*,* $x_{n+1}^k \in \mathcal{G}_1$*,* $u^k \in \mathcal{G}_1$*,* $v^k \in \mathcal{G}_2$*, compute*

$$\bar{u}^k = u^k - (x_{n+1}^k - \textstyle\sum_{i=1}^n R_i x_i^k + r)/\beta, \tag{5.6}$$

$$(\alpha_i I + \bar{A}_i)(\bar{x}_i^k) \ni \alpha_i x_i^k - R_i^* \bar{u}^k - Q_i^* v^k, \tag{5.7}$$

$$(\alpha_{n+1} I + A)(\bar{x}_{n+1}^k) \ni \alpha_{n+1} x_{n+1}^k + \bar{u}^k, \tag{5.8}$$

$$(I + \hat{\beta} B^{-1})(\bar{v}^k) \ni v^k + \hat{\beta}(\textstyle\sum_{i=1}^n Q_i \bar{x}_i^k - q). \tag{5.9}$$

If a prescribed stopping criterion is satisfied, the algorithm terminates. Otherwise, choose $\theta_k \in [\theta, \bar{\theta}]$*, compute*

$$\begin{aligned}
\phi_k &:= \textstyle\sum_{i=1}^{n+1}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \hat{\beta}^{-1}\|v^k - \bar{v}^k\|^2 + \langle \bar{x}_{n+1}^k - \sum_{i=1}^n R_i\bar{x}_i^k + r, u^k - \bar{u}^k\rangle - \sum_{i=1}^n\langle Q_i(x_i^k - \bar{x}_i^k), v^k - \bar{v}^k\rangle, \\
\psi_k &:= \textstyle\sum_{i=1}^{n+1}\|\alpha_i(x_i^k - \bar{x}_i^k) - Q_i^*(v^k - \bar{v}^k)\|^2 + \|\bar{x}_{n+1}^k - \sum_{i=1}^n R_i\bar{x}_i^k + r\|^2 + \|\hat{\beta}^{-1}(v^k - \bar{v}^k)\|^2, \\
\gamma_k &:= \theta_k\phi_k/\psi_k.
\end{aligned} \tag{5.10}$$

Step 2. For $i = 1, \ldots, n+1$*, compute in order*

$$\begin{aligned}
x_i^{k+1} &= x_i^k - \gamma_k(\alpha_i(x_i^k - \bar{x}_i^k) - Q_i^*(v^k - \bar{v}^k)), \\
u^{k+1} &= u^k - \gamma_k(\bar{x}_{n+1}^k - \textstyle\sum_{i=1}^n R_i\bar{x}_i^k + r), \\
v^{k+1} &= v^k - \gamma_k\hat{\beta}^{-1}(v^k - \bar{v}^k).
\end{aligned}$$

Set $k := k+1$*.*

Remark 1. *In the $n = 1$ case, when A, R, and r vanish, the problem* (1.1) *can be transformed into*

$$0 \in \bar{A}(x) + Q^* B(Qx - q),$$

and the corresponding iteration about the intermediate points can be simplified as

$$\begin{aligned}(\alpha I + \bar{A})(\bar{x}^k) &\ni \alpha x^k - Q^* v^k,\\ (I + \hat{\beta} B^{-1})(\bar{v}^k) &\ni v^k + \hat{\beta}(Q\bar{x}^k - q),\end{aligned}$$

which is a special case of the proximal point method. See [26–28] for more details.

Theorem 3. *Let $\{x_i^k\}_{i=1}^{n+1}$, $\{u^k\}$, and $\{v^k\}$ be the sequences generated by Algorithm 2. If Assumption 2 holds, define $\beta := \sum_{i=1}^{n+1}\beta_i$, $\hat{\beta} := (\sum_{i=1}^{n+1}\hat{\beta}_i)^{-1}$ and assume that*

$$\alpha_i > \|R_i\|^2/(4\beta_i) + \|Q_i\|^2/(4\hat{\beta}_i),\ \ i = 1, \ldots, n, \tag{5.11}$$

$$\alpha_{n+1} > 1/(4\beta_{n+1}), \tag{5.12}$$

then there exists some positive number $\hat{\gamma}$ satisfying

$$\|w^{k+1} - w^*\|^2 \le \|w^k - w^*\|^2 - \hat{\gamma}\left(\textstyle\sum_{i=1}^{n+1}(\|x_i^k - \bar{x}_i^k\|^2 + \beta_i^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2) + (n+1)\|u^k - \bar{u}^k\|^2\right). \tag{5.13}$$

Proof. For $i = 1, \ldots, n$, it can be deduced from (5.7) and (5.1) that

$$\bar{A}_i(\bar{x}_i^k) \ni \alpha_i(x_i^k - \bar{x}_i^k) - R_i^*\bar{u}^k - Q_i^* v^k,\ \ \bar{A}_i(x_i^*) \ni -R_i^* u^* - Q_i^* v^*.$$

According to the monotonicity of each $\bar{A}_i$, it follows that

$$\langle \bar{x}_i^k - x_i^*, \alpha_i(x_i^k - \bar{x}_i^k) - R_i^*(\bar{u}^k - u^*) - Q_i^*(v^k - v^*)\rangle \ge 0.$$

By utilizing $\bar{x}_i^k - x_i^* = x_i^k - x_i^* - (x_i^k - \bar{x}_i^k),\ i = 1, \ldots, n$, we obtain

$$\langle x_i^k - x_i^*, \alpha_i(x_i^k - \bar{x}_i^k)\rangle - \langle \bar{x}_i^k - x_i^*, R_i^*(\bar{u}^k - u^*) + Q_i^*(v^k - v^*)\rangle \ge \alpha_i\|x_i^k - \bar{x}_i^k\|^2. \tag{5.14}$$

Based on (5.8) and (5.2), we derive

$$A(\bar{x}_{n+1}^k) \ni \alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k,\ \ A(x_{n+1}^*) \ni u^*.$$

Then it follows from the monotonicity of A that

$$\langle \bar{x}_{n+1}^k - x_{n+1}^*, \alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k - u^*\rangle \ge 0,$$

which indicates

$$\langle x_{n+1}^k - x_{n+1}^*, \alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k)\rangle + \langle \bar{x}_{n+1}^k - x_{n+1}^*, \bar{u}^k - u^*\rangle \ge \alpha_{n+1}\|x_{n+1}^k - \bar{x}_{n+1}^k\|^2. \tag{5.15}$$

Similarly, we can derive from (5.9) and (5.4) that

$$B^{-1}(\bar{v}^k) \ni \hat{\beta}^{-1}(v^k - \bar{v}^k) + \textstyle\sum_{i=1}^n Q_i\bar{x}_i^k - q,\ \ B^{-1}(v^*) \ni \sum_{i=1}^n Q_i x_i^* - q.$$

From the monotonicity of B^{-1}, we have

$$\langle \bar{v}^k - v^*, \hat{\beta}^{-1}(v^k - \bar{v}^k) + \textstyle\sum_{i=1}^{n} Q_i(\bar{x}_i^k - x_i^*)\rangle \geq 0,$$

implying that

$$\langle v^k - v^*, \hat{\beta}^{-1}(v^k - \bar{v}^k)\rangle + \langle \bar{v}^k - v^*, \textstyle\sum_{i=1}^{n} Q_i(\bar{x}_i^k - x_i^*)\rangle \geq \hat{\beta}^{-1}\|v^k - \bar{v}^k\|^2. \tag{5.16}$$

Combining (5.14)–(5.16) with (5.5) results in,

$$\begin{aligned}
&\textstyle\sum_{i=1}^{n+1}\langle x_i^k - x_i^*, \alpha_i(x_i^k - \bar{x}_i^k)\rangle - \sum_{i=1}^{n+1}\langle \bar{x}_i^k - x_i^*, R_i^*(\bar{u}^k - u^*) + Q_i^*(v^k - v^*)\rangle \\
&\quad + \langle v^k - v^*, \hat{\beta}^{-1}(v^k - \bar{v}^k)\rangle + \langle \bar{v}^k - v^*, \textstyle\sum_{i=1}^{n+1} Q_i(\bar{x}_i^k - x_i^*)\rangle \\
&\geq \textstyle\sum_{i=1}^{n+1}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \hat{\beta}^{-1}\|v^k - \bar{v}^k\|^2.
\end{aligned}$$

Furthermore, based on (5.3) and (5.5), we derive that $\sum_{i=1}^{n+1} R_i x_i^* = r$, thereby

$$\begin{aligned}
&\textstyle\sum_{i=1}^{n+1}\langle x_i^k - x_i^*, \alpha_i(x_i^k - \bar{x}_i^k)\rangle - \langle \sum_{i=1}^{n+1} Q_i(\bar{x}_i^k - x_i^*), v^k - \bar{v}^k\rangle \\
&\quad + \langle \bar{u}^k - u^*, -\textstyle\sum_{i=1}^{n+1} R_i\bar{x}_i^k + r\rangle + \langle v^k - v^*, \hat{\beta}^{-1}(v^k - \bar{v}^k)\rangle \\
&\geq \textstyle\sum_{i=1}^{n+1}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \hat{\beta}^{-1}\|v^k - \bar{v}^k\|^2.
\end{aligned}$$

In terms of

$$\begin{aligned}
&\bar{x}_i^k - x_i^* = x_i^k - x_i^* - (x_i^k - \bar{x}_i^k),\ i = 1, \ldots, n, \\
&\bar{u}^k - u^* = u^k - u^* - (u^k - \bar{u}^k), \\
&\bar{v}^k - v^* = v^k - v^* - (v^k - \bar{v}^k),
\end{aligned}$$

we can further obtain

$$\begin{aligned}
&\textstyle\sum_{i=1}^{n+1}\langle x_i^k - x_i^*, \alpha_i(x_i^k - \bar{x}_i^k) - Q_i^*(v^k - \bar{v}^k)\rangle + \langle u^k - u^*, -\sum_{i=1}^{n+1} R_i\bar{x}_i^k + r\rangle + \langle v^k - v^*, \hat{\beta}^{-1}(v^k - \bar{v}^k)\rangle \\
&\geq \textstyle\sum_{i=1}^{n+1}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \hat{\beta}^{-1}\|v^k - \bar{v}^k\|^2 - \langle u^k - \bar{u}^k, \sum_{i=1}^{n+1} R_i\bar{x}_i^k - r\rangle - \sum_{i=1}^{n+1}\langle x_i^k - \bar{x}_i^k, Q_i^*(v^k - \bar{v}^k)\rangle.
\end{aligned}$$

Next, denote

$$w := \begin{pmatrix} x_1 \\ \vdots \\ x_{n+1} \\ u \\ v \end{pmatrix}, \quad d := \begin{pmatrix} \alpha_1(x_1 - \bar{x}_1) - Q_1^*(v - \bar{v}) \\ \vdots \\ \alpha_{n+1}(x_{n+1} - \bar{x}_{n+1}) - Q_{n+1}^*(v - \bar{v}) \\ -\sum_{i=1}^{n+1} R_i\bar{x}_i + r \\ \hat{\beta}^{-1}(v - \bar{v}) \end{pmatrix},$$

and

$$\phi_k = \textstyle\sum_{i=1}^{n+1}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \hat{\beta}^{-1}\|v^k - \bar{v}^k\|^2 - \langle u^k - \bar{u}^k, \sum_{i=1}^{n+1} R_i\bar{x}_i^k - r\rangle - \sum_{i=1}^{n+1}\langle x_i^k - \bar{x}_i^k, Q_i^*(v^k - \bar{v}^k)\rangle,$$

then the inequality above can be rewritten as $\langle w^k - w^*, d^k\rangle \geq \phi_k$, where w^* is the corresponding solution. Together with (5.10) and $\psi_k = \|d^k\|^2$, we obtain

$$\|w^{k+1} - w^*\|^2 = \|w^k - w^* - \gamma_k d^k\|^2$$

$$
\begin{aligned}
&= \|w^k - w^*\|^2 - 2\gamma_k\langle w^k - w^*, d^k\rangle + \gamma_k^2\|d^k\|^2 \\
&\le \|w^k - w^*\|^2 - 2\gamma_k\phi_k + \gamma_k^2\psi_k \\
&= \|w^k - w^*\|^2 - (2-\theta_k)\gamma_k\phi_k. \qquad (5.17)
\end{aligned}
$$

Note that

$$
\begin{aligned}
-\textstyle\sum_{i=1}^{n+1} R_i\bar{x}_i^k + r &= \bar{x}_{n+1}^k - \textstyle\sum_{i=1}^{n} R_i\bar{x}_i^k + r \\
&= x_{n+1}^k - \textstyle\sum_{i=1}^{n} R_i x_i^k + r + \bar{x}_{n+1}^k - x_{n+1}^k + \textstyle\sum_{i=1}^{n} R_i(x_i^k - \bar{x}_i^k) \\
&= \beta(u^k - \bar{u}^k) + \bar{x}_{n+1}^k - x_{n+1}^k + \textstyle\sum_{i=1}^{n} R_i(x_i^k - \bar{x}_i^k).
\end{aligned}
$$

From Lemma 2, $\beta := \sum_{i=1}^{n+1}\beta_i$ and $\hat{\beta}^{-1} := \sum_{i=1}^{n+1}\hat{\beta}_i$, we have

$$
\begin{aligned}
\phi_k &= \textstyle\sum_{i=1}^{n+1}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \beta\|u^k - \bar{u}^k\|^2 + \hat{\beta}^{-1}\|v^k - \bar{v}^k\|^2 + \sum_{i=1}^{n+1}\langle x_i^k - \bar{x}_i^k, R_i^*(u^k - \bar{u}^k) - Q_i^*(v^k - \bar{v}^k)\rangle \\
&= \textstyle\sum_{i=1}^{n+1}(\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \beta_i\|u^k - \bar{u}^k\|^2 + \hat{\beta}_i\|v^k - \bar{v}^k\|^2 + \langle x_i^k - \bar{x}_i^k, R_i^*(u^k - \bar{u}^k) - Q_i^*(v^k - \bar{v}^k)\rangle) \\
&= \textstyle\sum_{i=1}^{n+1}\hat{\beta}_i(\hat{\beta}_i^{-1}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \hat{\beta}_i^{-1}\beta_i\|\bar{u}^k - u^k\|^2 + \|v^k - \bar{v}^k\|^2 - \langle x_i^k - \bar{x}_i^k, \hat{\beta}_i^{-1}R_i^*(\bar{u}^k - u^k) + \hat{\beta}_i^{-1}Q_i^*(v^k - \bar{v}^k)\rangle) \\
&\ge \textstyle\sum_{i=1}^{n+1}\hat{\beta}_i\varphi(\hat{\beta}_i^{-1}\alpha_i, \hat{\beta}_i^{-1}\beta_i, \hat{\beta}_i^{-1}R_i^*, \hat{\beta}_i^{-1}Q_i^*)\left(\|x_i^k - \bar{x}_i^k\|^2 + \|u^k - \bar{u}^k\|^2 + \beta_i^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2\right) \\
&= \textstyle\sum_{i=1}^{n+1}\varphi(\alpha_i, \beta_i, R_i^*, Q_i^*)\left(\|x_i^k - \bar{x}_i^k\|^2 + \|u^k - \bar{u}^k\|^2 + \beta_i^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2\right),
\end{aligned}
$$

and the conditions (5.11) and (5.12) indicate

$$
4\alpha_i\beta_i > \|R_i^*\|^2 + \hat{\beta}_i^{-1}\beta_i\|Q_i^*\|^2 > \|(R_i^*, \sqrt{\hat{\beta}_i^{-1}\beta_i}Q_i^*)\|^2,\ i = 1, \ldots, n+1.
$$

So each

$$
\frac{1}{2}\left(\alpha_i + \beta_i - \sqrt{(\alpha_i - \beta_i)^2 + \|(R_i^*, \sqrt{\hat{\beta}_i^{-1}\beta_i}Q_i^*)\|^2}\right)
$$

must be positive. Let ρ be their minimum; thus, we further obtain

$$
\phi_k \ge \rho\left(\textstyle\sum_{i=1}^{n+1}\|x_i^k - \bar{x}_i^k\|^2 + (n+1)\|u^k - \bar{u}^k\|^2 + \sum_{i=1}^{n+1}\beta_i^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2\right). \qquad (5.18)
$$

On the other hand,

$$
\begin{aligned}
\psi_k &= \textstyle\sum_{i=1}^{n+1}\|\alpha_i(x_i^k - \bar{x}_i^k) - Q_i^*(v^k - \bar{v}^k)\|^2 + \| - \sum_{i=1}^{n+1}R_i\bar{x}_i^k + r\|^2 + \|\hat{\beta}^{-1}(v^k - \bar{v}^k)\|^2 \\
&= \textstyle\sum_{i=1}^{n+1}\|\alpha_i(x_i^k - \bar{x}_i^k) - \sqrt{\hat{\beta}_i^{-1}\beta_i}Q_i^*(\sqrt{\beta_i^{-1}\hat{\beta}_i}(v^k - \bar{v}^k))\|^2 \\
&\quad + \|\textstyle\sum_{i=1}^{n+1}\beta_i(u^k - \bar{u}^k) + \sum_{i=1}^{n+1}R_i(x_i^k - \bar{x}_i^k)\|^2 + \|\sum_{i=1}^{n+1}\sqrt{\beta_i\hat{\beta}_i^{-3}}(\sqrt{\beta_i^{-1}\hat{\beta}_i}(v^k - \bar{v}^k))\|^2 \\
&\le \textstyle\sum_{i=1}^{n+1}(\alpha_i^2 + \hat{\beta}_i^{-1}\beta_i\|Q_i\|^2)(\sum_{i=1}^{n+1}\|x_i^k - \bar{x}_i^k\|^2 + \sum_{i=1}^{n+1}\beta_i^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2) \\
&\quad + \textstyle\sum_{i=1}^{n+1}(\beta_i^2 + \|R_i\|^2)(\sum_{i=1}^{n+1}\|x_i^k - \bar{x}_i^k\|^2 + (n+1)\|u^k - \bar{u}^k\|^2) + \sum_{i=1}^{n+1}\beta_i\hat{\beta}_i^{-3}\sum_{i=1}^{n+1}\beta_i^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2 \\
&\le \textstyle\sum_{i=1}^{n+1}(\alpha_i^2 + \beta_i^2 + \beta_i\hat{\beta}_i^{-3} + \|R_i\|^2 + \hat{\beta}_i^{-1}\beta_i\|Q_i\|^2)\left(\sum_{i=1}^{n+1}\|x_i^k - \bar{x}_i^k\|^2 + (n+1)\|u^k - \bar{u}^k\|^2 + \sum_{i=1}^{n+1}\beta_i^{-1}\hat{\beta}_i\|v^k - \bar{v}^k\|^2\right).
\end{aligned}
$$

Thus, we can conclude that

$$
\gamma_k = \theta_k\phi_k/\psi_k \ge (\theta\rho)/\textstyle\sum_{i=1}^{n+1}(\alpha_i^2 + \beta_i^2 + \beta_i\hat{\beta}_i^{-3} + \|R_i\|^2 + \hat{\beta}_i^{-1}\beta_i\|Q_i\|^2) > 0.
$$

Combining this with (5.17) and (5.18) yields the desired result. □

Theorem 4. *Let* $\{x_i^k\}_{i=1}^{n+1}$, $\{u^k\}$, *and* $\{v^k\}$ *be the sequences generated by Algorithm 2. If Assumption 2 and conditions* (5.11) *and* (5.12) *hold, then the corresponding primal sequence* $\{x_i^k\}_{i=1}^{n}$ *weakly converges to a solution of the system of monotone inclusions (1.1).*

Proof. It follows from (5.13) that

$$(i)\ x_i^k - \bar{x}_i^k \to 0,\ u^k - \bar{u}^k \to 0,\ v^k - \bar{v}^k \to 0,\ i = 1, \ldots, n+1; \tag{5.19}$$

$$(ii)\ \{x_i^k\}_{i=1}^{n+1},\ \{u^k\},\ \{v^k\} \text{ are bounded in norm.} \tag{5.20}$$

Then, according to the definition of T, we can obtain

$$T(\bar{x}^k, \alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k, \bar{v}^k) = \begin{pmatrix} \vdots \\ \bar{A}_i(\bar{x}_i^k) + R_i^*(\alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k) + Q_i^*\bar{v}^k \\ \vdots \\ A^{-1}(\alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k) - \sum_{i=1}^n R_i\bar{x}_i^k + r \\ B^{-1}(\bar{v}^k) - \sum_{i=1}^n Q_i\bar{x}_i^k + q \end{pmatrix}.$$

Utilizing (5.7)–(5.9), we further have

$$\begin{aligned} &\bar{A}_i(\bar{x}_i^k) + R_i^*(\alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k) + Q_i^*\bar{v}^k \\ \ni\ &\alpha_i(x_i^k - \bar{x}_i^k) - R_i^*\bar{u}^k - Q_i^*v^k + \alpha_{n+1}R_i^*(x_{n+1}^k - \bar{x}_{n+1}^k) + R_i^*\bar{u}^k + Q_i^*\bar{v}^k \\ =\ &\alpha_i(x_i^k - \bar{x}_i^k) + \alpha_{n+1}R_i^*(x_{n+1}^k - \bar{x}_{n+1}^k) - Q_i^*(v^k - \bar{v}^k), \end{aligned}$$

$$\begin{aligned} &A^{-1}(\alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k) - \textstyle\sum_{i=1}^n R_i\bar{x}_i^k + r \\ \ni\ &\beta(u^k - \bar{u}^k) + \bar{x}_{n+1}^k - x_{n+1}^k + \textstyle\sum_{i=1}^n R_i(x_i^k - \bar{x}_i^k), \end{aligned}$$

$$\begin{aligned} &B^{-1}(\bar{v}^k) - \textstyle\sum_{i=1}^n Q_i\bar{x}_i^k + q \\ \ni\ &\hat{\beta}^{-1}(v^k - \bar{v}^k) + \textstyle\sum_{i=1}^n Q_i\bar{x}_i^k - q - \sum_{i=1}^n Q_i\bar{x}_i^k + q \\ =\ &\hat{\beta}^{-1}(v^k - \bar{v}^k). \end{aligned}$$

Therefore,

$$T(\bar{x}^k, \alpha_{n+1}(x_{n+1}^k - \bar{x}_{n+1}^k) + \bar{u}^k, \bar{v}^k) \ni \begin{pmatrix} \vdots \\ \alpha_i(x_i^k - \bar{x}_i^k) + \alpha_{n+1}R_i^*(x_{n+1}^k - \bar{x}_{n+1}^k) - Q_i^*(v^k - \bar{v}^k) \\ \vdots \\ \beta(u^k - \bar{u}^k) + \bar{x}_{n+1}^k - x_{n+1}^k + \sum_{i=1}^n R_i(x_i^k - \bar{x}_i^k) \\ \hat{\beta}^{-1}(v^k - \bar{v}^k) \end{pmatrix}. \tag{5.21}$$

Due to (5.19) and the boundedness of each R_i and Q_i, we can know that the right-hand term of (5.21) strongly converges to zero. Meanwhile, from (5.20), there exists at least one weak cluster point $(x^\infty,\ u^\infty,\ v^\infty)$ such that

$$x^{k_j} \rightharpoonup x^\infty,\ u^{k_j} \rightharpoonup u^\infty,\ v^{k_j} \rightharpoonup v^\infty,$$

which, together with (5.19), implies

$$\bar{x}^{k_j} \rightharpoonup x^\infty,\ \bar{u}^{k_j} \rightharpoonup u^\infty,\ \bar{v}^{k_j} \rightharpoonup v^\infty.$$

Finally, invoking [15, Lemma 3] or [24, Lemma 3.2], we can conclude that this weak cluster point is a solution point of $0 \in T(x, u, v)$, thereby also solving the problem (1.1). The proof of uniqueness of weak cluster point is standard [12, 25] and thus is omitted. □

6. The dual-first version of Algorithm 2

In this section, we introduce the dual-first version of Algorithm 2, providing a simple overview of its implementation.

Algorithm 3. *Step 0. Choose* $x_i^0 \in \mathcal{H}_i$, $i = 1, \ldots, n$, $x_{n+1}^0 \in \mathcal{G}_1$, $u^0 \in \mathcal{G}_1$, $v^0 \in \mathcal{G}_2$. *For* $i = 1, \ldots, n+1$, *choose* $\beta_i > 0$, $\hat{\beta}_i > 0$ *and* $0 < \theta \leq \bar{\theta} < 2$. *Set* $k := 0$.

Step 1. Choose α_i *satisfying*

$$\alpha_i > \|R_i\|^2/(4\beta_i) + \|Q_i\|^2/(4\hat{\beta}_i),\ i = 1, \ldots, n+1.$$

Calculate $\beta = \sum_{i=1}^{n+1}\beta_i$, $\hat{\beta} = (\sum_{i=1}^{n+1}\hat{\beta}_i)^{-1}$. *For* $x_i^k \in \mathcal{H}_i$, $i = 1, \ldots, n$, $x_{n+1}^k \in \mathcal{G}_1$, $u^k \in \mathcal{G}_1$, $v^k \in \mathcal{G}_2$, *compute*

$$\begin{aligned}
&\bar{u}^k = u^k - (x_{n+1}^k - \textstyle\sum_{i=1}^n R_i x_i^k + r)/\beta,\\
&(I + \hat{\beta}B^{-1})(\bar{v}^k) \ni v^k + \hat{\beta}(\textstyle\sum_{i=1}^n Q_i x_i^k - q),\\
&(\alpha_i I + \bar{A}_i)(\bar{x}_i^k) \ni \alpha_i x_i^k - R_i^* \bar{u}^k - Q_i^* \bar{v}^k,\\
&(\alpha_{n+1} I + A)(\bar{x}_{n+1}^k) \ni \alpha_{n+1} x_{n+1}^k + \bar{u}^k.
\end{aligned}$$

If a prescribed stopping criterion is satisfied, the algorithm terminates. Otherwise, choose $\theta_k \in [\theta, \bar{\theta}]$, *compute*

$$\begin{aligned}
&\phi_k := \textstyle\sum_{i=1}^{n+1}\alpha_i\|x_i^k - \bar{x}_i^k\|^2 + \hat{\beta}^{-1}\|v^k - \bar{v}^k\|^2 + \langle \bar{x}_{n+1}^k - \sum_{i=1}^n R_i\bar{x}_i^k + r, u^k - \bar{u}^k\rangle - \sum_{i=1}^n\langle Q_i(x_i^k - \bar{x}_i^k), v^k - \bar{v}^k\rangle,\\
&\psi_k := \textstyle\sum_{i=1}^{n+1}\|\alpha_i(x_i^k - \bar{x}_i^k) - Q_i^*(v^k - \bar{v}^k)\|^2 + \|\bar{x}_{n+1}^k - \sum_{i=1}^n R_i\bar{x}_i^k + r\|^2 + \|\hat{\beta}^{-1}(v^k - \bar{v}^k)\|^2,\\
&\gamma_k := \theta_k\phi_k/\psi_k.
\end{aligned}$$

Step 2. For $i = 1, \ldots, n+1$, *compute in order*

$$\begin{aligned}
x_i^{k+1} &= x_i^k - \gamma_k\left(\alpha_i(x_i^k - \bar{x}_i^k) - Q_i^*(v^k - \bar{v}^k)\right),\\
u^{k+1} &= u^k - \gamma_k(\bar{x}_{n+1}^k - \textstyle\sum_{i=1}^n R_i\bar{x}_i^k + r),\\
v^{k+1} &= v^k - \gamma_k\hat{\beta}^{-1}(v^k - \bar{v}^k).
\end{aligned}$$

Set $k := k+1$.

7. Numerical experiments

In this section, we performed numerical experiments via Python 3.9.2 to verify the practical usefulness of Algorithms 1–3 in solving large-scale rare feature selection in deep learning, compared with other state-of-the-art splitting algorithms, selected for their similarities in features, applicability, and implementation effort.

Vũ Splitting: The splitting method is due to Vũ [1] and also see [15, Algorithm 6].

JE Splitting: The splitting method of [9, Algorithm 1], originally proposed by Johnstone and Eckstein, is well suited for solving (1.2), and also see [15, Algorithm 7].

JE 2021: The splitting method of [9, Algorithm 1], originally proposed by Johnstone and Eckstein to solve the first test problem below.

The test problems we conducted were about rare feature selection [9, 10, 20]. In machine learning and data mining, datasets often contain a large number of features, but not all features can effectively assist prediction models. In fact, key information often only focuses on a few features. Therefore, feature selection techniques are particularly important as they can help to accurately identify valuable features for the model and eliminate irrelevant or redundant features. In this process, rare feature selection becomes a key step, aiming to screen out those features that appear less frequently but have significant value for model prediction or classification from numerous features. These rare features may be particularly sparse due to the sparsity of the dataset, but they may play a crucial role in the model construction process. Therefore, how to aggregate these features and extract them from numerous features is a challenge. Yan et al. [20] introduced a framework for aggregating rare features into denser features by making use of an auxiliary tree data structure and proposed a generalized regression model for the rare feature selection. Johnstone and Eckstein [9] improved the model to make it more suitable for splitting methods, and effectively solved the problem with their proposed algorithm.

Our first test problem comes from Johnstone and Eckstein [9]*, which can be stated as follows:

$$\min_{\beta_0,\gamma} \ \|\beta_0 e + XH\gamma - y\|_2^2/(2n) + \lambda\mu\|\gamma_{-r}\|_1 + \lambda(1-\mu)\|H\gamma\|_1, \tag{7.1}$$

where X is the n-by-d data matrix, H is the d-by-r coefficient matrix, $y \in \mathcal{R}^n$ is the target vector, $e \in \mathcal{R}^n$ is the all-ones vector, $\beta_0 \in \mathcal{R}$ is an offset, and $\gamma \in \mathcal{R}^r$. In [9, 20], the authors gave a detailed description of the relationship of $H,\ \gamma$ (see [9, Section 6.3] [20, Section 3] for more details). The ℓ_1 norm on γ enforces sparsity of γ , which in turn fuses together coefficients associated with similar features. The ℓ_1 norm on $H\gamma$ additionally enforces sparsity on these coefficients, which is also desirable.

As described in [9,20], we applied this model on the TripAdvisor hotel-review dataset. The response variable y was the overall rating of the hotel, in the set $\{1, 2, 3, 4, 5\}$. The features were the counts of certain adjectives in the review. Many adjectives were very rare, with 95% of the adjectives appearing in less than 5% of the reviews. There were 7573 adjectives from 169987 reviews, and the auxiliary similarity tree $\mathcal{T}$ had 15145 nodes. The 169987×7573 design matrix X and the 7573×15145 matrix H arising from the similarity tree $\mathcal{T}$ were both sparse, having 0.32% and 0.15% nonzero entries, respectively.

Obviously, if we let

$$d_1 = d + 1, \quad x = \begin{pmatrix} \beta_0 \\ \gamma \end{pmatrix} \in \mathcal{R}^{r+1}, \quad \mathbb{X} = (e, X) \in \mathcal{R}^{n\times d_1},$$

$$\mathbb{H} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0}^T & H \end{pmatrix} \in \mathcal{R}^{d_1\times(r+1)}, \quad \mathbb{M} = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0}^T & H \end{pmatrix} \in \mathcal{R}^{d_1\times(r+1)},$$

then the problem can be transformed into

$$\min_{x\in\mathcal{R}^{r+1}} \ \Phi(x) := \|\mathbb{X}\mathbb{H}x - y\|_2^2/(2n) + \lambda\mu(|x_2| + \ldots + |x_r|) + \lambda(1-\mu)\|\mathbb{M}x\|_1. \tag{7.2}$$

The corresponding optimality condition is

$$0 \in \lambda\mu\mathbb{J}x + (\mathbb{X}\mathbb{H})^T\left(\mathbb{X}\mathbb{H}x - y\right)/n + \lambda(1-\mu)\mathbb{M}^T\partial\|\cdot\|_1(\mathbb{M}x), \tag{7.3}$$

*TripAdvisor data are available at https://github.com/yanxht/TripAdvisorData.

where $\mathbb{J} = \text{diag}(0, \partial|\cdot|, \ldots, \partial|\cdot|, 0)$. This problem may be viewed as a special case of (1.2) with

$$\bar{A} = \lambda\mu\mathbb{J}, \quad A = (1/n)I, \quad B = \lambda(1-\mu)\partial\|\cdot\|_1,$$

$$R = \mathbb{X}\mathbb{H}, \quad r = y, \quad Q = \mathbb{M}, \quad q = 0.$$

Below we followed [9] to choose $n = 169987$, $d_1 = 7574$, and $r = 15145$. To verify the effectiveness of the methods, we fixed $\mu = 0.5$ and selected multiple values for $\lambda \in \{10^{-5}, 10^{-4}, 10^{-3}\}$.

For Algorithm 1 (in the $n = 1$ case), through trial and error, we tried $\alpha_1 = \alpha_2 = \alpha_3 = 0.01,\ 0.1,\ 1,\ 10,\ 100$, and so on, ultimately choosing

$$\theta_k \equiv 0.9, \quad \alpha_1 = 10, \quad \alpha_2 = 10, \quad \alpha_3 = 10,$$

$$\beta_1 = (1+10^{-9})(\|R\|^2 + \|Q\|^2)/(4\alpha_1), \quad \beta_2 = (1+10^{-9})/(4\alpha_2), \quad \beta = \beta_1 + \beta_2,$$
$$\hat{\beta}_1 = \beta_1, \quad \hat{\beta}_3 = (1+10^{-9})/(4\alpha_3), \quad \hat{\beta} = \hat{\beta}_1 + \hat{\beta}_3,$$

which satisfy (4.1) and (4.2) as required in Section 4. And for the starting points, we followed [9] to choose

$$x^0 = zeros(r+1, 1), \quad x_2^0 = zeros(n, 1), \quad x_3^0 = zeros(d_1, 1),$$
$$u^0 = zeros(n, 1), \quad v^0 = zeros(d_1, 1).$$

In the practical implementations of Algorithm 1, we set

$$\tilde{x}^k := x^k - (R^*\bar{u}^k + Q^*\bar{v}^k)/\alpha,$$

and got

$$(\alpha I + \bar{A})(\bar{x}^k) \ni \alpha\tilde{x}^k \quad \Rightarrow \quad \bar{x}^k = (I + \alpha^{-1}\bar{A})^{-1}(\tilde{x}^k).$$

Thus, we further obtained

$$\bar{x}^k = \begin{cases} (I + \alpha^{-1}\lambda\mu\partial|\cdot|)^{-1}(\tilde{x}[i]^k) = \text{sgn}(\tilde{x}[i]^k)\max\left\{|\tilde{x}[i]^k| - \alpha^{-1}\lambda\mu, 0\right\}, & i = 2, \ldots, n-1, \\ \tilde{x}[i]^k, & i = 1, n, \end{cases}$$

where the term on the right-hand side is the so-called soft shrinkage function, $\tilde{x}[i]$ represents the i-th component of $\tilde{x}$.

For Algorithms 2 and 3 (in the $n = 1$ case), by trial and error similar to Algorithm 1, we chose

$$\theta_k \equiv 0.9, \quad \alpha_1 = 10, \quad \alpha_2 = 10, \quad \beta_1 = (1+10^{-9})(\|R\|^2 + \|Q\|^2)/(4\alpha_1),$$

$$\beta_2 = (1+10^{-9})/(4\alpha_2), \quad \beta = \beta_1 + \beta_2, \quad \hat{\beta}_1 = \beta_1, \quad \hat{\beta} = 1/\hat{\beta}_1,$$

which satisfy the corresponding inequalities as required in Section 5. And for the starting points, we also chose

$$x^0 = zeros(r+1, 1), \quad x_2^0 = zeros(n, 1), \quad u^0 = zeros(n, 1), \quad v^0 = zeros(d_1, 1).$$

In the practical implementations of Algorithms 2 and 3, we set

$$\tilde{v}^k := v^k + \hat{\beta}(Q\bar{x}^k - q).$$

By utilizing Moreau identity, the resolvent of the operator B^{-1} can be computed as follows:

$$\bar{v}^k = (I + \hat{\beta}B^{-1})^{-1}(\tilde{v}^k) \equiv \tilde{v}^k - \hat{\beta}(I + \hat{\beta}^{-1}B)^{-1}(\hat{\beta}^{-1}\tilde{v}^k), \quad \hat{\beta} > 0.$$

For Vũ splitting, we chose $\tau = 2/n,\ \sigma_1 = 2/n,\ \sigma_2 = \sigma_1$ to satisfy the corresponding inequality as required in [1], and

$$A = \lambda\mu\mathbb{J}, \quad B_1 = (2/n)I, \quad B_2 = 2\lambda(1-\mu)\partial\|\cdot\|_1,$$

$$L_1 = \mathbb{XH}, \quad r_1 = y, \quad L_2 = \mathbb{M}, \quad r_2 = 0, \quad \lambda_k \equiv 1.$$

For the starting points, we chose

$$x^0 = zeros(r+1, 1), \quad v_1^0 = zeros(n, 1), \quad v_2^0 = zeros(d_1, 1).$$

For JE splitting, we chose

$$A_2 = \lambda(1-\mu)\partial\|\cdot\|_1, \quad A_3 = \lambda\mu\mathbb{J}, \quad G_1 = \mathbb{XH}, \quad G_2 = \mathbb{M}.$$

As to A_1, we explained its resolvent's evaluations in some details. Consider the problem of minimizing $f_1(G_1 z)$, where $f_1(\cdot) := \|\cdot - y\|^2/(2n)$ and $\nabla f_1 = A_1$. Be aware that, in [15, Algorithm 7], for the corresponding subproblem

$$(I + \rho_1 A_1)x_1 \ni t_1^k,$$

its solution x_1^k is equivalent to that of

$$\min\ \rho_1 f_1(x_1) + \|x_1 - t_1^k\|^2/2 \quad \Rightarrow \quad \min\ \|x_1 - y\|^2/(2n) + \|x_1 - t_1^k\|^2/(2\rho_1).$$

For the parameters, we chose

$$\rho_1 = \|G_1\|^2/n, \quad \rho_2 = \lambda(1-\mu), \quad \rho_3 = \lambda\mu,$$

$$\alpha_1 = 0.5, \quad \alpha_2 = 0.5, \quad \alpha_3 = 0.5, \quad \gamma = 1.0,$$

and for the starting points, we chose

$$z^1 = zeros(r+1, 1), \quad w_1^1 = zeros(n, 1), \quad w_2^1 = zeros(d_1, 1), \quad w_3^1 = zeros(r+1, 1),$$

$$x_1^0 = zeros(n, 1), \quad x_2^0 = zeros(d_1, 1), \quad x_3^0 = zeros(r+1, 1).$$

For JE 2021, we followed [9] to choose parameters.

Numerical results on the test problem were given in Figures 1–3, where $\epsilon_k = \lg(\|\mathbb{XH}x^k - y\|_1)$.

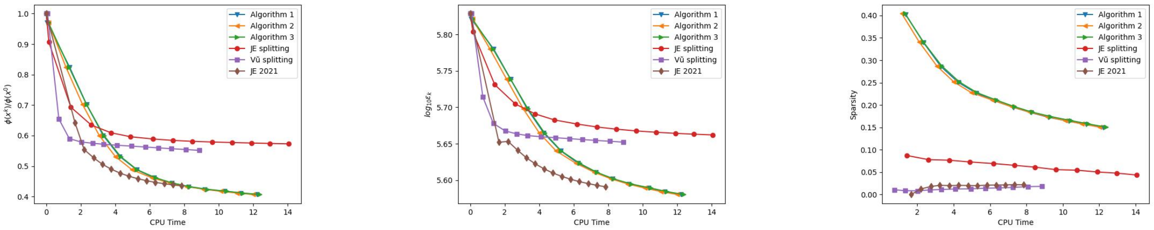

Figure 1. $\lambda = 10^{-5},\ \mu = 0.5$.

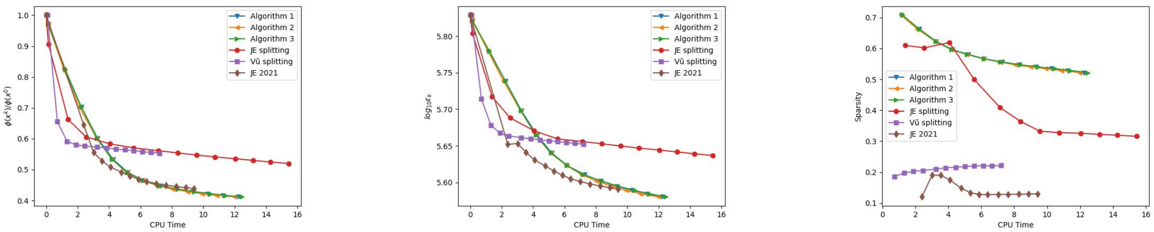

Figure 2. $\lambda = 10^{-4},\ \mu = 0.5$.

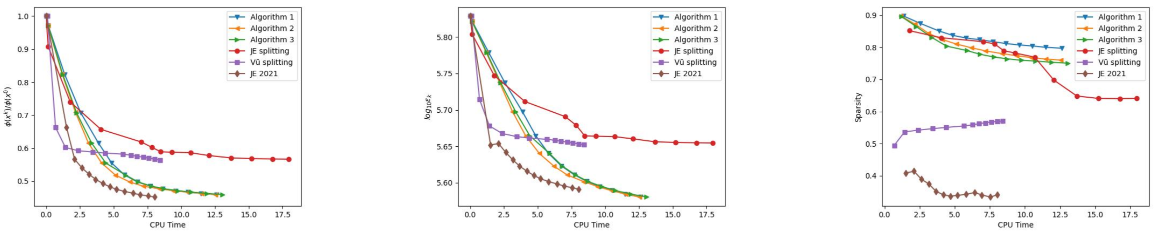

Figure 3. $\lambda = 10^{-3},\ \mu = 0.5$.

Figures 1–3 showed that, in each case, Algorithm 1 desirably achieved the highest sparsity among these algorithms and comparable accuracy to JE 2021. Unlike JE Splitting, JE 2021 was well suited for the case of $\bar{A}_i$ being further assumed to be Lipschitz continuous instead of (1.1) under consideration.

Our second test problem is based on the first test problem, replacing the least squares term with the ℓ_1 norm term:

$$\min_{x\in\mathcal{R}^{r+1}} \ \Phi(x) := \|\mathbb{XH}x - y\|_1/(n) + \lambda\mu(|x_2| + \ldots + |x_r|) + \lambda(1-\mu)\|\mathbb{M}x\|_1, \tag{7.4}$$

and the corresponding optimality condition is

$$0 \in \lambda\mu\mathbb{J}x + (\mathbb{XH})^T\partial\|\cdot\|_1\left(\mathbb{XH}x - y\right)/n + \lambda(1-\mu)\mathbb{M}^T\partial\|\cdot\|_1(\mathbb{M}x), \tag{7.5}$$

where $\mathbb{J} = \mathrm{diag}(0, \partial\,|\cdot|\,, \ldots, \partial\,|\cdot|\,, 0)$. We chose the same $n,\ d_1,\ r,\ \mu$, and λ as the first test problem. This problem may be viewed as a special case of (1.2) with

$$\bar{A} = \lambda\mu\mathbb{J}, \quad A = (1/n)\partial\|\cdot\|_1, \quad B = \lambda(1-\mu)\partial\|\cdot\|_1,$$

$$R = \mathbb{XH}, \quad r = y, \quad Q = \mathbb{M}, \quad q = 0.$$

For Algorithm 1 (in the $n = 1$ case), through trial and error, we first tried $\alpha_1 = \alpha_2 = \alpha_3 = 0.01,\ 0.1,\ 1,\ 10$, and 100 and then ascertained it to lie in the neighborhood of 1, ultimately choosing,

$$\theta_k \equiv 0.9, \quad \alpha_1 = 1.5, \quad \alpha_2 = 1.5, \quad \alpha_3 = 1.5,$$

$$\beta_1 = (1+10^{-9})(\|R\|^2+\|Q\|^2)/(4\alpha_1), \quad \beta_2 = (1+10^{-9})/(4\alpha_2), \quad \beta = \beta_1+\beta_2,$$
$$\hat\beta_1 = \beta_1, \quad \hat\beta_3 = (1+10^{-9})/(4\alpha_3), \quad \hat\beta = \hat\beta_1+\hat\beta_3,$$

which satisfy (4.1) and (4.2) as required in Section 4. And for the starting points, we followed [9] to choose

$$x^0 = zeros(r+1,1), \quad x_2^0 = zeros(n,1), \quad x_3^0 = zeros(d_1,1),$$
$$u^0 = zeros(n,1), \quad v^0 = zeros(d_1,1).$$

For Algorithms 2 and 3 (in the $n = 1$ case), through trial and error, similar to the process in Algorithm 1, we chose

$$\theta_k \equiv 0.9, \quad \alpha_1 = 0.5, \quad \alpha_2 = 0.5, \quad \beta_1 = (1+10^{-9})(\|R\|^2+\|Q\|^2)/(4\alpha_1),$$

$$\beta_2 = (1+10^{-9})/(4\alpha_2), \quad \beta = \beta_1+\beta_2, \quad \hat\beta_1 = \beta_1, \quad \hat\beta = 1/\hat\beta_1,$$

which satisfy the corresponding inequalities as required in Section 5. And for the starting points, we also chose

$$x^0 = zeros(r+1,1), \quad x_2^0 = zeros(n,1), \quad u^0 = zeros(n,1), \quad v^0 = zeros(d_1,1).$$

For Vũ splitting, we chose $\lambda_k \equiv 1$, $\sigma_1 = 1/n$, $\sigma_2 = \sigma_1$ and

$$A = \lambda\mu\mathbb{J}, \quad B_1 = (2/n)\partial\|\cdot\|_1, \quad B_2 = 2\lambda(1-\mu)\partial\|\cdot\|_1,$$

$$L_1 = \mathbb{XH}, \quad r_1 = y, \quad L_2 = \mathbb{M}, \quad r_2 = 0, \quad \tau = 2/(\sigma_1\|L_1\|^2+\sigma_2\|L_2\|^2) - 10^{-9},$$

which satisfy the corresponding inequality as required in [1]. For the starting points, we chose

$$x^0 = zeros(r+1,1), \quad v_1^0 = zeros(n,1), \quad v_2^0 = zeros(d_1,1).$$

For JE splitting, we chose

$$A_2 = \lambda(1-\mu)\partial\|\cdot\|_1, \quad A_3 = \lambda\mu\mathbb{J}, \quad G_1 = \mathbb{XH}, \quad G_2 = \mathbb{M}.$$

As to A_1, we explained its resolvent's evaluations in some details. Consider the problem of minimizing $f_1(G_1z)$, where $f_1(\cdot) := \|\cdot - y\|_1/(n)$ and $\nabla f_1 = A_1$. Be aware that, in [15, Algorithm 7], for the corresponding subproblem

$$(I+\rho_1A_1)x_1 \ni t_1^k,$$

its solution x_1^k is equivalent to that of

$$\min\ \rho_1 f_1(x_1) + \|x_1 - t_1^k\|^2/2 \quad \Rightarrow \quad \min\ \|x_1 - y\|_1/(n) + \|x_1 - t_1^k\|^2/(2\rho_1).$$

For the parameters, we chose

$$\rho_1 = \|G_1\|^2/n, \quad \rho_2 = \lambda(1-\mu), \quad \rho_3 = \lambda\mu,$$

$$\alpha_1 = 0.5, \quad \alpha_2 = 0.5, \quad \alpha_3 = 0.5, \quad \gamma = 0.1,$$

and for the starting points, we chose

$$z^1 = zeros(r+1,1), \quad w_1^1 = zeros(n,1), \quad w_2^1 = zeros(d_1,1), \quad w_3^1 = zeros(r+1,1),$$

$$x_1^0 = zeros(n,1), \quad x_2^0 = zeros(d_1,1), \quad x_3^0 = zeros(r+1,1).$$

Numerical results on the test problem were given in Figures 4–6, where $\epsilon_k = \lg(\|\mathbb{XH}x^k - y\|_1)$.

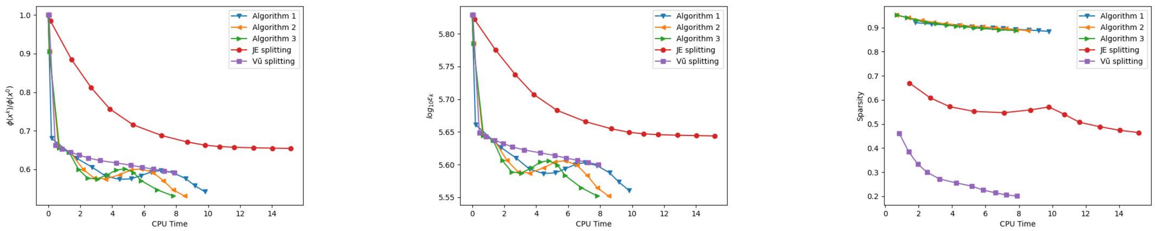

Figure 4. $\lambda = 10^{-5}$, $\mu = 0.5$.

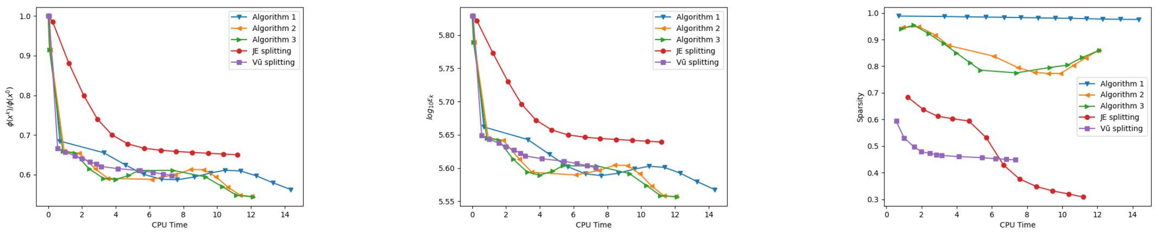

Figure 5. $\lambda = 10^{-4}$, $\mu = 0.5$.

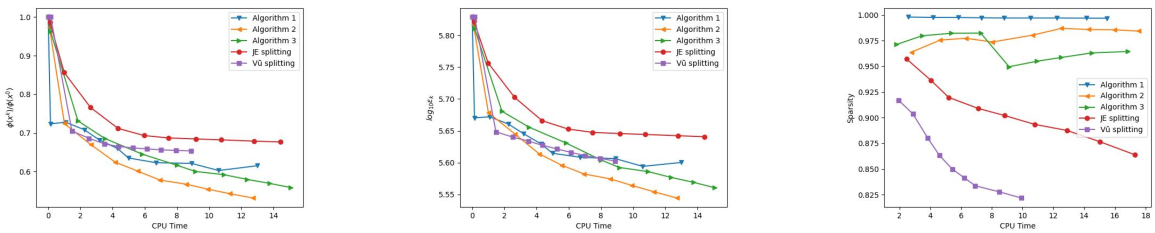

Figure 6. $\lambda = 10^{-3}$, $\mu = 0.5$.

From Figures 4–6, it can be observed that among these algorithms, Algorithms 1–3 demonstrated higher accuracy and sparsity. Since the initial condition $\Phi(x^0)$ was identical, the figures on the left-hand side revealed that Algorithms 1–3 achieved the highest accuracy within the same CPU time. The figures on the right-hand side indicated that, within the same CPU time, Algorithm 1 attained greater sparsity. Additionally, the figures in the middle showed that within the same CPU time, Algorithms 1–3 yielded smaller values of $\|\mathbb{XH}x^k - y\|_1$, which implied that the solutions obtained can better fit the model while ensuring sparsity.

8. Conclusions

In this article, we have introduced a new splitting method tailored for solving the system of three-operator monotone inclusions within Hilbert spaces, where the last two operators are linearly composed. Furthermore, by invoking a new inequality, we have analyzed weak convergence of this method under the weakest possible assumptions. To verify the practical effectiveness of our proposed splitting method, together with its variants, we have conducted rigorous numerical experiments, comparing their performance against other state-of-the-art methods in solving large-scale rare feature selection in deep learning. Finally, an interesting open question is whether or not it is possible to analyze the rate of convergence of this method, and we expect to explore it in the future.

Author contributions

Yunda Dong: Writing - review and editing, validation, supervision, methodology, conceptualization; Yiyi Li: Writing - original draft, validation, software, methodology, data curation. All authors have read and approved the final version of the manuscript for publication.

Acknowledgments

The authors are greatly indebted to the handling editor and the referees for their encouraging words and insightful suggestions, which improve the quality of this article. Special thanks go to Qiqi Luo and Yue Zhu for careful reading of the current version of this manuscript.

Conflict of interest

The authors declare no conflict of interest.

References

1. B. Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators, *Adv. Comput. Math.,* **38** (2013), 667–681. https://doi.org/10.1007/s10444-011-9254-8
2. D. Hieu, L. Vy, P. Quy, Three-operator splitting algorithm for a class of variational inclusion problems, *Bull. Iran. Math. Soc.,* **46** (2020), 1055–1071. https://doi.org/10.1007/s41980-019-00312-5
3. O. Iyiola, C. Enyi, Y. Shehu, Reflected three-operator splitting method for monotone inclusion problem, *Optim. Method. Softw.,* **37** (2022), 1527–1565. https://doi.org/10.1080/10556788.2021.1924715
4. E. Chouzenoux, M. Corbineau, J. Pesquet, A proximal interior point algorithm with applications to image processing, *J. Math. Imaging Vis.,* **62** (2020), 919–940. https://doi.org/10.1007/s10851-019-00916-w
5. Y. Tang, M. Wen, T. Zeng, Preconditioned three-operator splitting algorithm with applications to image restoration, *J. Sci. Comput.,* **92** (2022). https://doi.org/10.1007/s10915-022-01958-w
6. A. Padcharoen, D. Kitkuan, W. Kumam, P. Kumam, Tseng methods with inertial for solving inclusion problems and application to image deblurring and image recovery problems, *Comput. Math. Method. M.,* **3** (2021), e1088. https://doi.org/10.1002/cmm4.1088

7. V. Nguyen, N. Vinh, Two new splitting methods for three-operator monotone inclusions in Hilbert spaces, *Set-Valued Var. Anal.,* **32** (2024), 26. https://doi.org/10.1007/s11228-024-00730-6

8. B. Tan, X. Qin, J. Yao, Strong convergence of self-adaptive inertial algorithms for solving split variational inclusion problems with applications, *J. Sci. Comput.,* **87** (2021), 20. https://doi.org/10.1007/s10915-021-01428-9

9. P. Johnstone, J. Eckstein, Single-forward-step projective splitting: Exploiting cocoercivity, *Comput. Optim. Appl.,* **78** (2021), 125–166. https://doi.org/10.1007/s10589-020-00238-3

10. P. Johnstone, J. Eckstein, Projective splitting with forward steps, *Math. Program.,* **191** (2022), 631–670. https://doi.org/10.1007/s10107-020-01565-3

11. P. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.,* **16** (1979), 964–979. https://doi.org/10.1137/0716071

12. Y. Dong, A. Fischer, A family of operator splitting methods revisited, *Nonlinear Anal.,* **72** (2010), 4307–4315. https://doi.org/10.1016/j.na.2010.02.010

13. Y. Dong, Douglas-Rachford splitting method for semi-definite programming, *J. Appl. Math. Comput.,* **51** (2016), 569–591. https://doi.org/10.1007/s12190-015-0920-8

14. H. He, D. Han, A distributed Douglas-Rachford splitting method for multi-block convex minimization problems, *Adv. Comput. Math.,* **42** (2016), 27–53. https://doi.org/10.1007/s10444-015-9408-1

15. Y. Dong, A new splitting method for systems of monotone inclusions in Hilbert spaces, *Math. Comput. Simulat.,* **203** (2023), 518–537. https://doi.org/10.1016/j.matcom.2022.06.023

16. Y. Dong, X. Zhu, An inertial splitting method for monotone inclusions of three operators, *Int. J. Math. Stat. Oper. Res.,* **2** (2022), 43–60. https://doi.org/10.47509/IJMSOR.2022.v02i01.04

17. J. Eckstein, A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers, *J. Optim. Theory Appl.,* **173** (2017), 155–182. https://doi.org/10.1007/s10957-017-1074-7

18. X. Zhu, *Inertial splitting methods for monotone inclusions of three operators (Thesis)*, Zheng Zhou University, 2020.

19. Y. Dong, Extended splitting methods for systems of three-operator monotone inclusions with continuous operators, *Math. Comput. Simulat.,* **223** (2024), 86–107. https://doi.org/10.1016/j.matcom.2024.03.024

20. X. Yan, J. Bien, Rare feature selection in high dimensions, *J. Am. Stat. Assoc.,* **116** (2020), 887–900. https://doi.org/10.1080/01621459.2020.1796677

21. G. Minty, Monotone (nonlinear) operators in Hilbert space, *Duke Math. J.,* **29** (1962), 341–346. https://doi.org/10.1215/S0012-7094-62-02933-2

22. J. Eckstein, B. Svaiter, A family of projective splitting methods for the sum of two maximal monotone operators, *Math. Program.,* **111** (2008), 173–199. https://doi.org/10.1007/s10107-006-0070-8

23. P. Latafat, P. Patrinos, Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators, *Comput. Optim. Appl.,* **68** (2017), 57–93. https://doi.org/10.1007/s10589-017-9909-6

24. P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, *SIAM J. Control Optim.,* **38** (2000), 431–446. https://doi.org/10.1137/S0363012998338806

25. R. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.,* **14** (1976), 877–898. https://doi.org/10.1137/0314056

26. Y. Dong, Weak convergence of an extended splitting method for monotone inclusions, *J. Global Optim.,* **79** (2021), 257–277. https://doi.org/10.1007/s10898-020-00940-w

27. Q. Dong, M. Su, Y. Shehu, Three-operator reflected forward-backward splitting algorithm with double inertial effects, *Optim. Method. Softw.,* **39** (2024), 431–456. https://doi.org/10.1080/10556788.2024.2307470

28. K. Bredies, E. Chenchene, D. Lorenz, E. Naldi, Degenerate preconditioned proximal point algorithms, *SIAM J. Optim.,* **32** (2022), 2376–2401. https://doi.org/10.1137/21M1448112