manuscript No.
(will be inserted by the editor)

A Three-Operator Splitting Scheme Derived from Three-Block ADMM

Anshika Anshika · Debdas Ghosh · Xiangxiong Zhang*

Abstract In this paper, we derive a three-operator splitting scheme for solving monotone inclusion and convex optimization problems from the three-block ADMM method on the dual problem. The proposed scheme can be regarded as an extension of the Douglas-Rachford splitting to more operators. We also show an extension to multi-block models whose objective function is the sum of three or more functions. A numerical comparison with the Davis-Yin three-operator splitting method demonstrates that the new three-operator splitting scheme can still converge with a much larger step size.

Keywords Operator-splitting · Three-block model · Convex optimization · Douglas-Rachford splitting · Davis-Yin splitting · ADMM methods · Monotone inclusion.

1 Introduction

Operator splitting schemes reduce complex problems into a series of smaller subproblems that can be solved in parallel or sequentially. Although these techniques were introduced over 60 years ago, their importance has significantly increased in the past decade. Since then, they have been successfully applied to problems in PDEs and control, as well as large-scale applications in machine learning, signal processing, and imaging.

Initially, in 1955 [25] and 1956 [10], Peaceman–Rachford and Douglas–Rachford splitting methods were originally introduced as splitting methods to solve the heat equation. After that, Lions and Mercier [20] extended this technique to a sum of two maximal monotone operators. Thereafter, Raguet, Fadili, and Peyré [26,27] integrated the results of Douglas–Rachford and Forward–Backward

* Corresponding author: Xiangxiong Zhang (zhan1966@purdue.edu)

Anshika Anshika (anshika.rs.mat19@itbhu.ac.in)
Department of Mathematical Sciences, Indian Institute of Technology (Banaras Hindu University) Varanasi, Uttar Pradesh 221005, India

Debdas Ghosh (debdas.mat@itbhu.ac.in)
Department of Mathematical Sciences, Indian Institute of Technology (Banaras Hindu University) Varanasi, Uttar Pradesh 221005, India

Xiangxiong Zhang (zhan1966@purdue.edu)
Department of Mathematics, Purdue University, 150 North University Street, West Lafayette, Indiana 47907, USA

splitting schemes followed by its extension in [4]. These results were finalized by [8] by demonstrating their averagedness in the general cases involving two maximal monotone operators and one coercive operator. This combined Davis-Yin splitting is also referred to as Forward Douglas-Rachford splitting. Recently, certain operator-splitting methods such as alternating direction method of multipliers (ADMM) [11, 12] and split Bregman [13] have found new applications in image processing, statistical and machine learning, compressive sensing, matrix completion, finance and control. These methods have also been extended to handle distributed and decentralized optimization [3, 28, 29]. In convex optimization, operator splitting methods split constraint sets and objective functions into subproblems that are easier to solve than the original problem.

Consider solving a composite convex minimization problem in the form

$$\min_{x} d_1(x) + d_2(x) + d_3(x),$$

where $d_i(x)$ are proper closed convex functions with computable proximal operators as prox_{d_i}. When one of the functions has Lipschitz-continuous gradient, e.g., assume ∇d_2 is Lipschitz continuous with Lipschitz constant L, the Davis-Yin splitting (or Forward Douglas-Rachford splitting) scheme can be written as

$$\text{Davis-Yin Splitting :} \quad \begin{aligned} x^{k+\frac{1}{2}} &= \mathrm{prox}_{d_3}^{\gamma}(z^k) \\ x^{k+1} &= \mathrm{prox}_{d_1}^{\gamma}(2x^{k+\frac{1}{2}} - z^k - \gamma \nabla d_2(x^{k+\frac{1}{2}})) \\ z^{k+1} &= z^k + (x^{k+1} - x^{k+\frac{1}{2}}). \end{aligned} \tag{1}$$

In this paper, we will consider the following different splitting scheme:

$$\text{Proposed Splitting :} \quad \begin{aligned} x^{k+\frac{1}{2}} &= \mathrm{prox}_{d_3}^{\gamma}(z^k) \\ p^{k+1} &= \mathrm{prox}_{d_1}^{\gamma}(2x^{k+\frac{1}{2}} - z^k - \gamma \nabla d_2(x^{k+\frac{1}{2}})) \\ x^{k+1} &= \mathrm{prox}_{d_2}^{\gamma}(p^{k+1} + \gamma \nabla d_2(x^{k+\frac{1}{2}})) \\ z^{k+1} &= z^k + (x^{k+1} - x^{k+\frac{1}{2}}) \end{aligned} \tag{2}$$

As the first glance, the proposed splitting scheme is inferior to the Davis-Yin splitting since the new scheme needs to compute three proximal operators, while the Davis-Yin splitting only needs to compute two proximal operators. On the other hand, the extra computation of $\mathrm{prox}_{d_2}^{\gamma}$ might improve the robustness of the splitting. In particular, the Davis-Yin splitting can be proven to converge for any constant step size any $\gamma \in (0, \frac{2}{L})$ [8]. Numerically, when the step size γ is much larger $\frac{2}{L}$, the Davis-Yin splitting will not converge, and the new splitting method can still converge, as will be shown in numerical examples in Section 6.

The proposed splitting scheme can be derived from the classical three-block ADMM methods on the dual problem. In the literature, there are several studies of ADMM methods that address the generic case where the number of blocks $m \geq 3$. In [14], the strong convexity assumption has been assumed on all the given objective functions. Lin in [17] considered $(m-1)$ functions to be strongly convex and establish global convergence without imposing any restrictions on the penalty parameter. Moreover, in [19], the linear convergence was established under the assumption that the objective functions are Lipschitz continuous. In a sequel, the work in [15] showed that linear convergence is ensured if the step size in each updating step is sufficiently reduced and a certain error-bound condition holds. Moreover, it has been discussed in [5, 6, 14, 16, 19] that it is important to appropriately restrict the penalty parameter. The importance of restricting the penalty parameter to ensure faster convergence has been discussed extensively in [5, 6, 14, 16, 19]. Although the restriction can be conservative to ensure convergence, it may be relaxed to achieve faster rates. Recently, Davis and Yin [7] examined a variant of three-block ADMM that guarantees convergence under the condition that one of the objective functions is strongly convex and the stepsize parameter is bounded by a threshold value. In [18], Lin et al. proposed several alternative methods for three-block ADMM without any restriction on penalty parameters to solve regularized least square decomposition problems.

Consider solving an inclusion problem $0 \in (\mathbb{A} + \mathbb{B} + \mathbb{C})x$, where $\mathbb{A}, \mathbb{B}, \mathbb{C}$ are three maximal monotone operators defined on a Hilbert space $\mathcal{X}$ and the operator $\mathbb{C}$ is cocoerceive with parameter β. Let $\mathbb{J}_A = (I_{\mathcal{X}} + A)^{-1}$ denote the resolvent of a monotone operator A. Then the Davis-Yin splitting operator can be written as

$$T_{DY} = \mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}) + (I - \mathbb{J}_{\gamma\mathbb{B}}),$$

and the proposed splitting operator is given as

$$T = \mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}) + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}) + (I - \mathbb{J}_{\gamma\mathbb{B}}).$$

The Davis-Yin splitting operator T_{DY} can be proven to an averaged operator for any $\gamma \in (0, 2\beta)$. It is however nontrivial to prove averagedness of the operator T. Thus in this paper, we only discuss some properties of the proposed splitting under the assumption that it is an averaged operator.

The rest of this paper is organized as follows. In Section 2, we describe the notation and symbols used in this paper. In Section 3, we introduce the proposed splitting operator. The convergence of the proposed splitting operator will be discussed in Section 4. In Section 5, we show that the proposed splitting is equivalent to the three-block ADMM method on the dual problem and we also give an extension of splitting scheme to multiple operators. In Section 6, a numerical example will be shown to demonstrate the robustness of the proposed splitting method. Concluding remarks are given in Section 7.

2 Basic Notation and Fundamental Results

The following notation will be used throughout the paper.

- $\mathcal{X}$ denotes an infinite dimensional Hilbert space
- $\langle , \rangle$ denotes inner product associated to $\mathcal{X}$
- $(\lambda_j)_{j\geq 0} \subseteq \mathbb{R}_{++}$ denotes a stepsize sequence

The following definitions and facts are mostly standard.

Let S be a nonempty subset of $\mathcal{X}$. Then, for $L \geq 0$, a map $F : S \to \mathcal{X}$ is called L-Lipschitz if

$$\|F(x) - F(y)\| \leq L\|x - y\|, \text{ for all } x, y \in S.$$

A map F is called nonexpansive if it is 1-Lipschitz.

Let $I_{\mathcal{X}}$ be the identity map. A map $F_\alpha : S \to \mathcal{X}$ is called α-averaged if it can be written as

$$F_\alpha = (1 - \alpha)I_{\mathcal{X}} + \alpha F,$$

where F is some nonexpansive map. Moreover, an $(1/2)$-averged map is called firmly nonexpansive.

Let $2^{\mathcal{X}}$ denotes the power set of $\mathcal{X}$. A set-valued operator $A : \mathcal{X} \to 2^{\mathcal{X}}$ is called monotone if

$$\langle x-y, u-v\rangle \geq 0, \text{ for all } x, y \in \mathcal{X},\ u \in Ax,\ v \in Ay.$$

The set of zeroes of a monotone operator is defined by

$$\text{zer}(A) = \{x \in \mathcal{X} \mid 0 \in Ax\}.$$

An operator A is called β-strongly monotone if for all $x, y \in \mathcal{X}$, $u \in Ax,\ v \in Ay$, we have

$$\langle x-y, u-v\rangle \geq \beta\|x-y\|^2,\ \beta > 0.$$

An operator A is called β-cocoerceive (or β-inverse strongly monotone) if for any $\beta > 0$, we have

$$\langle u-v, x-y\rangle \geq \beta\|u-v\|^2, \text{ for all } x, y \in \mathcal{X},\ u \in Ax,\ v \in Ay.$$

Additionally, A is β-cocoerceive, the Cauchy-Schwarz inequality implies that

$$\|x-y\| \geq \beta\|u-v\|, \text{ for all } x, y \in \mathcal{X},\ u \in Ax,\ v \in Ay,\ \beta > 0. \tag{3}$$

Moreover, if f is a convex function and its gradient ∇f is L-Lipschitz, then ∇f is $1/L$-cocoerceive.

The graph of an operator A is denoted and defined by

$$\text{gra}(A) = \{(x, y) |\ x \in \mathcal{X}, y \in Ax\}.$$

A monotone operator is called maximal monotone provided the gra(A) is not contained in the graph of any other monotone set-valued operator.

The inverse of an operator A is denoted by A^{-1}, and is defined uniquely by its graph as follows

$$\text{gra}(A^{-1}) = \{(y, x) | x \in \mathcal{X}, y \in Ax\}.$$

The resolvent of a monotone operator A is denoted and defined by

$$\mathbb{J}_A = (I_{\mathcal{X}} + A)^{-1}.$$

The reflection of a monotone operator A is denoted and defined by

$$\text{refl}_A = 2\mathbb{J}_A - I_{\mathcal{X}}.$$

If A is maximal monotone, then refl$_A$ is nonexpansive.

Let $f : \mathcal{X} \to (-\infty, \infty]$ be a closed, proper, and convex function. The subdifferential set of f at x is denoted by a map $\partial f : \mathcal{X} \to 2^{\mathcal{X}}$, and is defined by

$$\partial f(x) = \{g \in \mathcal{X} | f(y) \geq f(x) + \langle y - x, g, \text{ for all } y \in \mathcal{X}\rangle\}.$$

For convenience we assume that $\nabla f(x) \in \partial f(x)$ denotes a subgradient of f at x.
The convex (or Frenchel) conjugate of a proper, closed, and convex function f is given by

$$f^*(y) = \sup_{x\in\mathcal{X}} \langle y, x\rangle - f(x).$$

The indicator function of a closed convex set $C \subseteq \mathcal{X}$ defined by

$$i_C(x) = \begin{cases} 0 & x \in C \\ +\infty & x \in \mathcal{X}/C, \end{cases}$$

is a proper closed convex function. For any $x \in \mathcal{X}$ and $\lambda > 0$, the proximal operator of a proper closed convex function f is defined by

$$\text{prox}_{\lambda f}(x) = \arg\min_{y \in \mathcal{X}} \left(f(y) + \tfrac{1}{2\lambda}\|y-x\|^2 \right) \text{ and } \text{refl}_{\lambda f} = 2\text{prox}_{\lambda f} - I_{\mathcal{X}}.$$

In this article, we use the following *cosine rule* and *Young's Inequality* given by:

$$2\langle y-x, z-x\rangle = \|y-x\|^2 + \|z-x\|^2 - \|y-z\|^2 \text{ for all } x,y,z \in \mathcal{X} \tag{4}$$

$$\text{and } ab \leq \frac{a^2}{2\varepsilon} + \frac{b^2\varepsilon}{2} \text{ for all } a,b \geq 0,\ \varepsilon > 0,\ \text{respectively.} \tag{5}$$

3 A Three-Operator Splitting Scheme

In this section, we introduce a three-operator splitting scheme, which can be used to solve nonsmooth and monotone inclusion optimization problems of many different forms. We consider the problem

$$\text{find } x \in \mathcal{X} \text{ such that } 0 \in (\mathbb{A}+\mathbb{B}+\mathbb{C})x, \tag{6}$$

where $\mathbb{A}, \mathbb{B}, \mathbb{C}$ are three maximal monotone operators defined on a Hilbert space $\mathcal{X}$ and the operator $\mathbb{C}$ is cocoerceive with parameter β. Then, for any $\gamma > 0$, we have

$$\begin{aligned}
& 0 \in \gamma(\mathbb{A}+\mathbb{B}+\mathbb{C})x \\
\iff\ & 0 \in (I+\gamma\mathbb{A})x - (I-\gamma\mathbb{B})x + \gamma\mathbb{C}x \\
\iff\ & 0 \in (I+\gamma\mathbb{A})x - \mathbb{R}_{\gamma\mathbb{B}}(I+\gamma\mathbb{B})x + \gamma\mathbb{C}x,\ \text{ since } \mathbb{R}_{\gamma\mathbb{B}}(I+\gamma\mathbb{B}) = (I-\gamma\mathbb{B}) \\
\iff\ & 0 \in (I+\gamma\mathbb{A})x - \mathbb{R}_{\gamma\mathbb{B}}z + \gamma\mathbb{C}x,\ \text{ assume } z \in (I+\gamma\mathbb{B})x \\
\iff\ & \mathbb{R}_{\gamma\mathbb{B}}z - \gamma\mathbb{C}x \in (I+\gamma\mathbb{A})x, \\
\iff\ & 2\mathbb{J}_{\gamma\mathbb{B}}z - z - \gamma\mathbb{C}x \in (I+\gamma\mathbb{A})x,\ \text{ since } \mathbb{R}_{\gamma\mathbb{B}}z = (2\mathbb{J}_{\gamma\mathbb{B}} - I)z \\
\iff\ & \mathbb{J}_{\gamma\mathbb{A}}(2\mathbb{J}_{\gamma\mathbb{B}}z - z - \gamma\mathbb{C}x) = x,\ \text{ since } \mathbb{J}_{\gamma\mathbb{A}} = (I+\gamma\mathbb{A})^{-1} \\
\iff\ & \mathbb{J}_{\gamma\mathbb{A}}(2\mathbb{J}_{\gamma\mathbb{B}}z - z - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}z) = x, \\
\iff\ & \mathbb{J}_{\gamma\mathbb{A}}(2\mathbb{J}_{\gamma\mathbb{B}}z - z - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}z) + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}z = x + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}z \\
\iff\ & \mathbb{J}_{\gamma\mathbb{A}}(2\mathbb{J}_{\gamma\mathbb{B}}z - z - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}z) + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}z = (I+\gamma\mathbb{C})x \\
\iff\ & \mathbb{J}_{\gamma\mathbb{C}}\left(\mathbb{J}_{\gamma\mathbb{A}}(2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}) + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}\right)z = x,\ \text{ since } \mathbb{J}_{\gamma\mathbb{C}} = (I+\gamma\mathbb{C})^{-1} \\
\iff\ & \mathbb{J}_{\gamma\mathbb{C}}\left(\mathbb{J}_{\gamma\mathbb{A}}(2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}) + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}\right)z = \mathbb{J}_{\gamma\mathbb{B}}(z),\ \text{ assumption } z \in (I+\gamma\mathbb{B})x \\
\iff\ & \left(\mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}) + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}) + (I - \mathbb{J}_{\gamma\mathbb{B}})\right)z = z,
\end{aligned} \tag{7}$$

where $T = \mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}) + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}) + (I - \mathbb{J}_{\gamma\mathbb{B}})$.

The proposed operator T in (7) splits the three-operator sum problem given in (6) into simpler sub-problems. In fact, forward backward splitting (FBS) and Douglas-Rachford splitting (DRS) are special cases of Algorithm discussed in 1. The proposed operator T can be reduced to the two-operator sum problem given by

$$\text{Find } x \in \mathcal{X} \text{ such that } 0 \in \mathbb{A}x + \mathbb{B}x.$$

The two special cases for the proposed operator T are as follows:

(i) If $\mathbb{A} = 0$, (7) reduces to

$$T = \mathbb{J}_{\gamma\mathbb{C}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I) + (I - \mathbb{J}_{\gamma\mathbb{B}}), \text{ which is DRS [20].}$$

(ii) If $\mathbb{C} = 0$, (7) reduces to

$$T = \mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I) + (I - \mathbb{J}_{\gamma\mathbb{B}}), \text{ which is DRS [20].}$$

(iii) If $\mathbb{B} = 0$, (7) reduces to

$$\mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (I - \gamma\mathbb{C}) + \gamma\mathbb{C})z = z.$$

The splitting above is not FBS [20, 24] but it is closely related to FBS since

$$\begin{aligned} & \mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (I - \gamma\mathbb{C}) + \gamma\mathbb{C})z = z \\ \iff & \mathbb{J}_{\gamma\mathbb{A}} \circ (I - \gamma\mathbb{C})z + \gamma\mathbb{C}z = (I + \gamma\mathbb{C})z \\ \iff & \mathbb{J}_{\gamma\mathbb{A}} \circ (I - \gamma\mathbb{C})z = z. \end{aligned}$$

For solving 6, consider the Krasnosel'skiĭ–Mann (KM) iteration with the operator T above:

$$\begin{aligned} & T_\lambda := (1-\lambda)I_{\mathcal{X}} + \lambda T \\ \text{and } & z^{k+1} = (1-\lambda_k)z^k + \lambda_k T z^k. \end{aligned} \tag{8}$$

If T is α-averaged, then the classical fixed point iteration theorem states that the iteration $z_{k+1} = T_\lambda(z_k)$ converges for any constant $\lambda \in (0, \frac{1}{\alpha}]$. See [2] and references therein for some latest development of strategies designing λ_k. The iteration can be implemented as follows:

Algorithm 1 Initialize $z^0 \in \mathcal{X}$, $\gamma \in (0, 2\beta]$, and sequence $(\lambda_k)_{k\geq 0} \in (0, (4\beta-\gamma)/2\beta)$. For $k = 0, 1, 2, \ldots$

1. Compute $x^k_{\mathbb{B}} = \mathbb{J}_{\gamma\mathbb{B}}(z^k)$;
2. Compute $x^k_{\mathbb{A}} = \mathbb{J}_{\gamma\mathbb{A}}(2x^k_{\mathbb{B}} - z^k - \gamma\mathbb{C}x^k_{\mathbb{B}})$ or $\mathbb{J}_{\gamma\mathbb{A}}(2\mathbb{J}_{\gamma\mathbb{B}}(z^k) - z^k - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z^k))$;
3. Compute $x^k_{\mathbb{C}} = \mathbb{J}_{\gamma\mathbb{C}}(x^k_{\mathbb{A}} + \gamma\mathbb{C}x^k_{\mathbb{B}})$ or $\mathbb{J}_{\gamma\mathbb{C}}(x^k_{\mathbb{A}} + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z^k))$;
4. Update $z^{k+1} = z^k + \lambda_k(x^k_{\mathbb{C}} - x^k_{\mathbb{B}})$.

4 Weak Convergence and Rates of Proposed Three-Block Operator Splitting Scheme

In this section, we discuss some properties of the operator T defined in (7). Figure 1 illustrates how proposed T is applied to a point $z \in \mathcal{X}$ corresponding to the points defined in Lemma 1.

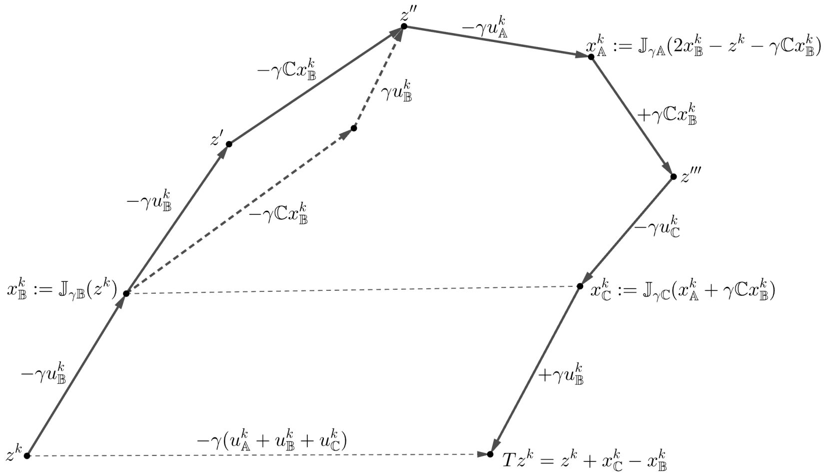


Fig. 1: The proposed mapping $T : z^k \to Tz^k$ and vectors $u_{\mathbb{B}}^k \in \mathbb{B}x_{\mathbb{B}}^k$, $u_{\mathbb{A}}^k \in \mathbb{A}x_{\mathbb{A}}^k$, and $u_{\mathbb{C}}^k \in \mathbb{C}x_{\mathbb{C}}^k$ as given in Lemma 1.

Lemma 1 *Let $z \in \mathcal{X}$ and define the points:*

$$\begin{array}{lll} x_{\mathbb{B}}^k = \mathbb{J}_{\gamma\mathbb{B}}(z^k), & z' = 2x_{\mathbb{B}}^k - z^k, & u_{\mathbb{B}}^k = \gamma^{-1}(z^k - x_{\mathbb{B}}^k) \in \mathbb{B}x_{\mathbb{B}}^k \\ x_{\mathbb{A}}^k = \mathbb{J}_{\gamma\mathbb{A}}(z''), & z''' = x_{\mathbb{A}}^k + \gamma\mathbb{C}x_{\mathbb{B}}^k, & u_{\mathbb{A}}^k = \gamma^{-1}(z'' - x_{\mathbb{A}}^k) \in \mathbb{A}x_{\mathbb{A}}^k \\ z'' = z' - \gamma\mathbb{C}x_{\mathbb{B}}^k, & x_{\mathbb{C}}^k = \mathbb{J}_{\gamma\mathbb{C}}(z'''), & u_{\mathbb{C}}^k = \gamma^{-1}(x_{\mathbb{A}}^k + \gamma\mathbb{C}x_{\mathbb{B}}^k - x_{\mathbb{C}}^k) \in \mathbb{C}x_{\mathbb{C}}^k. \end{array}$$

Then, the following identities hold:

$$Tz^k - z^k = x_{\mathbb{C}}^k - x_{\mathbb{B}}^k = -\gamma(u_{\mathbb{A}}^k + u_{\mathbb{B}}^k + u_{\mathbb{C}}^k) \textit{ and } Tz^k = x_{\mathbb{C}}^k + \gamma u_{\mathbb{B}}^k.$$

Proof In view of the definition of T, we observe that

$$Tz^k = z^k + x_{\mathbb{C}}^k - x_{\mathbb{B}}^k = x_{\mathbb{C}}^k + \gamma u_{\mathbb{B}}^k.$$

Now, we conclude that

$$\begin{aligned} Tz^k - z^k = x_{\mathbb{C}}^k - x_{\mathbb{B}}^k = x_{\mathbb{A}}^k + \gamma\mathbb{C}x_{\mathbb{B}}^k - \gamma u_{\mathbb{C}}^k - x_{\mathbb{B}}^k &= 2x_{\mathbb{B}}^k - z^k - \gamma u_{\mathbb{A}}^k - \gamma u_{\mathbb{C}}^k - x_{\mathbb{B}}^k \\ &= x_{\mathbb{B}}^k - z^k - \gamma u_{\mathbb{A}}^k - \gamma u_{\mathbb{C}}^k \\ &= -\gamma(u_{\mathbb{A}}^k + u_{\mathbb{B}}^k + u_{\mathbb{C}}^k). \end{aligned}$$

Next, we show that the fixed point identity for the operator T holds true. Moreover, with the help of any fixed point z^* of T and $\mathbb{J}_{\gamma\mathbb{B}}z^*$, a zero of $\mathbb{A} + \mathbb{B} + \mathbb{C}$ can be obtained.

Lemma 2 *Let $\mathbb{A}$, $\mathbb{B}$, and $\mathbb{C}$ be three operators. Then, the following set equality holds:*

$$\textit{zer}(\mathbb{A} + \mathbb{B} + \mathbb{C}) = \mathbb{J}_{\gamma\mathbb{B}}(\textit{Fix}\, T),$$

where Fix $T = \{x + \gamma u | 0 \in (\mathbb{A} + \mathbb{B} + \mathbb{C})x,\ u \in (\mathbb{B}x) \cap (-\mathbb{A}x - \mathbb{C}x)\}$.

Proof Let $x \in \text{zer}(\mathbb{A}+\mathbb{B}+\mathbb{C})$, that is, $0 \in (\mathbb{A}+\mathbb{B}+\mathbb{C})x$. Let $u_{\mathbb{A}} \in \mathbb{A}x$ and $u_{\mathbb{B}} \in \mathbb{B}x$ such that $\gamma(u_{\mathbb{A}} + u_{\mathbb{B}} + \mathbb{C}x) = 0$ and $z = x + \gamma u_{\mathbb{B}}$. We first show that z is a fixed point of T. Notice that $x = \mathbb{J}_{\gamma\mathbb{B}}(z)$. We have

$$\begin{aligned} & 2\mathbb{J}_{\gamma\mathbb{B}}(z) - z - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z) = 2x - z - \gamma\mathbb{C}x = x - \gamma u_{\mathbb{B}} - \gamma\mathbb{C}x = x + \gamma u_{\mathbb{A}} \\ \Rightarrow\ & \mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}}(z) - z - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z)) = x \\ \Rightarrow\ & \mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}}(z) - z - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z)) + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z)) = \mathbb{J}_{\gamma\mathbb{C}}(x + \gamma\mathbb{C}x) = x. \end{aligned}$$

With all the identities above, we conclude that

$$Tz = \mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}}(z) - z - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z)) + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z)) + \gamma u_{\mathbb{B}} = x + \gamma u_{\mathbb{B}} = z.$$

Next, assume that $z = x + \gamma u_{\mathbb{B}} \in \text{Fix}\, T$, then $x = \mathbb{J}_{\gamma\mathbb{B}}(z) \in \text{zer}(\mathbb{A}+\mathbb{B}+\mathbb{C})$ since

$$\begin{aligned} x + \gamma u_{\mathbb{B}} = Tz = \mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}}(z) - z - \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z)) + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z)) + \gamma u_{\mathbb{B}} \\ \Rightarrow x = \mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (2x - z - \gamma\mathbb{C}x) + \gamma\mathbb{C}x) \\ \Rightarrow x + \gamma\mathbb{C}x = \mathbb{J}_{\gamma\mathbb{A}} \circ (2x - z - \gamma\mathbb{C}x) + \gamma\mathbb{C}x \\ \Rightarrow x + \gamma u_{\mathbb{A}} = x - \gamma u_{\mathbb{B}} - \gamma\mathbb{C}x \\ \Rightarrow 0 = u_{\mathbb{A}} + u_{\mathbb{B}} + \mathbb{C}x. \end{aligned}$$

Next we discuss the convergence of the proposed Algorithm 1 under the assumption that T is an averaged operator. The following are some standard convergence results under the assumption that T is an averaged operator, following Corollary 2.1 and Theorem 2.1 in [8].

Theorem 1 *Assume that $T : \mathcal{X} \to \mathcal{X}$ be a-averaged with $a = \frac{2\beta}{4\beta-\gamma} < 1$. Let z^* be a fixed point of T. Assume $(\lambda_j)_{j\geq 0} \subseteq (0, \frac{1}{\alpha})$ be a sequence of relaxation parameters, where $\alpha = 1/(2-\varepsilon) < 2\beta/(4\beta-\gamma)$. Let $\sum_{j=0}^{\infty} \tau_j = \sum_{j=0}^{\infty} \lambda_j/\alpha(1-\lambda_j/\alpha) = \infty$. Let $z^0 \in \mathcal{X}$ and $(z^j)_{j\geq 0} \subseteq \mathcal{X}$ be the sequence generated by Algorithm 1. Then, the following results hold.*

(i) The sequence $\|z^j - z^*\|_{j\geq 0}$ is monotonically increasing for any $z^* \in \text{Fix}\, T$.
(ii) The fixed point residual sequence $\|Tz^j - z^j\|_{j\geq 0}$ is monotonically nonincreasing and converges to 0.
(iii) The sequence $(z^j)_{j\geq 0}$ converges weakly to a fixed point of T.
(iv) Assume that $\underline{\tau} = \inf_{j\geq 0} \tau_j > 0$ for $\tau > 0$. Then, for any $z^* \in \text{Fix}T$ and for all $k \geq 0$, the following convergence rates hold:

$$\|Tz^k - z^k\|^2 \leq \frac{\|z^0 - z^*\|^2}{\underline{\tau}(k+1)} \text{ and } \|Tz^k - z^k\|^2 = o\left(\tfrac{1}{k+1}\right).$$

Proof The proof for parts (i)-(iii) follows from Proposition 5.15 of [1], and the proof of part (iv) follows from [7].

Theorem 2 (Convergence theorem). *Assume that $T : \mathcal{X} \to \mathcal{X}$ be a-averaged with $a = \frac{2\beta}{4\beta-\gamma} < 1$. Assume $(\lambda_j)_{j\geq 0} \subseteq \left(0, \frac{1}{\alpha}\right)$ be a sequence of relaxation parameters, where $\alpha = 1/(2-\varepsilon) < 2\beta/(4\beta-\gamma)$. Let $\sum_{j=0}^{\infty} \tau_j = \sum_{j=0}^{\infty} \lambda_j/\alpha(1-\lambda_j/\alpha) = \infty$. Let $z^0 \in \mathcal{X}$ and $(z^j)_{j\geq 0} \subseteq \mathcal{X}$ be the sequence generated by Algorithm 1. Then, the following results hold.*

1. Suppose that $\inf_{j\geq 0} \lambda_j > 0$ and z^* be the weak limit point of z^k. Then, the following convergence results hold:
(a) $(\mathbb{C}x^j_{\mathbb{B}})_{j\geq 0}$ converges strongly to $\mathbb{C}x^*$ for any $x^* \in$ zer$(\mathbb{A}+\mathbb{B}+\mathbb{C})$,

(b) *the sequence* $\mathbb{J}_{\gamma\mathbb{B}}(z^j)_{j\geq 0}$ *weakly converges to* $\mathbb{J}_{\gamma\mathbb{B}}(z^*) \in zer(\mathbb{A}+\mathbb{B}+\mathbb{C})$.
(c) *the sequence* $(\mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C}\circ\mathbb{J}_{\gamma\mathbb{B}})(z^j)_{j\geq 0}$ *weakly converges to* $\mathbb{J}_{\gamma\mathbb{B}}(z^*) \in zer(\mathbb{A}+\mathbb{B}+\mathbb{C})$.
(d) *the sequence* $(\mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C}\circ\mathbb{J}_{\gamma\mathbb{B}}) + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}})(z^j)_{j\geq 0}$ *weakly converges to* $\mathbb{J}_{\gamma\mathbb{B}}(z^*) \in zer(\mathbb{A}+\mathbb{B}+\mathbb{C})$.

2. *The sequences* $\mathbb{J}_{\gamma\mathbb{B}}(z^j)_{j\geq 0}$ *and* $(\mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C}\circ\mathbb{J}_{\gamma\mathbb{B}}) + \gamma\mathbb{C}\mathbb{J}_{\gamma\mathbb{B}})(z^j)_{j\geq 0}$ *converges strongly to a point in* $zer(\mathbb{A}+\mathbb{B}+\mathbb{C})$ *if any of the following holds:*
(a) $\mathbb{A}$ *is uniformly monotone on every nonempty bounded subset of* $dom(\mathbb{A})$,
(b) $\mathbb{B}$ *is uniformly monotone on every nonempty bounded subset of* $dom(\mathbb{B})$,
(c) $\mathbb{C}$ *is demiregular at every point* $x \in zer(\mathbb{A}+\mathbb{B}+\mathbb{C})$.

Proof 1. (a) Let $k \geq 0$. Then, using Corollary 2.14 in [1], we observe that

$$\|z^{k+1} - z^*\|^2 = \|(1-\lambda_k)(z^k - z^*) + \lambda_k(Tz^k - z^*)\|^2$$
$$= (1-\lambda_k)\|z^k - z^*\|^2 + \lambda_k\|Tz^k - z^*\|^2 - \lambda_k(1-\lambda_k)\|Tz^k - z^k\|^2. \quad (9)$$

In view of Theorem 2.1 of [8], we get

$$\sum_{i=k}^{\infty}\|\mathbb{C}x_{\mathbb{B}}^k - \mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z^*)\|^2 \leq \frac{\|z^k - z^*\|^2}{\gamma\lambda_k\left(2\beta - \frac{\gamma}{\varepsilon}\right)},$$

which gives $\|\mathbb{C}x_{\mathbb{B}}^k - \mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z^*)\|^2 \to \infty$ as $k \to \infty$.

(b) We recall the notations from Lemma 1 given by

$$\begin{aligned} x_{\mathbb{B}}^k &= \mathbb{J}_{\gamma\mathbb{B}}(z^k), & u_{\mathbb{B}}^k &= \gamma^{-1}(z^k - x_{\mathbb{B}}^k) \in \mathbb{B}x_{\mathbb{B}}^k \\ x_{\mathbb{A}}^k &= \mathbb{J}_{\gamma\mathbb{A}}(2x_{\mathbb{B}}^k - z^k - \gamma\mathbb{C}x_{\mathbb{B}}^k), & u_{\mathbb{A}}^k &= \gamma^{-1}(2x_{\mathbb{B}}^k - z^k - \gamma\mathbb{C}x_{\mathbb{B}}^k - x_{\mathbb{A}}^k) \in \mathbb{A}x_{\mathbb{A}}^k \\ x_{\mathbb{C}}^k &= \mathbb{J}_{\gamma\mathbb{C}}(x_{\mathbb{A}}^k + \gamma\mathbb{C}x_{\mathbb{B}}^k), & u_{\mathbb{C}}^k &= \gamma^{-1}(x_{\mathbb{A}}^k + \gamma\mathbb{C}x_{\mathbb{B}}^k - x_{\mathbb{C}}^k) \in \mathbb{C}x_{\mathbb{C}}^k. \end{aligned}$$

Note that for all $k \geq 0$, we have

$$\|x_{\mathbb{B}}^k - \mathbb{J}_{\gamma\mathbb{B}}(z^*)\| = \|\mathbb{J}_{\gamma\mathbb{B}}(z^k) - \mathbb{J}_{\gamma\mathbb{B}}(z^*)\| \leq \|z^k - z^*\| \leq \|z^0 - z^*\|,$$

therefore the sequence $(x_{\mathbb{B}}^j)_{j\geq 0}$ is bounded and has a weak sequential cluster point $\bar{x}$.

Now, assume that there exists a subsequence $(k_j)_{j\geq 0}$ such that $x_{\mathbb{B}}^{k_j} \rightharpoonup \bar{x}$ as $j \to \infty$. Let $x^* \in \mathrm{zer}(\mathbb{A}+\mathbb{B}+\mathbb{C})$. Next, observe that $\mathbb{C}$ is maximal monotone and $\mathbb{C}x_{\mathbb{B}}^k \to \mathbb{C}x^*$, and $x_{\mathbb{B}}^{k_j} \to \bar{x}$, thus, in view of Proposition 20.33(ii) of [1] and weak-to-strong sequential closeness of $\mathbb{C}$, we have

$$\mathbb{C}\bar{x} = \mathbb{C}x^* \text{ and } \mathbb{C}x_{\mathbb{B}}^{k_j} = \mathbb{C}\bar{x}.$$

Further, in view of (ii) of Theorem 1 and Lemma 1, we have $x_{\mathbb{C}}^k - x_{\mathbb{B}}^k = Tz^k - z^k \to 0$ as $k \to \infty$. Thus, with $j \to \infty$, we obtain

$$x_{\mathbb{B}}^{k_j} \rightharpoonup \bar{x},\ x_{\mathbb{A}}^{k_j} \rightharpoonup \bar{x},\ x_{\mathbb{C}}^{k_j} \rightharpoonup \bar{x},\ \mathbb{C}x_{\mathbb{B}}^{k_j} \rightharpoonup \mathbb{C}\bar{x}$$
$$\text{and } u_{\mathbb{B}}^{k_j} \rightharpoonup \tfrac{1}{\gamma}(z^* - \bar{x}),\ u_{\mathbb{A}}^{k_j} \rightharpoonup \tfrac{1}{\gamma}(\bar{x} - z^* - \gamma\mathbb{C}\bar{x}),\ u_{\mathbb{C}}^{k_j} \rightharpoonup \tfrac{1}{\gamma}(\bar{x} + \gamma\mathbb{C}\bar{x}).$$

On applying Proposition, 25.5 of [1] to $(x_{\mathbb{A}}^{k_j}, u_{\mathbb{A}}^{k_j}) \in \mathrm{gra}\mathbb{A}$, $(x_{\mathbb{B}}^{k_j}, u_{\mathbb{B}}^{k_j}) \in \mathbb{B}$, and $(x_{\mathbb{B}}^{k_j}, \mathbb{C}x_{\mathbb{B}}^{k_j}) \in \mathbb{C}$, we observe that $\bar{x} \in \mathrm{zer}(\mathbb{A}+\mathbb{B}+\mathbb{C})$, $z^* - \bar{x} \in \gamma\mathbb{B}\bar{x}$, $\bar{x} - z^* - \gamma\mathbb{C}\bar{x} \in \gamma\mathbb{A}\bar{x}$, and $\bar{x} + \gamma\mathbb{C}\bar{x} \in \gamma\mathbb{C}\bar{x}$. Thus, we obtain $\bar{x} = \mathbb{J}_{\gamma\mathbb{B}}(z^*)$, and hence $\bar{x}$ is a unique weak sequential cluster point of $(x_{\mathbb{B}}^j)_{j\geq 0}$. Therefore, in view of Lemma 2.38 of [1], we conclude that $(x_{\mathbb{B}}^j)_{j\geq 0}$ converges weakly to $\mathbb{J}_{\gamma\mathbb{B}}(z^*)$.

(c) On proceeding in similar manner to part 1c and part 1d, we implies that

$$x_{\mathbb{C}}^k - x_{\mathbb{B}}^k = Tz^k - z^k \to 0 \text{ as } k \to \infty \text{ and } x_{\mathbb{B}}^k \rightharpoonup \mathbb{J}_{\gamma\mathbb{B}}(z^*) \implies x_{\mathbb{C}}^k \rightharpoonup \mathbb{J}_{\gamma\mathbb{B}}(z^*).$$

2. The proofs are in a similar manner to those given in Theorem 2.1 of [8].

Before we proceed to the analysis of convex optimization problems of using Algorithm 1 under several assumptions on the regularity of the problem, we discuss the following lemma:

Lemma 3 *Assume that $T : \mathcal{X} \to \mathcal{X}$ be a-averaged with $a = \frac{2\beta}{4\beta-\gamma} < 1$. Let $(z^j)_{j\geq 0}$ is generated by Algorithm 1 and $\gamma > 0$. Let z^* be a fixed point of T and $x^* = \mathbb{J}_{\gamma\beta}(z^*)$. Then, $(x_{\mathbb{A}}^j)_{j\geq 0}$ and $(x_{\mathbb{B}}^j)_{j\geq 0}$ are contained within the closed ball $\overline{B(x^*, (1+\gamma/\beta)\|z^0 - z^*\|)}$.*

Proof By (i) of Theorem 1, we have

$$\|x_{\mathbb{B}}^k - x^*\| = \|\mathbb{J}_{\gamma\mathbb{B}}(z^k) - \mathbb{J}_{\gamma\mathbb{B}}(z^*)\| \leq \|z^k - z^*\| \leq \|z^0 - z^*\|.$$

Similarly, we have

$$\begin{aligned} \|x_{\mathbb{A}}^k - x^*\| &= \|\mathbb{J}_{\gamma\mathbb{A}}(\text{refl}_{\gamma\mathbb{B}}(z^k) - \gamma\mathbb{C}x_{\mathbb{B}}^k) - \mathbb{J}_{\gamma\mathbb{A}}(\text{refl}_{\gamma\mathbb{B}}(z^*) - \gamma\mathbb{C}x^*)\| \\ &\leq \|\text{refl}_{\gamma\mathbb{B}}(z^k) - \text{refl}_{\gamma\mathbb{B}}(z^*) + \gamma\mathbb{C}x^* - \gamma\mathbb{C}x_{\mathbb{B}}^k\| \\ &\leq \|z^k - z^*\| + \frac{\gamma}{\beta}\|z^k - z^*\| \leq \left(1 + \frac{\gamma}{\beta}\right)\|z^0 - z^*\|. \end{aligned}$$

With the inequality above, we also have

$$\begin{aligned} \|x_{\mathbb{C}}^k - x^*\| &= \|\mathbb{J}_{\gamma\mathbb{C}}(x_{\mathbb{A}}^k + \gamma\mathbb{C}x_{\mathbb{B}}^k) - \mathbb{J}_{\gamma\mathbb{C}}(x^* + \gamma\mathbb{C}x^*)\| = \|x_{\mathbb{A}}^k - x^* + \gamma\mathbb{C}x_{\mathbb{B}}^k - \gamma\mathbb{C}x^*\| \\ &\leq \|z^k - z^*\| + \frac{2\gamma}{\beta}\|z^k - z^*\| \leq \left(1 + \frac{2\gamma}{\beta}\right)\|z^0 - z^*\|. \end{aligned}$$

5 The Multi-Block ADMM for Convex Minimization Problems

In this section, we first show that the three-operator splitting scheme is equivalent to the classical three-block ADMM method, then we also give a formal extension of the splitting scheme to multiple operators.

5.1 The classical three-block ADMM

We analyze the convex optimization problems under several assumptions on the regularity of the problem in this section:

- Every considered function is proper, closed, and convex.
- Every differentiable function is Frèchet differentiable.
- The functions $f_i : \mathcal{X} \to (-\infty, +\infty]$, $i = 1, 2, 3$, satisfy the existence of solution condition, i.e.,

$$\text{zer}(\partial f_1 + \partial f_2 + \partial f_3) \neq \emptyset.$$

Proposition 1 (Optimality conditions of prox). *Let $w \in \mathcal{X}$ and f be a proper, closed, and convex function. Then, the following identity holds.*

$$\tilde{w} = \textit{prox}_f^\gamma(w) \textit{ if and only if } \frac{1}{\gamma}(w - \tilde{w}) \in \partial f(\tilde{w}).$$

Proof In view of the definition of prox, we have

$$\begin{aligned}\tilde{w} = \text{prox}_f^\gamma(w) &\iff (I + \gamma\partial f)(\tilde{w}) = w \\ &\iff \tilde{w} = w - \gamma\partial f(\tilde{w}) \\ &\iff \frac{1}{\gamma}(w - \tilde{w}) \in \partial f(\tilde{w})\end{aligned}$$

Proposition 2 (Firm nonexpansiveness of prox). *Let $w, r \in \mathcal{X}$, and let $\tilde{w} = prox_f^\gamma(w)$ and $\tilde{r} = prox_f^\gamma(r)$. Then,*

$$\|\tilde{w} - \tilde{r}\|^2 \leq \langle \tilde{w} - \tilde{r}, w - r \rangle.$$

In particular, $prox_f^\gamma$ is nonexpansive.

Theorem 3 (Descent Lemma). *Let f be a differentiable function and ∇f is $\frac{1}{\gamma}$-Lipschitz. Then, for every $x, y \in \mathcal{X}$, we have*

$$f(x) \leq f(y) + \langle x - y, \nabla f(y) \rangle + \tfrac{1}{2\gamma}\|x - y\|^2.$$

Consider a convex minimization problem with linear constraints and a separable objective function given by :

$$\left.\begin{array}{ll}\min & f_1(x_1) + f_2(x_2) + f_3(x_3), \\ \text{subject to} & A_1x_1 + A_2x_2 + A_3x_3 = b, \\ & x_1 \in \mathcal{X}_1,\ x_2 \in \mathcal{X}_2,\ x_3 \in \mathcal{X}_3\end{array}\right\}, \tag{10}$$

where $f_i : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{+\infty\}$ are proper, closed, and convex functions (not necessarily smooth), $A_i \in \mathbb{R}^{m \times n_i}$ and $b \in \mathbb{R}^m$. Assume that f_2 is μ-strongly convex and $\mu > 0$.
We will show that the classical three-block ADMM algorithm is equivalent to the three-operator splitting in the previous section applying to the dual formulation of (10). Let f_i^* denote the convex conjugate of function f_i, and let

$$d_1(w) = f_1^*(A_1^*w),\ d_2(w) = f_2^*(A_2^*w),\ d_3(w) = f_3^*(A_1^*w) - \langle w, b \rangle.$$

The dual problem of (10) is given by

$$\min_w d_1(w) + d_2(w) + d_3(w).$$

The assumption that f_2 is μ-strongly convex implies ∇f_2^* is $\frac{1}{\mu}$-Lipschitz continuous. Since $\nabla(f_2^* \circ A_2^*) = A_2 \circ \nabla f_2 \circ A_2^*$, ∇d_2 is $\frac{\|A_2\|^2}{\mu}$-Lipschitz continuous, i.e., ∇d_2 is $\frac{\mu}{\|A_2\|^2}$-cocoerceive.

An extension of the original ADMM in [12] for (10) is given by

$$x_1^{k+1} = \underset{x_1 \in \mathcal{X}_1}{\text{argmin}}\, \{f_1(x_1) + \tfrac{\gamma}{2}\|(A_1x_1 + A_2x_2^k + A_3x_3^k - b) - \tfrac{1}{\gamma}w^k\|^2 \tag{11}$$

$$x_2^{k+1} = \underset{x_2 \in \mathcal{X}_2}{\text{argmin}}\, \{f_2(x_2) + \tfrac{\gamma}{2}\|(A_1x_1^{k+1} + A_2x_2 + A_3x_3^k - b) - \tfrac{1}{\gamma}w^k\|^2\} \tag{12}$$

$$x_3^{k+1} = \underset{x_3 \in \mathcal{X}_3}{\text{argmin}}\, \{f_3(x_3) + \tfrac{\gamma}{2}\|(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3 - b) - \tfrac{1}{\gamma}w^k\|^2\} \tag{13}$$

$$w^{k+1} = w^k - \gamma(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b), \tag{14}$$

where $w^k \in \mathbb{R}^m$ is the Lagrange multiplier, $\gamma > 0$ is the penalty parameter or step size.

5.2 The equivalence to the three-operator splitting

To obtain the suitable algorithm for finding the solution to (10), we start by considering the equation given in (13) as

$$
\begin{aligned}
& x_3^{k+1} = \underset{x_3 \in \mathcal{X}_3}{\operatorname{argmin}} \{f_3(x_3) + \tfrac{\gamma}{2}\|(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3 - b) - \tfrac{1}{\gamma} w^k\|^2\} \\
\iff & 0 \in \partial f_3(x_3^{k+1}) - A_3^\top (w^k - \gamma(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b) \\
\iff & 0 \in \partial f_3(x_3^{k+1}) - A_3^\top w^{k+1} \text{ from (14)} \\
\iff & A_3^\top w^{k+1} \in \partial f_3(x_3^{k+1}) \\
\iff & x_3^{k+1} \in \partial f_3^*(A_3^* w^{k+1}). \qquad (15)
\end{aligned}
$$

In view of (14), observe that

$$
\begin{aligned}
& w^{k+1} = w^k - \gamma(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b) \\
& w^{k+1} = z^k - \gamma(A_3 x_3^{k+1} - b) \\
\iff & w^{k+1} = z^k - \gamma(A_3 \partial f_3^*(A_3^* w^{k+1}) - b) \text{ assume } (z^k = w^k - \gamma(A_1 x_1^{k+1} + A_2 x_2^{k+1})) \\
\iff & w^{k+1} = z^k - \gamma \partial d_3(w^{k+1}) \qquad (16) \\
\iff & (I + \gamma \partial d_3)(w^{k+1}) = z^k \\
\iff & w^{k+1} = \operatorname{prox}_{d_3}^{\gamma}(z^k). \qquad (17)
\end{aligned}
$$

For convenience, let $\bar{x}^k := x^{k+1}$. Consider the equation given by (11), we obtain

$$
\begin{aligned}
& x_1^{k+1} = \underset{x_1 \in \mathcal{X}_1}{\operatorname{argmin}} \{f_1(x_1) + \tfrac{\gamma}{2}\|(A_1 x_1 + A_2 x_2^k + A_3 x_3^k - b) - \tfrac{1}{\gamma} w^k\|^2\} \\
\iff & 0 \in \partial f_1(x_1^{k+1}) - A_1^\top (w^k - \gamma(A_1 x_1^{k+1} + A_2 x_2^k + A_3 x_3^k - b)) \\
\iff & 0 \in \partial f_1(\bar{x}_1^k) - A_1^\top (w^k - \gamma(A_1 \bar{x}_1^k + A_2 x_2^k + A_3 x_3^k - b)) \\
\iff & 0 \in \partial f_1(\bar{x}_1^{k+1}) - A_1^\top (w^{k+1} - \gamma(A_1 \bar{x}_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b)) \text{ holds for any } k \in \mathbb{N} \cup \{0\} \\
\iff & 0 \in \partial f_1(\bar{x}_1^{k+1}) - A_1^\top p^{k+1} \\
\iff & A_1^\top p^{k+1} \in \partial f_1(\bar{x}_1^{k+1}) \text{ assume } (p^k = w^{k+1} - \gamma(A_1 \bar{x}_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1})) \\
\iff & \bar{x}_1^{k+1} \in \partial f_1^*(A_1^* p^{k+1}).
\end{aligned}
$$

We have

$$
\begin{aligned}
& p^{k+1} = w^{k+1} - \gamma(A_1 \bar{x}_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b) \\
\iff & p^{k+1} = w^{k+1} - \gamma(A_1 \partial f_1^*(A_1^* p^{k+1}) + A_2 \partial f_2^*(A_2^* w^{k+1}) + A_3 \partial f_3^*(A_3^* w^{k+1}) - b) \\
\iff & p^{k+1} = w^{k+1} - \gamma \partial d_1(p^{k+1}) - \gamma \partial d_2(w^{k+1}) - \gamma \partial d_3(w^{k+1}) \\
\iff & (I + \gamma \partial d_1)(p^{k+1}) = w^{k+1} - \gamma \partial d_2(w^{k+1}) - \gamma \partial d_3(w^{k+1}) \qquad (18) \\
\iff & (I + \gamma \partial d_1)(p^{k+1}) = 2w^{k+1} - z^k - \gamma \partial d_2(w^{k+1}) \text{ from (16)} \\
\iff & p^{k+1} = \operatorname{prox}_{d_1}^{\gamma}(2w^{k+1} - z^k - \gamma \partial d_2(w^{k+1})).
\end{aligned}
$$

By (12), we have

$$
\begin{aligned}
& x_2^{k+1} = \underset{x_2 \in \mathcal{X}_2}{\operatorname{argmin}} \{f_2(x_2) + \tfrac{\gamma}{2}\|(A_1x_1^{k+1} + A_2x_2 + A_3x_3^k - b) - \tfrac{1}{\gamma}w^k\|^2\} \\
\iff & 0 \in \partial f_2(x_2^{k+1}) - A_2^\top(w^k - \gamma(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^k - b) \\
\iff & 0 \in \partial f_2(\bar{x}_2^k) - A_2^\top(w^k - \gamma(A_1\bar{x}_1^k + A_2\bar{x}_2^k + A_3x_3^k - b) \\
\iff & 0 \in \partial f_2(\bar{x}_2^{k+1}) - A_1^\top(w^{k+1} - \gamma(A_1\bar{x}_1^{k+1} + A_2\bar{x}_2^{k+1} + A_3x_3^{k+1} - b) \text{ holds for any } k \in \mathbb{N} \cup \{0\} \\
\iff & 0 \in \partial f_2(\bar{x}_2^{k+1}) - A_2^\top v^{k+1} \text{ assume } (v^{k+1} = w^{k+1} - \gamma(A_1\bar{x}_1^{k+1} + A_2\bar{x}_2^{k+1} + A_3x_3^{k+1})) \\
\iff & A_2^\top v^{k+1} \in \partial f_2(\bar{x}_2^{k+1}) \\
\iff & \bar{x}_2^{k+1} \in \partial f_2^*(A_2^* v^{k+1}).
\end{aligned}
$$

Now, we observe that

$$
\begin{aligned}
& v^{k+1} = w^{k+1} - \gamma(A_1\bar{x}_1^{k+1} + A_2\bar{x}_2^{k+1} + A_3x_3^{k+1} - b) \\
\iff & v^{k+1} = w^{k+1} - \gamma(A_1\partial f_1^*(A_1^* p^{k+1}) + A_2\partial f_2^*(A_2^* v^{k+1}) + A_3\partial f_3^*(A_3^* w^{k+1}) - b) \\
\iff & v^{k+1} = w^{k+1} - \gamma\partial d_1(p^{k+1}) - \gamma\partial d_2(v^{k+1}) - \gamma\partial d_3(w^{k+1}) \\
\iff & (I + \gamma\partial d_2)(v^{k+1}) = w^{k+1} - \gamma\partial d_1(p^{k+1}) - \gamma\partial d_3(w^{k+1}) \text{ from (18)} \\
\iff & v^{k+1} = \operatorname{prox}_{d_2}^{\gamma}(p^{k+1} + \gamma\partial d_2(w^{k+1})).
\end{aligned}
$$

Now, in view of relations on p^{k+1}, v^{k+1}, w^{k+1}, and z^k, we have

$$
\begin{aligned}
\implies & z^k = w^k - \gamma(A_1x_1^{k+1} + A_2x_2^{k+1}) \\
\implies & z^k = w^k - \gamma(A_1\bar{x}_1^k + A_2\bar{x}_2^k) \\
\implies & z^{k+1} = w^{k+1} - \gamma(A_1\bar{x}_1^{k+1} + A_2\bar{x}_2^{k+1}) \text{ holds for any } k \in \mathbb{N} \cup \{0\}
\end{aligned} \tag{19}
$$

In view of relation on w^{k+1} and z^{k+1}, we have

$$
\begin{aligned}
w^{k+1} &= z^k - \gamma(A_3x_3^{k+1} - b) \\
v^{k+1} &= w^{k+1} - \gamma(A_1\bar{x}_1^{k+1} + A_2\bar{x}_2^{k+1} + A_3\bar{x}_3^{k+1} - b)
\end{aligned} \tag{20}
$$

Thus, in view of (19) and (20), we get

$$
z^{k+1} = z^k + v^{k+1} - w^{k+1}
$$

Hence, we derive a fixed point iteration given by

$$
\left.
\begin{aligned}
w^{k+1} &= \operatorname{prox}_{d_3}^{\gamma}(z^k) \\
p^{k+1} &= \operatorname{prox}_{d_1}^{\gamma}(2w^{k+1} - z^k - \gamma\partial d_2(w^{k+1})) \\
v^{k+1} &= \operatorname{prox}_{d_2}^{\gamma}(p^{k+1} + \gamma\partial d_2(w^{k+1})) \\
z^{k+1} &= z^k + v^{k+1} - w^{k+1}
\end{aligned}
\right\} \tag{21}
$$

which is consistent with the three-operator splitting scheme in the previous section. The splitting scheme can be implemented as in Algorithm 2 for minimizing $d_1(w) + d_2(w) + d_3(w)$ where d_i are proper closed convex functions and ∇d_2 is L-Lipshcitz continuous.

Algorithm 2 Initialize $z^0 \in \mathcal{X},\ \gamma \in (0, 2/L)$. For $k = 0, 1, 2, \ldots$

1. Compute $w^{k+1} = \text{prox}^{\gamma}_{d_3}(z^k)$;
2. Compute $p^{k+1} = \text{prox}^{\gamma}_{d_1}(2w^{k+1} - z^k - \gamma\nabla d_2(w^{k+1}))$;
3. Compute $v^{k+1} = \text{prox}^{\gamma}_{d_2}(p^{k+1} + \gamma\nabla d_2(w^{k+1}))$;
4. Update $z^{k+1} = z^k + (v^{k+1} - w^{k+1})$.

5.3 Convergence

For discussing the iteration (21), we introduce the variables w_{d_i} by writing (21) as follows:

$$\left.\begin{aligned} w^k_{d_3} &:= \lambda^{k+1} = \text{prox}^{\gamma}_{d_3}(z^k) \\ w^k_{d_1} &:= p^{k+1} = \text{prox}^{\gamma}_{d_1}(2\lambda^{k+1} - z^k - \gamma\partial d_2(\lambda^{k+1})) \\ w^k_{d_2} &:= v^{k+1} = \text{prox}^{\gamma}_{d_2}(p^{k+1} + \gamma\partial d_2(\lambda^{k+1})) \\ z^{k+1} &= z^k + w^k_{d_2} - w^k_{d_3}. \end{aligned}\right\} \tag{22}$$

For convenience we use $\nabla f(x) \in \partial f(x)$ to denote a subgradient of f at x.

Proposition 3 *Let $z^0 \in \mathcal{X}$ and $(z_j)_{j\geq 0}$ be the sequence generated by (22). Then the following identities hold:*

(i) $w^k_{d_3} = z^k - \gamma\nabla d_3(w^k_{d_3})$
(ii) $w^k_{d_1} - w^k_{d_3} = -\gamma(\nabla d_1(w^k_{d_1}) + \nabla d_2(w^k_{d_3}) + \nabla d_3(w^k_{d_3}))$
(iii) $w^k_{d_2} - w^k_{d_3} = -\gamma(\nabla d_1(w^k_{d_1}) + \nabla d_2(w^k_{d_2}) + \nabla d_3(w^k_{d_3}))$
(iv) $w^k_{d_1} - w^k_{d_3} = w^k_{d_2} - w^k_{d_3} + \gamma(\nabla d_2(w^k_{d_2}) - \nabla d_2(w^k_{d_3}))$
(v) $z^{k+1} - z^k = w^k_{d_2} - w^k_{d_3} = -\gamma(\nabla d_1(w^k_{d_1}) + \nabla d_2(w^k_{d_2}) + \nabla d_3(w^k_{d_3}))$.

Proof We start with the relation $w^k_{d_3} = \text{prox}^{\gamma}_{d_3}(z^k)$, we get

$$w^k_{d_3} = z^k - \gamma\nabla d_3(w^k_{d_3}). \tag{23}$$

Next, we have the following relation

$$\begin{aligned} & w^k_{d_1} = \text{prox}^{\gamma}_{d_1}(2w^k_{d_3} - z^k - \gamma\nabla d_2(w^k_{d_3})) \\ \iff\ & (I + \gamma\nabla d_1)(w^k_{d_1}) = 2w^k_{d_3} - z^k - \gamma\nabla d_2(w^k_{d_3}) \\ \iff\ & w^k_{d_1} - w^k_{d_3} = w^k_{d_3} - z^k - \gamma\nabla d_2(w^k_{d_3}) - \gamma\nabla d_1(w^k_{d_1}) \\ \iff\ & w^k_{d_1} - w^k_{d_3} = -\gamma(\nabla d_1(w^k_{d_1}) + \nabla d_2(w^k_{d_3}) + \nabla d_3(w^k_{d_3})). \end{aligned} \tag{24}$$

Again, we have the following relation

$$\begin{aligned} & w^k_{d_2} = \text{prox}^{\gamma}_{d_2}(w^k_{d_1} + \gamma\nabla d_2(w^k_{d_3})) \\ \iff\ & (I + \gamma\nabla d_2)(w^k_{d_2}) = w^k_{d_1} + \gamma\nabla d_2(w^k_{d_3}) - \gamma\nabla d_2(w^k_{d_2}) \\ \iff\ & w^k_{d_2} = w^k_{d_3} - \gamma\nabla d_2(w^k_{d_3}) - \gamma\nabla d_1(w^k_{d_1}) - \gamma\nabla d_2(w^k_{d_3}) - \gamma\nabla d_2(w^k_{d_2}) \\ \iff\ & w^k_{d_2} - w^k_{d_3} = -\gamma(\nabla d_1(w^k_{d_1}) + \nabla d_2(w^k_{d_2}) + \gamma\nabla d_3(w^k_{d_3})). \end{aligned} \tag{25}$$

Finally, we get

$$w^k_{d_1} - w^k_{d_3} = w^k_{d_2} - w^k_{d_3} + \gamma(\nabla d_2(w^k_{d_2}) - \nabla d_2(w^k_{d_3})) \text{ and } z^{k+1} - z^k = w^k_{d_2} - w^k_{d_3}. \tag{26}$$

Proposition 4 (Upper Inequality). *Let $w \in \mathcal{X}$ and w^* be the fixed point of the FPI algorithm given in (21). Then, the following inequality holds:*

$$
\begin{aligned}
&2\gamma(d_1(w_{d_1}^k) + d_2(w_{d_3}^k) + d_3(w_{d_3}^k) - (d_1 + d_2 + d_3)(w^*)) \\
&\leq \|z^k - w^*\|^2 - \|z^k - z^{k+1}\|^2 - \|z^{k+1} - w^*\|^2 + 2\gamma\langle \nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k), z^k - w^*\rangle \\
&-2\gamma\langle \nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k), z^k - z^{k+1}\rangle + 2\gamma\langle z^k - z^{k+1}, \nabla d_2(w_{d_2}^k)\rangle \\
&+2\gamma^2\langle \nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k), \nabla d_2(w_{d_2}^k)\rangle.
\end{aligned}
$$

Proof We will show that the required inequality holds for every $k \geq 0$. In observance of subgradient inequality, we get

$$
\begin{aligned}
&\quad 2\gamma(d_1(w_{d_1}^k) + d_2(w_{d_3}^k) + d_3(w_{d_3}^k) - (d_1 + d_2 + d_3)(w^*)) \\
&\leq 2\gamma\left(\langle w_{d_1}^k - w^*, \nabla d_1(w_{d_1}^k)\rangle + \langle w_{d_3}^k - w^*, \nabla d_2(w_{d_3}^k) + \nabla d_3(w_{d_3}^k)\rangle\right) \\
&= 2\gamma\langle w_{d_1}^k - w_{d_3}^k, \nabla d_1(w_{d_1}^k)\rangle + 2\gamma\langle w_{d_3}^k - w^*, \nabla d_1(w_{d_1}^k) + \nabla d_2(w_{d_3}^k) + \nabla d_3(w_{d_3}^k)\rangle \\
&= 2\langle w_{d_1}^k - w_{d_3}^k, \gamma\nabla d_1(w_{d_1}^k)\rangle + 2\langle w_{d_3}^k - w^*, w_{d_3}^k - w_{d_1}^k\rangle \text{ from (ii) of Proposition 3} \\
&= 2\langle w_{d_3}^k - w_{d_1}^k, w_{d_3}^k - w^* - \gamma\nabla d_1(w_{d_1}^k)\rangle \\
&= 2\langle w_{d_3}^k - w_{d_2}^k + \gamma(\nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k)), w_{d_3}^k - w^* - \gamma\nabla d_1(w_{d_1}^k)\rangle \text{ from (iv) of Proposition 3} \\
&= 2\langle z^k - z^{k+1} + \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), w_{d_3}^k - w^* - \gamma\nabla d_1(w_{d_1}^k)\rangle \text{ from (v) of Proposition 3} \\
&= 2\langle z^k - z^{k+1} + \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), z^k + w_{d_3}^k - z^k - w^* - \gamma\nabla d_1(w_{d_1}^k)\rangle \\
&= 2\langle z^k - z^{k+1} + \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), z^k - \gamma\nabla d_3(w_{d_3}^k) - w^* - \gamma\nabla d_1(w_{d_1}^k)\rangle \\
&= 2\langle z^k - z^{k+1} + \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), z^k - \gamma\nabla d_3(w_{d_3}^k) - w^* - \gamma\nabla d_1(w_{d_1}^k) - \gamma\nabla d_2(w_{d_2}^k) + \gamma\nabla d_2(w_{d_2}^k)\rangle \\
&= 2\langle z^k - z^{k+1} + \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), z^k - (z^k - z^{k+1}) - w^*\rangle \\
&\quad +2\langle z^k - z^{k+1} + \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), \gamma\nabla d_2(w_{d_2}^k)\rangle \text{ from (v) of Proposition 3} \\
&= 2\langle z^k - z^{k+1}, z^k - w^*\rangle - 2\langle z^k - z^{k+1}, z^k - z^{k+1}\rangle + 2\langle \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), z^k - w^*\rangle \\
&\quad -2\langle \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), z^k - z^{k+1}\rangle + 2\langle z^k - z^{k+1}, \gamma\nabla d_2(w_{d_2}^k)\rangle \\
&\quad +2\langle \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), \gamma\nabla d_2(w_{d_2}^k)\rangle \qquad (27) \\
&= \|z^k - z^{k+1}\|^2 + \|z^k - w^*\|^2 - \|z^{k+1} - w^*\|^2 - 2\|z^k - z^{k+1}\|^2 + 2\langle \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), z^k - w^*\rangle \\
&\quad -2\langle \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), z^k - z^{k+1}\rangle + 2\langle z^k - z^{k+1}, \gamma\nabla d_2(w_{d_2}^k)\rangle \\
&\quad +2\langle \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), \gamma\nabla d_2(w_{d_2}^k)\rangle \text{ from (4)} \\
&= \|z^k - w^*\|^2 - \|z^k - z^{k+1}\|^2 - \|z^{k+1} - w^*\|^2 + 2\gamma\langle \nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k), z^k - w^*\rangle \\
&\quad -2\gamma\langle \nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k), z^k - z^{k+1}\rangle + 2\gamma\langle z^k - z^{k+1}, \nabla d_2(w_{d_2}^k)\rangle \\
&\quad +2\gamma^2\langle \nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k), \nabla d_2(w_{d_2}^k)\rangle.
\end{aligned}
$$

Proposition 5 (Lower Inequality). *Let $w \in \mathcal{X}$ and w^* be the fixed point of the FPI algorithm given in (21). Then, the following inequality holds:*

$$
\begin{aligned}
&2\gamma(d_1(w_{d_1}^k) + d_2(w_{d_3}^k) + d_3(w_{d_3}^k) - (d_1 + d_2 + d_3)(w^*)) \\
&\geq \langle w_{d_2}^k - w_{d_3}^k, \nabla d_1(w^*)\rangle + \langle \gamma\nabla d_2(w_{d_2}^k) - \gamma\nabla d_2(w_{d_3}^k), \nabla d_1(w^*)\rangle.
\end{aligned}
$$

Proof By subgradient inequality and Proposition 3 (iii), we have

$$\begin{aligned}&d_1(w_{d_1}^k) - d_1(w^*)\\ \geq&\langle w_{d_1}^k - w^*, \nabla d_1(w^*)\rangle\\ =&\langle w_{d_1}^k - w_{d_3}^k, \nabla d_1(w^*)\rangle + \langle w_{d_3}^k - w^*, \nabla d_1(w^*)\rangle\\ =&\langle w_{d_2}^k - w_{d_3}^k + \gamma\nabla d_2(w_{d_2}^k) - \gamma\nabla d_2(w_{d_3}^k), \nabla d_1(w^*)\rangle + \langle w_{d_3}^k - w^*, \nabla d_1(w^*)\rangle.\end{aligned}$$

In a similar manner, we obtain

$$\begin{aligned}&d_2(w_{d_3}^k) - d_2(w^*) \geq \langle w_{d_3}^k - w^*, \nabla d_2(w^*)\rangle\\ \text{and}\quad &d_3(w_{d_3}^k) - d_3(w^*) \geq \langle w_{d_3}^k - w^*, \nabla d_3(w^*)\rangle.\end{aligned}$$

On adding the above three relations, we get

$$\begin{aligned}&d_1(w_{d_1}^k) + d_2(w_{d_3}^k) + d_3(w_{d_3}^k) - (d_1 + d_2 + d_3)(w^*)\\ &\geq \langle w_{d_2}^k - w_{d_3}^k, \nabla d_1(w^*)\rangle + \langle w_{d_3}^k - w^*, \nabla d_1(w^*) + \nabla d_2(w^*) + \nabla d_3(w^*)\rangle\\ &+\langle \gamma\nabla d_2(w_{d_2}^k) - \gamma\nabla d_2(w_{d_3}^k), \nabla d_1(w^*)\rangle. \end{aligned}\tag{28}$$

Now, from the optimality condition of a subdifferential set, we can assume that

$$\nabla d_1(w^*) + \nabla d_2(w^*) = -\nabla d_3(w^*) \in \partial d_3(w^*).$$

Therefore, on plugging the above relation in (28), we get

$$\begin{aligned}&d_1(w_{d_1}^k) + d_2(w_{d_3}^k) + d_3(w_{d_3}^k) - (d_1 + d_2 + d_3)(w^*)\\ &\geq \langle w_{d_2}^k - w_{d_3}^k, \nabla d_1(w^*)\rangle + \langle \gamma\nabla d_2(w_{d_2}^k) - \gamma\nabla d_2(w_{d_3}^k), \nabla d_1(w^*)\rangle.\end{aligned}$$

In the next theorem, we prove the convergence rate of the proposed Algorithm (21).

Theorem 4 *Consider the iteration* (21)*. Assume that the function* d_1 *is* L*-Lipschitz continuous on the closed ball* $\overline{B(0, (1 + \gamma/\beta)\|w^0 - w^*\|)}$*. Under the same assumptions in Theorem 1, we have the following convergence*

$$(d_1 + d_2 + d_3)(w_{d_3}^k) - (d_1 + d_2 + d_3)(w^*) = o\left(\frac{1}{\sqrt{k+1}}\right).$$

Proof Note that in view of Lemma 3, the sequences $w_{d_1}^k$ and $w_{d_3}^k$ are within the region where d_1 is L-Lipschitz continuous. Therefore, we have

$$\begin{aligned}&2\gamma\left(d_1(w_{d_1}^k) + d_2(w_{d_3}^k) + d_3(w_{d_3}^k) - (d_1 + d_2 + d_3)(w^*)\right)\\ &\leq 2\gamma\left(d_1(w_{d_3}^k) + d_2(w_{d_3}^k) + d_3(w_{d_3}^k) - (d_1 + d_2 + d_3)(w^*)\right) + 2\gamma L\|w_{d_1}^k - w_{d_3}^k\|.\end{aligned}$$

Now, employing relation (27) of Proposition 4 in above relation, we get

$$\begin{aligned}&2\gamma\left(d_1(w_{d_1}^k) + d_2(w_{d_3}^k) + d_3(w_{d_3}^k) - (d_1 + d_2 + d_3)(w^*)\right)\\ &\leq 2\langle z^k - z^{k+1}, z^k - w^*\rangle - 2\langle z^k - z^{k+1}, z^k - z^{k+1}\rangle + 2\langle \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), z^k - w^*\rangle\\ &-2\langle \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), z^k - z^{k+1}\rangle + 2\langle z^k - z^{k+1}, \gamma\nabla d_2(w_{d_2}^k)\rangle\\ &+2\langle \gamma\nabla d_2(w_{d_3}^k) - \gamma\nabla d_2(w_{d_2}^k), \gamma\nabla d_2(w_{d_2}^k)\rangle + 2\gamma L\|w_{d_1}^k - w_{d_3}^k\| \text{ from (4)}\\ &\leq 2\|z^k - z^{k+1}\|\|z^k - w^*\| + 2\|z^k - z^{k+1}\|\|z^k - z^{k+1}\| + 2\gamma\|\nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k)\|\|z^k - w^*\|\\ &+2\gamma\|\nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k)\|\|z^k - z^{k+1}\| + 2\gamma\|z^k - z^{k+1}\|\|\nabla d_2(w_{d_2}^k)\|\\ &+2\gamma^2\|\nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k)\|\|\nabla d_2(w_{d_2}^k)\| + 2\gamma L\|w_{d_1}^k - w_{d_3}^k\|\end{aligned}$$

Now, note that in view of Theorem 1, z^k is bounded. Therefore, in view of Lipschitz's continuity of $\text{prox}_{d_3}^{\gamma}$ and ∇d_2, we have that $w_{d_3}^k = \text{prox}_{d_3}^{\gamma}(z^k)$ and $\nabla d_2(w_{d_3}^k)$ are bounded. Similarly, $\nabla d_2(w_{d_2}^k)$ is also bounded.
By Theorem 1, $\|z^k - z^{k+1}\| = o\left(\frac{1}{\sqrt{k+1}}\right)$. By Proposition 3 (v), $\|w_{d_2}^k - w_{d_3}^k\| = o\left(\frac{1}{\sqrt{k+1}}\right)$. The Lipschitz continuity of ∇d_2, the gradients implies $\|\nabla d_2(w_{d_2}^k) - \nabla d_2(w_{d_3}^k)\| = o\left(\frac{1}{\sqrt{k+1}}\right)$. By Proposition 3 (iv), we also have $\|w_{d_1}^k - w_{d_3}^k\| = o\left(\frac{1}{\sqrt{k+1}}\right)$. Thus the order $o\left(\frac{1}{\sqrt{k+1}}\right)$ is proven.

5.4 Multi-operator splitting schemes

The Algorithm 2 can be extended to a 4-block problem with two of the objective functions being strongly convex. Consider the following problem given by

$$\left.\begin{array}{ll} \min & f_1(x_1) + f_2(x_2) + f_3(x_3) + f_4(x_4), \\ \text{subject to} & A_1x_1 + A_2x_2 + A_3x_3 + A_4x_4 = b, \\ & x_1 \in \mathcal{X}_1,\ x_2 \in \mathcal{X}_2, x_3 \in \mathcal{X}_3, x_4 \in \mathcal{X}_4 \end{array}\right\},$$

where for each $i = 1, 2, 3, 4$, the functions $f_i : \mathbb{R}^{n_i} \to \mathbb{R} \cup \{+\infty\}$ are proper, closed, and convex (not necessarily smooth) with the assumption that two of the functions are μ-strongly convex and $\mu > 0$. For each $i = 1, 2, 3, 4$, $A_i \in \mathbb{R}^{m \times n_i}$ and $b \in \mathbb{R}^4$. Then from the 4-block ADMM for the problem above, we can similarly derive the following 4-operator splitting for minimizing the dual problem $\sum_{i=1}^{4} d_i(x)$ where each d_i is a proper closed convex function, and at least two of them are strongly convex functions.

The extended 4-operator splitting algorithm is given as follows:

Algorithm 3 Initialize $z^0 \in \mathcal{X}$, $\gamma \in (0, \min\{L_2, L_3\})$. L_i is Lipschitz constant of ∇d_i.

1. Compute $w^{k+1} = \text{prox}_{d_4}^{\gamma}(z^k)$;
2. Compute $p^{k+1} = \text{prox}_{d_1}^{\gamma}(2w^{k+1} - z^k - \gamma\nabla d_2(w^{k+1}) - \gamma\nabla d_3(w^{k+1}))$;
3. Compute $v_1^{k+1} = \text{prox}_{d_2}^{\gamma}(p^{k+1} + \gamma\nabla d_2(w^{k+1}))$;
4. Compute $v_2^{k+1} = \text{prox}_{d_3}^{\gamma}(v_1^{k+1} + \gamma\nabla d_3(w^{k+1}))$;
5. Update $z^{k+1} = z^k + v_2^{k+1} - w^{k+1}$.

Remark 1 In general, a splitting scheme for m-operator can be similarly derived from m-block ADMM for minimizing $\sum_{i=1}^{m} d_i(x)$ where each d_i is a proper closed convex function, and at least $m-2$ of them have Lipschitz continuous gradients.

6 Applications and Numerical Results

In this section, we demonstrate the working of our proposed Algorithm 1 and 2. Next, we compare the behavior of the proposed Algorithm 1 with the method in [8].

Example 1 Consider the following problem given by

$$\min_{x\in\mathbb{R}^n} f(x)+g(Lx)+h(x), \tag{29}$$

where $f,\ g$, and h are proper, closed, and convex functions, and g is $(1/\beta)$-Lipschitz differentiable, and L is linear mapping. The proposed Algorithm 1 applies here with the following monotone operators given by:

$$\mathbb{A}=\partial f;\ \mathbb{B}=\nabla(g\circ L)=L^*\circ\nabla g\circ L;\ \mathbb{C}=\partial h$$

If zer $(\partial f+\nabla g+\partial h)\neq\emptyset$. Then, x^k is a weakly minimal solution to (29). The modified form of Algorithm 1 for problem (29) is discussed below:

Algorithm 4

Initialize an arbitrary $z^0\in\mathcal{X}$, stepsize $\gamma\in(0,2\beta/\|L\|^2)$ and $(\lambda_j)_{j\geq 0}\in(0,(4\beta-\gamma\|L\|^2)/2\beta)$.
For $k=0,1,\ldots$

1. Compute $x^{k+1}=\text{prox}_h^\gamma(z^k)$;
2. Compute $y^{k+1}=Lx^{k+1}$;
3. Compute $p^{k+1}=\text{prox}_f^\gamma(2x^{k+1}-z^k-\gamma L^*\nabla g(y^{k+1}))$;
4. Compute $v^{k+1}=\text{prox}_g^\gamma(p^{k+1}+\gamma\nabla g(y^{k+1}))$;
5. Update $z^{k+1}=z^k+\lambda_k(v^{k+1}-w^{k+1})$.

Example 2 Consider the following problem given by

$$\left.\begin{array}{ll}\min\limits_{x\in\mathbb{R}^n} & \frac{\alpha}{2}\|x-u\|_2^2+i_{\Lambda_1}(x)+i_{\Lambda_2}(x),\\ \text{subject to} & \Lambda_1=\{x: m\leq x_i\leq M,\quad\forall i\}\\ \text{and} & \Lambda_2=\{x: Ax=b\},\end{array}\right\} \tag{30}$$

where $A=[1,1,\ldots,1]$ and $b\in\mathbb{R}$. Such a simple problem can be used as a postprocessing step to enforce bounds for solving complicated PDEs [21–23].

Notice that this simple constrained minimization (30) can also be solved directly via the KKT system of the Lagrangian, which however might be less efficient than splitting methods for large problems, see a comparison of DRS with a direct solver of KKT system in [21, Appendix]. Moreover, following the analysis in [9], a sharp local linear convergence rate of Douglas-Rachford splitting for (30) can be derived, which can be further used to design optimal step size [23]. Though (30) can be solved by two-operator splitting, a more general version of (30) can no longer be easily solved by Douglas-Rachford splitting. For example, for stabilizing numerical schemes solving gas dynamics equations [30,31], the bound-preserving constraint $\Lambda_1=\{x: m\leq x_i\leq M,\quad\forall i\}$ in (30) would be replaced by the invariant domain preserving constraint $\Lambda_1=\{x\in\mathbb{R}^{n\times 3}: x_i\in G\subset\mathbb{R}^3\quad\forall i\}$ for some convex invariant domain set G, which a two-operator splitting cannot easily handle. Instead, a three operator splitting like three-operator Davis-Yin splitting or the proposed splitting can be used.

We consider a comparison of Davis-Yin splitting and the proposed three-operator splitting. Let $f(x)=\frac{\alpha}{2}\|x-u\|_2^2, g(x)=i_{\Lambda_1}(x)$, and $h(x)=i_{\Lambda_2}(x)$. The proposed Algorithm 1 applies here with the following operators given by:

$$\mathbb{A} = \partial g = \begin{cases} [0, +\infty], & \text{if } x_i = M \\ 0, & \text{if } x_i \in (m, M) \\ [-\infty, 0], & \text{if } x_i = m, \end{cases},$$

$$\text{and } \mathbb{B} = \partial h = \mathcal{R}(A^\top),$$

$$\text{and } \mathbb{C} = \partial f = \alpha(x - u).$$

Equivalently, we can set $d_1 = g,\ d_2 = f,\ d_3 = h$ in Algorithm 2 or the scheme (2) to obtain

$$\begin{aligned} x^{k+\frac{1}{2}} &= A^+(b - Az^k) + z^k \\ p^{k+1} &= \min(\max(2x^{k+\frac{1}{2}} - z^k + \alpha\gamma(x^{k+\frac{1}{2}} - u), m), M) \\ x^{k+1} &= \tfrac{1}{\alpha\gamma+1} p^{k+1} + \tfrac{\alpha\gamma}{\alpha\gamma+1} u \\ z^{k+1} &= z^k + x^{k+1} - x^{k+\frac{1}{2}}. \end{aligned}$$

The Davis-Yin scheme (1) with $d_1 = g,\ d_2 = f,\ d_3 = h$ becomes

$$\begin{aligned} x^{k+\frac{1}{2}} &= A^+(b - Az^k) + z^k \\ x^{k+1} &= \min(\max(2x^{k+\frac{1}{2}} - z^k + \alpha\gamma(x^{k+\frac{1}{2}} - u), m), M) \\ z^{k+1} &= z^k + x^{k+1} - x^{k+\frac{1}{2}}. \end{aligned}$$

We compare the the proposed splitting scheme (2) and Davis-Yin method (1) on the problem (30) with $\alpha = 1, n = 100, m = -1, M = 1$, and $b = Au, u \in \mathbb{R}^n$ where u is constructed by perturbing a sine profile by random noise:

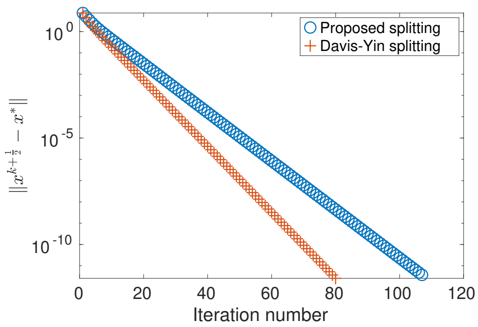
$$u_i = \sin(2\pi \frac{i}{n}) + 0.8 * \mathcal{N}(0, 1).$$

The number of entries in u greater than $M = 1$ is 17 and the number of entries in u less than $m = -1$ is 20. The minimizer x^* to (30) satisfies $x_i^* \in [m, M]$. The error measured by $\|x^{k+\frac{1}{2}} - x^*\|$ is shown in Figure 2.

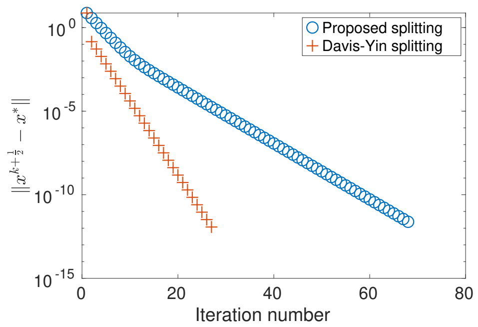
Let L be the Lipschitz constant of ∇f. As shown in Figure 2, the Davis-Yin method performs the best if using step size $\gamma = \frac{1}{L}$, and the proposed method is not faster than the Davis-Yin method if $\gamma < \frac{2}{L}$. On the other hand, Davis-Yin method will not converge if step size γ is much larger than $\frac{2}{L}$, but the proposed splitting method can still converge with such a large step size.

7 Concluding remarks

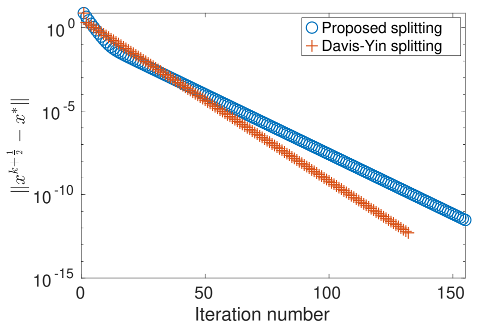
In this paper, we have considered a three-operator splitting scheme for solving monotone inclusion problems, which is an extension of the Douglas-Rachford splitting. In practice, it can allow larger range of step size, compared to the Davis-Yin Splitting. Some convergence properties are discussed under the assumption that the splitting operator is averaged.



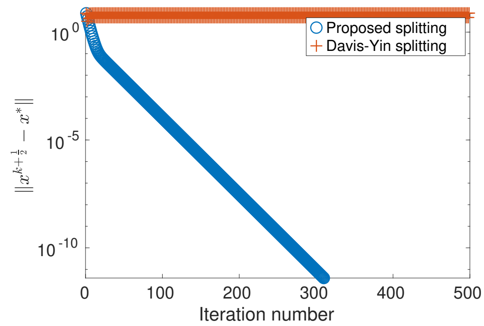
(a) $\gamma = \frac{0.3}{L}$



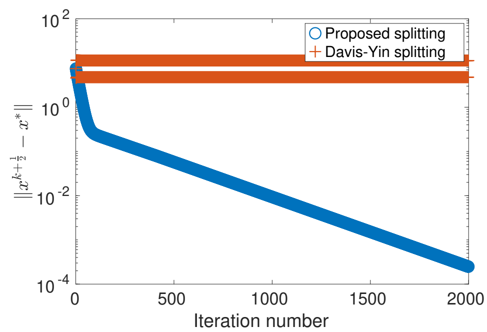
(b) $\gamma = \frac{1}{L}$



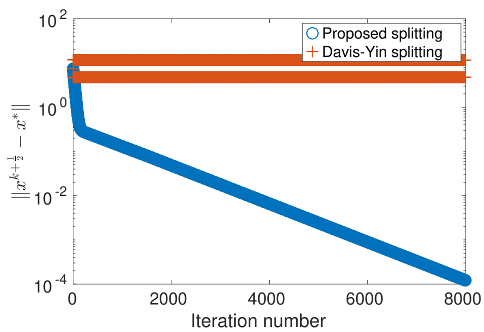
(c) $\gamma = \frac{1.8}{L}$



(d) $\gamma = \frac{3}{L}$



(e) $\gamma = \frac{20}{L}$



(f) $\gamma = \frac{40}{L}$

Fig. 2: Algorithm 2 on Example 2 for different values of step size γ. In this example, the Lipschitz constant of ∇f is $L = \alpha$

Acknowledgements The authors are grateful to the Purdue-India partnership program to make the collaboration possible.

Author Contributions All authors have contributed to the conception and design of the study, analysis, and numerical tests. The manuscript was mainly written and edited by AA and XZ. All authors read and approved the final manuscript.

Funding Anshika acknowledges a Science and Engineering Research Board-Overseas Visiting Doctoral Fellowship, Purdue-India partnership, with file number SB/S9/Z-03/2017-II (2022). X. Zhang is supported by NSF DMS-2208515. D. Ghosh acknowledges the financial support of the research grants MATRICS (MTR/2021/000696) and Core Research Grant (CRG/2022/001347) by the Science and Engineering Research Board, India.

Data availability Not applicable.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

References

1. Bauschke, H.H., Combettes, P.L., Bauschke, H.H., Combettes, P.L.: Correction to: Convex analysis and monotone operator theory in hilbert spaces. Springer (2017)
2. Boţ, R.I., Nguyen, D.K.: Fast Krasnosel'skiĭ–Mann Algorithm with a Convergence Rate of the Fixed Point Iteration of $o(\frac{1}{k})$. SIAM Journal on Numerical Analysis **61**(6), 2813–2843 (2023)
3. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., et al.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning **3**(1), 1–122 (2011)
4. Briceno-Arias, L.M.: Forward-douglas–rachford splitting and forward-partial inverse method for solving monotone inclusions. Optimization **64**(5), 1239–1261 (2015)
5. Cai, X., Han, D., Yuan, X.: The direct extension of ADMM for three-block separable convex minimization models is convergent when one function is strongly convex. Optimization Online **229**, 230 (2014)
6. Chen, C., Shen, Y., You, Y.: On the convergence analysis of the alternating direction method of multipliers with three blocks. In: Abstract and Applied Analysis, vol. 2013, p. 183961. Wiley Online Library (2013)
7. Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. Splitting methods in communication, imaging, science, and engineering pp. 115–163 (2016)
8. Davis, D., Yin, W.: A three-operator splitting scheme and its optimization applications. Set-valued and variational analysis **25**, 829–858 (2017)
9. Demanet, L., Zhang, X.: Eventual linear convergence of the Douglas-Rachford iteration for basis pursuit. Mathematics of Computation **85**(297), 209–238 (2016)
10. Douglas, J., Rachford, H.H.: On the numerical solution of heat conduction problems in two and three space variables. Transactions of the American mathematical Society **82**(2), 421–439 (1956)
11. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Computers & mathematics with applications **2**(1), 17–40 (1976)
12. Glowinski, R., Marroco, A.: Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique **9**(R2), 41–76 (1975)
13. Goldstein, T., Osher, S.: The split bregman method for l1-regularized problems. SIAM journal on imaging sciences **2**(2), 323–343 (2009)
14. Han, D., Yuan, X.: A note on the alternating direction method of multipliers. Journal of Optimization Theory and Applications **155**, 227–238 (2012)
15. Hong, M., Luo, Z.Q.: On the linear convergence of the alternating direction method of multipliers. Mathematical Programming **162**(1), 165–199 (2017)
16. Li, M., Sun, D., Toh, K.C.: A convergent 3-block semi-proximal ADMM for convex minimization problems with one strongly convex block. Asia-Pacific Journal of Operational Research **32**(04), 1550024 (2015)

17. Lin, T., Ma, S., Zhang, S.: On the global linear convergence of the ADMM with multiblock variables. SIAM Journal on Optimization **25**(3), 1478–1497 (2015)
18. Lin, T., Ma, S., Zhang, S.: Iteration complexity analysis of multi-block ADMM for a family of convex minimization without strong convexity. Journal of Scientific Computing **69**(1), 52–81 (2016)
19. Lin, T.Y., Ma, S.Q., Zhang, S.Z.: On the sublinear convergence rate of multi-block ADMM. Journal of the Operations Research Society of China **3**, 251–274 (2015)
20. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM Journal on Numerical Analysis **16**(6), 964–979 (1979)
21. Liu, C., Buzzard, G.T., Zhang, X.: An optimization based limiter for enforcing positivity in a semi-implicit discontinuous Galerkin scheme for compressible Navier–Stokes equations. Journal of Computational Physics **519**, 113440 (2024)
22. Liu, C., Hu, J., Taitano, W.T., Zhang, X.: An optimization-based positivity-preserving limiter in semi-implicit discontinuous Galerkin schemes solving Fokker-Planck equations (2024). URL https://arxiv.org/abs/2410.19143
23. Liu, C., Riviere, B., Shen, J., Zhang, X.: A simple and efficient convex optimization based bound-preserving high order accurate limiter for Cahn–Hilliard–Navier–Stokes system. SIAM Journal on Scientific Computing **46**(3), A1923–A1948 (2024)
24. Passty, G.B.: Ergodic convergence to a zero of the sum of monotone operators in hilbert space. Journal of Mathematical Analysis and Applications **72**(2), 383–390 (1979)
25. Peaceman, D.W., Rachford Jr, H.H.: The numerical solution of parabolic and elliptic differential equations. Journal of the Society for industrial and Applied Mathematics **3**(1), 28–41 (1955)
26. Raguet, H.: A note on the forward-douglas–rachford splitting for monotone inclusion and convex optimization. Optimization Letters **13**, 717–740 (2019)
27. Raguet, H., Fadili, J., Peyré, G.: A generalized forward-backward splitting. SIAM Journal on Imaging Sciences **6**(3), 1199–1226 (2013)
28. Shi, W., Ling, Q., Yuan, K., Wu, G., Yin, W.: On the linear convergence of the ADMM in decentralized consensus optimization. IEEE Transactions on Signal Processing **62**(7), 1750–1761 (2014)
29. Wei, E., Ozdaglar, A.: Distributed alternating direction method of multipliers. In: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), pp. 5445–5450. IEEE (2012)
30. Zhang, X.: On positivity-preserving high order discontinuous Galerkin schemes for compressible Navier–Stokes equations. Journal of Computational Physics **328**, 301–343 (2017)
31. Zhang, X., Shu, C.W.: A minimum entropy principle of high order schemes for gas dynamics equations. Numerische Mathematik **121**(3), 545–563 (2012)