



A three-operator splitting scheme derived from three-block ADMM

Anshika Anshika¹ · Jiaxing Li² · Debdas Ghosh¹ · Xiangxiong Zhang²

Received: 30 October 2024 / Revised: 2 October 2025 / Accepted: 15 November 2025
© The Author(s) 2025

Abstract

This work presents a new three-operator splitting method to handle monotone inclusion and convex optimization problems. The proposed splitting serves as another natural extension of the Douglas-Rachford splitting technique to problems involving three operators. For solving a composite convex minimization of a sum of three functions, its formula resembles but is different from Davis-Yin splitting and the dual formulation of the classical three-block ADMM. Numerical tests suggest that such a splitting scheme is robust in the sense of allowing larger step sizes. When two functions have orthogonal domains, the splitting operator can be proven $1/2$ -averaged, which implies convergence of the iteration scheme using any positive step size.

Keywords Three-operator splitting · Convex optimization · Forward-Douglas-Rachford splitting · Davis-Yin splitting · ADMM methods · Monotone inclusion

1 Introduction

Operator splitting methods break down complex problems into smaller, more manageable subproblems that can be tackled either one after another or at the same time. While these methods have been around for over six decades, their relevance has grown

✉ Xiangxiong Zhang
zhan1966@purdue.edu

Anshika Anshika
anshika.rs.mat19@itbhu.ac.in

Jiaxing Li
li4944@purdue.edu

Debdas Ghosh
debdas.mat@itbhu.ac.in

¹ Department of Mathematical Sciences, Indian Institute of Technology (Banaras Hindu University), Varanasi, Uttar Pradesh 221005, India

² Department of Mathematics, Purdue University, 150 North University Street, West Lafayette, Indiana 47907, USA

considerably in recent years. They have proven to be highly effective across multiple fields ranging—from partial differential equations, control theory, and high dimensional problems in machine learning, image analysis, signal processing, statistical estimation, image processing, compressive sensing, finance, matrix completion, and control (Shi et al. 2014; Wei and Ozdaglar 2012; Boyd et al. 2011).

In 1950s, (Peaceman and Rachford 1955) and (Douglas and Rachford 1956) operator splitting techniques were first introduced for solving the heat equation in two dimensions. In Lions and Mercier (1979), Lions and Mercier extended these methods to address the formulation involving the sum of two operators, each being maximal and monotone, and the method in Lions and Mercier (1979) is named Douglas-Rachford splitting (DRS). In Raguét et al. (2013); Raguét (2019), it was shown that Douglas-Rachford and Forward-Backward splitting can be integrated, followed by its extension in Briceno-Arias (2015). Such a Forward-Douglas-Rachford (FDR) splitting was proven convergent by Davis and Yin in Davis and Yin (2017), and also referred to as Davis-Yin splitting. It is also well known that DRS is equivalent to the Alternating Direction Method of Multipliers (ADMM), which was studied in the dual setting by Gabay and Mercier (1976), Glowinski and Marroco (1975). DRS and ADMM have been a popular tool in many applications. In particular, (Goldstein and Osher 2009) later advanced the application of ADMM through the development of the split Bregman method, particularly in imaging and sparse recovery contexts.

In this paper, we propose a splitting scheme which can be used for solving an inclusion problem $0 \in (\mathbb{A} + \mathbb{B} + \mathbb{C})x$, with operators $\mathbb{A}, \mathbb{B}, \mathbb{C}$ being three maximal monotone mappings over a real Hilbert space \mathcal{X} and the operator \mathbb{C} is cocoerceive with parameter β . Let $\mathbb{J}_{\mathbb{A}} = (I + \mathbb{A})^{-1}$ denote the resolvent of a monotone operator \mathbb{A} . The proposed splitting operator is given as

$$T = \mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C} \circ \mathbb{J}_{\gamma\mathbb{B}}) + \gamma\mathbb{C} \circ \mathbb{J}_{\gamma\mathbb{B}}) + (I - \mathbb{J}_{\gamma\mathbb{B}}). \quad (1)$$

Such a splitting is inspired by the Davis-Yin splitting (Davis and Yin 2017), which can be written as

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

The main application of such a splitting is to solve a composite convex minimization in the form of

$$\min_x d_1(x) + d_2(x) + d_3(x),$$

where $d_i(x)$ are convex closed and proper functions, and their proximal mapping

$$\text{prox}_{\gamma d_i}(y) := \text{argmin}_x d_i(x) + \frac{1}{2\gamma} \|x - y\|^2$$

can be efficiently computed. When one of the functions has Lipschitz-continuous gradient, e.g., assume ∇d_2 is Lipschitz continuous, characterized by a Lipschitz constant

L , our proposed splitting scheme with a step size $\gamma > 0$ is written as

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

The Davis-Yin splitting (or Forward-Douglas-Rachford splitting) (Davis and Yin 2017; Raguet 2019) can be written as

$$\begin{aligned}
 & x^{k+\frac{1}{2}} = \text{prox}_{\gamma d_3}(z^k) \\
 \text{Davis-Yin Splitting : } & x^{k+1} = \text{prox}_{\gamma d_1}(2x^{k+\frac{1}{2}} - z^k - \gamma \nabla d_2(x^{k+\frac{1}{2}})) \\
 & z^{k+1} = z^k + (x^{k+1} - x^{k+\frac{1}{2}}). \tag{3}
 \end{aligned}$$

As will be shown in Section 5.2, the dual form of the classical three-block ADMM method with a penalty parameter or step size $\gamma > 0$ can be written as

$$\begin{aligned}
 & x^{k+\frac{1}{2}} = \text{prox}_{\gamma d_3}(z^k) \\
 \text{ADMM Dual Form : } & p^{k+1} = \text{prox}_{\gamma d_1}(2x^{k+\frac{1}{2}} - z^k - \gamma \nabla d_2(x^k)) \\
 & x^{k+1} = \text{prox}_{\gamma d_2}(p^{k+1} + \gamma \nabla d_2(x^k)) \\
 & z^{k+1} = z^k + (x^{k+1} - x^{k+\frac{1}{2}}). \tag{4}
 \end{aligned}$$

Our proposed scheme is motivated by the dual form of the empirically efficient, classical three-block ADMM, from which we derive a new method to solve (8). In numerical experiments, the original dual form converges only when the step size is constrained within a very limited range. By introducing a simple modification to this dual form, our new scheme significantly widens the permissible range of the step size, thereby improving its robustness and convergence properties. We can see that the proposed new scheme (2) is similar to but different from the Davis-Yin scheme (3) and the ADMM dual form (4). The proposed splitting method may appear less efficient than the Davis-Yin scheme, as it involves evaluating three proximal operators instead of two. On the other hand, the extra computation of $\text{prox}_{\gamma d_2}$ might improve the robustness of the splitting. In Davis and Yin (2017), the Davis-Yin splitting was proven to converge for any constant step size $\gamma \in (0, \frac{2}{L})$. Numerically, when the step size γ is much larger than $\frac{2}{L}$, the Davis-Yin splitting (3) and the ADMM dual form (4) will not converge, and the proposed splitting method (2) can still converge, as will be shown in numerical examples in Section 6.

Notice that the proposed new scheme (2) can be regarded as a modified version of the ADMM dual form (4). Thus the proposed splitting scheme (2) can be also be formally extended to multiple operators by similarly modifying the dual form of

multiple-block ADMM methods as will be shown in Section 5. In the literature, there are several studies of ADMM methods that address the generic case where the number of blocks $m \geq 3$. Despite ADMM's widespread success in two-block convex optimization, the direct extension of ADMM to three or more blocks was shown to be not necessarily convergent (Chen et al. 2016). A key result in this direction is the provision of a concrete example demonstrating divergence of the direct three-block ADMM, marking a fundamental limitation of naive extensions. To address this, Cai et al. (2017) proposed a convergent three-block ADMM, assuming strong convexity of one function and establishing global convergence under additional structural conditions. Later, He and Yuan introduced a generalized symmetric ADMM (Bai et al. 2018) which permits larger dual step sizes while retaining convergence guarantees for separable multi-block convex programs. Beyond ADMM, notable progress has also been made in Forward-Backward-type primal-dual algorithms. The classical Chambolle-Pock (CP) algorithm (Chambolle and Pock 2011) was widely used for imaging problems and saddle-point formulations, but its behavior under inertial parameters in the range $(0, 1)$ was not well understood. Recent works such as He et al. (2022) and Bai et al. (2025) addressed this by introducing corrected Forward-Backward-Adjoint frameworks. These not only clarified convergence in the inertial regime but also resolved long-standing questions about the robustness of primal-dual methods when extrapolation or momentum steps are introduced. These developments form a critical part of the ongoing effort to restore convergence in multi-block settings—an area of direct relevance to our work. In Han and Yuan (2012), the strong convexity assumption has been assumed on all the given objective functions. Lin et al. in Lin et al. (2015) considered $(m - 1)$ functions to be strongly convex and established global convergence without imposing any restrictions on the penalty parameter. Furthermore, Lin et al. (2015) established linear convergence assuming that the objective functions satisfy Lipschitz continuity. In a sequel, the work in Hong and Luo (2017) showed that linear convergence is ensured if the step size in each updating step is sufficiently reduced and an appropriate error-bound assumption holds. Moreover, it has been discussed in Cai et al. (2017), Chen et al. (2013), Han and Yuan (2012), Li et al. (2015), Lin et al. (2015) that the penalty parameter must be suitably controlled for favourable convergence behaviour. The importance of restricting the penalty parameter to ensure faster convergence has been discussed extensively in Cai et al. (2017), Chen et al. (2013), Han and Yuan (2012), Li et al. (2015), Lin et al. (2015). Although the restriction can be conservative to ensure convergence, it may be relaxed to achieve faster rates. In Davis and Yin (2016), Davis and Yin proposed a convergent three-block ADMM variant by assuming strong convexity in one objective component and the step size parameter is bounded by a threshold value. In Lin et al. (2016), Lin et al. proposed several alternative methods for three-block ADMM without any restriction on penalty parameters to solve regularized least square decomposition problems.

The Davis-Yin splitting operator T_{DY} is an averaged operator for any $\gamma \in (0, 2\beta)$ thus converges under suitable assumptions, as proven in Davis and Yin (2017). It is however nontrivial to prove the averagedness of the operator T defined in (1) under the same assumptions, due to its more complicated structure. In Section 5.3, we consider a special assumption that $d_1(x)$ and $d_2(x)$ have orthogonal domains, under which the new splitting operator can be proven averaged thus the proposed scheme (2) converges

with any step size $\gamma > 0$. Such a special assumption is similar to but weaker than the sufficient condition $A_1^\top A_2 = 0$ to ensure convergence of three-block ADMM in Chen et al. (2016), as will be explained in Remark 5.4. On the other hand, the Davis-Yin splitting still needs the step size constraint $\gamma < \frac{2}{L}$ for convergence even under this special assumption, as will be explained in Remark 5.9.

The main contributions of this paper include the introduction to a new three-operator splitting scheme. As will be reviewed in Section 3.1, there are other three-operator splitting schemes. To the best of our knowledge, (1) is different from the existing three-operator splitting schemes. As will be shown in Section 6, the scheme (2) uses only one parameter γ and it is numerically more robust when $\gamma > \frac{2}{L}$ than existing one parameter three-operator splitting schemes such as Davis-Yin splitting. Though there are other three-operator splitting schemes which can also allow much larger step sizes, these schemes usually involve tuning another parameter for convergence. In other words, the main practical advantage of (2) is its robustness without further tuning parameters. Allowing large step sizes could be useful when the Lipschitz constant L is unknown and hard to estimate. We also prove its convergence using any step size $\gamma > 0$ for the composite minimization under one special assumption that d_1 and d_2 have orthogonal domains. In the Appendix, we derive the primal form of the proposed splitting scheme (2), which gives a 3-block ADMM type scheme. Compared to many existing 3-block ADMM variants, this particular variant does not seem to have any advantage, and the Appendix only serves the purpose of comparing it with existing methods for interested readers. On the other hand, the proposed operator splitting (1) can also be used to solve an operator inclusion problem like $0 \in (\mathbb{A} + \mathbb{B} + \mathbb{C})x$, which however ADMM type methods do not directly solve. We emphasize that there are some important modern data science applications which can only be formulated as such a monotone operator inclusion problem but not a convex minimization problem, e.g., (Nurbekyan et al. 2024).

The rest of this paper is organized as follows. Section 2 describes the notation and symbols. In Section 3, we first review some existing three-operator splitting schemes then introduce the proposed new splitting operator. Section 4 includes standard results under the assumption that the splitting operator is averaged. In Section 5, we prove the splitting operator is averaged for the composite convex minimization under the assumption that two functions have orthogonal domains. Section 6 presents numerical examples, followed by concluding remarks in Section 7.

2 Basic notation and fundamental results

We adopt the following notation for use throughout the paper.

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

We now present some standard definitions and basic facts that will be used throughout.

A map $F : S \rightarrow \mathcal{X}$, defined on a nonempty subset $S \subseteq \mathcal{X}$ is said to be L -Lipschitz for $L \geq 0$ if for every $x, y \in S$

$$\|F(x) - F(y)\| \leq L\|x - y\|.$$

F is said to be nonexpansive if the above inequality holds for $L = 1$.

Let $I_{\mathcal{X}}$ be the identity map. A map $F_{\alpha} : S \rightarrow \mathcal{X}$ 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, if F is $(1/2)$ -averaged, then it is said to be firmly nonexpansive. For convenience, we may use I in place of $I_{\mathcal{X}}$.

Let $2^{\mathcal{X}}$ denotes the power set of \mathcal{X} . An operator $\mathbb{A} : \mathcal{X} \rightarrow 2^{\mathcal{X}}$ is said to be monotone if for every $x, y \in \mathcal{X}$, $u \in \mathbb{A}x$, $v \in \mathbb{A}y$, the inequality

$$\langle x - y, u - v \rangle \geq 0 \text{ holds.}$$

The solution set (also called the zeroes set) of a monotone operator \mathbb{A} is defined as

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

An operator \mathbb{A} is strongly monotone with parameter $\beta > 0$ if

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

for all $x, y \in \mathcal{X}$ and $u \in \mathbb{A}x$, $v \in \mathbb{A}y$.

An operator \mathbb{A} is called cocoerceive with constant $\beta > 0$ if

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

Additionally, from the definition of \mathbb{A} , Cauchy-Schwarz inequality and β -cocoerceive of \mathbb{A} , it follows that

$$\|x - y\| \geq \beta\|u - v\|, \text{ for all } x, y \in \mathcal{X}, u \in \mathbb{A}x, v \in \mathbb{A}y, \beta > 0. \quad (5)$$

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

The operator \mathbb{A} has a graph defined as

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

This $\text{gra}(\mathbb{A})$ is called maximal monotone if it is not a proper subset of the graph of any other monotone operator.

The graph of the inverse operator is defined by

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

The resolvent and reflection of a monotone operator \mathbb{A} is denoted by $\mathbb{J}_{\mathbb{A}}$ and $\mathbf{R}_{\mathbb{A}}$, respectively, and defined by

$$\mathbb{J}_{\mathbb{A}} = (I + \mathbb{A})^{-1} \text{ and } \mathbf{R}_{\mathbb{A}} = 2\mathbb{J}_{\mathbb{A}} - I.$$

For maximal monotone \mathbb{A} , $\mathbf{R}_{\mathbb{A}}$ is nonexpansive.

For an extended function $f : \mathcal{X} \rightarrow (-\infty, \infty]$ that is proper, closed (or equivalently *lower semi-continuous*), and convex, its subdifferential set at x is denoted by a map $\partial f : \mathcal{X} \rightarrow 2^{\mathcal{X}}$,

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

Moreover, if f is differentiable at x , then $\nabla f(x) \in \partial f(x)$.

The convex (or Fenchel) conjugate of f is given by

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

The indicator function $i_C(x)$ of a closed convex set $C \subseteq \mathcal{X}$ is

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

The proximal and reflection operators for f with a step size $\lambda > 0$ are

$$\text{prox}_{\lambda f}(x) = \arg \min_{y \in \mathcal{X}} (f(y) + \frac{1}{2\lambda} \|y - x\|^2) \text{ and } \mathbf{R}_{\lambda f} = 2 \text{prox}_{\lambda f} - I.$$

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

$$2\langle y - x, z - x \rangle = \|y - x\|^2 + \|z - x\|^2 - \|y - z\|^2 \text{ for all } x, y, z \in \mathcal{X} \quad (6)$$

$$\text{and } ab \leq \frac{a^2}{2\varepsilon} + \frac{b^2\varepsilon}{2} \text{ for all } a, b \geq 0, \varepsilon > 0, \text{ respectively.} \quad (7)$$

3 A three-operator splitting scheme

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

$$\text{find } x \in \mathcal{X} \text{ such that } 0 \in (\mathbb{A} + \mathbb{B} + \mathbb{C})x, \quad (8)$$

where $\mathbb{A}, \mathbb{B}, \mathbb{C}$ are three maximal monotone operators defined on a real Hilbert space \mathcal{X} and the operator \mathbb{C} is cocoerceive with parameter β .

3.1 Existing three-operator splitting schemes

We first review a few existing three-operator splitting schemes. For solving the problem (8), there are at least the following three-operator splitting methods:

1. The Forward-Douglas-Rachford splitting is also known as Davis-Yin splitting T_{DY} , and its convergence was proven by Davis and Yin in Davis and Yin (2017). See also (Raguet et al. 2013; Raguet 2019) for the extension of Forward-Douglas-Rachford splitting to multiple operators.
2. In Malitsky and Tam (2020), Malitsky and Tam introduced a Forward-Reflected-Backward splitting method with a generalization for three operators. This method requires \mathbb{A} to be maximal monotone, \mathbb{B} to be monotone and Lipschitz, and \mathbb{C} to be cocoerceive. The iteration scheme can be written as

$$x_{k+1} = \mathbb{J}_{\gamma\mathbb{A}}(x_k - 2\gamma\mathbb{B}(x_k) + \gamma\mathbb{B}(x_{k-1}) - \gamma\mathbb{C}(x_k)). \quad (9)$$

3. In Ryu and Vũ (2020), Ryu and Vu combined Douglas-Rachford splitting with Forward-Backward-Forward splitting (FDRF) under the assumption that two operators are maximal monotone and the other is monotone and Lipschitz,

$$T_{\text{FDRF}} := (I - \gamma\mathbb{C}) \circ \mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C} \circ \mathbb{J}_{\gamma\mathbb{B}}) + I - (I - \gamma\mathbb{C}) \circ \mathbb{J}_{\gamma\mathbb{B}}. \quad (10)$$

The FDRF iteration can be written as

$$\begin{aligned} x_{n+1} &= \mathbb{J}_{\gamma\mathbb{B}}z_n \\ y_{n+1} &= \mathbb{J}_{\gamma\mathbb{A}}(2x_{n+1} - z_n - \gamma\mathbb{C}x_{n+1}) \\ z_{n+1} &= z_n + y_{n+1} - x_{n+1} - \gamma(\mathbb{C}y_{n+1} - \mathbb{C}x_{n+1}) \quad (\text{FDRF}). \end{aligned} \quad (11)$$

For relaxing assumptions needed for convergence, in Ryu and Vũ (2020) Ryu and Vu also proposed a method that combines DR with Forward-Reflected-Backward Splitting (FRDR) with an extra parameter β , which can be written as

$$\begin{aligned} x_{n+1} &= \mathbb{J}_{\gamma\mathbb{B}}(x_n - \gamma u_n - \gamma(2\mathbb{C}x_n - \mathbb{C}x_{n-1})) \\ y_{n+1} &= \mathbb{J}_{\beta\mathbb{A}}(2x_{n+1} - x_n + \beta u_n) \\ u_{n+1} &= u_n + \frac{1}{\beta}(2x_{n+1} - x_n - y_{n+1}) \quad (\text{FRDR}). \end{aligned} \quad (12)$$

One application for these three-operator splitting schemes is to solve a composite convex minimization problem in the following form

$$x^* \in \arg \min_{x \in \mathcal{X}} \{d_1(x) + d_2(x) + D_3(x)\}, \quad (13)$$

where d_1 , d_2 , and D_3 are convex, closed and proper. Operator splitting schemes apply to such a problem with $\mathbb{A} = \partial d_1$, $\mathbb{B} = \partial d_2$, $\mathbb{C} = \partial D_3$, and $\mathbb{J}_{\gamma \partial D_3}$ is the proximal operator $\text{prox}_{\gamma D_3}$ for the function D_3 . For the case $D_3(x) = d_3(Bx)$ where B is a matrix, Yan introduced the following primal-dual algorithm (PD3O) in Yan (2018):

$$\begin{aligned} x_k &= \text{prox}_{\gamma d_1}(z_k) \\ s_{k+1} &= \text{prox}_{\delta d_3^*} \left((I - \gamma \delta B B^\top) s_k + \delta B^\top (2x_k - z_k - \gamma \nabla d_2(x_k)) \right) \\ z_{k+1} &= x_k - \gamma \nabla d_2(x_k) - \gamma B^\top s_{k+1} \quad (\text{PD3O}). \end{aligned} \tag{14}$$

The PD3O method reduces to Davis-Yin splitting when $B = I$, and it is preferred when it is much easier to compute $\mathbb{J}_{\partial d_3}$ than $\mathbb{J}_{\partial D_3}$, e.g., D_3 is the total variation (TV) norm function and d_3 is the ℓ^1 -norm function.

3.2 A new three-operator splitting

Next, we derive a new three-operator splitting. To the best of our knowledge, it is different from all existing three-operator splitting methods in the literature. 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 - \mathbf{R}_{\gamma \mathbb{B}}(I + \gamma \mathbb{B})x \\ &\quad + \gamma \mathbb{C}x \quad [\text{since } \mathbf{R}_{\gamma \mathbb{B}}(I + \gamma \mathbb{B}) = (I - \gamma \mathbb{B})] \\ &\iff 0 \in (I + \gamma \mathbb{A})x - \mathbf{R}_{\gamma \mathbb{B}z} + \gamma \mathbb{C}x \quad [\text{assume } z \in (I + \gamma \mathbb{B})x] \\ &\iff \mathbf{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 \quad [\text{since } \mathbf{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 \quad [\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} \circ \mathbb{J}_{\gamma \mathbb{B}z}) = x \\ &\iff \mathbb{J}_{\gamma \mathbb{A}}(2\mathbb{J}_{\gamma \mathbb{B}z} - z - \gamma \mathbb{C} \circ \mathbb{J}_{\gamma \mathbb{B}z}) + \gamma \mathbb{C} \circ \mathbb{J}_{\gamma \mathbb{B}z} = x + \gamma \mathbb{C} \circ \mathbb{J}_{\gamma \mathbb{B}z} \\ &\iff \mathbb{J}_{\gamma \mathbb{A}}(2\mathbb{J}_{\gamma \mathbb{B}z} - z - \gamma \mathbb{C} \circ \mathbb{J}_{\gamma \mathbb{B}z}) + \gamma \mathbb{C} \circ \mathbb{J}_{\gamma \mathbb{B}z} = (I + \gamma \mathbb{C})x \\ &\iff \mathbb{J}_{\gamma \mathbb{C}}(\mathbb{J}_{\gamma \mathbb{A}}(2\mathbb{J}_{\gamma \mathbb{B}} - I - \gamma \mathbb{C} \circ \mathbb{J}_{\gamma \mathbb{B}}) + \gamma \mathbb{C} \circ \mathbb{J}_{\gamma \mathbb{B}})z \\ &= x \quad [\text{since } \mathbb{J}_{\gamma \mathbb{C}} = (I + \gamma \mathbb{C})^{-1}] \\ &\iff \mathbb{J}_{\gamma \mathbb{C}}(\mathbb{J}_{\gamma \mathbb{A}}(2\mathbb{J}_{\gamma \mathbb{B}} - I - \gamma \mathbb{C} \circ \mathbb{J}_{\gamma \mathbb{B}}) + \gamma \mathbb{C} \circ \mathbb{J}_{\gamma \mathbb{B}})z = \mathbb{J}_{\gamma \mathbb{B}z} \\ &[\text{assumption } z \in (I + \gamma \mathbb{B})x] \\ &\iff (\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} \circ \mathbb{J}_{\gamma \mathbb{B}}) + (I - \mathbb{J}_{\gamma \mathbb{B}}))z = z \\ &\iff Tz = z, \quad T = \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}}) \\ &\quad + \gamma \mathbb{C} \circ \mathbb{J}_{\gamma \mathbb{B}}) + (I - \mathbb{J}_{\gamma \mathbb{B}}). \end{aligned} \tag{15}$$

The proposed operator T in (15) splits the three-operator sum problem given in (8) into simpler sub-problems. For special cases, Algorithm 1 reduces to Douglas-

Rachford splitting (DRS) and a splitting closely related to Forward-Backward splitting (FBS). The proposed operator T reduces to two-operator splittings for 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 $\mathbb{A} = 0$, (15) reduces to

$$T = \mathbb{J}_{\gamma\mathbb{C}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I) + (I - \mathbb{J}_{\gamma\mathbb{B}}), \text{ which is DRS [31].}$$

(ii) If $\mathbb{C} = 0$, (15) reduces to

$$T = \mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I) + (I - \mathbb{J}_{\gamma\mathbb{B}}), \text{ which is DRS [31].}$$

(iii) If $\mathbb{B} = 0$, (15) 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 (Lions and Mercier 1979; Passty 1979) 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 (8), consider the Krasnosel’skiĭ–Mann (KM) iteration with the operator T above:

$$\begin{aligned} T_\lambda &:= (1 - \lambda)I + \lambda T \\ \text{and } z^{k+1} &= (1 - \lambda_k)z^k + \lambda_k T z^k. \end{aligned} \tag{16}$$

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 (Boţ and Nguyen 2023) and references therein for some of the latest developments of strategies for designing λ_k . The steps of the iterative procedure are given in Algorithm 1:

As a remark, for solving an inclusion problem like

$$x^* \in \arg \min_{x \in \mathcal{X}} \{d_1(x) + d_2(x) + d_3(Bx)\},$$

when $B = I$, PD3O and other primal–dual algorithms such as Condat (2013), Vũ (2011), the Primal–Dual Fixed–Point algorithm (Chen et al. 2016), and the Asymmetric Forward–Backward–Adjoint splitting (Latafat and Patrinos 2017) naturally give algorithms for a general inclusion problem (8). On the other hand, these primal–dual

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 $k = 0, 1, 2, \dots$

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

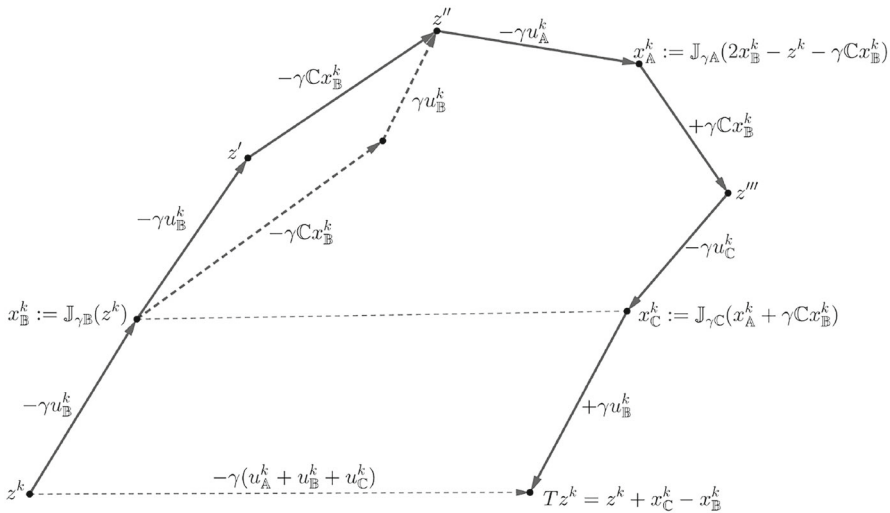


Fig. 1 The proposed mapping $T : z^k \rightarrow Tz^k$ and vectors $u_{\mathbb{B}}^k \in \mathbb{B}x_{\mathbb{B}}^k$, $u_{\mathbb{A}}^k \in \mathbb{A}x_{\mathbb{A}}^k$, and $u_{\mathbb{C}}^k \in \mathbb{C}x_{\mathbb{C}}^k$ as given in Lemma 4.1

methods were designed with motivations to exploit the explicit separation between B and d_3 , allowing the proximal operator of d_3 and the linear mapping B to be handled efficiently. Nonetheless, for solving a general inclusion problem (8), the proposed new splitting is indeed different from these existing methods. Numerical tests in Section 6 suggest that the proposed new splitting for solving composite convex minimization problems is quite robust with any positive step size. Furthermore, it can be regarded as a variant of the Davis-Yin splitting, and this variant has the numerical advantage of allowing much larger step sizes, which can be proven for a special problem in Section 5.3.

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 (15). Figure 1 provides a geometric illustration of how T acts on a point $z \in \mathcal{X}$ corresponding to the points defined in Lemma 4.1.

Lemma 4.1 For $z \in \mathcal{X}$, define the points:

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

Based on the above relations, the following equations hold:

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

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 relation associated with 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 4.2 Let \mathbb{A} , \mathbb{B} , and \mathbb{C} be three operators. Then, we have the following set equality:

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

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

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

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

With all the identities above, we conclude that

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

$$= x + \gamma u_{\mathbb{B}} = z.$$

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

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

□

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 Davis and Yin (2017).

Theorem 4.3 Assume that $T : \mathcal{X} \rightarrow \mathcal{X}$ is α -averaged with $\alpha = \frac{2\beta}{4\beta - \gamma} < 1$ and suppose that $z^* \in \text{Fix } T$. Consider a sequence of relaxation parameters $(\lambda_j)_{j \geq 0} \subset (0, \frac{1}{\alpha})$, where $\alpha = 1/(2 - \varepsilon)$ and $\alpha < 2\beta/(4\beta - \gamma)$. Let $\sum_{j=0}^{\infty} \tau_j = \sum_{j=0}^{\infty} (1 - \lambda_j/\alpha)\lambda_j/\alpha = \infty$. 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 hold:

- (i) The norm sequence $\|z^j - z^*\|_{j \geq 0}$ is monotonically nonincreasing for any $z^* \in \text{Fix } T$.
- (ii) The residuals $\|Tz^j - z^j\|_{j \geq 0}$ form a monotonically nonincreasing sequence which 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 $k \geq 0$, we have the following convergence rate:

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

Proof The proof for parts (i)-(iii) follows from Proposition 5.15 of Bauschke and Combettes (2017), and the proof of part (iv) follows from Davis and Yin (2016). □

Theorem 4.4 (Convergence theorem). Let $T : \mathcal{X} \rightarrow \mathcal{X}$ be an α -averaged operator with $\alpha = \frac{2\beta}{4\beta - \gamma} < 1$. Let $(\lambda_j)_{j \geq 0} \subseteq (0, \frac{1}{\alpha})$ be a sequence of relaxation parameters, where $\alpha = 1/(2 - \varepsilon) < 2\beta/(4\beta - \gamma)$. Suppose $\sum_{j=0}^{\infty} \tau_j = \sum_{j=0}^{\infty} (1 - \lambda_j/\alpha)\lambda_j/\alpha = \infty$ is satisfied. 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. Assume $\inf_{j \geq 0} \lambda_j > 0$ and z^* denote a weak limit point of z^k . Then:

- (a) The sequence $(\mathbb{C}x_{\mathbb{B}}^j)_{j \geq 0}$ converges strongly to $\mathbb{C}x^*$, where $x^* \in \text{zer}(\mathbb{A} + \mathbb{B} + \mathbb{C})$.

- (b) The iterates $\mathbb{J}_{\gamma\mathbb{B}}(z^j)_{j \geq 0}$ converges weakly to $\mathbb{J}_{\gamma\mathbb{B}}(z^*) \in \text{zer}(\mathbb{A} + \mathbb{B} + \mathbb{C})$.
 - (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 $\mathbb{J}_{\gamma\mathbb{B}}(z^*) \in \text{zer}(\mathbb{A} + \mathbb{B} + \mathbb{C})$.
 - (d) the sequence $(\mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C} \circ \mathbb{J}_{\gamma\mathbb{B}}) + \gamma\mathbb{C} \circ \mathbb{J}_{\gamma\mathbb{B}}))(z^j)_{j \geq 0}$ weakly converges to $\mathbb{J}_{\gamma\mathbb{B}}(z^*) \in \text{zer}(\mathbb{A} + \mathbb{B} + \mathbb{C})$.
2. The sequences $\mathbb{J}_{\gamma\mathbb{B}}(z^j)_{j \geq 0}$ and $(\mathbb{J}_{\gamma\mathbb{C}} \circ (\mathbb{J}_{\gamma\mathbb{A}} \circ (2\mathbb{J}_{\gamma\mathbb{B}} - I - \gamma\mathbb{C} \circ \mathbb{J}_{\gamma\mathbb{B}}) + \gamma\mathbb{C} \circ \mathbb{J}_{\gamma\mathbb{B}}))(z^j)_{j \geq 0}$ converges strongly to a solution in $\text{zer}(\mathbb{A} + \mathbb{B} + \mathbb{C})$ if at least one of the following holds:
- (a) Operator \mathbb{A} is uniformly monotone on every nonempty bounded subset of its domain.
 - (b) Operator \mathbb{B} is uniformly monotone on every nonempty bounded subset of its domain.
 - (c) Operatoe \mathbb{C} is demiregular at each solution point in $\text{zer}(\mathbb{A} + \mathbb{B} + \mathbb{C})$.

Proof 1. (a) Let $k \geq 0$. Then, using Corollary 2.14 in Bauschke and Combettes (2017), we observe that

$$\begin{aligned} \|z^{k+1} - z^*\|^2 &= \|(1 - \lambda_k)(z^k - z^*) + \lambda_k(Tz^k - z^*)\|^2 \\ &= (1 - \lambda_k)\|z^k - z^*\|^2 + \lambda_k\|Tz^k - z^*\|^2 - \lambda_k(1 - \lambda_k)\|Tz^k - z^k\|^2 \end{aligned} \tag{17}$$

In view of Theorem 2.1 of Davis and Yin (2017), we get

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

which gives $\|\mathbb{C}x_{\mathbb{B}}^k - \mathbb{C}\mathbb{J}_{\gamma\mathbb{B}}(z^*)\|^2 \rightarrow 0$ as $k \rightarrow \infty$.

(b) We recall the notations from Lemma 4.1 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 $(x_{\mathbb{B}}^j)_{j \geq 0}$ is bounded and admits a weak sequential limit labelled as \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 $j \rightarrow \infty$. Let $x^* \in \text{zer}(\mathbb{A} + \mathbb{B} + \mathbb{C})$. Next, observe that \mathbb{C} is maximal monotone and $\mathbb{C}x_{\mathbb{B}}^k \rightarrow \mathbb{C}x^*$, and $x_{\mathbb{B}}^{k_j} \rightarrow \bar{x}$, thus, in view of Proposition 20.33(ii) of Bauschke and Combettes (2017) 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 4.3 and Lemma 4.1, we have $x_{\mathbb{C}}^k - x_{\mathbb{B}}^k = Tz^k - z^k \rightarrow 0$ as $k \rightarrow \infty$. Thus, with $j \rightarrow \infty$, we obtain

$$x_{\mathbb{B}}^{k_j} \rightharpoonup \bar{x}, x_{\mathbb{A}}^{k_j} \rightharpoonup \bar{x}, x_{\mathbb{C}}^{k_j} \rightharpoonup \bar{x}, \mathbb{C}x_{\mathbb{B}}^{k_j} \rightharpoonup \mathbb{C}\bar{x}$$

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

On applying Proposition, 25.5 of Bauschke and Combettes (2017) to $(x_{\mathbb{A}}^{k_j}, u_{\mathbb{A}}^{k_j}) \in \text{gra}(\mathbb{A}), (x_{\mathbb{B}}^{k_j}, u_{\mathbb{B}}^{k_j}) \in \mathbb{B}$, and $(x_{\mathbb{B}}^{k_j}, \mathbb{C}x_{\mathbb{B}}^{k_j}) \in \mathbb{C}$, we observe that $\bar{x} \in \text{zer}(\mathbb{A} + \mathbb{B} + \mathbb{C}), z^* - \bar{x} \in \gamma\mathbb{B}\bar{x}, \bar{x} - z^* - \gamma\mathbb{C}\bar{x} \in \gamma\mathbb{A}\bar{x}$, 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 Bauschke and Combettes (2017), 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

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

2. The proofs are in a similar manner to those given in Theorem 2.1 of Davis and Yin (2017). □

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 4.5 *Assume that $T : \mathcal{X} \rightarrow \mathcal{X}$ is α -averaged with $\alpha = \frac{2\beta}{4\beta - \gamma} < 1$. 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 4.3, 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

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

With the inequality above, we also have

$$\|x_{\mathbb{C}}^k - x^*\| = \|\mathbb{J}_{\gamma\mathbb{C}}(x_{\mathbb{A}}^k + \gamma\mathbb{C}x_{\mathbb{B}}^k) - \mathbb{J}_{\gamma\mathbb{C}}(x^* + \gamma\mathbb{C}x^*)\| = \|x_{\mathbb{A}}^k - x^* + \gamma\mathbb{C}x_{\mathbb{B}}^k - \gamma\mathbb{C}x^*\|$$

$$\leq \|z^k - z^*\| + \frac{2\gamma}{\beta} \|z^k - z^*\| \leq \left(1 + \frac{2\gamma}{\beta}\right) \|z^0 - z^*\|.$$

□

5 Application to composite convex minimization problems

In this section, we consider the application of the proposed three-operator splitting for composite convex minimization problems. We first review the classical three-block ADMM method, then in Section 5.2 we derive its dual form. We will also discuss the convergence of the proposed new splitting scheme.

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 Fréchet differentiable.
- The functions $f_i : \mathcal{X} \rightarrow (-\infty, +\infty]$, $i = 1, 2, 3$, satisfy the existence of solution condition, i.e.,

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

The following are some well known facts.

Proposition 5.1 (Optimality conditions of prox). *Let $w \in \mathcal{X}$ and f be a proper, closed, and convex function. Then, the following identity holds.*

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

Proposition 5.2 (Firm nonexpansiveness of prox). *Let $w, r \in \mathcal{X}$, and let $\tilde{w} = \text{prox}_{\gamma f}(w)$ and $\tilde{r} = \text{prox}_{\gamma f}(r)$. Then,*

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

In particular, $\text{prox}_{\gamma f}$ is nonexpansive.

Theorem 5.3 (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*

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

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

$$\left. \begin{aligned} \min \quad & f_1(x_1) + f_2(x_2) + f_3(x_3), \\ \text{subject to} \quad & A_1x_1 + A_2x_2 + A_3x_3 = b, \\ & x_1 \in \mathcal{X}_1 = \mathbb{R}^{n_1}, x_2 \in \mathcal{X}_2 = \mathbb{R}^{n_2}, x_3 \in \mathcal{X}_3 = \mathbb{R}^{n_3} \end{aligned} \right\}, \quad (18)$$

where $f_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper, closed, and convex functions (not necessarily smooth), $A_i \in \mathbb{R}^{m \times n_i}$ and $b \in \mathbb{R}^m$. Let f_i^* denote the convex conjugate of function f_i , and let

$$d_1(w) = f_1^*(A_1^\top w), \quad d_2(w) = f_2^*(A_2^\top w), \quad d_3(w) = f_3^*(A_3^\top w) - \langle w, b \rangle.$$

The dual problem of (18) is given by

$$\min_{w \in \mathbb{R}^m} d_1(w) + d_2(w) + d_3(w) \quad (19)$$

The original two-block ADMM method in Glowinski and Marroco (1975) has two parameters γ and σ , and it converges with certain parameter constraints for convex problems. If using only one parameter, i.e., taking $\sigma = \gamma$, the classical two-block ADMM method in Glowinski and Marroco (1975) is equivalent to the Douglas-Rachford splitting, thus such a single parameter two-block ADMM converges with any step size $\gamma > 0$ for convex problems, implied by the averagedness of the DRS operator. One could consider multiple-block ADMM methods with multiple parameters, but we only consider ADMM methods using a single step size here to compare its dual form to the proposed splitting scheme in this paper. A direct extension of the original ADMM in Glowinski and Marroco (1975) with only one step size for (18) is given by

$$x_1^{k+1} = \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \left\{ f_1(x_1) + \frac{\gamma}{2} \|(A_1x_1 + A_2x_2^k + A_3x_3^k - b) - \frac{1}{\gamma}w^k\|^2 \right\} \quad (20a)$$

$$x_2^{k+1} = \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \left\{ f_2(x_2) + \frac{\gamma}{2} \|(A_1x_1^{k+1} + A_2x_2 + A_3x_3^k - b) - \frac{1}{\gamma}w^k\|^2 \right\} \quad (20b)$$

$$x_3^{k+1} = \operatorname{argmin}_{x_3 \in \mathcal{X}_3} \left\{ f_3(x_3) + \frac{\gamma}{2} \|(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3 - b) - \frac{1}{\gamma}w^k\|^2 \right\} \quad (20c)$$

$$w^{k+1} = w^k - \gamma(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b), \quad (20d)$$

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

5.2 The similarity compared with the dual formulation of the three-block ADMM method

In this subsection, we derive the dual formulation of (20), which will be a scheme solving the dual problem (19). For convenience, define $\bar{x}^k := x^{k+1}$. We first list the change of variables needed in this subsection:

$$z^k = w^k - \gamma(A_1x_1^{k+1} + A_2x_2^{k+1}) \tag{21}$$

$$p^{k+1} = w^{k+1} - \gamma(A_1\bar{x}_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b) \tag{22}$$

$$v^{k+1} = w^{k+1} - \gamma(A_1\bar{x}_1^{k+1} + A_2\bar{x}_2^{k+1} + A_3x_3^{k+1} - b). \tag{23}$$

We start the derivation with (20c) as follows

$$\begin{aligned} x_3^{k+1} &= \operatorname{argmin}_{x_3 \in \mathcal{X}_3} \{f_3(x_3) + \frac{\gamma}{2} \|(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3 - b) - \frac{1}{\gamma}w^k\|^2\} \\ &\iff 0 \in \partial f_3(x_3^{k+1}) - A_3^\top(w^k - \gamma(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b)) \\ &\iff 0 \in \partial f_3(x_3^{k+1}) - A_3^\top w^{k+1} \quad [\text{from (20d)}] \\ &\iff A_3^\top w^{k+1} \in \partial f_3(x_3^{k+1}) \\ &\iff x_3^{k+1} \in \partial f_3^*(A_3^\top w^{k+1}). \end{aligned} \tag{24}$$

In view of (20d), observe that

$$\begin{aligned} w^{k+1} &= w^k - \gamma(A_1x_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b) \\ w^{k+1} &= z^k - \gamma(A_3x_3^{k+1} - b) \quad [\text{by definition of } z^k \text{ in (21)}] \end{aligned} \tag{25}$$

$$\begin{aligned} \iff w^{k+1} &= z^k - \gamma(A_3\partial f_3^*(A_3^\top w^{k+1}) - b) \quad [\text{from (24)}] \\ \iff w^{k+1} &= z^k - \gamma\partial d_3(w^{k+1}) \end{aligned} \tag{26}$$

$$\begin{aligned} \iff (I + \gamma\partial d_3)(w^{k+1}) &= z^k \\ \iff w^{k+1} &= \operatorname{prox}_{\gamma d_3}(z^k). \end{aligned} \tag{27}$$

Consider the equation given by (20a), we have

$$\begin{aligned} x_1^{k+1} &= \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \{f_1(x_1) + \frac{\gamma}{2} \|(A_1x_1 + A_2x_2^k + A_3x_3^k - b) - \frac{1}{\gamma}w^k\|^2\} \\ &\iff 0 \in \partial f_1(x_1^{k+1}) - A_1^\top(w^k - \gamma(A_1x_1^{k+1} + A_2x_2^k + A_3x_3^k - b)) \\ &\iff 0 \in \partial f_1(\bar{x}_1^k) - A_1^\top(w^k - \gamma(A_1\bar{x}_1^k + A_2x_2^k + A_3x_3^k - b)) \\ &\iff 0 \in \partial f_1(\bar{x}_1^{k+1}) \\ &\quad - A_1^\top(w^{k+1} - \gamma(A_1\bar{x}_1^{k+1} + A_2x_2^{k+1} + A_3x_3^{k+1} - b)) \text{ holds for any } k \in \mathbb{N} \cup \{0\} \\ &\iff 0 \in \partial f_1(\bar{x}_1^{k+1}) - A_1^\top p^{k+1} \quad [\text{by definition of } p^{k+1} \text{ in (22)}] \\ &\iff A_1^\top p^{k+1} \in \partial f_1(\bar{x}_1^{k+1}) \\ &\iff \bar{x}_1^{k+1} \in \partial f_1^*(A_1^\top p^{k+1}). \end{aligned} \tag{28}$$

Next, from (20b), we have

$$\begin{aligned}
 x_2^{k+1} &= \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \{f_2(x_2) + \frac{\gamma}{2} \|(A_1 x_1^{k+1} + A_2 x_2 + A_3 x_3^k - b) - \frac{1}{\gamma} w^k\|^2\} \\
 &\iff 0 \in \partial f_2(x_2^{k+1}) - A_2^\top (w^k - \gamma(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^k - b)) \\
 &\iff 0 \in \partial f_2(\bar{x}_2^k) - A_2^\top (w^k - \gamma(A_1 \bar{x}_1^k + A_2 \bar{x}_2^k + A_3 x_3^k - b)) \\
 &\iff 0 \in \partial f_2(\bar{x}_2^{k+1}) \\
 &\quad - A_2^\top (w^{k+1} - \gamma(A_1 \bar{x}_1^{k+1} + A_2 \bar{x}_2^{k+1} + A_3 x_3^{k+1} - b)) \text{ holds for any } k \in \mathbb{N} \cup \{0\} \\
 &\iff 0 \in \partial f_2(\bar{x}_2^{k+1}) - A_2^\top v^{k+1} \text{ [by definition of } v^{k+1} \text{ in (23)]} \\
 &\iff A_2^\top v^{k+1} \in \partial f_2(\bar{x}_2^{k+1}) \\
 &\iff \bar{x}_2^{k+1} \in \partial f_2^*(A_2^\top v^{k+1}). \tag{29}
 \end{aligned}$$

Now, from (24), (28), and (29), observe that

$$\begin{aligned}
 v^{k+1} &= w^{k+1} - \gamma(A_1 \bar{x}_1^{k+1} + A_2 \bar{x}_2^{k+1} + A_3 x_3^{k+1} - b) \\
 &\iff v^{k+1} = w^{k+1} - \gamma(A_1 \partial f_1^*(A_1^\top p^{k+1}) + A_2 \partial f_2^*(A_2^\top v^{k+1}) \\
 &\quad + A_3 \partial f_3^*(A_3^\top w^{k+1}) - b) \\
 &\iff v^{k+1} = w^{k+1} - \gamma \partial d_1(p^{k+1}) - \gamma \partial d_2(v^{k+1}) - \gamma \partial d_3(w^{k+1}). \tag{30}
 \end{aligned}$$

Note that,

$$\begin{aligned}
 p^{k+1} &= w^{k+1} - \gamma(A_1 \bar{x}_1^{k+1} + A_2 \bar{x}_2^{k+1} + A_3 x_3^{k+1} - b) \\
 &\iff p^{k+1} = w^{k+1} - \gamma(A_1 \partial f_1^*(A_1^\top p^{k+1}) + A_2 \partial f_2^*(A_2^\top v^k) \\
 &\quad + A_3 \partial f_3^*(A_3^\top w^{k+1}) - b) \\
 &\iff p^{k+1} = w^{k+1} - \gamma \partial d_1(p^{k+1}) - \gamma \partial d_2(v^k) - \gamma \partial d_3(w^{k+1}) \tag{31} \\
 &\iff (I + \gamma \partial d_1)(p^{k+1}) = w^{k+1} - \gamma \partial d_2(v^k) - \gamma \partial d_3(w^{k+1}) \\
 &\iff (I + \gamma \partial d_1)(p^{k+1}) = 2w^{k+1} - z^k - \gamma \partial d_2(v^k) \text{ [from (26)]} \\
 &\iff p^{k+1} = \operatorname{prox}_{\gamma d_1}(2w^{k+1} - z^k - \gamma \partial d_2(v^k)). \tag{32}
 \end{aligned}$$

Now by (30), we have

$$\begin{aligned}
 v^{k+1} &= w^{k+1} - \gamma \partial d_1(p^{k+1}) - \gamma \partial d_2(v^{k+1}) - \gamma \partial d_3(w^{k+1}) \\
 &\iff (I + \gamma \partial d_2)(v^{k+1}) = w^{k+1} - \gamma \partial d_1(p^{k+1}) - \gamma \partial d_3(w^{k+1}) \\
 &\iff (I + \gamma \partial d_2)(v^{k+1}) = p^{k+1} + \gamma \partial d_2(v^k) \text{ [from (31)]} \\
 &\iff v^{k+1} = \operatorname{prox}_{\gamma d_2}(p^{k+1} + \gamma \partial d_2(v^k)).
 \end{aligned}$$

Last, we have

$$\begin{aligned}
 z^k &= w^k - \gamma(A_1 x_1^{k+1} + A_2 x_2^{k+1}) \text{ [by definition of } z^k \text{ in (21)]} \\
 &\implies z^{k+1} = w^{k+1} - \gamma(A_1 \bar{x}_1^{k+1} + A_2 \bar{x}_2^{k+1})
 \end{aligned}$$

$$\begin{aligned} \implies z^{k+1} &= z^k - \gamma(A_1\bar{x}_1^{k+1} + A_2\bar{x}_2^{k+1} + A_3x_3^{k+1} - b) \quad [\text{by (25)}] \\ \implies z^{k+1} &= z^k + v^{k+1} - w^{k+1} \quad [\text{by definition of } v^{k+1} \text{ in (23)}]. \end{aligned}$$

Hence, we derive the dual formulation of the three-block ADMM as

$$\left. \begin{aligned} w^{k+1} &= \text{prox}_{\gamma d_3}(z^k) \\ p^{k+1} &= \text{prox}_{\gamma d_1}(2w^{k+1} - z^k - \gamma \partial d_2(v^k)) \\ v^{k+1} &= \text{prox}_{\gamma d_2}(p^{k+1} + \gamma \partial d_2(v^k)) \\ z^{k+1} &= z^k + v^{k+1} - w^{k+1} \end{aligned} \right\}, \tag{33}$$

where we have abused notation to use $\partial d_2(v^k)$ to denote any element in the subdifferential set $\partial d_2(v^k)$. As a comparison, the proposed new splitting in the previous section for minimizing $d_1(w) + d_2(w) + d_3(w)$ is written as

$$\left. \begin{aligned} w^{k+1} &= \text{prox}_{\gamma d_3}(z^k) \\ p^{k+1} &= \text{prox}_{\gamma d_1}(2w^{k+1} - z^k - \gamma \partial d_2(w^{k+1})) \\ v^{k+1} &= \text{prox}_{\gamma d_2}(p^{k+1} + \gamma \partial d_2(w^{k+1})) \\ z^{k+1} &= z^k + v^{k+1} - w^{k+1} \end{aligned} \right\}. \tag{34}$$

We can observe that the dual formulation to three-block ADMM (33) is similar to but different from the proposed scheme (34).

The splitting scheme (34) can be implemented as in Algorithm 2 for solving (19) 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 $k = 0, 1, 2, \dots$

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 The averagedness of the splitting operator for a special case

In this subsection, in addition to the basic convexity assumption of functions d_i for solving (19), we show that the proposed splitting operator T corresponding to iteration (34) is $\frac{1}{2}$ -averaged under the following special assumption:

Assumption 1 Assume $\mathcal{X} = X \oplus Y$ where X, Y are orthogonal subspaces, the domain of $d_1(w)$ is inside X and the domain of $d_2(w)$ is inside Y .

Remark 5.4 In the literature, in addition to basic convexity assumptions of f_i , the convergence of the three-block ADMM iteration (20) (or equivalently (33)) has been well studied. In Cai et al. (2017), the convergence of (20) to a saddle point can be proven if assuming strong convexity of f_3 . In Chen et al. (2016), a sufficient condition for the convergence of the 3-block ADMM iteration (20) is $A_1^\top A_2 = 0$, or $A_2^\top A_3 = 0$, or $A_1^\top A_3 = 0$. For $d_1(w) = f_1^*(A_1^\top w)$ and $d_2(w) = f_2^*(A_2^\top w)$, $A_1^\top A_2 = 0$ implies Assumption 1, thus the assumption we consider here is weaker than $A_1^\top A_2 = 0$.

Lemma 5.5 Under Assumption 1, if using the same subderivative $\tilde{\nabla} d_2(w^{k+1}) \in \partial d_2(w^{k+1})$ in the second line and third line in (34), the iteration (34) is equivalent to

$$z^{k+1} = z^k + \text{prox}_{\gamma d_2}(\text{prox}_{\gamma d_1}(2 \text{prox}_{\gamma d_3}(z^k) - z^k)) - \text{prox}_{\gamma d_3}(z^k). \tag{35}$$

Proof Without loss of generality, we consider the following special case of the domains of d_1 and d_2 being orthogonal complement to each other. Let any $w \in \mathcal{X}$ be written as $w = (x, y) \in X \oplus Y$ with $x \in X$ and $y \in Y$, and assume $d_1(x, y) = d_1(x, 0)$, $d_2(x, y) = d_2(0, y)$. Then we have $\tilde{\nabla}_w d_2(x, y) = \tilde{\nabla}_y d_2(0, y)$ for all $x \in X, y \in Y$. Let x_w, y_w be the X and Y components of w . Let x_z, y_z be the X and Y components of z . Let ∂_x, ∂_y be the x subdifferential and y subdifferential, respectively. The second line of (34) can be written as

$$\begin{aligned} p^{k+1} &= \underset{h=(x,y)}{\text{argmin}} \left\{ d_1(h) + \frac{1}{2\gamma} \|h - (2w^{k+1} - z^k - \gamma \tilde{\nabla} d_2(w^{k+1}))\|^2 \right\} \\ &= \underset{(x,y)}{\text{argmin}} \left\{ d_1(x, 0) + \frac{1}{2\gamma} \|(x, y) - [2(x_w^{k+1}, y_w^{k+1}) - (x_z^k, y_z^k) - \gamma(0, \tilde{\nabla}_y d_2(0, y_w^{k+1}))]\|^2 \right\} \\ &= \underset{x}{\text{argmin}} \left\{ d_1(x, 0) + \frac{1}{2\gamma} \|(x, 0) - (2(x_w^{k+1}, 0) - (x_z^k, 0))\|^2 \right\} \\ &\quad \oplus \underset{y}{\text{argmin}} \left\{ \frac{1}{2\gamma} \|(0, y) - [2(x_w^{k+1}, y_w^{k+1}) - (0, y_z^k) - \gamma(0, \tilde{\nabla}_y d_2(0, y_w^{k+1}))]\|^2 \right\} \\ &= \underset{x \in X}{\text{argmin}} \left\{ d_1(x, 0) + \frac{1}{2\gamma} \|(x, 0) - (2(x_w^{k+1}, 0) - (x_z^k, 0))\|^2 \right\} \\ &\quad \oplus \underset{y \in Y}{\text{argmin}} \left\{ \frac{1}{2\gamma} \|(0, y) - (2(x_w^{k+1}, y_w^{k+1}) - (0, y_z^k))\|^2 \right\} - \gamma(0, \tilde{\nabla}_y d_2(0, y_w^{k+1})) \\ &= \underset{(x,y)}{\text{argmin}} \left\{ d_1(x, y) + \frac{1}{2\gamma} \|(x, y) - (2(x_w^{k+1}, y_w^{k+1}) - (x_z^k, y_z^k))\|^2 \right\} \\ &\quad - \gamma(0, \tilde{\nabla}_y d_2(0, y_w^{k+1})) \\ &= \text{prox}_{\gamma d_1} \left(2w^{k+1} - z^k \right) - \gamma \tilde{\nabla} d_2(w^{k+1}). \end{aligned}$$

Thus under Assumption 1, if $\tilde{\nabla}d_2(w^{k+1})$ in the second line and $\tilde{\nabla}d_2(w^{k+1})$ in the third line of (34) are taken to be the same subderivative, then (34) becomes:

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

which simplifies to (35) and the proof is concluded. □

In general, the composition of two proximal operators of convex functions is 2/3-averaged and is not 1/2-averaged (i.e., firmly nonexpansive), see Bauschke and Combettes (2017).

Lemma 5.6 *Under Assumption 1, the composition $\text{prox}_{\gamma d_2} \circ \text{prox}_{\gamma d_1}$ is firmly nonexpansive.*

Proof Without loss of generality, we consider the following special case of the domains of d_1 and d_2 being orthogonal complement to each other. Let any $w \in \mathcal{X}$ be written as $w = (x, y) \in X \oplus Y$ with $x \in X$ and $y \in Y$, and assume $d_1(x, y) = d_1(x, 0)$, $d_2(x, y) = d_2(0, y)$. Let $w \in \mathcal{X}$ be given and x_w, y_w be the X and Y components of w , then

$$\begin{aligned} \text{prox}_{\gamma d_2} \circ \text{prox}_{\gamma d_1}(w) &= \text{prox}_{\gamma d_2}\left(\underset{h=(x,y)}{\text{argmin}}\left\{d_1(h) + \frac{1}{2\gamma}\|h - w\|^2\right\}\right) \\ &= \text{prox}_{\gamma d_2}\left(\underset{x}{\text{argmin}}\{d_1(x, 0) + \frac{1}{2\gamma}\|(x, 0) - (x_w, 0)\|^2\}, \right. \\ &\quad \left. \underset{y}{\text{argmin}}\left\{\frac{1}{2\gamma}\|(0, y) - (0, y_w)\|^2\right\}\right) \\ &= \text{prox}_{\gamma d_2}\left(\underset{x}{\text{argmin}}\left\{d_1(x, 0) + \frac{1}{2\gamma}\|(x, 0) - (x_w, 0)\|^2\right\}, y_w\right) \\ &= \underset{\hat{h}=(\hat{x}, \hat{y})}{\text{argmin}}\left\{d_2(\hat{h}) + \frac{1}{2\gamma}\|\hat{h} - \left(\underset{x}{\text{argmin}}\{d_1(x, 0) + \frac{1}{2\gamma}\|(x, 0) - (x_w, 0)\|^2\}, y_w\right)\|^2\right\} \\ &= \underset{\hat{x}}{\text{argmin}}\left\{d_2(0, \hat{y}) + \frac{1}{2\gamma}\|(\hat{x}, 0) - \left(\underset{x}{\text{argmin}}\{d_1(x, 0) + \frac{1}{2\gamma}\|(x, 0) - (x_w, 0)\|^2\}, 0\right)\|^2\right\} \\ &\quad \oplus \underset{\hat{y}}{\text{argmin}}\left\{d_2(0, \hat{y}) + \frac{1}{2\gamma}\|(0, \hat{y}) - (0, y_w)\|^2\right\} \\ &= \underset{\hat{x}}{\text{argmin}}\left\{\frac{1}{2\gamma}\|(\hat{x}, 0) - \left(\underset{x}{\text{argmin}}\{d_1(x, 0) + \frac{1}{2\gamma}\|(x, 0) - (x_w, 0)\|^2\}, 0\right)\|^2\right\} \\ &\quad \oplus \underset{\hat{y}}{\text{argmin}}\left\{d_2(0, \hat{y}) + \frac{1}{2\gamma}\|(0, \hat{y}) - (0, y_w)\|^2\right\} \\ &= \underset{x}{\text{argmin}}\left\{d_1(x, y) + \frac{1}{2\gamma}\|(x, y) - (x_w, y_w)\|^2\right\} \end{aligned}$$

$$\begin{aligned} & \bigoplus_y \operatorname{argmin} \left\{ d_2(x, y) + \frac{1}{2\gamma} \|(x, y) - (x_w, y_w)\|^2 \right\} \\ &= \operatorname{argmin}_{h=(x,y)} \{ (d_1 + d_2)(x, y) + \frac{1}{2\gamma} \|(x, y) - (x_w, y_w)\|^2 \} = \operatorname{prox}_{\gamma(d_1+d_2)}(w), \end{aligned}$$

which is firmly nonexpansive since it is the proximal operator of a convex closed proper function $d_1 + d_2$. □

Lemma 5.7 (Lemma 2.3 in Davis and Yin (2017)). *Let \mathcal{X} be a Hilbert space. Let $S := U + T_1 \circ V$, where $U, T_1 : \mathcal{X} \rightarrow \mathcal{X}$ are both firmly nonexpansive and $V : \mathcal{X} \rightarrow \mathcal{X}$. Let $W = I - (2U + V)$. Then we have for all $z, w \in \mathcal{X}$:*

$$\|Sz - Sw\|^2 \leq \|z - w\|^2 - \|(I - S)z - (I - S)w\|^2 - 2\langle T_1 \circ Vz - T_1 \circ Vw, Wz - Ww \rangle. \tag{36}$$

Theorem 5.8 *Under Assumption 1 and the assumption that d_1, d_2, d_3 are closed convex proper functions, if using the same subderivative $\tilde{\nabla}d_2(w^{k+1})$ in the second line and third line in (34), the splitting operator for the iteration (34) is $\frac{1}{2}$ -averaged thus (34) converges to the minimizer with any step size $\gamma > 0$.*

Proof Notice that (35) derived in Lemma 5.5 corresponds to the following splitting operator:

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

where $\mathbb{A} = \partial d_1, \mathbb{B} = \partial d_2$ and $\mathbb{C} = \partial d_3$. For any closed convex proper function f , its proximal operator $\operatorname{prox}_{\gamma f}$ is firmly nonexpansive, and so is $I - \operatorname{prox}_{\gamma f}$. Lemma 5.6 implies $\mathbb{J}_{\gamma\mathbb{C}} \circ \mathbb{J}_{\gamma\mathbb{A}}$ is firmly nonexpansive. Let $U = (I - \mathbb{J}_{\gamma\mathbb{B}})$, $T_1 = \mathbb{J}_{\gamma\mathbb{C}} \circ \mathbb{J}_{\gamma\mathbb{A}}$, $V = 2\mathbb{J}_{\gamma\mathbb{B}} - I$, then we have $S = T$ and $W = 0$ in Lemma 5.7, thus (36) gives

$$\|Tz - Tw\|^2 \leq \|z - w\|^2 - \frac{1 - \frac{1}{2}}{\frac{1}{2}} \|(I - T)z - (I - T)w\|^2, \tag{37}$$

which implies T is firmly nonexpansive thus $\frac{1}{2}$ -averaged. Therefore, results in Section 4, apply to establish the convergence of the splitting scheme (34). □

Remark 5.9 For a convex minimization $d_1(x) + d_2(x) + d_3(x)$ with differentiable d_2 , Theorem 5.8 shows that the proposed splitting scheme (2) converges for any step size $\gamma > 0$ under Assumption 1. Even under Assumption 1, the Davis-Yin splitting (3) still needs the step size to be small enough $\gamma < \frac{2}{L}$, since ∇d_2 cannot be cancelled out like in Lemma 5.5 for the proposed splitting scheme (2). Thus for special problems satisfying Assumption 1, the proposed new scheme is provably much more robust than Davis-Yin splitting in the sense of provable convergence using any positive step size.

5.4 Some basic properties

In this subsection, we discuss some basic properties of the proposed splitting scheme (34) under certain assumptions. We introduce the variables w_{d_i} by writing (34) as follows:

$$\left. \begin{aligned} w_{d_3}^k &:= \lambda^{k+1} = \text{prox}_{\gamma d_3}(z^k) \\ w_{d_1}^k &:= p^{k+1} = \text{prox}_{\gamma d_1}(2\lambda^{k+1} - z^k - \gamma \partial d_2(\lambda^{k+1})) \\ w_{d_2}^k &:= v^{k+1} = \text{prox}_{\gamma d_2}(p^{k+1} + \gamma \partial d_2(\lambda^{k+1})) \\ z^{k+1} &= z^k + w_{d_2}^k - w_{d_3}^k \end{aligned} \right\}. \quad (38)$$

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

Proposition 5.10 *Let $z^0 \in \mathcal{X}$ and $(z_j)_{j \geq 0}$ be the sequence generated by (38). Then the following identities hold:*

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

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

$$w_{d_3}^k = z^k - \gamma \tilde{\nabla} d_3(w_{d_3}^k). \quad (39)$$

Next, we have the following relation

$$\begin{aligned} w_{d_1}^k &= \text{prox}_{\gamma d_1}(2w_{d_3}^k - z^k - \gamma \nabla d_2(w_{d_3}^k)) \\ &\iff (I + \gamma \tilde{\nabla} d_1)(w_{d_1}^k) = 2w_{d_3}^k - z^k - \gamma \nabla d_2(w_{d_3}^k) \\ &\iff w_{d_1}^k - w_{d_3}^k = w_{d_3}^k - z^k - \gamma \nabla d_2(w_{d_3}^k) - \gamma \tilde{\nabla} d_1(w_{d_1}^k) \\ &\iff w_{d_1}^k - w_{d_3}^k = -\gamma(\tilde{\nabla} d_1(w_{d_1}^k) + \nabla d_2(w_{d_3}^k) + \tilde{\nabla} d_3(w_{d_3}^k)). \end{aligned} \quad (40)$$

Again, we have the following relation

$$\begin{aligned} w_{d_2}^k &= \text{prox}_{\gamma d_2}(w_{d_1}^k + \gamma \nabla d_2(w_{d_3}^k)) \\ &\iff (I + \gamma \nabla d_2)(w_{d_2}^k) = w_{d_1}^k + \gamma \nabla d_2(w_{d_3}^k) \\ &\iff w_{d_2}^k = (w_{d_3}^k - \gamma \tilde{\nabla} d_1(w_{d_1}^k) - \gamma \nabla d_2(w_{d_3}^k) - \gamma \tilde{\nabla} d_3(w_{d_3}^k)) + \gamma \nabla d_2(w_{d_3}^k) \\ &\quad - \gamma \nabla d_2(w_{d_2}^k) \text{ from (40)} \\ &\iff w_{d_2}^k - w_{d_3}^k = -\gamma(\tilde{\nabla} d_1(w_{d_1}^k) + \nabla d_2(w_{d_2}^k) + \tilde{\nabla} d_3(w_{d_3}^k)). \end{aligned} \quad (41)$$

Finally, we get

$$w_{d_1}^k - w_{d_3}^k = w_{d_2}^k - w_{d_3}^k + \gamma(\nabla d_2(w_{d_2}^k) - \nabla d_2(w_{d_3}^k)) \text{ and } z^{k+1} - z^k = w_{d_2}^k - w_{d_3}^k \tag{42}$$

□

Proposition 5.11 (Upper Inequality). *Let $w \in \mathcal{X}$ and w^* be the fixed point of the FPI algorithm given in (34). Then, the following inequality holds:*

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

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

$$\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 2\gamma \left(\langle w_{d_1}^k - w^*, \tilde{\nabla} d_1(w_{d_1}^k) \rangle + \langle w_{d_3}^k - w^*, \nabla d_2(w_{d_3}^k) + \tilde{\nabla} d_3(w_{d_3}^k) \rangle \right) \\ & = 2\gamma \langle w_{d_1}^k - w_{d_3}^k, \tilde{\nabla} d_1(w_{d_1}^k) \rangle + 2\gamma \langle w_{d_3}^k - w^*, \tilde{\nabla} d_1(w_{d_1}^k) + \nabla d_2(w_{d_3}^k) + \tilde{\nabla} d_3(w_{d_3}^k) \rangle \\ & = 2 \langle w_{d_1}^k - w_{d_3}^k, \gamma \tilde{\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 5.10} \\ & = 2 \langle w_{d_3}^k - w_{d_1}^k, w_{d_3}^k - w^* - \gamma \tilde{\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^* \\ & \quad - \gamma \tilde{\nabla} d_1(w_{d_1}^k) \rangle \text{ from (iv) of Proposition 5.10} \\ & = 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^* \\ & \quad - \gamma \tilde{\nabla} d_1(w_{d_1}^k) \rangle \text{ from (v) of Proposition 5.10} \\ & = 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 \tilde{\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 \tilde{\nabla} d_3(w_{d_3}^k) - w^* - \gamma \tilde{\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 \tilde{\nabla} d_3(w_{d_3}^k) \\ & \quad - w^* - \gamma \tilde{\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 5.10} \\ & = 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) \\ & \quad - \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 \\ & + 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 \\ & = \|z^k - z^{k+1}\|^2 + \|z^k - w^*\|^2 - \|z^{k+1} - w^*\|^2 - 2\|z^k - z^{k+1}\|^2 \end{aligned}$$

$$\begin{aligned}
& +2\langle \gamma \nabla d_2(w_{d_3}^k) - \gamma \nabla d_2(w_{d_2}^k), z^k - w^* \rangle \\
& -2\langle \gamma \nabla d_2(w_{d_3}^k) - \gamma \nabla d_2(w_{d_2}^k), z^k - z^{k+1} \rangle + 2\langle z^k - z^{k+1}, \gamma \nabla d_2(w_{d_2}^k) \rangle \\
& +2\langle \gamma \nabla d_2(w_{d_3}^k) - \gamma \nabla d_2(w_{d_2}^k), \gamma \nabla d_2(w_{d_2}^k) \rangle \text{ from (6)} \\
= & \|z^k - w^*\|^2 - \|z^k - z^{k+1}\|^2 - \|z^{k+1} - w^*\|^2 + 2\gamma \langle \nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k), z^k - w^* \rangle \\
& -2\gamma \langle \nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k), z^k - z^{k+1} \rangle + 2\gamma \langle z^k - z^{k+1}, \nabla d_2(w_{d_2}^k) \rangle \\
& +2\gamma^2 \langle \nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k), \nabla d_2(w_{d_2}^k) \rangle. \tag{43}
\end{aligned}$$

□

Proposition 5.12 (Lower Inequality). *Let $w \in \mathcal{X}$ and w^* be the fixed point of the FPI algorithm given in (34). Then, the following inequality holds:*

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

Proof By subgradient inequality and Proposition 5.10 (iii), we have

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

In a similar manner, we obtain

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

On adding the above three relations, we get

$$\begin{aligned}
& d_1(w_{d_1}^k) + d_2(w_{d_3}^k) + d_3(w_{d_3}^k) - (d_1 + d_2 + d_3)(w^*) \\
& \geq \langle w_{d_2}^k - w_{d_3}^k, \tilde{\nabla} d_1(w^*) \rangle + \langle w_{d_3}^k - w^*, \tilde{\nabla} d_1(w^*) + \nabla d_2(w^*) + \tilde{\nabla} d_3(w^*) \rangle \\
& \quad + \langle \gamma \nabla d_2(w_{d_2}^k) - \gamma \nabla d_2(w_{d_3}^k), \tilde{\nabla} d_1(w^*) \rangle. \tag{44}
\end{aligned}$$

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

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

Therefore, on plugging the above relation in (44), we get

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

□

In the next theorem, we prove the convergence rate of the proposed Algorithm (34).

Theorem 5.13 Consider the iteration (34). Assume that the function d_1 is L -Lipschitz continuous on the closed ball $B(0, (1 + \gamma/\beta)\|w^0 - w^*\|)$. Under the same assumptions in Theorem 4.3, 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 4.5, 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 (43) of Proposition 5.11 in above relation, we get

$$\begin{aligned} & 2\gamma \left(d_1(w_{d_1}^k) + d_2(w_{d_3}^k) + d_3(w_{d_3}^k) - (d_1 + d_2 + d_3)(w^*) \right) \\ & \leq 2\langle z^k - z^{k+1}, z^k - w^* \rangle - 2\langle z^k - z^{k+1}, z^k - z^{k+1} \rangle \\ & \quad + 2\langle \gamma \nabla d_2(w_{d_3}^k) - \gamma \nabla d_2(w_{d_2}^k), z^k - w^* \rangle \\ & \quad - 2\langle \gamma \nabla d_2(w_{d_3}^k) - \gamma \nabla d_2(w_{d_2}^k), z^k - z^{k+1} \rangle + 2\langle z^k - z^{k+1}, \gamma \nabla d_2(w_{d_2}^k) \rangle \\ & \quad + 2\langle \gamma \nabla d_2(w_{d_3}^k) - \gamma \nabla d_2(w_{d_2}^k), \gamma \nabla d_2(w_{d_2}^k) \rangle + 2\gamma L \|w_{d_1}^k - w_{d_3}^k\| \text{ from (6)} \\ & \leq 2\|z^k - z^{k+1}\| \|z^k - w^*\| + 2\|z^k - z^{k+1}\| \|z^k - z^{k+1}\| \\ & \quad + 2\gamma \|\nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k)\| \|z^k - w^*\| \\ & \quad + 2\gamma \|\nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k)\| \|z^k - z^{k+1}\| + 2\gamma \|z^k - z^{k+1}\| \|\nabla d_2(w_{d_2}^k)\| \\ & \quad + 2\gamma^2 \|\nabla d_2(w_{d_3}^k) - \nabla d_2(w_{d_2}^k)\| \|\nabla d_2(w_{d_2}^k)\| + 2\gamma L \|w_{d_1}^k - w_{d_3}^k\| \end{aligned}$$

Now, note that in view of Theorem 4.3, z^k is bounded. Therefore, in view of Lipschitz continuity of $\text{prox}_{\gamma d_3}$ and ∇d_2 , we have that $w_{d_3}^k = \text{prox}_{\gamma d_3}(z^k)$ and $\nabla d_2(w_{d_3}^k)$ are bounded. Similarly, $\nabla d_2(w_{d_2}^k)$ is also bounded.

By Theorem 4.3, $\|z^k - z^{k+1}\| = o\left(\frac{1}{\sqrt{k+1}}\right)$. By Proposition 5.10 (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 5.10 (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. □

5.5 Multi-operator splitting schemes

Algorithm 2 can be formally extended to a four-block problem. Consider the following problem given by

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

where for each $i = 1, 2, 3, 4$, the functions $f_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper, closed, and convex (not necessarily smooth). For each $i = 1, 2, 3, 4$, $A_i \in \mathbb{R}^{m \times n_i}$ and $b \in \mathbb{R}^m$. Notice that the proposed three-operator splitting scheme (34) can be regarded as a modification to the dual form of three-block ADMM (33). Thus similar to derivations in Section 5.2, we can first derive the dual form of the classical four-block ADMM for the problem above, then perform similar modifications to obtain a formal extension of the proposed scheme (34) for minimizing the dual problem $\sum_{i=1}^4 d_i(x)$ where each d_i is a proper closed convex function.

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 the Lipschitz constant of ∇d_i .

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

Remark 5.14 In general, a splitting scheme for m -operator can be similarly derived from modifying the dual form of 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 Davis and Yin (2017) and Ryu and Vü (2020).

Example 1 Consider the following problem given by

$$\min_{x \in \mathbb{R}^n} f(x) + g(Lx) + h(x), \tag{45}$$

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 $\text{zer}(\partial f + \nabla g + \partial h) \neq \emptyset$. Then, x^k is a weakly minimal solution to (45). The modified form of Algorithm 1 for problem (45) 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 $k = 0, 1, \dots$

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

Example 2 Consider the following problem given by

$$\left. \begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{\alpha}{2} \|x - u\|_2^2 + i_{\Lambda_1}(x) + i_{\Lambda_2}(x), \\ \text{subject to} \quad & \Lambda_1 = \{x : m \leq x_i \leq M, \quad \forall i\} \\ \text{and} \quad & \Lambda_2 = \{x : Ax = b\} \end{aligned} \right\}, \tag{46}$$

where $A = [1, 1, \dots, 1]$, $b \in \mathbb{R}$, $u \in \mathbb{R}^n$ are given. Here $\alpha > 0$ is a fixed constant. Such a simple problem can be used as a postprocessing step to enforce bounds for solving complicated PDEs (Liu et al. 2024, ?, 2025).

Notice that this simple constrained minimization (46) 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 (Liu et al. 2024, Appendix). Moreover, following the analysis in Demanet and Zhang (2016), a sharp local linear convergence rate of Douglas-Rachford splitting for (46) can be derived, which can be further used to design optimal step size (Liu et al. 2024). Though (46) can be solved by two-operator splitting, a more general version of (46) can no longer be easily solved by Douglas-Rachford splitting. For example, for stabilizing numerical schemes solving gas dynamics equations (Zhang and Shu 2012; Zhang 2017), the bound-preserving constraint $\Lambda_1 = \{x : m \leq x_i \leq M, \quad \forall i\}$ in (46) 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 conduct a comprehensive comparison of the major splitting schemes surveyed in (3.1) that are applicable to this problem. This includes Davis-Yin splitting, Forward-Douglas-Rachford-Forward (FDRF) (11), Forward-Reflected-Douglas-Rachford (FRDR) (12), ADMM Dual Form (4), and our proposed three-operator splitting (2). Let $f(x) = \frac{\alpha}{2}\|x - u\|_2^2$, $g(x) = i_{\Delta_1}(x)$, and $h(x) = i_{\Delta_2}(x)$. The proposed scheme (2) applies here with the following operators given by:

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

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

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

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

The Davis-Yin scheme (3) with $d_1 = g$, $d_2 = f$, $d_3 = h$ becomes

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

The FDRF scheme with $d_1 = g$, $d_2 = f$, $d_3 = h$ becomes

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

The FRDR scheme with $d_1 = g$, $d_2 = f$, $d_3 = h$ becomes

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

We compare the proposed splitting scheme (2), Davis-Yin method (3) FDRF method, and FRDR method, ADMM dual form on the problem (46) 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 (46) 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 true Lipschitz constant of ∇f . Let μ be an estimate of the Lipschitz constant L , and we use the following step sizes:

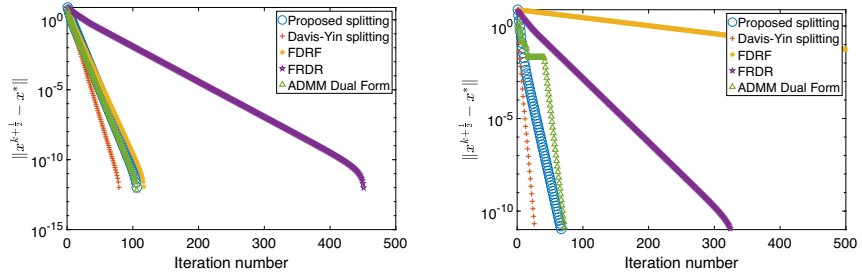
1. Davis-Yin splitting, FDRF, ADMM dual form and the new splitting method all use the same step size $\gamma = \frac{1}{\mu}$.
2. The FRDR method has two parameters and it converges for any $\beta > 0$ and $\gamma_{FRDR} < \frac{\beta}{1+2L\beta}$. We use a small enough $\beta > 0$ and $\gamma_{FRDR} = \frac{\beta}{1+2\mu\beta}$.

For this simple test, we have $L = \alpha$, but in practice usually one has to estimate the Lipschitz constant. To this end, we show the performance using different estimates μ in Figure 2, in which the optimal solution x^* is generated by running Davis-Yin method enough number of iterations with $\mu = L$. The Davis-Yin method performs the best if using step size $\gamma = \frac{1}{L}$, i.e., when there is an accurate estimate of L , and the proposed method is not faster than the Davis-Yin method if $\gamma < \frac{2}{L}$. The Davis-Yin method, FDRF and ADMM dual form will not converge using an underestimated Lipschitz constant like $\mu = \frac{L}{3}$ (or equivalently using a large step size like $\gamma = \frac{3}{L}$). Even with a significantly underestimated Lipschitz constant like $\mu = \frac{L}{40}$ (or equivalently using a large step size like $\gamma = \frac{40}{L}$ in the proposed new method), the proposed new method and FRDR converge. If using a significantly underestimated Lipschitz constant, FRDR is much faster with a properly tuned parameter β , but FRDR does not converge without tuning the extra parameter β . Though the proposed new method is slower than FRDR, it does not need an extra parameter thus it needs no further parameter tuning, which is its advantage. In other words, when there is no accurate estimate of the Lipschitz constant, this simple test suggests that the proposed new splitting is robust and easy to use.

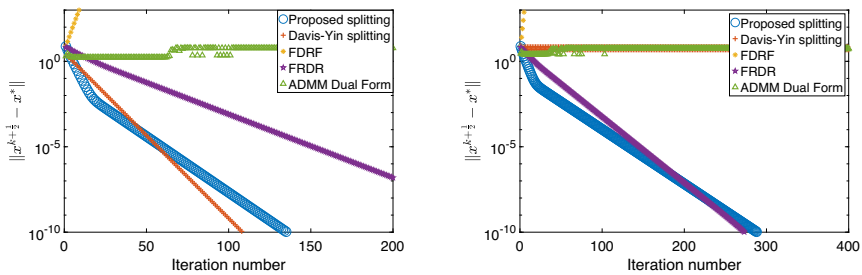
Example 3 Next we test the performance of the proposed new splitting on a fused lasso in the following form

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|_2^2 + \mu_1 \|x\|_1 + \mu_2 \|Bx\|_1, \tag{47}$$

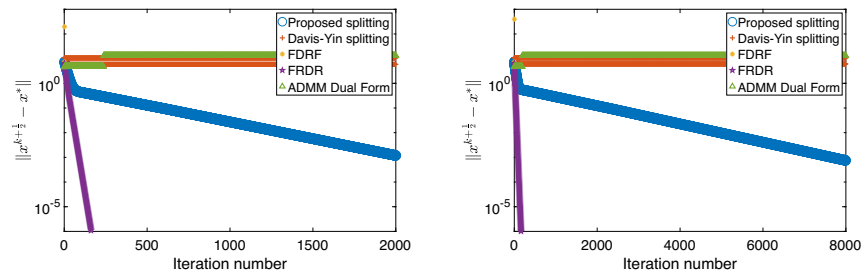
where $A \in \mathbb{R}^{r \times n}, b \in \mathbb{R}^r$, and $B = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \dots & \dots & \\ & & & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(n-1) \times n}$. The fused lasso has been used in signal processing, genomics, and machine learning to produce sparse



(a) Lipschitz constant estimate $\mu = \frac{L}{0.3}$. The step size is $\gamma = \frac{0.3}{L}$. (b) Lipschitz constant estimate $\mu = 0.99L$. The step size is $\gamma = \frac{0.99}{L}$.



(c) Lipschitz constant estimate $\mu = \frac{L}{1.8}$. The step size is $\gamma = \frac{1.8}{L}$. ADMM Dual Form and FDRF fail to converge. (d) Lipschitz constant estimate $\mu = \frac{L}{3}$. The step size is $\gamma = \frac{3}{L}$. ADMM Dual Form, Davis-Yin and FDRF fail to converge.



(e) Lipschitz constant estimate $\mu = \frac{L}{20}$. The step size is $\gamma = \frac{20}{L}$. ADMM Dual Form, Davis-Yin and FDRF fail to converge. (f) Lipschitz constant estimate $\mu = \frac{L}{40}$. The step size is $\gamma = \frac{40}{L}$. ADMM Dual Form, Davis-Yin and FDRF fail to converge.

Fig. 2 Results in Example 2 for performance of using different values of μ , which is an estimate of the Lipschitz constant $L = \alpha$. For ADMM Dual Form, Davis-Yin splitting, FDRF and the proposed new splitting method, we use the same step size $\gamma = \frac{1}{\mu}$. For the FRDR method, we use a small enough $\beta = 0.1$ and $\gamma_{FRDR} = \frac{\beta}{1+2\mu\beta}$. FRDR method will diverge without using small enough β for $\mu \ll L$

and locally smoothness models. When using a splitting method like Davis-Yin and the proposed new method in this paper for (47), one needs the proximal operator for the TV-norm function $\|Bx\|_1$, which is expensive to approximate. The PD30 in Yan (2018) avoids using the proximal operator of $\|Bx\|_1$. For a very large scale fused lasso problem, e.g., TV-norm for 2D and 3D data, a method like PD30 should be used.

We emphasize that we compare the proposed method with Davis-Yin splitting and PD3O for the fused lasso problem only for the purpose of validating the numerical performance of the new operator splitting. Let $f(x) = \frac{1}{2}\|Ax - b\|_2^2$, $g(x) = \mu_1\|x\|_1$, and $h(x) = \mu_2\|Bx\|_1$. It is well known that the proximal operator for $\|x\|_1$ with a step size $\gamma > 0$ is the shrinkage operator $\text{Shrinkage}_{\gamma}$. Set $d_1 = g$, $d_2 = f$, $d_3 = h$ in Algorithm 2 or the scheme (2) to obtain

$$x^{k+\frac{1}{2}} = \text{prox}_{\gamma\mu_2\|Bx\|_1}(z^k) = \underset{x}{\text{argmin}} \left(\mu_2\|Bx\|_1 + \frac{1}{2\gamma}\|x - z^k\|^2 \right) \tag{48a}$$

$$p^{k+1} = \text{Shrinkage}_{\gamma\mu_1} \left[2x^{k+\frac{1}{2}} - z^k - \gamma \left(A^\top (Ax^{k+\frac{1}{2}} - b) \right) \right] \tag{48b}$$

$$x^{k+1} = \left(A^\top A + \frac{1}{\gamma}I \right)^{-1} \left(A^\top b + \frac{1}{\gamma} \left(p^{k+1} + \gamma \left(A^\top (Ax^{k+\frac{1}{2}}) - b \right) \right) \right) \tag{48c}$$

$$z^{k+1} = z^k + x^{k+1} - x^{k+\frac{1}{2}}. \tag{48d}$$

The Davis-Yin splitting with $d_1 = g$, $d_2 = f$, $d_3 = h$ becomes

$$x^{k+\frac{1}{2}} = \text{prox}_{\gamma\mu_2\|Bx\|_1}(z^k) = \underset{x \in \mathcal{X}}{\text{argmin}} \left(\mu_2\|Bx\|_1 + \frac{1}{2\gamma}\|x - z^k\|^2 \right) \tag{49a}$$

$$x^{k+1} = \text{Shrinkage}_{\gamma\mu_1} \left[2x^{k+\frac{1}{2}} - z^k - \gamma \left(A^\top (Ax^{k+\frac{1}{2}} - b) \right) \right] \tag{49b}$$

$$z^{k+1} = z^k + x^{k+1} - x^{k+\frac{1}{2}}. \tag{49c}$$

Notice that the subproblem (48a) is the classical TV norm minimization problem and it has been known that such a problem can be efficiently approximated by the classical 2-block ADMM method. See (Torres et al. 2025) for a recent result of local linear convergence of ADMM method solving TV norm minimization problems. In each iteration of 2-block ADMM solving (48a), the main computation cost is the inversion $B^\top B + aI$ where a is a positive constant. For 1D TV-norm, $B^\top B + aI$ is a tridiagonal matrix, and we use backslash in MATLAB to efficiently invert such a tridiagonal matrix in our implementation.

For the PD3O scheme (14), we set $d_1(x) = \mu_1\|x\|_1$, $d_2 = f$, $d_3 = \mu_2\|x\|_1$, then only matrix multiplication and shrinkage operator are needed thus it is much cheaper and easier to implement.

We compare the the proposed splitting scheme (48) with Davis-Yin method (49) and PD3O (14) on the problem (47) with A being a random matrix whose elements follow the standard Gaussian distribution, and b is obtained by adding independent and identically distributed Gaussian noise with variance 0.01 onto Ax , $\mu_1 = 20$ and $\mu_2 = 200$.

We first consider a problem of size $A \in \mathbb{R}^{100 \times 1000}$ shown in Figure 3. In the 2-block ADMM iteration for the subproblem (48a), we set the step size as 1000 with stopping thresholds as either the difference between two iterations being less than 1×10^{-8} or the total number of iterations exceeding 10000. As we can observe in Figure 3, the more expensive methods (48) and (49) converge faster than PD3O for certain precision regime, and the proposed new method (48) is more robust with larger step size γ than

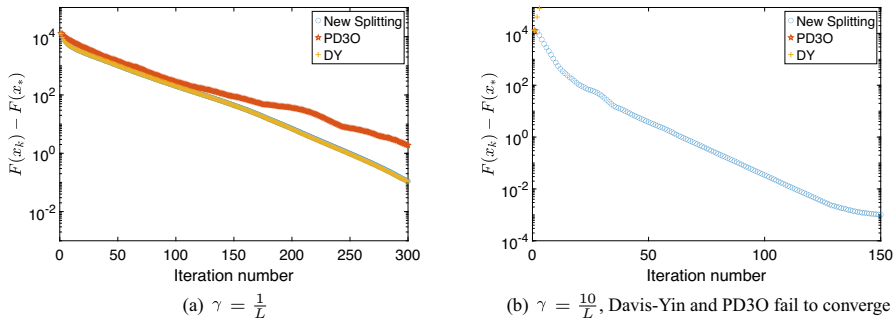


Fig. 3 Cost function value v.s. iteration numbers for a fused lasso problem with $r = 100$, $n = 1000$, $\mu_1 = 20$, $\mu = 200$. Performance of the three methods using the same step size γ and the extra parameter in PD3O is taken as $\delta = \frac{0.9}{\|BB^T\|} \frac{1}{\gamma}$. The reference solution x_* is approximated by 40, 000 iterations of PD3O with $\gamma = \frac{1}{L}$

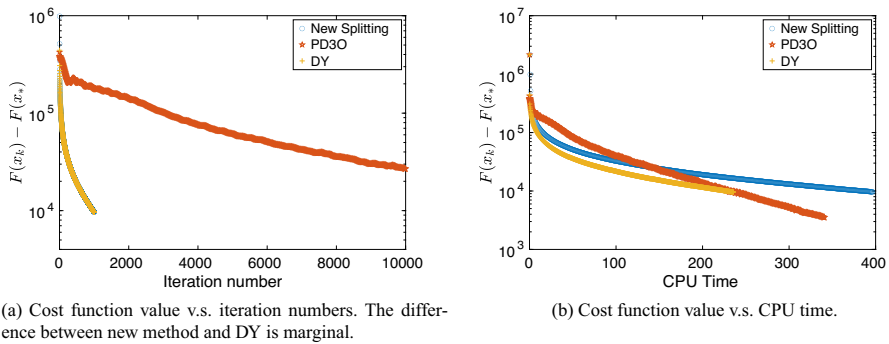


Fig. 4 A fused Lasso problem with $r = 400$, $n = 20000$, $\mu_1 = 20$, $\mu_2 = 200$. Step size is $\gamma = \frac{1}{L}$. The extra parameter in PD3O is taken as $\delta = 1$. The reference solution x_* is approximated by 40, 000 iterations of PD3O

Davis-Yin splitting (49). PD3O and Davis-Yin, and ADMM Dual Form splitting fail to converge when the step size γ is significantly larger than $\frac{10}{L}$.

Next we consider a problem of larger size with $A \in \mathbb{R}^{400 \times 20000}$. Figure 4 shows the performance. For the PD3O method, we use parameters $\delta = 1$, $\gamma = \frac{1}{L}$, which may not be the optimal ones, but it is already faster in CPU time than the two expensive methods (48) and (49). For the subproblem (48a) in this test, we set the step size as 1000 with stopping thresholds as either the difference between two iterations being less than 1×10^{-6} or the total number of iterations exceeding 500 for the 2-block ADMM iteration. Though the iteration numbers of the proposed new method and the Davis-Yin splitting are almost the same as shown in Figure 4(a), the proposed new method is slower in CPU time due to the extra step (48c). Nonetheless, the proposed scheme (48) is more robust for larger step size as shown in Figure 3.

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. It is similar to but different from Davis-Yin splitting, and the dual form of the standard three-block ADMM. In our numerical tests, it can allow a larger range of step size, compared to the Davis-Yin Splitting and the dual form of the standard three-block ADMM. For solving a composite convex minimization problem, the splitting operator can be proven 1/2-averaged if two functions have orthogonal domains. Future work includes exploration of nonsmooth problems as those recently studied in Yin et al. (2024), Bai et al. (2025).

Appendix: Derivation of a 3-block ADMM-like algorithm

Since the proposed splitting scheme is different from the classical three-block ADMM, in this section we derive the ADMM-like form of the proposed scheme. For the proposed algorithm

$$w^{k+1} = \text{prox}_{\gamma d_3}(z^k) \tag{50}$$

$$p^{k+1} = \text{prox}_{\gamma d_1}(2w^{k+1} - z^k - \gamma \partial d_2(v^k)) \tag{51}$$

$$v^{k+1} = \text{prox}_{\gamma d_2}(p^{k+1} + \gamma \partial d_2(v^k)) \tag{52}$$

$$z^{k+1} = z^k + v^{k+1} - w^{k+1}, \tag{53}$$

we start with relation (50)

$$\begin{aligned} w^{k+1} = \text{prox}_{\gamma d_3}(z^k) &\iff (I + \gamma \partial d_3)(w^{k+1}) = z^k \\ &\iff w^{k+1} = z^k - \gamma \partial d_3(w^{k+1}) \\ &\iff w^{k+1} = z^k - \gamma (A_3 \partial f_3^*(A_3^T w^{k+1}) - b) \\ &\iff A_3^T w^{k+1} = A_3^T (z^k - \gamma (A_3 \partial f_3^*(A_3^T w^{k+1}) - b)) \\ &\iff \partial f_3(\bar{w}_{k+1}) = A_3^T (z^k - \gamma A_3(\bar{w}_{k+1}) + \gamma b), \end{aligned}$$

where $\bar{w}^{k+1} = \partial f_3^*(A_3^T w^{k+1})$, and f^* denotes the convex conjugate of the proper, closed, and convex function f . By duality, they satisfy $A_3^T w^{k+1} = \partial f_3(\bar{w}^{k+1})$. From relation (51), we have

$$\begin{aligned} p^{k+1} &= \text{prox}_{\gamma d_1}(2w^{k+1} - z^k - \gamma \partial d_2(w^{k+1})) \\ &\iff (I + \gamma \partial d_1)(p^{k+1}) = 2w^{k+1} - z^k - \gamma \partial d_2(w^{k+1}) \\ &\iff p^{k+1} = 2w^{k+1} - z^k - \gamma \partial d_2(w^{k+1}) - \gamma \partial d_1(p^{k+1}) \\ &\iff p^{k+1} = 2w^{k+1} - z^k - \gamma A_2 \partial f_2^*(A_2^T w^{k+1}) - \gamma A_1 \partial f_1^*(A_1^T p^{k+1}) \\ &\iff A_1^T p^{k+1} = A_1^T (2w^{k+1} - z^k - \gamma A_2 \partial f_2^*(A_2^T w^{k+1}) - \gamma A_1 \partial f_1^*(A_1^T p^{k+1})) \\ &\iff \partial f_1(\bar{p}_{k+1}) = A_1^T (2w^{k+1} - z^k - \gamma A_2 \partial f_2^*(A_2^T w^{k+1}) - \gamma A_1 \bar{p}^{k+1}) \end{aligned}$$

$$\begin{aligned} \iff \partial f_1(\bar{p}_{k+1}) &= A_1^T(2(z^k - \gamma A_3 \bar{w}_{k+1} + \gamma b)) - z^k \\ &\quad - \gamma A_2 \partial f_2^*(A_2^T w^{k+1}) - \gamma A_1 \bar{p}^{k+1} \\ \iff \partial f_1(\bar{p}_{k+1}) &= A_1^T(z^k - 2\gamma A_3 \bar{w}_{k+1} - \gamma A_2 \tilde{w}^{k+1} - \gamma A_1 \bar{p}^{k+1} + 2\gamma b), \end{aligned}$$

where $\bar{p}^{k+1} = \partial f_1^*(A_1^T p^{k+1})$ and $\tilde{w}^{k+1} = \partial f_2^*(A_2^T w^{k+1})$. From relation (52), we have

$$\begin{aligned} v^{k+1} &= \text{prox}_{\gamma d_2}(p^{k+1} + \gamma \partial d_2(w^k)) \\ \iff (I + \gamma \partial d_2)(v^{k+1}) &= p^{k+1} + \gamma \partial d_2(w^k) \\ \iff v^{k+1} &= p^{k+1} + \gamma \partial d_2(w^k) - \gamma \partial d_2(v^{k+1}) \\ \iff v^{k+1} &= p^{k+1} + \gamma A_2 \partial f_2^*(A_2^T w^{k+1}) - \gamma A_2 \partial d_2(A_2^T v^{k+1}) \\ \iff A_2^T v^{k+1} &= A_2^T(p^{k+1} + \gamma A_2 \partial f_2^*(A_2^T w^{k+1}) - \gamma A_2 \partial d_2(A_2^T v^{k+1})) \\ \iff \partial f_2(\bar{v}^{k+1}) &= A_2^T(p^{k+1} + \gamma A_2 \partial f_2^*(A_2^T w^{k+1}) - \gamma A_2 \bar{v}^{k+1}) \\ \iff \partial f_2(\bar{v}^{k+1}) &= A_2^T(2w^{k+1} - z^k - \gamma A_1 \bar{p}^{k+1} - \gamma A_2 \bar{v}^{k+1}) \\ \iff \partial f_2(\bar{v}^{k+1}) &= A_2^T(2(z^k - \gamma A_3(\bar{w}_{k+1}) + \gamma b) - z^k - \gamma A_1 \bar{p}^{k+1} - \gamma A_2 \bar{v}^{k+1}) \\ \iff \partial f_2(\bar{v}^{k+1}) &= A_2^T(z^k - \gamma A_1 \bar{p}^{k+1} - \gamma A_2 \bar{v}^{k+1} - 2\gamma A_3(\bar{w}_{k+1}) + 2\gamma b), \end{aligned}$$

where $\bar{v}^{k+1} = \partial f_2^*(A_2^T v^{k+1})$. Finally, from (53) we have

$$\begin{aligned} z^{k+1} &= z^k + v^{k+1} - w^{k+1} \\ &= z^k + p^{k+1} + \gamma A_2 \tilde{w}^{k+1} - \gamma A_2 \bar{v}^{k+1} - z^k + \gamma A_3 \bar{w}^{k+1} - \gamma b \\ &= z^k - \gamma(A_1 \bar{p}^{k+1} + A_2 \bar{v}^{k+1} + A_3 \bar{w}^{k+1} - b) \end{aligned}$$

To summarize, after renaming the variables, we have

$$\begin{aligned} \partial f_3(\bar{w}_{k+1}) &= A_3^T(z^k - \gamma A_3(\bar{w}_{k+1}) + \gamma b) \\ \iff x_3^{k+1} = \bar{w}^{k+1} &= \underset{x_3 \in \mathcal{X}_3}{\operatorname{argmin}} \left\{ f_3(x_3) + \frac{\gamma}{2} \|A_3 x_3 - b - \frac{z^k}{\gamma}\|^2 \right\} \\ \partial f_1(\bar{p}_{k+1}) &= A_1^T(z^k - 2\gamma A_3 \bar{w}_{k+1} - \gamma A_2 \bar{w}^{k+1} - \gamma A_1 \bar{p}^{k+1} + 2\gamma b) \\ \iff x_1^{k+1} = \bar{p}^{k+1} &= \underset{x_1 \in \mathcal{X}_1}{\operatorname{argmin}} \left\{ f_1(x_1) + \frac{\gamma}{2} \|A_1 x_1 + A_2 \bar{w}^{k+1} - 2A_3 x_3^{k+1} - 2b - \frac{z^k}{\gamma}\|^2 \right\} \\ \partial f_2(\bar{v}^{k+1}) &= A_2^T(z^k - \gamma A_1 \bar{p}^{k+1} - \gamma A_2 \bar{v}^{k+1} - 2\gamma A_3(\bar{w}_{k+1}) + 2\gamma b) \\ \iff x_2^{k+1} = \bar{v}^{k+1} &= \underset{x_2 \in \mathcal{X}_2}{\operatorname{argmin}} \left\{ f_2(x_2) + \frac{\gamma}{2} \|A_1 x_1^{k+1} + A_2 x_2 - 2A_3 x_3^{k+1} - 2b - \frac{z^k}{\gamma}\|^2 \right\} \\ z^{k+1} &= z^k - \gamma(A_1 \bar{p}^{k+1} + A_2 \bar{v}^{k+1} + A_3 \bar{w}^{k+1} - b) \\ \iff z^{k+1} &= z^k - \gamma(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b). \end{aligned}$$

So the derived algorithm is

$$\left. \begin{aligned} x_3^{k+1} &= \operatorname{argmin}_{x_3 \in \mathcal{X}_3} \left\{ f_3(x_3) + \frac{\gamma}{2} \|A_3 x_3 - b - \frac{z^k}{\gamma}\|^2 \right\} \\ x_1^{k+1} &= \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \left\{ f_1(x_1) + \frac{\gamma}{2} \|A_1 x_1 + A_2 \tilde{w}^{k+1} - 2A_3 x_3^{k+1} - 2b - \frac{z^k}{\gamma}\|^2 \right\} \\ x_2^{k+1} &= \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \left\{ f_2(x_2) + \frac{\gamma}{2} \|A_1 x_1^{k+1} + A_2 x_2 - 2A_3 x_3^{k+1} - 2b - \frac{z^k}{\gamma}\|^2 \right\} \\ z^{k+1} &= z^k - \gamma(A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b) \end{aligned} \right\}, \tag{54}$$

where $\tilde{w}^{k+1} = \partial f_2^*(A_2^T w^{k+1}) = \partial f_2^*(A_2^T (z^k - \gamma(A_3 x_3^{k+1} - b)))$.

Since there are quite a few variants of 3-block ADMM methods, we consider a numerical comparison of some existing variants to the derived 3-block ADMM-like algorithm (54), as a demonstration of its practical efficiency. We consider an example transformed from the problem (46). Let $f_2(x_2) = \frac{\alpha}{2} \|x_2 - u\|_2^2$, $f_1(x_1) = i_{\Lambda_1}(x_1)$, and $f_3(x_3) = i_{\Lambda_2}(x_3)$, where the two indicator functions are for the same two sets as in (46). The derived scheme applies here with the following constraint:

$$A_1 x_1 + A_2 x_2 + A_3 x_3 = d,$$

where $A_1 = \begin{bmatrix} I \\ 0 \end{bmatrix} \in \mathbb{R}^{2n \times n}$, $A_2 = \begin{bmatrix} -I \\ I \end{bmatrix} \in \mathbb{R}^{2n \times n}$, $A_3 = \begin{bmatrix} 0 \\ -I \end{bmatrix} \in \mathbb{R}^{2n \times n}$, and $d = \mathbf{0}_{2n} \in \mathbb{R}^{2n}$. For this problem, we obtain from (54):

$$\begin{aligned} x_3^{k+1} &= -z_2^k/\gamma + \left(\sum_{i=1}^n z_2^k(i)/\gamma + b \right) / n \cdot \mathbf{1}_n; \\ \tilde{w}^{k+1} &= (-z_1^k + z_2^k + \gamma x_3^{k+1})/\alpha + u; \\ x_1^{k+1} &= \min(\max(\tilde{w}^{k+1} + z_1/\gamma), m), M) \\ x_2^{k+1} &= (\alpha u + \gamma x_1^{k+1} - 2\gamma x_3^{k+1} - z_1^k + z_2^k)/(\alpha + 2\gamma) \\ z_1^{k+1} &= z_1^k - \gamma(x_1^{k+1} - x_2^{k+1}) \\ z_2^{k+1} &= z_2^k - \gamma(x_2^{k+1} - x_3^{k+1}). \end{aligned}$$

The direct extension of 3-block ADMM (33) applied to this problem is:

$$\begin{aligned} x_1^{k+1} &= \min(\max(\tilde{x}_2^{k+1} + w_1/\gamma), m), M); \\ x_2^{k+1} &= (\alpha u + \gamma x_1^{k+1} + \gamma x_3^{k+1} - w_1^k + w_2^k)/(\alpha + 2\gamma) \\ x_3^{k+1} &= x_2 - w_2/\gamma - \left(\sum_{i=1}^n (x_2(i) - w_2^k(i)/\gamma) + b \right) / n \cdot \mathbf{1}_n \end{aligned}$$

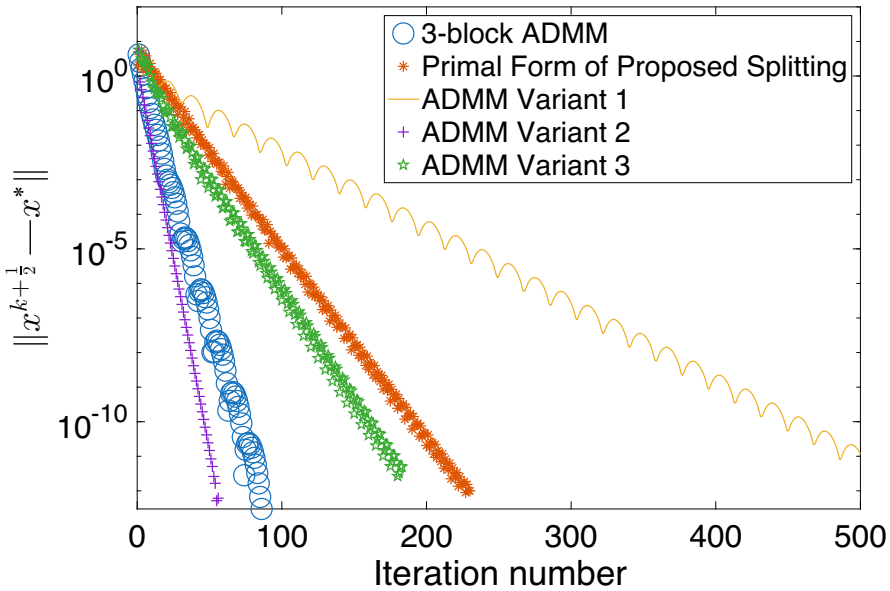


Fig. 5 Comparison of the error $\|x^{k+\frac{1}{2}} - x^*\|$ between the derived algorithm (54) (the corresponding legend is *Primal Form of Proposed Splitting*) and other 3-block ADMM variants. The convergence rate of derived algorithm is between the ADMM variants.

$$\begin{aligned} w_1^{k+1} &= w_1^k - \gamma(x_1^{k+1} - x_2^{k+1}) \\ w_2^{k+1} &= w_2^k - \gamma(x_2^{k+1} - x_3^{k+1}), \end{aligned}$$

where $\mathbf{1}_n$ is the n -dimensional vector with all components equal to 1. We compare the derived algorithm with the direct extension of 3-block extension of ADMM with variant 1 (He and Yuan 2018) that uses prediction correction steps with parameter $\tau = \frac{1}{2}$ in the correction step, variant 2 in Davis and Yin (2017) that only changes the x_1^{k+1} update, and variant 3 in He et al. (2012) that uses prediction-correction steps with Gaussian back substitution with $\alpha = \frac{1}{2}$ in the correction step. For simplicity we use step size $\gamma = 1$ for algorithms. All methods are tested on the problem (46) with the same parameters $\alpha = 1, n = 100, m = -1, M = 1$, and $b = Au$, where u is constructed by perturbing a sine profile by random noise:

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

The comparison is shown in Figure 5.

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. The numerical tests were mainly performed by JL 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.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bai J, Chen Y, Yu X, Zhang H (2025) Generalized asymmetric forward-backward-adjoint algorithms for convex-concave saddle-point problem. *J Sci Comput* 102(3):80
- Bai J, Li J, Xu F, Zhang H (2018) Generalized symmetric ADMM for separable convex optimization. *Comput Optim Appl* 70:129–170
- Bai J, Zhang M, Zhang H (2025) An inexact ADMM for separable nonconvex and nonsmooth optimization. *Comput Optim Appl* 1–35
- Bauschke HH, Combettes PL (2017) *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, Berlin
- Boğ RI, Nguyen DK (2023) Fast Krasnosel'skiĭ-Mann algorithm with a convergence rate of the fixed point iteration of $o(\frac{1}{k})$. *SIAM J Numer Anal* 61(6):2813–2843
- Boyd S, Parikh N, Chu E, Peleato B, Eckstein J et al (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found Trends@ Machine Learn* 3(1):1–122
- Briceno-Arias LM (2015) Forward-Douglas-Rachford splitting and forward-partial inverse method for solving monotone inclusions. *Optimization* 64(5):1239–1261
- Cai X, Han D, Yuan X (2017) On the convergence of the direct extension of ADMM for three-block separable convex minimization models with one strongly convex function. *Comput Optim Appl* 66:39–73
- Chambolle A, Pock T (2011) A first-order primal-dual algorithm for convex problems with applications to imaging. *J Math Imag Vision* 40(1):120–145
- Chen C, He B, Ye Y, Yuan X (2016) The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Math Program* 155(1–2):57–79
- Chen C, Shen Y, You Y (2013) 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

- Chen P, Huang J, Zhang X (2016) A primal-dual fixed point algorithm for minimization of the sum of three convex separable functions. *Fixed Point Theory Appl*. <https://doi.org/10.1186/s13663-016-0543-2>
- Condat L (2013) A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J Optim Theory Appl* 158:460–479
- Davis D, Yin W (2016) Convergence rate analysis of several splitting schemes. *Splitting Methods Commun Imag Sci Eng* 115–163. DOI:10.1007/978-3-319-41589-5_4
- Davis D, Yin W (2017) A three-operator splitting scheme and its optimization applications. *Set-Valued Var Anal* 25:829–858
- Demanet L, Zhang X (2016) Eventual linear convergence of the Douglas-Rachford iteration for basis pursuit. *Math Comput* 85(297):209–238
- Douglas J, Rachford HH (1956) On the numerical solution of heat conduction problems in two and three space variables. *Trans Am Math Soc* 82(2):421–439
- Gabay D, Mercier B (1976) A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Comput Math Appl* 2(1):17–40
- Glowinski R, Marroco A (1975) 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
- Goldstein T, Osher S (2009) The split Bregman method for l_1 -regularized problems. *SIAM J Imag Sci* 2(2):323–343
- Han D, Yuan X (2012) A note on the alternating direction method of multipliers. *J Optim Theory Appl* 155:227–238
- He B, Ma F, Xu S, Yuan X (2022) A generalized primal-dual algorithm with improved convergence condition for saddle point problems. *SIAM J Imag Sci* 15(3):1157–1183
- He B, Tao M, Yuan X (2012) Alternating direction method with gaussian back substitution for separable convex programming. *SIAM J Optim* 22:313–340
- He B, Yuan X (2018) A class of ADMM-based algorithms for three-block separable convex programming. *Comput Optim Appl* 70:791–826
- Hong M, Luo ZQ (2017) On the linear convergence of the alternating direction method of multipliers. *Math Program* 162(1):165–199
- Lataf P, Patrinos P (2017) Asymmetric forward–backward–adjoint splitting for solving monotone inclusions involving three operators. *Comput Optim Appl* 68:57–93
- Li M, Sun D, Toh KC (2015) A convergent 3-block semi-proximal ADMM for convex minimization problems with one strongly convex block. *Asia-Pacific J Oper Res* 32(04):1550024
- Lin T, Ma S, Zhang S (2015) On the global linear convergence of the ADMM with multiblock variables. *SIAM J Optim* 25(3):1478–1497
- Lin T, Ma S, Zhang S (2016) Iteration complexity analysis of multi-block ADMM for a family of convex minimization without strong convexity. *J Sci Comput* 69(1):52–81
- Lin TY, Ma SQ, Zhang SZ (2015) On the sublinear convergence rate of multi-block ADMM. *J Oper Res Soc China* 3:251–274
- Lions PL, Mercier B (1979) Splitting algorithms for the sum of two nonlinear operators. *SIAM J Numer Anal* 16(6):964–979
- Liu C, Buzzard GT, Zhang X (2024) An optimization based limiter for enforcing positivity in a semi-implicit discontinuous Galerkin scheme for compressible Navier-Stokes equations. *J Comput Phys* 519:113440
- Liu C, Hu J, Taitano WT, Zhang X (2025) An optimization-based positivity-preserving limiter in semi-implicit discontinuous Galerkin schemes solving Fokker-Planck equations. *Comput Math Appl* 192:54–71
- Liu C, Riviere B, Shen J, Zhang X (2024) A simple and efficient convex optimization based bound-preserving high order accurate limiter for Cahn-Hilliard-Navier-Stokes system. *SIAM J Sci Comput* 46(3):A1923–A1948
- Malitsky Y, Tam MK (2020) A forward-backward splitting method for monotone inclusions without cocoercivity. *SIAM J Optim* 30(2):1451–1472
- Nurbekyan L, Liu S, Chow YT (2024) Monotone inclusion methods for a class of second-order non-potential mean-field games. *arXiv preprint arXiv:2403.20290*
- Passty GB (1979) Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J Math Anal Appl* 72(2):383–390

- Peaceman DW, Rachford HH Jr (1955) The numerical solution of parabolic and elliptic differential equations. *J Soc Ind Appl Math* 3(1):28–41
- Raguet H (2019) A note on the forward-Douglas-Rachford splitting for monotone inclusion and convex optimization. *Optim Lett* 13:717–740
- Raguet H, Fadili J, Peyré G (2013) A generalized forward-backward splitting. *SIAM J Imag Sci* 6(3):1199–1226
- Ryu EK, Vũ BC (2020) Finding the forward-Douglas–Rachford-forward method. *J Optim Theory Appl* 184(3):858–876
- Shi W, Ling Q, Yuan K, Wu G, Yin W (2014) On the linear convergence of the ADMM in decentralized consensus optimization. *IEEE Trans Signal Process* 62(7):1750–1761
- Torres EG, Jacobs M, Zhang X (2025) Asymptotic linear convergence of ADMM for isotropic TV norm compressed sensing. arXiv preprint [arXiv:2505.01240](https://arxiv.org/abs/2505.01240)
- Vũ BC (2011) A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv Comput Math* 38:667–681
- Wei E, Ozdaglar A (2012) Distributed alternating direction method of multipliers. In: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC), pp. 5445–5450. IEEE
- Yan M (2018) A new primal-dual algorithm for minimizing the sum of three functions with a linear operator. *J Sci Comput* 76:1698–1717
- Yin J, Tang C, Jian J, Huang Q (2024) A partial Bregman ADMM with a general relaxation factor for structured nonconvex and nonsmooth optimization. *J Global Optim* 89(4):899–926
- Zhang X (2017) On positivity-preserving high order discontinuous Galerkin schemes for compressible Navier-Stokes equations. *J Comput Phys* 328:301–343
- Zhang X, Shu CW (2012) A minimum entropy principle of high order schemes for gas dynamics equations. *Numer Math* 121(3):545–563

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.