

Extending Douglas–Rachford Splitting for Convex Optimization

Max Nilsson[†], Anton Åkerman[†], and Pontus Giselsson

Department of Automatic Control, Lund University, Lund, Sweden

{max.nilsson, anton.akerman, pontus.giselsson}@control.lth.se

[†]These authors contributed equally to this work.

Abstract

The Douglas–Rachford splitting method is a classical and widely used algorithm for solving monotone inclusions involving the sum of two maximally monotone operators. It was recently shown to be the unique frugal, no-lifting resolvent-splitting method that is unconditionally convergent in the general two-operator setting. In this work, we show that this uniqueness does not hold in the convex optimization case: when the operators are subdifferentials of proper, closed, convex functions, a strictly larger class of frugal, no-lifting resolvent-splitting methods is unconditionally convergent. We provide a complete characterization of all such methods in the convex optimization setting and prove that this characterization is sharp: unconditional convergence holds exactly on the identified parameter regions. These results immediately yield new families of convergent ADMM-type and Chambolle–Pock-type methods obtained through their Douglas–Rachford reformulations.

1 Introduction

The Douglas–Rachford splitting method is a celebrated algorithm for solving monotone inclusion problems of the form

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in A(x) + B(x), \quad (1)$$

where $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ are maximally monotone operators and \mathcal{H} is a real Hilbert space. While originally developed for solving heat conduction problems [1], it has since Lions and Mercier in [2] proved its applicability to maximal monotone inclusion problems been extensively used and analyzed. A particularly important case of study is the convex optimization setting with problems of the form

$$\underset{x \in \mathcal{H}}{\text{minimize}} \quad f(x) + g(x), \quad (2)$$

where $f, g : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper, closed, and convex functions. Under mild assumptions (see e.g., [3]), the problem (2) is a specialization of (1) by letting $A = \partial f$ and $B = \partial g$.

An early interpretation of the Douglas–Rachford method was provided in [4, 5], where it was shown to be an instance of the proximal point algorithm [6, 7], offering a powerful framework for its analysis. Many subsequent works have provided (local) linear convergence for the Douglas–Rachford splitting method under various assumptions. For instance: [8, 9, 10, 11] for

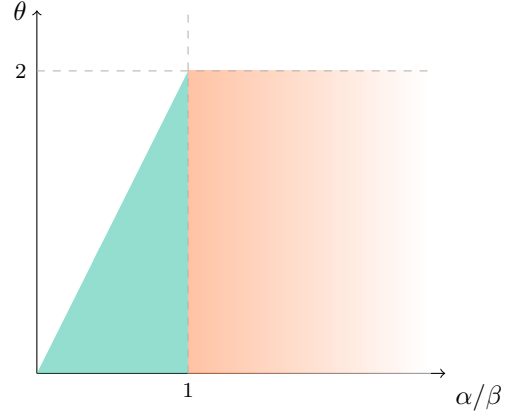


Figure 1: Parameter region of unconditional convergence for the algorithm in (4). The region of unconditional convergence for the general monotone inclusion setting (i.e., for the algorithm in (3)) is the line segment $\alpha/\beta = 1$ and $\theta \in (0, 2)$.

convex and nonconvex feasibility problems, [12, 13, 14] for convex optimization problems under more or less general partial smoothness assumptions, [2, 15, 16, 17, 18] under different combinations of strong monotonicity and smoothness assumptions in monotone inclusion and convex optimization settings, and [19, 20] for nonconvex optimization. Many of these analyses are tight in the sense that they are not improvable without changing the assumptions.

The remarkable success of the Douglas–Rachford splitting method can perhaps in part be explained by the result of [21], which shows that it is the unique “simple” resolvent splitting method for problems of the form (1) that both possesses the *fixed-point encoding* property and is *unconditionally convergent*. Here, “simple” refers to methods that use *minimal lifting* and are *frugal*. Loosely speaking, a resolvent splitting method is an algorithm that accesses the operators A and B only through their resolvents (denoted $J_{\alpha A}$ and $J_{\beta B}$ respectively, where $\alpha, \beta \in \mathbb{R}_{++}$ are step sizes) and forms iterates from pre-defined linear combinations of their inputs and outputs. Such a method is *frugal* if it evaluates each resolvent only once per iteration, has *minimal lifting* if the dimension d (the lifting number) of the product space vector in \mathcal{H}^d that carries information between algorithm iterations is as small as possible (in the Douglas–Rachford case, we have no lifting, i.e., $d = 1$), is a *fixed-point encoding* if the fixed-point set of the algorithm and the solution set of the problem are simultaneously nonempty, and is *unconditionally convergent* if there exist parameter choices guaranteeing convergence for every problem in the problem class and every algorithm starting point.

The uniqueness of the Douglas–Rachford method under these conditions for problems of the form (1) is established in [21] through two results: one parameterizing the class of frugal, no-lifting resolvent methods with the fixed-point encoding property, and one characterizing its subset of methods with unconditional convergence. It was shown that each frugal, no-lifting resolvent method for solving (1)—up to equivalence—satisfies the fixed-point encoding property if and only if it has the form

$$\begin{cases} x_1 = J_{\alpha A}(z) \\ x_2 = J_{\beta B} \left(\left(1 + \frac{\beta}{\alpha}\right) x_1 - \frac{\beta}{\alpha} z \right), \\ z^+ = z + \theta(x_2 - x_1) \end{cases}, \quad (3)$$

with parameters $\alpha, \beta \in \mathbb{R}_{++}$ and $\theta \in \mathbb{R} \setminus \{0\}$. The second result in [21] establishes that, among algorithms of the form (3), the restrictions $\alpha = \beta$ and $\theta \in (0, 2)$ are necessary and sufficient for unconditional convergence whenever $\dim(\mathcal{H}) \geq 2$. The results in [21] have sparked much research on minimal lifting methods, see, e.g., [22, 23, 24, 25], which address monotone inclusions involving sums of arbitrarily many operators.

In this work, we address the following question in the two-operator setting:

Can we characterize all frugal, no-lifting, unconditionally convergent resolvent splitting methods with fixed-point encoding for solving (1) in the convex optimization setting $A = \partial f$, $B = \partial g$?

We show that the answer is *yes*. Moreover, the restriction to the optimization setting leads to a significantly larger class of admissible algorithms than in the general monotone inclusion setting of [21]. In particular, we show that the convex-optimization specialization of (3)

$$\begin{cases} x_1 = \text{prox}_{\alpha f}(z) \\ x_2 = \text{prox}_{\beta g} \left(\left(1 + \frac{\beta}{\alpha}\right) x_1 - \frac{\beta}{\alpha} z \right), \\ z^+ = z + \theta(x_2 - x_1) \end{cases}, \quad (4)$$

completely characterizes all frugal, no-lifting, unconditionally convergent resolvent splittings with fixed-point encoding in the convex optimization setting (2), when the parameters (α, β, θ) satisfy

$$(\alpha, \beta, \theta) \in \left\{ (\alpha, \beta, \theta) \in \mathbb{R}_{++}^3 \mid \theta < \min\{2, 2\alpha/\beta\} \right\}. \quad (5)$$

Figure 1 illustrates how the unconditional-convergence region expands in the convex optimization setting, shown in terms of the ratio α/β , which in the monotone-inclusion setting is constrained to satisfy $\alpha/\beta = 1$.

Unlike the Douglas–Rachford case, our sufficiency analysis does not rely on the conceptually simple proximal point algorithm or nonexpansive operators. Instead, we partition the region of unconditionally convergent parameters in (5) as $(\alpha, \beta, \theta) \in S^{(1)} \cup S^{(2)}$, where

$$\begin{aligned} S^{(1)} &:= \left\{ (\alpha, \beta, \theta) \in \mathbb{R}_{++}^3 \mid \alpha/\beta \in (0, 1], \theta \in (0, 2\alpha/\beta) \right\}, \\ S^{(2)} &:= \left\{ (\alpha, \beta, \theta) \in \mathbb{R}_{++}^3 \mid \alpha/\beta \in [1, \infty), \theta \in (0, 2) \right\}, \end{aligned}$$

and develop Lyapunov analyses for these newly identified parameter regions whose intersection $(S^{(1)} \cap S^{(2)} = \mathbb{R}_{++} \times \{(1, 1)\} \times (0, 2))$ represents the unconditional-convergence region for monotone inclusions established in [21]. The discovery of the Lyapunov inequalities has been assisted by the automatic Lyapunov methodologies and tools of [26, 27]. Necessity is demonstrated using simple one-dimensional counterexamples involving the zero function and the indicator function of $\{0\}$.

While we primarily characterize the entire class of frugal, minimal lifting, unconditionally convergent resolvent splittings with fixed-point encoding in the optimization setting, we emphasize that all our convergence guarantees for methods with $\alpha/\beta \neq 1$ appear to be new. Because the alternating direction method of multipliers (ADMM) [28, 29, 30] arises as a special case of Douglas–Rachford splitting [31, 5], our results yield a family of completely new convergent ADMM-type methods for convex optimization. Likewise, since the Chambolle–Pock algorithm [32] is also a Douglas–Rachford instance [33], we obtain an expanded class of convergent Chambolle–Pock-type methods, some of which are new. In particular, we present the first doubly relaxed Chambolle–Pock-type method (see Algorithm 4 for the specific relaxations) with freedom in both relaxation parameters. Previous works, such as [34, 35], have one of them fixed.

The rest of the paper is organized as follows. Section 2 introduces notation and definitions and presents a few preliminary results. Section 3 presents the problem setting and our main result; the complete characterization of all frugal, no-lifting resolvent splittings in the convex optimization setting that are both unconditionally convergent and has the fixed-point encoding property. Section 4 is devoted to proving the main result. Section 5 shows ergodic $\mathcal{O}(1/K)$ -convergence for two primal-dual gap functions for the characterized methods. Section 6 applies our method to specific problem instances to derive new convergent ADMM-, Chambolle–Pock-, and parallel splitting-type algorithms. Section 7 presents an operator composition perspective of the algorithm in (3) and Section 8 concludes the paper.

2 Preliminaries

2.1 Notation

Our notation closely follows [36, 3]. We let $\mathbb{N} = \{0, 1, \dots\}$ denote the natural numbers, \mathbb{R} denote the real numbers, \mathbb{R}_+ the nonnegative numbers, and \mathbb{R}_{++} the positive numbers. Moreover \mathbb{R}^n denotes the n -dimensional real vectors, $\mathbb{R}^{n \times m}$ denotes the real $n \times m$ matrices, and \mathbb{S}^n the real symmetric $n \times n$ matrices.

We let $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ be a real Hilbert space. The notation $A : \mathcal{H} \rightrightarrows \mathcal{H}$ means that A is set-valued and maps \mathcal{H} into subsets of \mathcal{H} . Its graph is defined as $\text{gra}(A) := \{(x, u) \mid u \in Ax\}$. An operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is monotone if $\langle u - v, x - y \rangle \geq 0$ for all $x, y \in \mathcal{H}$, $u \in A(x)$, and $v \in A(y)$. It is maximally monotone if there exists no monotone operator $B : \mathcal{H} \rightrightarrows \mathcal{H}$ such that $\text{gra}(B)$ properly contains $\text{gra}(A)$. The resolvent of $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is given by $J_A := (A + \text{Id})^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ and is single-valued for all maximally monotone operators, see [3, Corollary 23.9]. Since we work exclusively with resolvents of maximally monotone operators A , we regard J_A as a mapping $J_A : \mathcal{H} \rightarrow \mathcal{H}$. For a proper, closed, and convex function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$, its subdifferential is the set-valued operator $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$ defined by

$$\partial f(x) := \{u \in \mathcal{H} \mid f(y) \geq f(x) + \langle u, y - x \rangle \text{ for all } y \in \mathcal{H}\}$$

for each $x \in \mathcal{H}$. The subdifferential ∂f is maximally monotone [3, Theorem 20.25], and its resolvent coincides with the proximal operator of f , i.e., $\text{prox}_f := J_{\partial f}$; see [3, Chapter 24] for details. For a function $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ we let $\text{Argmin}_{x \in \mathcal{H}}(f(x)) = \{x \in \mathcal{H} \mid f(x) = \inf_{z \in \mathcal{H}} f(z)\}$

Let C be a subset of \mathcal{H} . We let $\iota_C : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ denote the indicator function of C , which equals 0 for $x \in C$ and $+\infty$ for $x \notin C$.

Given a matrix $Q = (q_{ij})_{i,j} \in \mathbb{R}^{n \times m}$, we associate it with the bounded linear operator $Q \otimes \text{Id} : \mathcal{H}^m \rightarrow \mathcal{H}^n$. Throughout, we use the same notation for the matrix and the operator in the following sense: given a matrix $Q \in \mathbb{R}^{n \times m}$ and a vector $v = (v_1, \dots, v_m) \in \mathcal{H}^m$, then $Qv := (Q \otimes \text{Id})v = (\sum_{j=1}^m q_{1j}v_j, \dots, \sum_{j=1}^m q_{nj}v_j) \in \mathcal{H}^n$.

Moreover, we will overload $\langle \cdot, \cdot \rangle$ to also denote the induced inner product on \mathcal{H}^m , i.e., for all $(v, w) \in \mathcal{H}^m \times \mathcal{H}^m$, we let $\langle v, w \rangle := \sum_{i=1}^m \langle v_i, w_i \rangle$. For any symmetric matrix $Q \in \mathbb{S}^m$ and any $v \in \mathcal{H}^m$, we denote $\mathcal{Q}(Q, v) := \langle v, Qv \rangle$.

We will use the following notations extensively throughout the paper. The set of maximally monotone operators is denoted by

$$\mathcal{A} := \{A : \mathcal{H} \rightrightarrows \mathcal{H} \mid A \text{ is maximally monotone}\}, \quad (6)$$

the set of proper, closed, and convex functions by

$$\mathcal{F} := \{f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\} \mid f \text{ is proper, closed, and convex}\},$$

and the corresponding subdifferential set by

$$\partial \mathcal{F} := \{A : \mathcal{H} \rightrightarrows \mathcal{H} \mid A = \partial f \text{ with } f \in \mathcal{F}\}.$$

We note that $\partial \mathcal{F} \subset \mathcal{A}$ by [3, Theorem 20.25]. Throughout this paper, the Hilbert space \mathcal{H} is assumed to satisfy $\dim \mathcal{H} \geq 1$, i.e., \mathcal{H} should not be a singleton.

2.2 Definitions and Preliminary Results

Definition 2.1 (D_f): Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. Then $D_f : (\text{dom } f)^2 \times \mathcal{H} \rightarrow \mathbb{R}$ is defined by

$$D_f(x, y, u) := f(x) - f(y) - \langle u, x - y \rangle$$

for each $(x, y) \in (\text{dom } f)^2$ and $u \in \mathcal{H}$.

A function D_f satisfying Definition 2.1 will be called a *Bregman-type distance*.

Lemma 2.1: Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. Then

$$D_f(z, y, v) - D_f(x, y, v) = D_f(z, x, u) + \langle u - v, z - x \rangle,$$

for each $(x, y, z) \in (\text{dom } f)^3$ and $(u, v) \in \mathcal{H}^2$.

Proof. By Definition 2.1 we have that

$$\begin{aligned} D_f(z, y, v) - D_f(x, y, v) &= (f(z) - f(y) - \langle v, z - y \rangle) - (f(x) - f(y) - \langle v, x - y \rangle) \\ &= f(z) - f(x) - \langle v, z - x \rangle \\ &= D_f(z, x, u) + \langle u - v, z - x \rangle. \end{aligned}$$

□

Lemma 2.2: *Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. Then*

$$D_f(x, y, v) + D_f(y, x, u) = \langle u - v, x - y \rangle$$

for each $(x, y) \in (\text{dom } f)^2$ and $(u, v) \in \mathcal{H}^2$.

Proof. Let $z = y$ in Lemma 2.1 and note that $D_f(y, y, v) = 0$. □

Lemma 2.3: *Let $f : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$. Then $D_f(x, y, u) \geq 0$ for each $(x, y) \in (\text{dom } f)^2$ and $u \in \partial f(y)$.*

Proof. This follows directly from the subdifferential definition [3, Definition 16.1]. □

3 Problem Setting and Main Results

In this section, we present our main result: a complete characterization of the frugal, no-lifting resolvent-splitting methods that are unconditionally convergent and satisfy the fixed-point encoding property for problems of the form (1) with $A = \partial f$ and $B = \partial g$, where $f, g : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper, closed, and convex. The starting point for our analysis is Theorem 1 of [21], which gives an explicit parameterization of all frugal, no-lifting resolvent-splitting methods satisfying the fixed-point encoding property, under the assumption that the two operators are maximally monotone. Since the proof of that result relies only on subdifferential operators of proper, closed, convex functions, the same parameterization applies in our setting. This parameterization is given explicitly by Algorithm 1, and hence the task of obtaining our characterization reduces to determining which choices of (α, β, θ) in Algorithm 1 lead to unconditional convergence in the convex optimization setting.

To avoid unnecessary notational overhead, we refer the reader to [21] for the formal definition of the class of frugal, no-lifting resolvent-splitting methods and for the proof of its equivalence to the parameterization in Algorithm 1. For convenience, we reproduce the characterization result here in a form relevant to our convex optimization setting.

Algorithm 1 Extended Douglas–Rachford Splitting

1: **Input:** $(A, B, z^0, \alpha, \beta, \theta) \in \mathcal{A} \times \mathcal{A} \times \mathcal{H} \times \mathbb{R}_{++} \times \mathbb{R}_{++} \times (\mathbb{R} \setminus \{0\})$

2: **for** $k = 0, 1, 2, \dots$ **do**

3: **Update:**

$$\left| \begin{aligned} x_1^k &= J_{\alpha A}(z^k) \\ x_2^k &= J_{\beta B} \left(\left(1 + \frac{\beta}{\alpha} \right) x_1^k - \frac{\beta}{\alpha} z^k \right) \\ z^{k+1} &= z^k + \theta(x_2^k - x_1^k) \end{aligned} \right.$$

4: **end for**

Theorem 3.1 (From [21] Theorem 1): *Let $f, g : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper, closed, and convex. Up to equivalence, a method is a frugal, no-lifting resolvent-splitting satisfying the fixed-point encoding property for solving (1) with $A = \partial f$ and $B = \partial g$ if and only if it is of the form Algorithm 1.*

The fixed-point encoding property ensures that solutions of the inclusion problem (1) correspond exactly to fixed points of Algorithm 1. Since unconditional convergence is defined in terms of convergence to such fixed points, we begin by making the fixed-point set of Algorithm 1 explicit.

Definition 3.1 ($\text{Fix}(A, B, \alpha, \beta)$): *For any $(A, B, \alpha, \beta) \in \mathcal{A}^2 \times \mathbb{R}_{++}^2$, let*

$$\text{Fix}(A, B, \alpha, \beta) := \left\{ (x^*, x^*, z^*) \in \mathcal{H}^3 \mid x^* = J_{\alpha A}(z^*) = J_{\beta B} \left(\left(1 + \frac{\beta}{\alpha}\right) x^* - \frac{\beta}{\alpha} z^* \right) \right\}.$$

For ease of reference, we also explicitly state the equivalence between the solution set of the inclusion (1) and the fixed-point set of Algorithm 1. In particular, this shows that whenever $(x^*, x^*, z^*) \in \text{Fix}(A, B, \alpha, \beta)$, the point x^* must be a solution of the corresponding inclusion (1).

Proposition 3.1: *Let $(A, B, \alpha, \beta) \in \mathcal{A}^2 \times \mathbb{R}_{++}^2$ and $x^* \in \mathcal{H}$. Then, $x^* \in \text{zer}(A + B)$ if and only if there exists $z^* \in \mathcal{H}$ such that $(x^*, x^*, z^*) \in \text{Fix}(A, B, \alpha, \beta)$.*

Proof. We have that $x^* \in \text{zer}(A + B)$ if and only if there exists some $z^* \in \mathcal{H}$ such that

$$\begin{aligned} & \begin{cases} \frac{1}{\alpha}(z^* - x^*) & \in A(x^*) \\ \frac{1}{\alpha}(x^* - z^*) & \in B(x^*) \end{cases} \\ \iff & \begin{cases} \frac{1}{\alpha}(z^* - x^*) & \in A(x^*) \\ \frac{1}{\beta} \left(\left(1 + \frac{\beta}{\alpha}\right) x^* - \frac{\beta}{\alpha} z^* - x^* \right) & \in B(x^*) \end{cases} \\ \iff & \begin{cases} x^* & = J_{\alpha A}(z^*) \\ x^* & = J_{\beta B} \left(\left(1 + \frac{\beta}{\alpha}\right) x^* - \frac{\beta}{\alpha} z^* \right) \end{cases} \\ \iff & (x^*, x^*, z^*) \in \text{Fix}(A, B, \alpha, \beta). \end{aligned}$$

□

We now formalize what it means for Algorithm 1 to converge unconditionally for a given parameter choice over a subset of operator pairs from \mathcal{A} , the class of maximally monotone operators introduced in (6).

Definition 3.2: *Let $\mathcal{C} \subset \mathcal{A} \times \mathcal{A}$ and $(\alpha, \beta, \theta) \in \mathbb{R}_{++}^2 \times (\mathbb{R} \setminus \{0\})$. We say that Algorithm 1 **converges unconditionally for** (α, β, θ) **over** \mathcal{C} if, for every $z^0 \in \mathcal{H}$ and every $(A, B) \in \mathcal{C}$ with $\text{zer}(A + B) \neq \emptyset$, the sequence $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ generated by Algorithm 1 applied to $(A, B, z^0, \alpha, \beta, \theta)$ converges weakly to a point $(x^*, x^*, z^*) \in \text{Fix}(A, B, \alpha, \beta)$.*

We now introduce a convenient notation for describing the set of parameters in Algorithm 1 that guarantee unconditional convergence for a given class of operator pairs. This will allow us to state our results in a compact form.

Definition 3.3 ($S(\mathcal{C})$): *Let $\mathcal{C} \subset \mathcal{A} \times \mathcal{A}$. Define*

$$S(\mathcal{C}) := \{(\alpha, \beta, \theta) \in \mathbb{R}_{++}^2 \times (\mathbb{R} \setminus \{0\}) \mid \text{Algorithm 1 converges unconditionally for } (\alpha, \beta, \theta) \text{ over } \mathcal{C}\}.$$

With this notation in place, we can state the following key result from [21, Theorem 2].

Theorem 3.2 ([21] Theorem 2): *If $\dim \mathcal{H} \geq 2$, then Algorithm 1 converges unconditionally for $(\alpha, \beta, \theta) \in \mathbb{R}_{++}^2 \times (\mathbb{R} \setminus \{0\})$ over $\mathcal{A} \times \mathcal{A}$ if and only if $\alpha = \beta$ and $\theta \in (0, 2)$, i.e.,*

$$S(\mathcal{A} \times \mathcal{A}) = \mathbb{R}_{++}\{(1, 1)\} \times (0, 2).$$

Before stating our main result in Theorem 3.3, we introduce the following two parameter sets, whose union will be shown to constitute the full region of unconditional convergence in the convex optimization setting. This division is used because our proof treats the two cases separately. The extent to which this enlarges the admissible parameter region compared with the maximal monotone setting of [21] is illustrated in Figure 1.

Definition 3.4 ($S^{(i)}$): *Let $S^{(1)}, S^{(2)} \subset \mathbb{R}^3$ be defined by*

$$\begin{aligned} S^{(1)} &:= \left\{ (\alpha, \beta, \theta) \in \mathbb{R}_{++}^3 \mid \alpha/\beta \in (0, 1], \theta \in (0, 2\alpha/\beta) \right\}, \\ S^{(2)} &:= \left\{ (\alpha, \beta, \theta) \in \mathbb{R}_{++}^3 \mid \alpha/\beta \in [1, \infty), \theta \in (0, 2) \right\}. \end{aligned}$$

We are now ready to state our main result.

Theorem 3.3: *Algorithm 1 converges unconditionally for $(\alpha, \beta, \theta) \in \mathbb{R}_{++}^2 \times (\mathbb{R} \setminus \{0\})$ over $\partial\mathcal{F} \times \partial\mathcal{F}$ if and only if $\theta \in (0, \min\{2, 2\alpha/\beta\})$. Moreover,*

$$S(\partial\mathcal{F} \times \partial\mathcal{F}) = S(\partial\mathcal{F} \times \mathcal{A}) = S^{(1)} \cup S^{(2)}.$$

Note that this result applies not only to the convex optimization setting but also to the mixed case in which A is a subdifferential operator of a proper, closed, and convex function while B is an arbitrary maximally monotone operator.

In contrast to the maximal monotone setting, no additional assumptions on the dimension of \mathcal{H} are needed when restricting to $\partial\mathcal{F} \times \partial\mathcal{F}$ or $\partial\mathcal{F} \times \mathcal{A}$. In particular, Theorem 3.3 also covers the one-dimensional case. In light of the fact that every maximally monotone operator on \mathbb{R} is the subdifferential of a proper, closed, and convex function (see [3, Corollary 22.23]) and since any one-dimensional \mathcal{H} is isometrically isomorphic to \mathbb{R} which preserves maximal monotonicity and subdifferentials, Theorem 3.3 allows us to revisit [21, Theorem 2] and close the gap left open in the case $\dim \mathcal{H} = 1$.

Corollary 3.1: *If $\mathcal{H} = \mathbb{R}$, then $S(\mathcal{A} \times \mathcal{A}) = S(\partial\mathcal{F} \times \partial\mathcal{F}) = S^{(1)} \cup S^{(2)}$.*

Proof. Combine Theorem 3.3 with the fact that when $\mathcal{H} = \mathbb{R}$, then $\mathcal{A} = \partial\mathcal{F}$, see [3, Corollary 22.23]. \square

4 Convergence Analysis

This section is devoted to the proof of Theorem 3.3. In Section 4.1, we establish sufficiency of the parameter conditions by means of a Lyapunov analysis, showing that

$$S^{(1)} \cup S^{(2)} \subset S(\partial\mathcal{F} \times \mathcal{A}) \subset S(\partial\mathcal{F} \times \partial\mathcal{F}).$$

In Section 4.2, we prove necessity by constructing explicit counterexamples that do not converge outside these regions, thereby showing that

$$S(\partial\mathcal{F} \times \partial\mathcal{F}) \subset S^{(1)} \cup S^{(2)}.$$

Finally, in Section 4.3, we combine these results to conclude the proof of Theorem 3.3.

4.1 Lyapunov Analysis

The Lyapunov analysis will be carried out over the class $\partial\mathcal{F} \times \mathcal{A}$; that is, we consider inclusion problems of the form

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in \partial f(x) + B(x),$$

where $f \in \mathcal{F}$ and $B \in \mathcal{A}$. This includes inclusion problems over $\partial\mathcal{F} \times \partial\mathcal{F}$ as a special case.

We will consider sequences $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$, generated by Algorithm 1 applied to $(\partial f, B, \alpha, \beta, \theta) \in \partial\mathcal{F} \times \mathcal{A} \times \mathbb{R}_{++}^2 \times (\mathbb{R} \setminus \{0\})$. We first introduce some auxiliary variables in the following definition to simplify notation.

Definition 4.1 (u^*, u_i^k): Let $(f, B) \in \mathcal{F} \times \mathcal{A}$ such that $\text{zer}(\partial f + B) \neq \emptyset$ and $(\alpha, \beta, \theta) \in \mathbb{R}_{++}^2 \times (\mathbb{R} \setminus \{0\})$. Let $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ be generated by Algorithm 1 applied to $(\partial f, B, \alpha, \beta, \theta)$. Let $(x^*, x^*, z^*) \in \text{Fix}(\partial f, B, \alpha, \beta)$. Then, define

$$u^* := \frac{1}{\alpha}(z^* - x^*) \in \partial f(x^*)$$

and define the sequences $(u_1^k)_{k \in \mathbb{N}}$ and $(u_2^k)_{k \in \mathbb{N}}$ by

$$\begin{aligned} u_1^k &:= \frac{1}{\alpha}(z^k - x_1^k) \in \partial f(x_1^k) \\ u_2^k &:= \frac{1}{\beta} \left(\left(1 + \frac{\beta}{\alpha}\right) x_1^k - \frac{\beta}{\alpha} z^k - x_2^k \right) \in B(x_2^k) \end{aligned}$$

for every $k \in \mathbb{N}$.

We now introduce the *Lyapunov function* V_k and the *residual function* R_k . Since the analysis separates into the two parameter regions $S^{(1)}$ and $S^{(2)}$, we define two versions $V_k^{(i)}$ and $R_k^{(i)}$ for $i \in \{1, 2\}$, each tailored to its respective region. When $(\alpha, \beta, \theta) \in S^{(i)}$, the pair $(V_k^{(i)}, R_k^{(i)})$ is the one used in the convergence analysis.

Definition 4.2 ($Q_k^{(i)}, V_k^{(i)}, R_k^{(i)}$): Let $(f, B) \in \mathcal{F} \times \mathcal{A}$ such that $\text{zer}(\partial f + B) \neq \emptyset$ and $(\alpha, \beta, \theta) \in \mathbb{R}_{++}^2 \times (\mathbb{R} \setminus \{0\})$ such that $\theta \neq 4\alpha/\beta$. Let $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ be generated by Algorithm 1 applied to $(\partial f, B, \alpha, \beta, \theta)$. Let $(x^*, x^*, z^*) \in \text{Fix}(\partial f, B, \alpha, \beta)$ and $\tau := \alpha/\beta \in \mathbb{R}_{++}$. Then, define

$$\begin{aligned} Q_k^{(1)} &:= \left\| z^k - z^* + \frac{\theta}{2}(x_2^k - x_1^k) + (\tau - 1)(x_1^k - x^*) \right\|^2 + \frac{\theta(4\tau - \theta)}{4} \|x_2^k - x_1^k\|^2 \\ &\quad + \tau(1 - \tau) \|x_1^k - x^*\|^2, \\ V_k^{(1)} &:= Q_k^{(1)} + 2\alpha(1 - \tau)D_f(x_1^k, x^*, u^*), \\ R_k^{(1)} &:= \theta(2\tau - \theta) \|x_1^k - x_2^k\|^2 + (1 - \tau) \|x_1^{k+1} - x_1^k\|^2 + 2\alpha(1 - \tau)D_f(x_1^k, x_1^{k+1}, u_1^{k+1}) \end{aligned}$$

and

$$\begin{aligned} Q_k^{(2)} &:= \left\| z^k - z^* + \frac{\theta}{2}(x_2^k - x_1^k) \right\|^2 + \frac{\theta(4\tau - \theta)}{4} \left\| x_2^k - x_1^k + \frac{2(\tau - 1)}{4\tau - \theta}(x_1^k - x^*) \right\|^2 \\ &\quad + \frac{\tau(\tau - 1)(4 - \theta)}{4\tau - \theta} \|x_1^k - x^*\|^2, \\ V_k^{(2)} &:= Q_k^{(2)} + 2\alpha(\tau - 1)D_f(x^*, x_1^k, u_1^k), \\ R_k^{(2)} &:= (\tau - 1) \|x_1^{k+1} - x_1^k + \theta(x_1^k - x_2^k)\|^2 + \tau\theta(2 - \theta) \|x_1^k - x_2^k\|^2 + 2\alpha(\tau - 1)D_f(x_1^{k+1}, x_1^k, u_1^k), \end{aligned}$$

for all $k \in \mathbb{N}$.

Remark 4.1: These Lyapunov and residual functions are in fact the same on the region we consider, i.e., $V_k^{(1)} = V_k^{(2)}$ and $R_k^{(1)} = R_k^{(2)}$ for all $(\alpha, \beta, \theta) \in S^{(1)} \cup S^{(2)}$. We nevertheless use these two representations since it, for instance, is not immediately clear why $V_k^{(1)}$ and $R_k^{(1)}$ are nonnegative for $(\alpha, \beta, \theta) \in S^{(2)}$ and conversely for $V_k^{(2)}$ and $R_k^{(2)}$ on $(\alpha, \beta, \theta) \in S^{(1)}$. Working with both representations makes the analysis more transparent. The identities $V_k^{(1)} = V_k^{(2)}$ and $R_k^{(1)} = R_k^{(2)}$ are verified symbolically in [37], but the analysis to come does not rely on these identities.

Remark 4.2: Note that if $(\alpha, \beta, \theta) \in S^{(1)} \cup S^{(2)}$, then $4\alpha/\beta > \theta$. This implies, in particular, that the definition of $Q_k^{(2)}$ never involves division by 0.

Remark 4.3: Note that, since $\text{ran } J_{\alpha\partial f} = \text{dom } \partial f \subset \text{dom } f$, by [3, Proposition 23.2(i), Proposition 16.4(i)], the Bregman-type distances D_f in $V_k^{(i)}$ and $R_k^{(i)}$ are well-defined and nonnegative according to Definitions 2.1 and 4.1 and Lemma 2.3 for all $i \in \{1, 2\}$ and all $k \in \mathbb{N}$.

Remark 4.4: Note that there for each $i \in \{1, 2\}$ exists a unique matrix $Q_V^{(i)} \in \mathbb{S}^3$ such that

$$Q_k^{(i)} = \mathcal{Q}(Q_V^{(i)}, (x_1^k - x^*, x_2^k - x^*, z^k - z^*))$$

holds for all $k \in \mathbb{N}$ and all $(x^*, x^*, z^*) \in \text{Fix}(\partial f, B, \alpha, \beta)$. Viewing $Q_k^{(i)}$ as a quadratic form in this way will simplify the proof of Proposition 4.3, see Appendix A for the explicit expression for $Q_V^{(i)}$, lifted to $\mathbb{R}^5 \times \mathbb{R}^5$.

Besides the Lyapunov and residual functions, we need the following quantity in order to state the Lyapunov equality.

Definition 4.3 (I_k): With $u_i^k, u^* \in \mathcal{H}$ as in Definition 4.1, let

$$I_k := \langle u_1^k - u^*, x_1^k - x^* \rangle + \langle u_2^k + u^*, x_2^k - x^* \rangle + \langle u_1^{k+1} - u^*, x_1^{k+1} - x^* \rangle + \langle u_2^{k+1} + u^*, x_2^{k+1} - x^* \rangle.$$

for each $k \in \mathbb{N}$.

Lemma 4.1: The quantity I_k in Definition 4.3 satisfies $I_k \geq 0$ for all $k \in \mathbb{N}$.

Proof. Let $k \in \mathbb{N}$. By Definition 4.1 we get that each of the four inner products in I_k is a monotonicity term; the first and third arise from the monotonicity of ∂f evaluated at (x_1^k, x^*) and (x_1^{k+1}, x^*) , while the second and fourth arise from the monotonicity of B evaluated at (x_2^k, x^*) and (x_2^{k+1}, x^*) . Therefore, $I_k \geq 0$. \square

The following Lyapunov equality is the key component for establishing unconditional convergence of Algorithm 1.

Proposition 4.1: With $V_k^{(i)}$ and $R_k^{(i)}$ as in Definition 4.2 and I_k as in Definition 4.3, if $(\alpha, \beta, \theta) \in S^{(i)}$, then $V_k^{(i)} \geq 0$, $R_k^{(i)} \geq 0$, and

$$V_{k+1}^{(i)} = V_k^{(i)} - R_k^{(i)} - \theta\alpha I_k, \tag{7}$$

holds for all $k \in \mathbb{N}$ and $i \in \{1, 2\}$.

Proof. The fact that $V_k^{(i)}$ and $R_k^{(i)}$ for both $i \in \{1, 2\}$ are nonnegative follows immediately from their respective definitions, since all coefficients in front of the norms and the Bregman-type distances are nonnegative when $(\alpha, \beta, \theta) \in S^{(i)}$ and that the Bregman-type distances are

nonnegative by Lemma 2.3. The equality follows after expanding every norm expression of $V_k^{(i)}$, $V_{k+1}^{(i)}$, and $R_k^{(i)}$, using the update step $z^{k+1} = z^k + \theta(x_2^k - x_1^k)$, using Lemma 2.1, and verifying that

$$V_k^{(i)} - V_{k+1}^{(i)} - R_k^{(i)} = \theta\alpha I_k \quad (\mathbf{LE})$$

holds for all $k \in \mathbb{N}$ and $i \in \{1, 2\}$. The Lyapunov equality in (\mathbf{LE}) is verified for both $i \in \{1, 2\}$ in Appendix A and symbolically in [37]. \square

The Lyapunov equality in Proposition 4.1 allows us to draw the following conclusions.

Proposition 4.2: *With $(x_i^k)_{k \in \mathbb{N}}$ and $(z^k)_{k \in \mathbb{N}}$ generated by Algorithm 1, $(u_i^k)_{k \in \mathbb{N}}$ as in Definition 4.1, $V_k^{(i)}$, $Q_k^{(i)}$, $R_k^{(i)}$ as in Definition 4.2, and I_k as in Definition 4.3, we have that if $(\alpha, \beta, \theta) \in S^{(i)}$, then*

- (i) $(R_k^{(i)})_{k \in \mathbb{N}}$ and $(I_k)_{k \in \mathbb{N}}$ are summable,
- (ii) $x_1^k - x_2^k \rightarrow 0$ and $u_1^k + u_2^k \rightarrow 0$ as $k \rightarrow \infty$,
- (iii) $(x_1^k)_{k \in \mathbb{N}}$, $(x_2^k)_{k \in \mathbb{N}}$, $(u_1^k)_{k \in \mathbb{N}}$, $(u_2^k)_{k \in \mathbb{N}}$ and $(z^k)_{k \in \mathbb{N}}$ are bounded,
- (iv) $(Q_k^{(i)})_{k \in \mathbb{N}}$ converges,

holds for all $i \in \{1, 2\}$.

Proof. Unless stated otherwise, the following is valid for both $i = 1$ and $i = 2$. From Proposition 4.1 and Lemma 4.1 we conclude after telescoping (\mathbf{LE}) that, for all $K \in \mathbb{N}$:

$$0 \leq \sum_{k=0}^K (R_k^{(i)} + \theta\alpha I_k) \leq V_0^{(i)} - V_{K+1}^{(i)} \leq V_0^{(i)},$$

where we also used $\alpha, \theta > 0$ for all $(\alpha, \beta, \theta) \in S^{(i)}$. By letting $K \rightarrow \infty$, we conclude that (i) holds.

Moreover, since all coefficients in front of the terms in $R_k^{(i)}$ are nonnegative for $(\alpha, \beta, \theta) \in S^{(i)}$, since the Bregman-type terms are nonnegative by Lemma 2.3, and since $\tau = \alpha/\beta$ and $\theta(2 - \theta)$ are strictly positive, we get that $R_k^{(i)} \rightarrow 0$ as $k \rightarrow \infty$ implies that $x_1^k - x_2^k \rightarrow 0$ as $k \rightarrow \infty$. By Definition 4.1, we have that

$$u_1^k + u_2^k = \frac{1}{\beta}(x_1^k - x_2^k) \rightarrow 0$$

as $k \rightarrow \infty$ and so (ii) holds.

Since

$$0 \leq V_{k+1}^{(i)} \leq V_k^{(i)} - R_k^{(i)} \leq V_k^{(i)}$$

for all $k \in \mathbb{N}$, we conclude that $(V_k^{(i)})_{k \in \mathbb{N}}$ is a sequence of nonincreasing nonnegative numbers and so $(V_k^{(i)})_{k \in \mathbb{N}}$ converges. Since all terms of $V_k^{(i)}$ are nonnegative, we have in particular if $i = 1$ that $\tau(1 - \tau) \|x_1^k - x^*\|^2$ is, so $((1 - \tau)(x_1^k - x^*))_{k \in \mathbb{N}}$ is bounded since $\tau > 0$. Using this, since also

$$\left(\left\| z^k - z^* + \frac{\theta}{2}(x_2^k - x_1^k) + (\tau - 1)(x_1^k - x^*) \right\|^2 \right)_{k \in \mathbb{N}}$$

is bounded and combining with (ii), this shows boundedness of $(z^k - z^*)_{k \in \mathbb{N}}$. If instead $i = 2$, by a similar argument, we have that

$$\left(\left\| z^k - z^* + \frac{\theta}{2}(x_2^k - x_1^k) \right\|^2 \right)_{k \in \mathbb{N}}$$

is bounded, which combined with (ii) shows that $(z^k - z^*)_{k \in \mathbb{N}}$ is bounded.

By nonexpansiveness of resolvents, [3, Proposition 2.38], the update of Algorithm 1, and the fixed-point set in Definition 3.1, this implies that $(x_1^k - x^*)_{k \in \mathbb{N}}$ and $(x_2^k - x^*)_{k \in \mathbb{N}}$ are bounded. Therefore, $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ is bounded and by Definition 4.1 we conclude that $(u_1^k)_{k \in \mathbb{N}}$ and $(u_2^k)_{k \in \mathbb{N}}$ are bounded, so (iii) holds.

Using $u^* \in \mathcal{H}$ as in Definition 4.1, we conclude that

$$\begin{aligned} 0 &\leq D_f(x^*, x_1^k, u_1^k) + D_f(x_1^k, x^*, u^*) \\ &= \langle u_1^k - u^*, x_1^k - x^* \rangle \\ &\leq \langle u_1^k - u^*, x_1^k - x^* \rangle + \langle u_2^k + u^*, x_2^k - x^* \rangle \\ &= -\langle u_1^k + u_2^k, x^* \rangle + \langle u^*, x_2^k - x_1^k \rangle + \langle u_1^k + u_2^k, x_1^k \rangle + \langle u_2^k, x_2^k - x_1^k \rangle \rightarrow 0 \end{aligned}$$

as $k \rightarrow \infty$, where the first equality is due to Lemma 2.2, the inequality is due to the monotonicity of B since $u_2^k \in B(x_2^k)$ and $-u^* = B(x^*)$, and the limit is a consequence of boundedness of $(x_1^k)_{k \in \mathbb{N}}$ and $(u_2^k)_{k \in \mathbb{N}}$ combined with (ii). We conclude that $D_f(x^*, x_1^k, u_1^k) \rightarrow 0$ and $D_f(x_1^k, x^*, u^*) \rightarrow 0$ as $k \rightarrow \infty$. Combining this with the fact that $(V_k^{(i)})_{k \in \mathbb{N}}$ converges, we get in light of the definition of $V_k^{(i)}$ in Definition 4.2 that (iv) holds. \square

Proposition 4.3: *Algorithm 1 converges unconditionally for $(\alpha, \beta, \theta) \in S^{(1)} \cup S^{(2)}$ over $\partial\mathcal{F} \times \mathcal{A}$.*

Proof. In this proof, we use the notation introduced in Definitions 4.1 and 4.2.

By Proposition 4.2(iii), the sequence $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ is bounded, therefore the sequence $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ has a nonempty set of weak sequential cluster points, see [3, Lemma 2.45].

We begin by showing that all weak sequential cluster points of $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ are elements of $\text{Fix}(\partial f, B, \alpha, \beta)$. To that end, suppose that $(\bar{x}_1, \bar{x}_2, \bar{z}) \in \mathcal{H}^3$ is a weak sequential cluster point of $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$. By Proposition 4.2(ii), we get that $\bar{x}_1 = \bar{x}_2$. Let $\bar{x} := \bar{x}_1$. With the definition of u_2^k in Definition 4.1, we get that the subsequence $(x_1^{k_n}, u_2^{k_n}) \rightharpoonup (\bar{x}, \frac{1}{\alpha}(\bar{x} - \bar{z}))$ as $n \rightarrow \infty$. Moreover

$$\begin{aligned} \begin{cases} x_1^k &= J_{\alpha\partial f}(z^k) \\ x_2^k &= J_{\beta B}\left(\left(1 + \frac{\beta}{\alpha}\right)x_1^k - \frac{\beta}{\alpha}z^k\right) \end{cases} \iff \begin{cases} u_1^k &\in \partial f(x_1^k) \\ u_2^k &\in B(x_2^k) \end{cases} \\ &\iff \begin{cases} u_1^k + u_2^k &\in \partial f(x_1^k) + u_2^k \\ x_2^k - x_1^k &\in -x_1^k + B^{-1}(u_2^k) \end{cases} \\ &\iff \begin{bmatrix} u_1^k + u_2^k \\ x_2^k - x_1^k \end{bmatrix} \in \underbrace{\begin{bmatrix} \partial f & \text{Id} \\ -\text{Id} & B^{-1} \end{bmatrix}}_{=:C} \begin{bmatrix} x_1^k \\ u_2^k \end{bmatrix}, \end{aligned}$$

where from Proposition 4.2(ii) we conclude that $(u_1^k + u_2^k, x_2^k - x_1^k) \rightarrow 0$ as $k \rightarrow \infty$. Since C equals the sum of two maximally monotone operators

$$C = \begin{bmatrix} \partial f & 0 \\ 0 & B^{-1} \end{bmatrix} + \begin{bmatrix} 0 & \text{Id} \\ -\text{Id} & 0 \end{bmatrix},$$

the latter of which has full domain, we conclude that C is maximally monotone [3, Corollary 25.5(i)] and that $\text{gra}(C)$ is sequentially closed in the weak-strong topology of $\mathcal{H} \times \mathcal{H}$ [3, Proposition 20.38], implying, since $(u_1^{k_n} + u_2^{k_n}, x_2^{k_n} - x_1^{k_n}) \in C(x_1^{k_n}, u_2^{k_n})$, that $(\bar{x}, \frac{1}{\alpha}(\bar{x} - \bar{z})) \in \text{zer } C$. This implies that

$$\begin{aligned}
\begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} \partial f & \text{Id} \\ -\text{Id} & B^{-1} \end{bmatrix} \begin{bmatrix} \bar{x} \\ \frac{1}{\alpha}(\bar{x} - \bar{z}) \end{bmatrix} &\iff \begin{cases} 0 & \in \partial f(\bar{x}) + \frac{1}{\alpha}(\bar{x} - \bar{z}) \\ 0 & \in -\bar{x} + B^{-1} \left(\frac{1}{\alpha}(\bar{x} - \bar{z}) \right) \end{cases} \\
&\iff \begin{cases} \frac{1}{\alpha}(\bar{z} - \bar{x}) & \in \partial f(\bar{x}) \\ \frac{1}{\alpha}(\bar{x} - \bar{z}) & \in B(\bar{x}) \end{cases} \\
&\iff \begin{cases} \frac{1}{\alpha}(\bar{z} - \bar{x}) & \in \partial f(\bar{x}) \\ \frac{1}{\beta} \left(\left(1 + \frac{\beta}{\alpha}\right) \bar{x} - \frac{\beta}{\alpha} \bar{z} - \bar{x} \right) & \in B(\bar{x}) \end{cases} \\
&\iff \begin{cases} \bar{x} & = J_{\alpha \partial f}(\bar{z}) \\ \bar{x} & = J_{\beta B} \left(\left(1 + \frac{\beta}{\alpha}\right) \bar{x} - \frac{\beta}{\alpha} \bar{z} \right) \end{cases} \\
&\iff (\bar{x}, \bar{x}, \bar{z}) \in \text{Fix}(\partial f, B, \alpha, \beta).
\end{aligned}$$

We have shown that all weak sequential cluster points of $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ are elements of $\text{Fix}(\partial f, B, \alpha, \beta)$.

Next, we show that there exists only one weak sequential cluster point of the sequence $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$. Let $v^k := (x_1^k, x_2^k, z^k)$ for all $k \in \mathbb{N}$. Suppose that $\bar{v} := (\bar{x}_1, \bar{x}_2, \bar{z}) \in \mathcal{H}^3$ and $\hat{v} := (\hat{x}_1, \hat{x}_2, \hat{z}) \in \mathcal{H}^3$ are two weak sequential cluster points of $(v^k)_{k \in \mathbb{N}}$, say $v^{k_n} \rightharpoonup \bar{v}$ and $v^{\ell_n} \rightharpoonup \hat{v}$ as $n \rightarrow \infty$. We get from before that $\bar{x}_1 = \bar{x}_2$ and $\hat{x}_1 = \hat{x}_2$. Define $\bar{x} := \bar{x}_1$ and $\hat{x} := \hat{x}_1$. Moreover, both \bar{v} and \hat{v} lie in $\text{Fix}(\partial f, B, \alpha, \beta)$.

Let $i \in \{1, 2\}$ be such that $(\alpha, \beta, \theta) \in S^{(i)}$. Let $Q_V^{(i)} \in \mathbb{S}^3$ be the matrix from Remark 4.4, see also Appendix A for the explicit form. By Proposition 4.2(iv), both $(\mathcal{Q}(Q_V^{(i)}, v^k - \bar{v}))_{k \in \mathbb{N}}$ and $(\mathcal{Q}(Q_V^{(i)}, v^k - \hat{v}))_{k \in \mathbb{N}}$ converge. Since

$$2 \langle v^k, Q_V^{(i)}(\bar{v} - \hat{v}) \rangle = \mathcal{Q}(Q_V^{(i)}, v^k - \bar{v}) - \mathcal{Q}(Q_V^{(i)}, v^k - \hat{v}) + \mathcal{Q}(Q_V^{(i)}, \bar{v}) + \mathcal{Q}(Q_V^{(i)}, \hat{v})$$

we get that $(\langle v^k, Q_V^{(i)}(\bar{v} - \hat{v}) \rangle)_{k \in \mathbb{N}}$ converges as well. Therefore, passing along the subsequences $(v^{k_n})_{n \in \mathbb{N}}$ and $(v^{\ell_n})_{n \in \mathbb{N}}$ gives that

$$\langle \bar{v}, Q_V^{(i)}(\bar{v} - \hat{v}) \rangle = \langle \hat{v}, Q_V^{(i)}(\bar{v} - \hat{v}) \rangle,$$

which implies that

$$\mathcal{Q}(Q_V^{(i)}, \bar{v} - \hat{v}) = 0. \tag{8}$$

If $i = 1$, then (8) equals

$$\|\bar{z} - \hat{z} + (\tau - 1)(\bar{x} - \hat{x})\|^2 + \tau(1 - \tau) \|\bar{x} - \hat{x}\|^2 = 0,$$

where $1 - \tau \geq 0$ for $(\alpha, \beta, \theta) \in S^{(1)}$. If $i = 2$, then (8) equals

$$\|\bar{z} - \hat{z}\|^2 + \frac{\theta(4\tau - \theta)}{4} \left\| \frac{2(\tau - 1)}{4\tau - \theta} (\bar{x} - \hat{x}) \right\|^2 + \frac{\tau(\tau - 1)(4 - \theta)}{4\tau - \theta} \|\bar{x} - \hat{x}\|^2 = 0,$$

where $4\tau - \theta$, $4 - \theta$ and $\tau - 1$ are nonnegative for $(\alpha, \beta, \theta) \in S^{(2)}$.

This implies that $\bar{z} = \hat{z}$ in both cases. Since

$$\bar{x} = J_{\alpha\partial f}(\bar{z}) = J_{\alpha\partial f}(\hat{z}) = \hat{x}$$

we find that $\bar{v} = \hat{v}$, proving uniqueness of weak sequential cluster points.

Therefore, $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ is bounded and every weak sequential cluster point of $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ belongs to $\text{Fix}(\partial f, B, \alpha, \beta)$, and so by [3, Lemma 2.47], the sequence $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ converges weakly to a point in $\text{Fix}(\partial f, B, \alpha, \beta)$. \square

4.2 Counterexamples

In the previous section, we have shown convergence of Algorithm 1 over the parameter region $S^{(1)} \cup S^{(2)}$. With the following counterexamples we show that this region is tight, in the sense that Algorithm 1 fails to be unconditionally convergent over $\partial\mathcal{F} \times \partial\mathcal{F}$ for parameters (α, β, θ) outside of $S^{(1)} \cup S^{(2)}$.

Proposition 4.4: *If Algorithm 1 is unconditionally convergent for $(\alpha, \beta, \theta) \in \mathbb{R}_{++}^2 \times (\mathbb{R} \setminus \{0\})$ over $\partial\mathcal{F} \times \partial\mathcal{F}$ then $(\alpha, \beta, \theta) \in S^{(1)} \cup S^{(2)}$, i.e., $S(\partial\mathcal{F} \times \partial\mathcal{F}) \subset S^{(1)} \cup S^{(2)}$.*

Proof. Let $(\alpha, \beta, \theta) \in \mathbb{R}_{++}^2 \times (\mathbb{R} \setminus \{0\})$ be such that Algorithm 1 is unconditionally convergent for (α, β, θ) over $\partial\mathcal{F} \times \partial\mathcal{F}$.

First, let $f = 0 \in \mathcal{F}$ and $g = \iota_{\{0\}} \in \mathcal{F}$. Note that $\text{zer}(\partial f + \partial g) = \{0\}$. Let $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ be generated by Algorithm 1 applied to $(\partial f, \partial g, z^0, \alpha, \beta, \theta)$ given some initial point $z^0 \in \mathcal{H} \setminus \{0\}$. Note that such a z^0 exists since $\dim \mathcal{H} \geq 1$. Then,

$$\begin{aligned} x_1^k &= \text{prox}_{\alpha f}(z^k) = z^k \\ x_2^k &= \text{prox}_{\beta g} \left(\left(1 + \frac{\beta}{\alpha}\right) x_1^k - \frac{\beta}{\alpha} z^k \right) = 0 \\ z^{k+1} &= z^k + \theta(x_2^k - x_1^k) = (1 - \theta)z^k \end{aligned}$$

for every $k \in \mathbb{N}$. Therefore, we conclude in particular that $\theta \in (0, 2)$.

Likewise, let us repeat the above argument applied for $f = \iota_{\{0\}} \in \mathcal{F}$ and $g = 0 \in \mathcal{F}$ with again $\text{zer}(\partial f + \partial g) = \{0\}$. Then

$$\begin{aligned} x_1^k &= \text{prox}_{\alpha f}(z^k) = 0 \\ x_2^k &= \text{prox}_{\beta g} \left(\left(1 + \frac{\beta}{\alpha}\right) x_1^k - \frac{\beta}{\alpha} z^k \right) = -\frac{\beta}{\alpha} z^k \\ z^{k+1} &= z^k + \theta(x_2^k - x_1^k) = \left(1 - \theta \frac{\beta}{\alpha}\right) z^k \end{aligned}$$

Therefore, we conclude in particular that $\theta \in (0, 2\alpha/\beta)$.

Put together, if Algorithm 1 is unconditionally convergent for $(\alpha, \beta, \theta) \in \mathbb{R}_{++}^2 \times (\mathbb{R} \setminus \{0\})$ over $\partial\mathcal{F} \times \partial\mathcal{F}$ then $\theta \in (0, 2) \cap (0, 2\alpha/\beta)$, i.e., $(\alpha, \beta, \theta) \in S^{(1)} \cup S^{(2)}$. \square

4.3 Proof of Theorem 3.3

Combining the results in Sections 4.1 and 4.2 we can now prove our main result.

Proof of Theorem 3.3. Proposition 4.3 gives that

$$S^{(1)} \cup S^{(2)} \subset S(\partial\mathcal{F} \times \mathcal{A}) \subset S(\partial\mathcal{F} \times \partial\mathcal{F}).$$

Proposition 4.4 shows that $S(\partial\mathcal{F} \times \partial\mathcal{F}) \subset S^{(1)} \cup S^{(2)}$. Therefore,

$$S^{(1)} \cup S^{(2)} = S(\partial\mathcal{F} \times \mathcal{A}) = S(\partial\mathcal{F} \times \partial\mathcal{F}).$$

\square

5 Convergence Rates

In this section, we show ergodic convergence of the primal-dual gap for the extended Douglas–Rachford method in Algorithm 1, with convergence rate $\mathcal{O}(1/K)$. We consider inclusion problems of the form

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in \partial f(x) + \partial g(x),$$

where $f, g : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper, closed, and convex, such that $\text{zer}(\partial f + \partial g) \neq \emptyset$. Let $(x^*, u^*) \in \mathcal{H}^2$ be such that

$$\begin{cases} u^* & \in \partial f(x^*) \\ x^* & \in \partial g^*(-u^*) \end{cases}. \quad (9)$$

We define the Lagrangian functions $\mathcal{L} : \mathcal{H}^2 \rightarrow \mathbb{R} \cup \{-\infty, +\infty\}$ and $\bar{\mathcal{L}} : \mathcal{H}^2 \rightarrow \mathbb{R} \cup \{-\infty, +\infty\}$ as

$$\begin{aligned} \mathcal{L}(x, u) &:= f(x) + \langle x, u \rangle - g^*(u) \\ \bar{\mathcal{L}}(x, u) &:= g(x) + \langle x, u \rangle - f^*(u) \end{aligned}$$

and the primal-dual gap functions $\mathcal{D}_{x^*, -u^*} : \mathcal{H}^2 \rightarrow \mathbb{R} \cup \{+\infty\}$ and $\bar{\mathcal{D}}_{x^*, u^*} : \mathcal{H}^2 \rightarrow \mathbb{R} \cup \{+\infty\}$ as

$$\begin{aligned} \mathcal{D}_{x^*, -u^*}(x, u) &:= \mathcal{L}(x, -u^*) - \mathcal{L}(x^*, u) \\ \bar{\mathcal{D}}_{x^*, u^*}(x, u) &:= \bar{\mathcal{L}}(x, u^*) - \bar{\mathcal{L}}(x^*, u), \end{aligned}$$

for all $(x, u) \in \mathcal{H}^2$. In particular, since

$$\begin{aligned} \mathcal{D}_{x^*, -u^*}(x, u) &= f(x) - \langle x, u^* \rangle - f(x^*) + g^*(u) - \langle x^*, u \rangle - g^*(-u^*) \\ &= f(x) - \langle x - x^*, u^* \rangle - f(x^*) + g^*(u) - \langle x^*, u + u^* \rangle - g^*(-u^*) \\ &= D_f(x, x^*, u^*) + D_{g^*}(u, -u^*, x^*), \end{aligned} \quad (10)$$

and similarly

$$\bar{\mathcal{D}}_{x^*, u^*}(x, u) = D_{f^*}(u, u^*, x^*) + D_g(x, x^*, -u^*), \quad (11)$$

for all $(x, u) \in \mathcal{H}^2$, we have by convexity of f and g and by (9) combined with Lemma 2.3 that $\mathcal{D}_{x^*, -u^*}$ and $\bar{\mathcal{D}}_{x^*, u^*}$ are convex and nonnegative.

Before presenting the ergodic convergence rate result, let us define the ergodic sequences of Algorithm 1.

Definition 5.1 (Ergodic iterates of Algorithm 1): *Let $(f, g) \in \mathcal{F}^2$ such that $\text{zer}(\partial f + \partial g) \neq \emptyset$ and let $(x^*, u^*) \in \mathcal{H}^2$ satisfy (9). Let $(\alpha, \beta, \theta) \in S^{(i)}$ for $i \in \{1, 2\}$. For $((x_1^k, x_2^k, z^k))_{k \in \mathbb{N}}$ generated by Algorithm 1 applied to $(\partial f, \partial g, \alpha, \beta, \theta)$ and $(u_1^k)_{k \in \mathbb{N}}$ and $(u_2^k)_{k \in \mathbb{N}}$ as in Definition 4.1 we define the ergodic iterates*

$$\bar{x}_j^K := \frac{1}{K+1} \sum_{k=0}^K x_j^k \quad \text{and} \quad \bar{u}_j^K := \frac{1}{K+1} \sum_{k=0}^K u_j^k$$

for each $j \in \{1, 2\}$ and $K \in \mathbb{N}$.

Proposition 5.1: *For the ergodic iterates in Definition 5.1, the nonnegative primal-dual gap sequences*

$$(\mathcal{D}_{x^*, -u^*}(\bar{x}_1^K, \bar{u}_2^K))_{K \in \mathbb{N}} \quad \text{and} \quad (\bar{\mathcal{D}}_{x^*, u^*}(\bar{x}_2^K, \bar{u}_1^K))_{K \in \mathbb{N}}$$

converge to zero with rate $\mathcal{O}(1/K)$.

Proof. We will use the following identity. If $h : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}$ is proper, closed and convex, $x, y \in \mathcal{H}$, $u \in \partial h(x)$, and $v \in \partial h(y)$, then

$$\begin{aligned} D_h(y, x, u) + D_{h^*}(v, u, x) &= h(y) - h(x) - \langle u, y - x \rangle + h^*(v) - h^*(u) - \langle x, v - u \rangle \\ &= (h(y) + h^*(v) - \langle v, y \rangle) + \langle v, y \rangle \\ &\quad - (h(x) + h^*(u) - \langle x, u \rangle) - \langle u, y \rangle + \langle x, v - u \rangle \\ &= \langle v - u, y - x \rangle, \end{aligned} \tag{12}$$

since the expressions in the parentheses are zero by the Fenchel-Young equality.

By Proposition 3.1, there exists some $z^* \in \mathcal{H}$ such that $(x^*, x^*, z^*) \in \text{Fix}(\partial f, \partial g, \alpha, \beta)$. Combining (10) with (11), and then using (12) for f and g we get that

$$\begin{aligned} \mathcal{D}_{x^*, -u^*}(x_1^k, u_2^k) + \bar{\mathcal{D}}_{x^*, u^*}(x_2^k, u_1^k) &= D_f(x_1^k, x^*, u^*) + D_{f^*}(u_1^k, u^*, x^*) \\ &\quad + D_g(x_2^k, x^*, -u^*) + D_{g^*}(u_2^k, -u^*, x^*) \\ &= \langle u_1^k - u^*, x_1^k - x^* \rangle + \langle u_2^k + u^*, x_2^k - x^* \rangle \end{aligned} \tag{13}$$

for all $k \in \mathbb{N}$.

Since $\mathcal{D}_{x^*, -u^*}$ and $\bar{\mathcal{D}}_{x^*, u^*}$ are convex, we have by Jensen's inequality that

$$\begin{aligned} \mathcal{D}_{x^*, -u^*}(\bar{x}_1^K, \bar{u}_2^K) + \bar{\mathcal{D}}_{x^*, u^*}(\bar{x}_2^K, \bar{u}_1^K) &\leq \frac{1}{K+1} \sum_{k=0}^K (\mathcal{D}_{x^*, -u^*}(x_1^k, u_2^k) + \bar{\mathcal{D}}_{x^*, u^*}(x_2^k, u_1^k)) \\ &= \frac{1}{K+1} \sum_{k=0}^K (\langle u_1^k - u^*, x_1^k - x^* \rangle + \langle u_2^k + u^*, x_2^k - x^* \rangle) \\ &= \frac{1}{K+1} \sum_{k=0}^K (I_k - \langle u_1^{k+1} - u^*, x_1^{k+1} - x^* \rangle - \langle u_2^{k+1} + u^*, x_2^{k+1} - x^* \rangle) \\ &\leq \frac{1}{K+1} \sum_{k=0}^K I_k \leq \frac{1}{\alpha\theta(K+1)} \sum_{k=0}^K (\alpha\theta I_k + R_k^{(i)}) \leq \frac{V_0^{(i)}}{\alpha\theta(K+1)} \end{aligned}$$

for all $K \in \mathbb{N}$ and for $V_k^{(i)}$ and $R_k^{(i)}$ defined in Definition 4.2 and I_k defined in Definition 4.3, where the first equality follows from (13), the second equality from Definition 4.3, the second inequality from monotonicity of ∂f and ∂g since $(x_1^{k+1}, u_1^{k+1}), (x^*, u^*) \in \text{gra } \partial f$ and $(x_2^{k+1}, u_2^{k+1}), (x^*, -u^*) \in \text{gra } \partial g$ by Definition 4.1, the third inequality from nonnegativity of $R_k^{(i)}$ (Proposition 4.2), and the fourth inequality from telescope summation of the Lyapunov equality (7) in Proposition 4.1 and nonnegativity of $V_k^{(i)}$ (Proposition 4.2). By nonnegativity of the primal-dual gap functions this shows ergodic convergence with rate $\mathcal{O}(1/K)$. \square

6 Extending ADMM, Chambolle–Pock, and Parallel Splitting

In this section, we apply Algorithm 1 to different problem formulations and, by invoking Theorem 3.3, arrive at new convergent variants of ADMM, the Chambolle–Pock method, and Douglas–Rachford splitting applied to a consensus problem formulation, sometimes referred to as parallel splitting.

6.1 Extending ADMM

To arrive at an extended version of ADMM, we follow the derivation and notation of [38, Section 3.1]. We specialize to the finite-dimensional case $\mathcal{H} = \mathbb{R}^p$, though we note that, under

mild assumptions, we also have weak convergence in the general real Hilbert setting. Let $f : \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^q \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper, closed, and convex functions. Let $A \in \mathbb{R}^{n \times p}$, $B \in \mathbb{R}^{n \times q}$, and $c \in \mathbb{R}^n$. Consider the following convex optimization problem:

$$\begin{aligned} & \underset{(x,y) \in \mathbb{R}^p \times \mathbb{R}^q}{\text{minimize}} && f(x) + g(y) \\ & \text{subject to} && Ax + By = c. \end{aligned} \quad (14)$$

which can be reformulated as the following unconstrained problem

$$\underset{z \in \mathbb{R}^n}{\text{minimize}} (A \triangleright f)(z) + (B \triangleright g)(c - z), \quad (15)$$

using the infimal postcomposition, defined as

$$(A \triangleright f)(z) := \inf\{f(x) \mid Ax = z\}$$

for each $z \in \mathbb{R}^n$ and analogously for $B \triangleright g$. The classical ADMM method alternately minimizes the augmented Lagrangian of (14)

$$L_\alpha(x, y, u) := f(x) + g(y) + \langle u, Ax + By - c \rangle + \frac{\alpha}{2} \|Ax + By - c\|^2$$

and updates the dual variable u according to

$$\begin{cases} x^{k+1} \in \underset{x \in \mathbb{R}^p}{\text{Argmin}} L_\alpha(x, y^k, u^k) \\ y^{k+1} \in \underset{y \in \mathbb{R}^q}{\text{Argmin}} L_\alpha(x^{k+1}, y, u^k) \\ u^{k+1} = u^k + \alpha(Ax^{k+1} + By^{k+1} - c) \end{cases} \quad (16)$$

This method can be derived by applying unrelaxed Douglas–Rachford splitting (i.e., with $\theta = 1$) either to (15) or to the corresponding dual problem

$$\underset{u \in \mathcal{H}}{\text{minimize}} \underbrace{f^*(-A^\top u)}_{=: \tilde{f}(u)} + \underbrace{g^*(-B^\top u) + c^\top u}_{=: \tilde{g}(u)}. \quad (17)$$

By instead applying Algorithm 1 to (17), and invoking the convergence result in Theorem 3.3, we obtain new convergent variants of ADMM. With relaxation parameter $\theta = 1$, we arrive at Algorithm 2, which extends the classical ADMM in (16) by allowing for different penalty parameters α and β in the two augmented Lagrangians. From Theorem 3.1, we conclude, under mild additional assumptions, that Algorithm 2 converges if $0 < \alpha < 2\beta$.

Algorithm 2 Extended ADMM

- 1: **Input:** (f, g, A, B) as in (14), $u^0 \in \mathbb{R}^n$, $(\alpha, \beta) \in \mathbb{R}_{++}^2$ such that $\alpha < 2\beta$.
- 2: **for** $k = 0, 1, 2, \dots$ **do**
- 3: **Update:**

$$\begin{cases} x^{k+1} \in \underset{x \in \mathbb{R}^p}{\text{Argmin}} L_\beta(x, y^k, u^k) \\ y^{k+1} \in \underset{y \in \mathbb{R}^q}{\text{Argmin}} L_\alpha(x^{k+1}, y, u^k) \\ u^{k+1} = u^k + \alpha(Ax^{k+1} + By^{k+1} - c) \end{cases}$$

- 4: **end for**
-

Algorithm 3 Extended ADMM (general)

1: **Input:** (f, g, A, B) as in (14), $u^0 \in \mathbb{R}^n$, $(\alpha, \beta, \theta) \in \mathbb{R}_{++}^3$ such that $\theta < \min\{2, 2\beta/\alpha\}$.

2: **for** $k = 0, 1, 2, \dots$ **do**

3: **Update:**

$$\begin{cases} x^{k+1} \in \underset{x \in \mathbb{R}^p}{\text{Argmin}} L_\beta(x, y^k, u^k + \alpha(1 - \theta)(By^k - c)) \\ y^{k+1} \in \underset{y \in \mathbb{R}^q}{\text{Argmin}} L_\alpha(\theta x^{k+1}, y, u^k) \\ u^{k+1} = u^k + \theta\alpha(Ax^{k+1} + By^{k+1} - c) \end{cases}$$

4: **end for**

For a general relaxation parameter $\theta \in (0, \min\{2, 2\beta/\alpha\})$ we instead arrive at the following relaxed version of extended ADMM, see Algorithm 3. For the derivation of the extended ADMM, we use the following identity (see [38, Equation (2.6)]). Let $h : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper, closed and convex and let $C \in \mathbb{R}^{n \times m}$. If $\text{relint dom } h^* \cap \text{ran}(-C^\top) \neq \emptyset$ then

$$v = \text{prox}_{\alpha h^* \circ (-C^\top)}(u) \iff \begin{cases} x \in \underset{x \in \mathbb{R}^n}{\text{Argmin}} \left(h(x) + \langle u, Cx \rangle + \frac{\alpha}{2} \|Cx\|^2 \right) \\ v = u + \alpha Cx. \end{cases} \quad (18)$$

Applying the extended Douglas–Rachford method in Algorithm 1 to (17) with parameters $(\alpha, \beta, \theta\alpha/\beta)$ for $(\alpha, \beta, \theta) \in \mathbb{R}_{++}^3$ gives the update

$$\begin{aligned} \mu^{k+1/2} &= J_{\alpha\partial\tilde{g}}(\psi^k) \\ \mu^{k+1} &= J_{\beta\partial\tilde{f}}\left(\left(1 + \frac{\beta}{\alpha}\right)\mu^{k+1/2} - \frac{\beta}{\alpha}\psi^k\right) \\ \psi^{k+1} &= \psi^k + \theta\frac{\alpha}{\beta}(\mu^{k+1} - \mu^{k+1/2}). \end{aligned} \quad (19)$$

We have that $J_{\alpha\partial\tilde{g}}(\cdot) = J_{\alpha\partial(g^* \circ (-B^\top))}(\cdot - \alpha c)$ by [3, Proposition 23.17]. Then by using (18) and defining auxiliary variables $\tilde{y}^{k+1} \in \mathbb{R}^q$ and $\tilde{x}^{k+1} \in \mathbb{R}^p$, the update of (19) becomes

$$\begin{aligned} \tilde{y}^{k+1} &\in \underset{y \in \mathbb{R}^q}{\text{Argmin}} \left(g(y) + \langle \psi^k - \alpha c, By \rangle + \frac{\alpha}{2} \|By\|^2 \right) \\ &= \underset{y \in \mathbb{R}^q}{\text{Argmin}} \left(g(y) + \langle \psi^k - \theta\alpha A\tilde{x}^k, By \rangle + \frac{\alpha}{2} \|\theta A\tilde{x}^k + By - c\|^2 \right) \\ \mu^{k+1/2} &= \psi^k + \alpha(B\tilde{y}^{k+1} - c) \\ \tilde{x}^{k+1} &\in \underset{x \in \mathbb{R}^p}{\text{Argmin}} \left(f(x) + \left\langle \left(1 + \frac{\beta}{\alpha}\right)\mu^{k+1/2} - \frac{\beta}{\alpha}\psi^k, Ax \right\rangle + \frac{\beta}{2} \|Ax\|^2 \right) \\ &= \underset{x \in \mathbb{R}^p}{\text{Argmin}} \left(f(x) + \langle \psi^k + (\alpha + \beta)(B\tilde{y}^{k+1} - c), Ax \rangle + \frac{\beta}{2} \|Ax\|^2 \right) \\ &= \underset{x \in \mathbb{R}^p}{\text{Argmin}} \left(f(x) + \langle \psi^k + \alpha(B\tilde{y}^{k+1} - c), Ax \rangle + \frac{\beta}{2} \|Ax + B\tilde{y}^{k+1} - c\|^2 \right) \\ \mu^{k+1} &= \psi^k + (\alpha + \beta)(B\tilde{y}^{k+1} - c) + \beta A\tilde{x}^{k+1} \\ \psi^{k+1} &= \psi^k + \frac{\theta\alpha}{\beta}(\mu^{k+1} - \mu^{k+1/2}) = \psi^k + \theta\alpha(A\tilde{x}^{k+1} + B\tilde{y}^{k+1} - c). \end{aligned}$$

We can now remove μ^{k+1} and $\mu^{k+1/2}$ as they are redundant, make the substitution $u^k = \psi^k -$

$\theta\alpha A\tilde{x}^k$, and reorganize the iterates to arrive at the update

$$\begin{aligned}\tilde{y}^{k+1} &\in \underset{y \in \mathbb{R}^q}{\text{Argmin}} \left(g(y) + \langle u^k, By \rangle + \frac{\alpha}{2} \|\theta A\tilde{x}^k + By - c\|^2 \right) \\ \tilde{x}^{k+1} &\in \underset{x \in \mathbb{R}^p}{\text{Argmin}} \left(f(x) + \langle u^{k+1} + (1-\theta)\alpha(B\tilde{y}^{k+1} - c), Ax \rangle + \frac{\beta}{2} \|Ax + B\tilde{y}^{k+1} - c\|^2 \right) \\ u^{k+1} &= u^k + \theta\alpha (A\tilde{x}^k + B\tilde{y}^{k+1} - c).\end{aligned}$$

Using the definition of the augmented Lagrangian and reordering, we get the updates of Algorithm 3 and by Theorem 3.3 we have convergence if $\theta\alpha/\beta \in (0, \min\{2, 2\alpha/\beta\}) \iff \theta \in (0, \min\{2\beta/\alpha, 2\})$. We then get Algorithm 2 by letting $\theta = 1$.

Remark 6.1: If $(\alpha, \beta, \theta\alpha/\beta) \in S^{(1)} \cup S^{(2)}$ and the inclusion problem obtained from (17) is solvable and involves operators in $\partial\mathcal{F} \times \partial\mathcal{F}$ we have convergence by Theorem 3.3. For the derivation, we furthermore require that the right implication of (18) holds both for \tilde{f} and for \tilde{g} , that both functions are proper, closed and convex and that the minimization problems involved are solvable (see, for example, [38] and [39] for more discussion on regularity assumptions).

6.2 Extending Chambolle–Pock

To arrive at an extended version of the Chambolle–Pock method, we follow the derivation and notation of [33]. We specialize to the finite-dimensional case $\mathcal{H} = \mathbb{R}^n$, though we note that, under mild assumptions, we also have weak convergence in the general real Hilbert setting. Let $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper, closed, and convex functions. Let $A \in \mathbb{R}^{m \times n}$. Consider the optimization problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) + g(Ax). \quad (20)$$

This optimization problem can be solved with the Primal-Dual Hybrid Gradient algorithm (PDHG), also known as the Chambolle–Pock algorithm [32], which is given by the update step

$$\begin{cases} x^{k+1} = \text{prox}_{\tau f} (x^k - \tau A^\top z^k) \\ z^{k+1} = \text{prox}_{\sigma g^*} (z^k + \sigma A (x^{k+1} + \theta(x^{k+1} - x^k))) \end{cases}. \quad (21)$$

This algorithm was originally shown to converge for $\theta = 1$ and $\tau\sigma \|A\|^2 < 1$, see [32], where $\|A\|$ denotes the spectral norm of A . Recent analyses of this Chambolle–Pock algorithm allow for larger parameter regions than in [32]. In [40, 41, 42], the step-size convergence region is extended to $\tau\sigma \|A\|^2 < 4/3$, still with $\theta = 1$. These results have been generalized in [35] to the $\theta \neq 1$ setting. In particular, they show convergence whenever $\theta > 1/2$ and $\tau\sigma \|A\|^2 < 4/(1+2\theta)$. Moreover, the bound for $\tau\sigma \|A\|^2$ is tight, see [35].

In [34, Algorithm 4], they propose the following relaxation for the Chambolle–Pock method with $\theta = 1$:

$$\begin{cases} \bar{x}^k = \text{prox}_{\tau f} (x^k - \tau A^\top z^k) \\ \bar{z}^k = \text{prox}_{\sigma g^*} (z^k + \sigma A (2\bar{x}^k - x^k)) \\ x^{k+1} = x^k + \rho (\bar{x}^k - x^k) \\ z^{k+1} = z^k + \rho (\bar{z}^k - z^k) \end{cases} \quad (22)$$

and show convergence whenever $\tau\sigma \|A\|^2 < 1$ and $\rho \in (0, 2)$, which was later extended in [43] to $\tau\sigma \|A\|^2 \leq 1$ and $\rho \in (0, 2)$. This method has been shown in [33] to be a special case of the

relaxed Douglas–Rachford algorithm when applied to a specific problem formulation. This route results in the same convergent parameter region as in [43], i.e., $\tau\sigma\|A\|^2 \leq 1$ and $\rho \in (0, 2)$.

When the extended Douglas–Rachford method proposed in this paper is combined with the approach of [33], we obtain the doubly relaxed Chambolle–Pock method of Algorithm 4, combining both types of relaxation. See Section 6.2.1 for its derivation.

Algorithm 4 Doubly Relaxed Chambolle–Pock

```

1: Input:  $(f, g, A)$  as in (20),  $(x^0, z^0) \in \mathbb{R}^n \times \mathbb{R}^m$ ,
2:  $(\tau, \sigma, \theta, \rho) \in \mathbb{R}_{++}^4$  such that  $\rho \in (0, \min\{2, 2\theta\})$  and  $\tau\sigma\|A\|^2 \leq 1/\theta$ .
3: for  $k = 0, 1, 2 \dots$  do
4:   Update:
      
$$\begin{cases} \bar{x}^k = \text{prox}_{\tau f}(x^k - \tau A^\top z^k) \\ \bar{z}^k = \text{prox}_{\sigma g^*}(z^k + \sigma A(\bar{x}^k + \theta(\bar{x}^k - x^k))) \\ x^{k+1} = x^k + \rho(\bar{x}^k - x^k) \\ z^{k+1} = z^k + \rho(\bar{z}^k - z^k) \end{cases}$$

5: end for

```

This algorithm is guaranteed to converge on a strictly larger parameter region—namely $(\tau, \sigma, \theta, \rho) \in \mathbb{R}_{++}^4$ such that $\rho < \min\{2, 2\theta\}$ and $\tau\sigma\|A\|^2 \leq 1/\theta$ —than that established in [33]. In fact, our analysis of Algorithm 4 is tight in the sense that none of these two conditions can be relaxed without restricting the other one. To see this, let $n = m = 1$, $A = 1$, $f = g^* = 0$. This defines an optimization problem of the form (20) with the unique minimizer 0, and gives the iteration

$$v^{k+1} = \begin{bmatrix} 1 & -\rho\tau \\ \rho\sigma & 1 - \rho\sigma\tau(1 + \theta) \end{bmatrix} v^k, \quad (23)$$

where $v^k = (x^k, z^k)$. With the variable $\omega := \tau\sigma\|A\|^2$, we let λ_- denote the eigenvalue of (23) given by the expression

$$\lambda_-(\rho, \omega, \theta) := 1 - \frac{\rho\omega(1 + \theta)}{2} - \frac{\rho\sqrt{\omega}}{2} \sqrt{\omega(1 + \theta)^2 - 4}$$

for each $(\rho, \omega, \theta) \in \mathbb{R}_{++}^3$. For all $\omega \geq 1/\theta$ with $\theta, \rho > 0$, the discriminant of $\lambda_-(\rho, \omega, \theta)$ is nonnegative and $\lambda_-(\rho, \omega, \theta) \in \mathbb{R}$. Note that $\lambda_-(\rho, \omega, \theta)$ is continuous and decreasing in the variables ρ, ω for $\omega \geq 1/\theta$, i.e., $\lambda_-(\bar{\rho}, \bar{\omega}, \theta) \leq \lambda_-(\rho, \omega, \theta)$ holds for all $\bar{\rho} \geq \rho$ and $\bar{\omega} \geq \omega \geq 1/\theta$. At the bound $\rho = \min\{2, 2\theta\}$ and $\omega = 1/\theta$ we have after a short calculation that $\lambda_-(\min\{2, 2\theta\}, 1/\theta, \theta) = -1$ for all $\theta \in \mathbb{R}_{++}$, which gives, by continuity, that none of the upper bounds can be relaxed without simultaneously restricting the other.

This result reveals a trade-off between the two bounds: the conditions $\rho < \min\{2, 2\theta\}$ and $\tau\sigma\|A\|^2 \leq 1/\theta$ cannot be relaxed independently—making one upper bound larger necessarily forces the other to become more restrictive. However, if we, for instance, fix $\rho \in (0, \min\{2, 2\theta\})$ to a specific value, the admissible step-size region can be enlarged. For example, for $\rho = 1$ and $\theta > \frac{1}{2}$, the analysis in [35] allows $\tau\sigma\|A\|^2 < 4/(1 + 2\theta)$, which is strictly less restrictive than our bound in this regime. It remains to be seen if the Chambolle–Pock method can be derived from the Douglas–Rachford method without conservatism in the step-size restriction for each individual choice of ρ .

Let us summarize our contributions in the doubly relaxed Chambolle–Pock setting. We present, to the best of the authors’ knowledge, the first convergence guarantee for the doubly

relaxed Chambolle–Pock method, i.e., with freedom in both relaxation parameters ρ and θ . Interestingly, the step sizes τ, σ can be chosen arbitrarily large by taking the relaxation parameters ρ and θ small enough. We also provide the first convergence analysis for the case $\theta \in (0, 1/2)$.

6.2.1 Derivation of Double-Relaxed Chambolle–Pock

We will now derive Algorithm 4, following the steps in [33], and prove its parameter convergence region. Let $\gamma > 0$ be such that $\gamma \|A\| \leq 1$. Let $C := (\gamma^{-2}I - AA^\top)^{1/2} \in \mathbb{R}^{m \times m}$ and $B := \begin{bmatrix} A & C \end{bmatrix} \in \mathbb{R}^{m \times (n+m)}$, implying that $BB^\top = \gamma^{-2}I$. The matrices C and B are introduced only for the derivation and the final algorithm will be independent of these, see Algorithm 4. Consider the lifted proper, closed, and convex functions $\tilde{f}, \tilde{g} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ defined by

$$\begin{aligned}\tilde{f}(u_1, u_2) &:= f(u_1) + \iota_{\{0\}}(u_2), \\ \tilde{g}(u_1, u_2) &:= g(Au_1 + Cu_2),\end{aligned}$$

for each $u = (u_1, u_2) \in \mathbb{R}^n \times \mathbb{R}^m$. Consider the optimization problem

$$\underset{u \in \mathbb{R}^n \times \mathbb{R}^m}{\text{minimize}} \quad \tilde{f}(u) + \tilde{g}(u), \quad (24)$$

which is equivalent to (20). The corresponding optimality condition to (24) is

$$0 \in \partial \tilde{f}(u) + \tilde{G}(u), \quad (25)$$

where $\tilde{G} := B^\top \circ \partial g \circ B : \mathbb{R}^{n+m} \rightrightarrows \mathbb{R}^{n+m}$ is a maximally monotone operator.

Under mild assumptions, the monotone operator \tilde{G} is the subdifferential of a proper, closed, and convex function. We will not need to assume that and instead apply Theorem 3.3 to the pair $(\partial \tilde{f}, \tilde{G}) \in \partial \mathcal{F} \times \mathcal{A}$. Since

$$\begin{aligned}(\exists u^* \in \mathbb{R}^{n+m}) \quad u^* &\in \text{zer}(\partial \tilde{f} + \tilde{G}) \\ \implies (\exists u^* \in \mathbb{R}^{n+m}) \quad u^* &\in \text{zer}(\partial \tilde{f} + \partial \tilde{g}) \\ \implies (\exists u^* \in \mathbb{R}^{n+m}) \quad u^* &\in \text{Argmin}(\tilde{f} + \tilde{g}) \\ \iff (\exists u_1^* \in \mathbb{R}^n) \quad u_1^* &\in \text{Argmin}(f + g \circ A),\end{aligned} \quad (26)$$

solving the inclusion (25) implies solving the optimization problem (20). Note that the first implication in (26) is an equivalence under some regularity condition on (g, B) , see [36, Theorem 23.9], and that the second implication in (26) is an equivalence under some regularity condition on (\tilde{f}, \tilde{g}) , see [36, Theorem 23.8]. When both these regularity conditions hold, then solving the inclusion problem (25) is equivalent with solving the optimization problem (20).

Let $(\tau, \eta, \rho) \in S^{(1)} \cup S^{(2)}$ and $y^0 = (y_1^0, y_2^0) \in \mathbb{R}^n \times \mathbb{R}^m$, then Algorithm 1 applied to (25) gives

$$\begin{aligned}\bar{u}^k &= \text{prox}_{\tau \tilde{f}}(y^k) \\ w^k &= J_{\eta \tilde{G}} \left(\left(1 + \frac{\eta}{\tau}\right) \bar{u}^k - \frac{\eta}{\tau} y^k \right) \\ y^{k+1} &= y^k + \rho (w^k - \bar{u}^k).\end{aligned} \quad (27)$$

Using $\bar{w}^k := \left(\left(1 + \frac{\eta}{\tau}\right) \bar{u}^k - \frac{\eta}{\tau} y^k \right) - w^k$ and the Moreau decomposition we get that (27) becomes

$$\begin{aligned}\bar{u}^k &= \text{prox}_{\tau \tilde{f}}(y^k) \\ \bar{w}^k &= J_{(\eta \tilde{G})^{-1}} \left(\left(1 + \frac{\eta}{\tau}\right) \bar{u}^k - \frac{\eta}{\tau} y^k \right) \\ y^{k+1} &= \left(1 - \rho \frac{\eta}{\tau}\right) y^k + \rho \left(\frac{\eta}{\tau} \bar{u}^k - \bar{w}^k \right).\end{aligned} \quad (28)$$

Introduce $\sigma := \gamma^2/\eta$ and use the identity

$$J_{(\eta\tilde{G})^{-1}}(u) = \eta B^\top J_{\sigma\partial g^*}(\sigma Bu)$$

that holds for all $u \in \mathbb{R}^{n+m}$ (see [33, Equation (34)]), to reformulate (28) as

$$\begin{aligned}\bar{u}^k &= (\text{prox}_{\tau f}(y_1^k), 0) \\ \bar{w}^k &= \eta B^\top \text{prox}_{\sigma g^*} \left(\sigma B \left(\left(1 + \frac{\eta}{\tau}\right) \bar{u}^k - \frac{\eta}{\tau} y^k \right) \right) \\ y^{k+1} &= \left(1 - \rho \frac{\eta}{\tau}\right) y^k + \rho \left(\frac{\eta}{\tau} \bar{u}^k - \bar{w}^k \right).\end{aligned}\tag{29}$$

Introduce the variables $u^k, w^k \in \mathbb{R}^{n+m}$ for all $k \in \mathbb{N}$ such that

$$\begin{aligned}u^{k+1} &= \left(1 - \rho \frac{\eta}{\tau}\right) u^k + \rho \frac{\eta}{\tau} \bar{u}^k \\ w^k &= u^k - y^k\end{aligned}$$

with $u^0 = (u_1^0, 0)$ so that

$$\begin{cases} y_1^0 &= u_1^0 - w_1^0 \\ y_2^0 &= -w_2^0 \end{cases}.\tag{30}$$

Moreover,

$$\begin{aligned}w^{k+1} &= u^{k+1} - y^{k+1} \\ &= \left(1 - \rho \frac{\eta}{\tau}\right) (u^k - y^k) + \rho \bar{w}^k \\ &= \left(1 - \rho \frac{\eta}{\tau}\right) w^k + \rho \bar{w}^k\end{aligned}$$

and so we reformulate (29) as

$$\begin{aligned}\bar{u}^k &= (\text{prox}_{\tau f}(u_1^k - w_1^k), 0) \\ \bar{w}^k &= \eta B^\top \text{prox}_{\sigma g^*} \left(\sigma \frac{\eta}{\tau} B w^k + \sigma B \left(\left(1 + \frac{\eta}{\tau}\right) \bar{u}^k - \frac{\eta}{\tau} u^k \right) \right) \\ u^{k+1} &= \left(1 - \rho \frac{\eta}{\tau}\right) u^k + \rho \frac{\eta}{\tau} \bar{u}^k \\ w^{k+1} &= \left(1 - \rho \frac{\eta}{\tau}\right) w^k + \rho \bar{w}^k.\end{aligned}\tag{31}$$

Since $\bar{u}_2^k = 0$ for all $k \in \mathbb{N}$ and $u_2^0 = 0$, we conclude that $u_2^k = 0$ for all $k \in \mathbb{N}$. Let

$$\bar{z}^k := \text{prox}_{\sigma g^*} \left(\sigma \frac{\eta}{\tau} B w^k + \sigma B \left(\left(1 + \frac{\eta}{\tau}\right) \bar{u}^k - \frac{\eta}{\tau} u^k \right) \right),$$

so that $\bar{w}^k = \eta B^\top \bar{z}^k$ holds for all $k \in \mathbb{N}$. Moreover,

$$\begin{aligned}B \bar{w}^k &= \eta B B^\top \bar{z}^k \\ &= \frac{\eta}{\gamma^2} \bar{z}^k.\end{aligned}\tag{32}$$

Therefore, by choosing w^0 to be in the range of B^\top also w^k will be in the range of B^\top for all $k \in \mathbb{N}$, i.e., $w^k = \tau B^\top z^k$ for some $z^k \in \mathbb{R}^m$. Moreover,

$$\begin{aligned} Bw^k &= \tau BB^\top z^k \\ &= \frac{\tau}{\gamma^2} z^k. \end{aligned} \tag{33}$$

Let $x^k := u_1^k$ and $\bar{x}^k := \bar{u}_1^k$ for all $k \in \mathbb{N}$. Then,

$$\begin{aligned} x^{k+1} &= \left(1 - \rho \frac{\eta}{\tau}\right) x^k + \rho \frac{\eta}{\tau} \bar{x}^k \\ z^{k+1} &= \frac{\gamma^2}{\tau} Bw^{k+1} \\ &= \left(1 - \rho \frac{\eta}{\tau}\right) \frac{\gamma^2}{\tau} Bw^k + \rho \frac{\gamma^2}{\tau} B\bar{w}^k \\ &= \left(1 - \rho \frac{\eta}{\tau}\right) z^k + \rho \frac{\eta}{\tau} \bar{z}^k, \end{aligned}$$

where we in the last equality used (32) and (33). Given initial points $x^0 \in \mathbb{R}^n$ and $z^0 \in \mathbb{R}^m$, let $w^0 = \tau B^\top z^0$ and $u^0 = (x^0, 0)$ and set $y^0 \in \mathbb{R}^n \times \mathbb{R}^m$ according to (30). Note that when C is non-singular, i.e., when $\gamma \|A\| < 1$, then this mapping between (x^0, z^0) and y^0 is a bijection.

Now let's perform a final change of variables. At this point, we have that (31) becomes

$$\begin{aligned} \bar{x}^k &= \text{prox}_{\tau f} \left(x^k - \tau A^\top z^k \right) \\ \bar{z}^k &= \text{prox}_{\sigma g^*} \left(z^k + \sigma A \left(\left(1 + \frac{\eta}{\tau}\right) \bar{x}^k - \frac{\eta}{\tau} x^k \right) \right) \\ x^{k+1} &= x^k + \rho \frac{\eta}{\tau} (\bar{x}^k - x^k) \\ z^{k+1} &= z^k + \rho \frac{\eta}{\tau} (\bar{z}^k - z^k) \end{aligned} \tag{34}$$

with the bounds $(\tau, \eta, \sigma, \rho) \in \mathbb{R}_{++}^4$ such that $\rho \in (0, \min\{2, 2\tau/\eta\})$ and $\eta\sigma \|A\|^2 \leq 1$. Let $\theta = \eta/\tau$ and $\bar{\rho} = \rho\theta$, then (34) becomes

$$\begin{cases} \bar{x}^k = \text{prox}_{\tau f} \left(x^k - \tau A^\top z^k \right) \\ \bar{z}^k = \text{prox}_{\sigma g^*} \left(z^k + \sigma A \left((1 + \theta) \bar{x}^k - \theta x^k \right) \right) \\ x^{k+1} = x^k + \bar{\rho} (\bar{x}^k - x^k) \\ z^{k+1} = z^k + \bar{\rho} (\bar{z}^k - z^k) \end{cases}$$

with bounds $\bar{\rho} \in (0, \min\{2, 2\theta\})$ and $\tau\sigma \|A\|^2 \leq 1/\theta$, which is exactly Algorithm 4.

6.3 Extending Parallel Splitting

In this section, we extend the so-called Parallel Splitting algorithm, see [3, Proposition 28.7], which is the Douglas–Rachford algorithm applied to a consensus problem formulation. Consider a family of functions $\{f_i\}_{i=1}^n$ where $f_i \in \mathcal{F}$ for all $i \in \{1, \dots, n\}$ and the optimization problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \sum_{i=1}^n f_i(x). \tag{35}$$

Let $\Delta := \{y \in \mathcal{H}^n \mid y_1 = y_2 = \dots = y_n\}$ and let $f : \mathcal{H}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be given by $f(y) := \sum_{i=1}^n f_i(y_i)$ for each $y \in \mathcal{H}^n$. Consider the lifted consensus problem

$$\underset{y \in \mathcal{H}^n}{\text{minimize}} f(y) + \iota_\Delta(y) \quad (36)$$

with sufficient optimality condition

$$0 \in \partial f(y) + \partial \iota_\Delta(y). \quad (37)$$

Then,

$$(J_{\alpha \partial f}(y))_i = \text{prox}_{\alpha f_i}(y_i) \quad (\forall i \in \{1, \dots, n\}),$$

and

$$\left(\text{prox}_{\beta \iota_\Delta}(y)\right)_i = \frac{1}{n} \sum_{j=1}^n y_j \quad (\forall i \in \{1, \dots, n\}),$$

holds for all $y \in \mathcal{H}^n$. Moreover, the projection $\text{prox}_{\beta \iota_\Delta}$ is a linear mapping independent of β . We can therefore let $\gamma = \beta/\alpha$ range freely when Algorithm 1 is applied to (37), as long as $\theta < \min\{2, 2/\gamma\}$. We state Algorithm 1 applied to (37) in Algorithm 5. This generalizes the parallel splitting method in [3, Proposition 28.7] that is obtained by letting $\gamma = 1$. Algorithm 5 converges for these parameters if (37) has a solution.

Algorithm 5 Extended Parallel Splitting

- 1: **Input:** $f : \mathcal{H}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ as in Section 6.3, $y^0 \in \mathcal{H}^n$,
- 2: $(\alpha, \gamma, \theta) \in \mathbb{R}_{++}^3$ such that $\theta < \min\{2, 2/\gamma\}$.
- 3: **for** $k = 0, 1, 2 \dots$ **do**
- 4: **Update:**

$$\left\{ \begin{array}{l} p^k = \frac{1}{n} \sum_{j=1}^n y_j^k \\ x_i^k = \text{prox}_{\alpha f_i}(y_i^k) \quad (\forall i \in \{1, \dots, n\}) \\ q^k = \frac{1}{n} \sum_{j=1}^n x_j^k \\ y_i^{k+1} = y_i^k + \theta \left((1 + \gamma) q^k - \gamma p^k - x_i^k \right) \quad (\forall i \in \{1, \dots, n\}) \end{array} \right.$$

- 5: **end for**
-

7 Operator Composition Interpretation

In this section, we reinterpret our results through the lens of averaged, nonexpansive, and conic operators. We show that the algorithm operator defined by (4) need not be averaged when $\alpha \neq \beta$, and its convergence therefore lies outside the scope of classical averaged operator theory and the associated (preconditioned) proximal point framework. Readers not familiar with these operator-theoretic notions may safely skip this section, as it contains no new results.

We begin by recalling some basic definitions and properties of nonexpansive, averaged, and conic operators. An operator $N : \mathcal{H} \rightarrow \mathcal{H}$ is nonexpansive if $\|Nx - Ny\| \leq \|x - y\|$ for each $x, y \in \mathcal{H}$. For $\lambda \geq 0$, an operator $C : \mathcal{H} \rightarrow \mathcal{H}$ is called λ -conic if there exists a nonexpansive mapping $N : \mathcal{H} \rightarrow \mathcal{H}$ such that $C = (1 - \lambda)\text{Id} + \lambda N$. If $\lambda \in [0, 1)$ then C is λ -averaged, if $\lambda = 1$

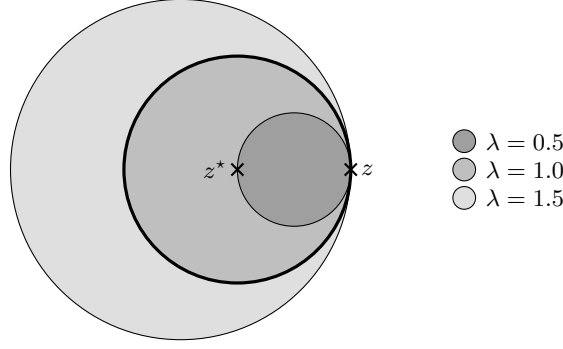


Figure 2: Two-dimensional representations of three λ -conic operators. When $\lambda = 0.5$, the operator is averaged (and therefore nonexpansive); when $\lambda = 1.0$, the operator is nonexpansive but not averaged; and when $\lambda = 1.5$, the operator is conic but not necessarily nonexpansive. For each operator $C : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and any fixed point $z^* \in \mathbb{R}^2$, evaluating C at $z \in \mathbb{R}^2$ produces $C(z) \in \mathbb{R}^2$ lying within the corresponding shaded region. For averaged operators, the distance to all fixed points strictly decreases unless $C(z) = z$; for nonexpansive operators, the distance never increases; and for conic operators with $\lambda > 1$, the distance may increase.

then $C = N$ is nonexpansive, and if $\lambda > 1$ then C is not necessarily nonexpansive. Figure 2 provides a graphical illustration of these classes. Given a mapping $R : \mathcal{H} \rightarrow \mathcal{H}$ and an iterate $z \in \mathcal{H}$, we say that the update $z^+ := (1 - \mu)z + \mu R(z)$ with $\mu \in [0, 1]$ is a μ -averaged iteration of R . We also say that $T := (1 - \mu)\text{Id} + \mu R$ is a μ -averaging of R . Since the iterates of an averaged operator converge weakly to a fixed point, this framework—equivalent to the relaxed proximal point framework [6, 7]—provides a powerful tool for analyzing fixed-point iterations, see, e.g., [44, 45, 46, 47, 3]. We also rely on the following result on compositions of conic operators.

Fact 7.1 ([45, Theorem 4.2, Proposition 4.5]): *Let $\lambda_1, \lambda_2 \geq 0$ and let $R_1 : \mathcal{H} \rightarrow \mathcal{H}$ and $R_2 : \mathcal{H} \rightarrow \mathcal{H}$ be λ_1 -conic and λ_2 -conic respectively.*

- i) *If $\lambda_1 \lambda_2 < 1$, then $R_2 R_1$ is $\frac{\lambda_1 + \lambda_2 - 2\lambda_1 \lambda_2}{1 - \lambda_1 \lambda_2}$ -conic.*
- ii) *If $\max(\lambda_1, \lambda_2) = 1$, then $R_2 R_1$ is nonexpansive.*
- iii) *If $\lambda_1 \lambda_2 > 1$, then there exists $\tilde{\lambda}_1$ -conic \tilde{R}_1 and $\tilde{\lambda}_2$ -conic \tilde{R}_2 such that $\tilde{R}_2 \tilde{R}_1$ is not conic.*

The algorithm in (3) with parameters $(\alpha, \beta, \theta) \in \mathbb{R}_{++}^2 \times (\mathbb{R} \setminus \{0\})$, repeated here for convenience

$$\begin{cases} x_1 = J_{\alpha A}(z) \\ x_2 = J_{\beta B} \left(\left(1 + \frac{\beta}{\alpha}\right) x_1 - \frac{\beta}{\alpha} z \right), \\ z^+ = z + \theta(x_2 - x_1) \end{cases} \quad (38)$$

can equivalently be written as an averaged iteration of a composition of two conic operators. To see this, let us introduce the *relaxed resolvents*

$$R_A^{\alpha, \beta} := \left(1 + \frac{\beta}{\alpha}\right) J_{\alpha A} - \frac{\beta}{\alpha} \text{Id} \quad \text{and} \quad R_B^{\beta, \alpha} := \left(1 + \frac{\alpha}{\beta}\right) J_{\beta B} - \frac{\alpha}{\beta} \text{Id}, \quad (39)$$

which allow us to rewrite the update in (38) as

$$z^+ = T_{A,B}^{\alpha, \beta, \theta}(z) \quad \text{where} \quad T_{A,B}^{\alpha, \beta, \theta} := \left(1 - \frac{\theta\beta}{\alpha + \beta}\right) \text{Id} + \frac{\theta\beta}{\alpha + \beta} R_B^{\beta, \alpha} R_A^{\alpha, \beta}. \quad (40)$$

Under the parameter restrictions in (5), we have $\frac{\theta\beta}{\alpha+\beta} \in [0, 1]$ so (40) is a $\frac{\theta\beta}{\alpha+\beta}$ -averaged iteration of the composition $R_B^{\beta,\alpha} R_A^{\alpha,\beta}$. Each resolvent $J_{\alpha A}$ of a maximally monotone operator is firmly nonexpansive, equivalently $\frac{1}{2}$ -averaged [3, Remark 4.34]. Therefore, for each $A \in \mathcal{A}$, there exists a nonexpansive $N : \mathcal{H} \rightarrow \mathcal{H}$ such that $J_{\alpha A} = \frac{1}{2}(\text{Id} + N)$, implying that

$$R_A^{\alpha,\beta} = \left(1 + \frac{\beta}{\alpha}\right) \frac{1}{2} (\text{Id} + N) - \frac{\beta}{\alpha} \text{Id} = \left(1 - \frac{1}{2} \left(1 + \frac{\beta}{\alpha}\right)\right) \text{Id} + \frac{1}{2} \left(1 + \frac{\beta}{\alpha}\right) N. \quad (41)$$

An identical expression holds for $R_B^{\beta,\alpha}$ with β/α replaced by α/β . It follows that $R_A^{\alpha,\beta}$ and $R_B^{\beta,\alpha}$ are $\lambda_{\alpha,\beta}$ -conic and $\lambda_{\beta,\alpha}$ -conic operators, respectively, where

$$\lambda_{\alpha,\beta} := \frac{1}{2} \left(1 + \frac{\beta}{\alpha}\right) \quad \text{and} \quad \lambda_{\beta,\alpha} := \frac{1}{2} \left(1 + \frac{\alpha}{\beta}\right), \quad (42)$$

and we conclude that (40) indeed is an averaged iteration of a composition of conic operators. Using that an operator is firmly nonexpansive *if and only if* it is the resolvent of a maximally monotone operator [3, Corollary 23.9], we obtain the following characterization of the relaxed resolvents.

Proposition 7.1: *Let \mathcal{C}_λ be the class of λ -conic operators $C : \mathcal{H} \rightarrow \mathcal{H}$. Then*

$$\begin{aligned} \mathcal{C}_{\lambda_{\alpha,\beta}} &= \{R : \mathcal{H} \rightarrow \mathcal{H} \mid R = R_A^{\alpha,\beta} \text{ for some } A \in \mathcal{A}\}, \\ \mathcal{C}_{\lambda_{\beta,\alpha}} &= \{R : \mathcal{H} \rightarrow \mathcal{H} \mid R = R_B^{\beta,\alpha} \text{ for some } B \in \mathcal{A}\}, \end{aligned}$$

where $\lambda_{\alpha,\beta}$ and $\lambda_{\beta,\alpha}$ are given in (42), and $R_A^{\alpha,\beta}$ and $R_B^{\beta,\alpha}$ are defined in (39).

Let us examine the conic constants $\lambda_{\alpha,\beta}$ and $\lambda_{\beta,\alpha}$ more closely. Since $\alpha, \beta > 0$, we immediately have $\lambda_{\alpha,\beta}, \lambda_{\beta,\alpha} > \frac{1}{2}$. The product of the two constants satisfies

$$\lambda_{\alpha,\beta} \lambda_{\beta,\alpha} = \frac{1}{4} \left(1 + \frac{\beta}{\alpha}\right) \left(1 + \frac{\alpha}{\beta}\right) = \frac{1}{4} \left(2 + \frac{\beta}{\alpha} + \frac{\alpha}{\beta}\right) \geq 1$$

with equality if and only if $\alpha = \beta$. Moreover, $\lambda_{\alpha,\beta} < 1$ if and only if $\lambda_{\beta,\alpha} > 1$, and vice versa. Together with Fact 7.1, this implies that exactly one of the following holds:

- i) $R_A^{\alpha,\beta}$ and $R_B^{\beta,\alpha}$ are nonexpansive implying that $R_B^{\beta,\alpha} R_A^{\alpha,\beta}$ is nonexpansive (if $\alpha = \beta$),
- ii) $R_A^{\alpha,\beta}$ is $\lambda_{\alpha,\beta}$ -conic, $R_B^{\beta,\alpha}$ is $\lambda_{\beta,\alpha}$ -averaged, and there exists $(A, B) \in \mathcal{A} \times \mathcal{A}$ such that $R_B^{\beta,\alpha} R_A^{\alpha,\beta}$ is not conic (if $\alpha < \beta$),
- iii) $R_A^{\alpha,\beta}$ is $\lambda_{\alpha,\beta}$ -averaged, $R_B^{\beta,\alpha}$ is $\lambda_{\beta,\alpha}$ -conic, and there exists $(A, B) \in \mathcal{A} \times \mathcal{A}$ such that $R_B^{\beta,\alpha} R_A^{\alpha,\beta}$ is not conic (if $\alpha > \beta$).

When $\alpha = \beta$ and $\theta \in (0, 2)$, as in the unconditionally convergent case for maximally monotone A and B , the relaxed resolvents are nonexpansive and algorithm (40) reduces to a $\frac{\theta}{2}$ -averaged fixed-point iteration of the composition of the two nonexpansive operators. Consequently, the algorithm operator $T_{A,B}^{\alpha,\beta,\theta}$ is averaged. It is straightforward to verify that an operator is averaged if and only if it can be written as an averaging of a conic operator. The above discussion therefore lets us paraphrase [21, Theorem 2] as follows.

The algorithm in (38) converges unconditionally over $\mathcal{A} \times \mathcal{A}$ if and only if its associated algorithm operator $T_{A,B}^{\alpha,\beta,\theta}$ in (40) is an averaged operator for all $(A, B) \in \mathcal{A} \times \mathcal{A}$.

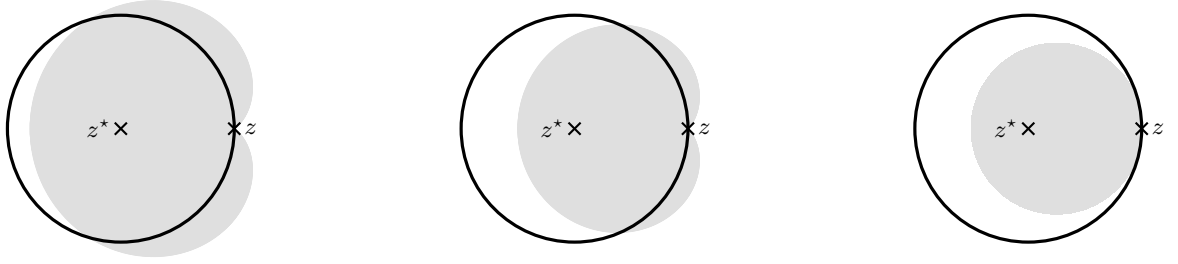


Figure 3: Two-dimensional representations of the possible locations of $z^+ \in \mathbb{R}^2$ in (40) relative to an arbitrary fixed point $z^* \in \mathbb{R}^2$, for a given starting point $z \in \mathbb{R}^2$. *Left figure:* $(\alpha, \beta, \theta) = (1, 9, 0.2)$ implying $(\lambda_{\alpha, \beta}, \lambda_{\beta, \alpha}) = (5, 0.56)$, i.e., that z^+ is obtained by applying to z a 0.18-averaging of a composition of a 5-conic and a 0.56-conic operator. *Middle figure:* $(\alpha, \beta, \theta) = (1, 0.2, 1.5)$ implying $(\lambda_{\alpha, \beta}, \lambda_{\beta, \alpha}) = (0.6, 3)$, i.e., that z^+ is obtained by applying to z a 0.25-averaging of a composition of a 0.6-conic and a 3-conic operator. *Right figure:* $(\alpha, \beta, \theta) = (1, 1, 1.5)$ implying $(\lambda_{\alpha, \beta}, \lambda_{\beta, \alpha}) = (1, 1)$, i.e., that z^+ is obtained by applying to z a 0.75-averaging of a composition of two nonexpansive operators. The left and middle figures illustrate that, in these regimes, the associated algorithm operators may fail to be averaged.

In the optimization setting, our analysis shows that the monotone-operator parameter requirement $\alpha = \beta$ can be relaxed to the inequality $\frac{\alpha}{\beta} > \frac{\theta}{2}$ (see (5)), meaning that α must be sufficiently large relative to β . Whenever $\alpha \neq \beta$, there exist operators $(A, B) \in \mathcal{A} \times \mathcal{A}$ —and even $(A, B) \in \partial\mathcal{F} \times \partial\mathcal{F}$, as we will demonstrate later—for which the algorithm operator $T_{A,B}^{\alpha, \beta, \theta}$ is not averaged although the corresponding algorithm in (38) converges unconditionally over $\partial\mathcal{F} \times \partial\mathcal{F}$. This implies that alternative proof techniques are required—such as the Lyapunov analysis developed in this paper—to establish convergence in the convex optimization setting. See Figure 3 for two-dimensional graphical illustrations of the potential lack of averagedness and Figure 4 for similar illustrations overlaid with an example. Illustrations similar to those in Figures 2 to 4 have appeared before in the literature as a tool for understanding algorithm dynamics [4, 45, 48].

Let us present the problem behind the illustration in Figure 4. Let $c \in \mathbb{R} \setminus \{0\}$ and consider $f = \iota_{C_1}$ and $g = \iota_{C_2}$, where

$$C_1 = \{x \in \mathbb{R}^2 \mid x = (ct, t) \text{ for } t \in \mathbb{R}\} \quad \text{and} \quad C_2 = \{x \in \mathbb{R}^2 \mid x = (0, t) \text{ for } t \in \mathbb{R}\}.$$

The problem of finding an $x \in \mathbb{R}^2$ such that

$$0 \in \partial\iota_{C_1}(x) + \partial\iota_{C_2}(x) \tag{43}$$

(where $\partial\iota_{C_i}$ is the normal cone of $C_i \subset \mathbb{R}^2$) is a convex feasibility problem whose unique solution is $x = (0, 0)$. Figure 4 shows that the algorithm operator $T_{A,B}^{\alpha, \beta, \theta}$ can fail to be averaged for this problem by visualizing parameter choices for which $T_{\partial\iota_{C_1}, \partial\iota_{C_2}}^{\alpha, \beta, \theta}$ is convergent yet neither averaged nor nonexpansive. The example uses $c = \frac{1}{2}$. In this case, the algorithm operator is represented by the matrix:

$$T := T_{\partial\iota_{C_1}, \partial\iota_{C_2}}^{\alpha, \beta, \theta} = \begin{bmatrix} 1 - \frac{\theta}{5} & -\frac{2\theta}{5} \\ \frac{2\theta\beta}{5\alpha} & 1 - \frac{\theta\beta}{5\alpha} \end{bmatrix}.$$

Because (43) is a feasibility problem, the resolvent steps reduce to projections, so the algorithm depends on α and β only through the ratio β/α . The iteration matrix T is invertible

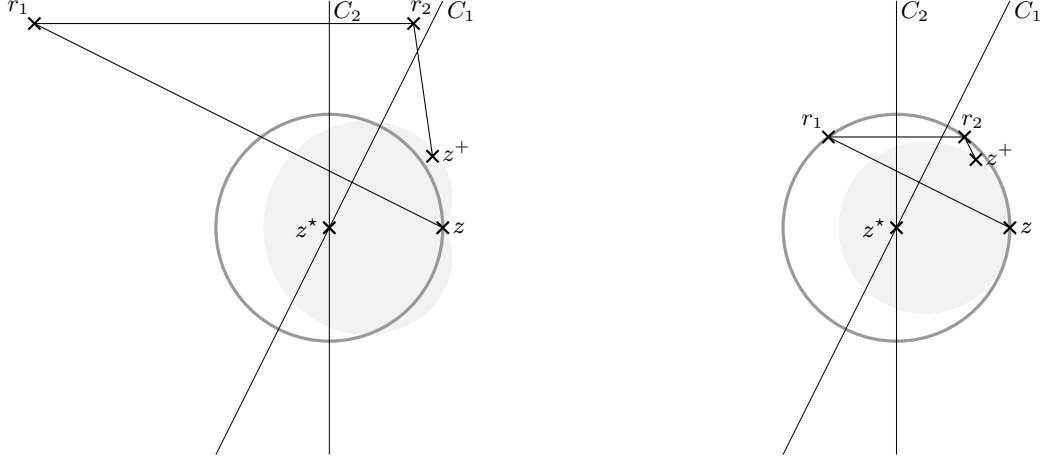


Figure 4: Two-dimensional illustration of one step of (40) applied to the feasibility problem (43), with the regions from Figure 3 indicating where z^+ may lie shown in the background. We denote $r_1 = R_{\partial_{C_1}}^{\alpha, \beta}(z)$, $r_2 = R_{\partial_{C_2}}^{\beta, \alpha}(r_1)$, and $z^+ = \left(1 - \frac{\theta\beta}{\alpha+\beta}\right)z + \frac{\theta\beta}{\alpha+\beta}r_2$. *Left figure:* $(\alpha, \beta, \theta) = (1, 3.5, 0.45)$ implying $\frac{\theta\beta}{\alpha+\beta} = 0.35$ and $(\lambda_{\alpha, \beta}, \lambda_{\beta, \alpha}) = (2.25, 0.64)$. Thus, the first relaxed resolvent is conic but not nonexpansive, and the second is averaged. The resulting z^+ lies outside the nonexpansiveness circle. *Right figure:* $(\alpha, \beta, \theta) = (1, 1, 1.5)$ implying $\frac{\theta\beta}{\alpha+\beta} = 0.75$ and $(\lambda_{\alpha, \beta}, \lambda_{\beta, \alpha}) = (1, 1)$, i.e., the Douglas–Rachford setting. In this case, both relaxed resolvents are nonexpansive reflected resolvents. Consequently, the point z^+ remains inside the nonexpansiveness circle. Both these algorithms converge, although the distance to a fixed point may temporarily increase between iterations in the left example.

for every choice of (α, β, θ) satisfying the parameter restrictions in (5), and convergence to the unique fixed point at $(0, 0)$ follows from an eigenvalue analysis. The asymptotic rate of convergence is governed by the magnitude of the largest eigenvalue of T . Moreover, the largest singular value of T is the (smallest) Lipschitz constant (in \mathbb{R}^2), which, for problems over $\mathcal{A} \times \mathcal{A}$, is illustrated in Figures 3 and 4 as the point in the shaded region that lies the farthest from the fixed point.

In the Douglas–Rachford case (with $\theta = 1$ corresponding to $\frac{1}{2}$ -averaging), the magnitude of the largest eigenvalue is known to equal the cosine of the Friedrichs angle [10], which here is $2/\sqrt{5} \approx 0.89$. This is the best possible rate in the Douglas–Rachford setting with $\theta \in (0, 2)$ and $\alpha = \beta$. Since the matrix T is normal in this case, the largest singular value coincides with the magnitude of the largest eigenvalue. Indeed, the right-most illustrations in Figures 3 and 4 indicate that this is not a coincidence as the set of possible algorithm operators must have Lipschitz constant at most 1.

We can optimize the magnitude of the largest eigenvalue of T with two feasible parameter choices

$$(\alpha, \beta, \theta) = \left(1, \frac{1-\epsilon}{\sigma}, 2\sigma\right) \quad \text{and} \quad (\alpha, \beta, \theta) = \left(1, \frac{\sigma}{1-\epsilon}, 2(1-\epsilon)\right),$$

where $\sigma = 9 - 4\sqrt{5} \approx 0.056$ and $\epsilon \in (0, 1)$. We note that the resulting matrices are unitarily similar with unitary matrix $P \in \mathbb{R}^{2 \times 2}$ that has zero on the diagonal and $P_{12} = 1$ and $P_{21} = -1$. They therefore share the same eigenvalues and singular values. As $\epsilon \downarrow 0$, we approach the boundary of the feasible parameter region. For small enough $\epsilon > 0$, the largest eigenvalue magnitude of T becomes arbitrarily close to $\sqrt{(3+2\sigma)/5} \approx 0.79$ in both cases, which is a strictly better rate than in the Douglas–Rachford case. The largest singular values become in both cases

arbitrarily close to $\sqrt{1 + 2(1 - \sigma)(\sqrt{\sigma^2 + 4} - \sigma)}/5 \approx 1.32$. A largest singular value exceeding 1 is equivalent to the operator being not nonexpansive, and therefore not averaged. This shows that indeed, the algorithm operator can fail to be averaged and still converge unconditionally. In the left illustration of Figure 4, we showcase one parameter choice for which this is the case.

To characterize the broader class of unconditionally convergent methods in the convex optimization setting, we had to step outside the usual framework of averaged operators. While allowing iterations that are not averaged is not, by itself, the fundamental source of the improved convergence observed in our example, it does suggest that the additional freedom in parameter selection can translate into meaningful performance gains.

8 Conclusions

In this work, we have shown that, in the convex optimization setting, the Douglas–Rachford splitting method is not unique as an unconditionally convergent, frugal, no-lifting resolvent-splitting scheme. When the operators are subdifferentials of proper, closed, convex functions, we identify a strictly larger class of such methods and provide a complete, sharp characterization of all parameter choices that yield unconditional convergence. This characterization, in turn, immediately yields new, provably convergent families of ADMM- and Chambolle–Pock-type algorithms via their Douglas–Rachford reformulations, including parameter regimes that were previously not known to converge.

Our Lyapunov functions explicitly involve function values, placing the analysis beyond the standard averaged-operator framework. Because the Lyapunov construction uses function values of only the function whose subdifferential is applied first in each iteration, our results extend to mixed problems combining one subdifferential operator with one general maximally monotone operator, provided the algorithm respects this order. Preliminary numerical evidence from [27] suggests that the same parameter region may remain valid when the order of the operators is reversed, although establishing this rigorously will require a new Lyapunov analysis. We formalize this in the following conjecture.

Conjecture 8.1: *Algorithm 1 converges unconditionally for $(\alpha, \beta, \theta) \in \mathbb{R}_{++}^2 \times (\mathbb{R} \setminus \{0\})$ over $\mathcal{A} \times \partial\mathcal{F}$ if and only if $(\alpha, \beta, \theta) \in S^{(1)} \cup S^{(2)}$, i.e.,*

$$S(\mathcal{A} \times \partial\mathcal{F}) = S(\partial\mathcal{F} \times \partial\mathcal{F}) = S^{(1)} \cup S^{(2)}.$$

Beyond the Douglas–Rachford setting, we expect that many fixed-parameter methods for monotone inclusions may admit strictly larger regions of guaranteed convergence when their analysis is specialized to the convex optimization setting.

Appendices

A Verification of Equality (LE) in Proposition 4.1

The equality in (LE) is also verified symbolically in [37].

We will occasionally place $*$ in the lower-triangular part of a matrix to avoid repeating entries. The resulting matrix should always be interpreted as symmetric; for example,

$$\begin{bmatrix} a & b & c \\ * & d & e \\ * & * & f \end{bmatrix} = \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}.$$

Expanding the norms of the quantities in Definition 4.2, we obtain the following expressions in terms of quadratic forms for all $k \in \mathbb{N}$:

$$\begin{aligned} V_k^{(1)} &= \mathcal{Q} \left(\underbrace{\begin{bmatrix} \theta - \tau + 1 & -\frac{\theta(1+\tau)}{2} & 0 & 0 & \tau - 1 - \frac{\theta}{2} \\ * & \theta\tau & 0 & 0 & \frac{\theta}{2} \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 1 \end{bmatrix}}_{=:Q_V^{(1)}}, \begin{bmatrix} x_1^k - x^* \\ x_2^k - x^* \\ x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ z^k - z^* \end{bmatrix} \right) + 2\alpha(1 - \tau)D_f(x_1^k, x^*, u^*), \\ \\ V_k^{(2)} &= \mathcal{Q} \left(\underbrace{\begin{bmatrix} \theta + \tau - 1 & -\frac{\theta(1+\tau)}{2} & 0 & 0 & -\frac{\theta}{2} \\ * & \theta\tau & 0 & 0 & \frac{\theta}{2} \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 1 \end{bmatrix}}_{=:Q_V^{(2)}}, \begin{bmatrix} x_1^k - x^* \\ x_2^k - x^* \\ x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ z^k - z^* \end{bmatrix} \right) + 2\alpha(\tau - 1)D_f(x^*, x_1^k, u_1^k), \\ \\ R_k^{(1)} &= \mathcal{Q} \left(\underbrace{\begin{bmatrix} \theta(2\tau - \theta) - \tau + 1 & -\theta(2\tau - \theta) & -(1 - \tau) & 0 & 0 \\ * & \theta(2\tau - \theta) & 0 & 0 & 0 \\ * & * & 1 - \tau & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix}}_{=:Q_R^{(1)}}, \begin{bmatrix} x_1^k - x^* \\ x_2^k - x^* \\ x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ z^k - z^* \end{bmatrix} \right) + 2\alpha(1 - \tau)D_f(x_1^k, x_1^{k+1}, u_1^{k+1}), \end{aligned}$$

$$R_k^{(2)} = \mathcal{Q} \left(\underbrace{\begin{bmatrix} \theta(2-\theta) + \tau - 1 & \theta(\theta - 1 - \tau) & (\theta - 1)(\tau - 1) & 0 & 0 \\ * & \theta(2\tau - \theta) & -\theta(\tau - 1) & 0 & 0 \\ * & * & \tau - 1 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix}}_{=:Q_R^{(2)}} \begin{bmatrix} x_1^k - x^* \\ x_2^k - x^* \\ x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ z^k - z^* \end{bmatrix} \right) \\ + 2\alpha(\tau - 1)D_f(x_1^{k+1}, x_1^k, u_1^k),$$

and

$$V_{k+1}^{(1)} = \mathcal{Q} \left(Q_V^{(1)}, \begin{bmatrix} x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ x_1^{k+2} - x^* \\ x_2^{k+2} - x^* \\ z^{k+1} - z^* \end{bmatrix} \right) + 2\alpha(1 - \tau)D_f(x_1^{k+1}, x^*, u^*) \\ V_{k+1}^{(2)} = \mathcal{Q} \left(Q_V^{(2)}, \begin{bmatrix} x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ x_1^{k+2} - x^* \\ x_2^{k+2} - x^* \\ z^{k+1} - z^* \end{bmatrix} \right) + 2\alpha(\tau - 1)D_f(x^*, x_1^{k+1}, u_1^{k+1}).$$

For a matrix $Q \in \mathbb{S}^5$ with structure $Q = \begin{bmatrix} q_{11} & q_{12} & 0 & 0 & q_{13} \\ * & q_{22} & 0 & 0 & q_{23} \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 1 \end{bmatrix}$ we have that

$$\mathcal{Q} \left(Q, \begin{bmatrix} x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ x_1^{k+2} - x^* \\ x_2^{k+2} - x^* \\ z^{k+1} - z^* \end{bmatrix} \right) = \mathcal{Q} \left(Q, \begin{bmatrix} x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ 0 \\ 0 \\ z^{k+1} - z^* \end{bmatrix} \right) = \mathcal{Q} \left(Q, \underbrace{\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\theta & \theta & 0 & 0 & 1 \end{bmatrix}}_{=:M} \begin{bmatrix} x_1^k - x^* \\ x_2^k - x^* \\ x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ z^k - z^* \end{bmatrix} \right) \\ = \mathcal{Q} \left(\begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\theta & \theta & 0 & 0 & 1 \end{bmatrix}^\top \begin{bmatrix} q_{11} & q_{12} & 0 & 0 & q_{13} \\ * & q_{22} & 0 & 0 & q_{23} \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\theta & \theta & 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} x_1^k - x^* \\ x_2^k - x^* \\ x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ z^k - z^* \end{bmatrix} \right)$$

$$= \mathcal{Q} \left(\underbrace{\begin{bmatrix} \theta^2 & -\theta^2 & -\theta q_{13} & -\theta q_{23} & -\theta \\ * & \theta^2 & \theta q_{13} & \theta q_{23} & \theta \\ * & * & q_{11} & q_{12} & q_{13} \\ * & * & * & q_{22} & q_{23} \\ * & * & * & * & 1 \end{bmatrix}}_{=M^\top Q M}, \begin{bmatrix} x_1^k - x^* \\ x_2^k - x^* \\ x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ z^k - z^* \end{bmatrix} \right),$$

where the first equality is due to the zero columns of Q and the second equality follows from the update step of Algorithm 1. Therefore

$$V_{k+1}^{(1)} = \mathcal{Q} \left(\underbrace{\begin{bmatrix} \theta^2 & -\theta^2 & -\theta(\tau - 1 - \frac{\theta}{2}) & -\frac{\theta^2}{2} & -\theta \\ * & \theta^2 & \theta(\tau - 1 - \frac{\theta}{2}) & \frac{\theta^2}{2} & \theta \\ * & * & 1 - \tau + \theta & -\frac{\theta(1+\tau)}{2} & \tau - 1 - \frac{\theta}{2} \\ * & * & * & \theta\tau & \frac{\theta}{2} \\ * & * & * & * & 1 \end{bmatrix}}_{=M^\top Q_V^{(1)} M}, \begin{bmatrix} x_1^k - x^* \\ x_2^k - x^* \\ x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ z^k - z^* \end{bmatrix} \right) + 2\alpha(1 - \tau)D_f(x_1^{k+1}, x^*, u^*)$$

and

$$V_{k+1}^{(2)} = \mathcal{Q} \left(\underbrace{\begin{bmatrix} \theta^2 & -\theta^2 & \frac{\theta^2}{2} & -\frac{\theta^2}{2} & -\theta \\ * & \theta^2 & -\frac{\theta^2}{2} & \frac{\theta^2}{2} & \theta \\ * & * & \theta + \tau - 1 & -\frac{\theta(1+\tau)}{2} & -\frac{\theta}{2} \\ * & * & * & \theta\tau & \frac{\theta}{2} \\ * & * & * & * & 1 \end{bmatrix}}_{=M^\top Q_V^{(2)} M}, \begin{bmatrix} x_1^k - x^* \\ x_2^k - x^* \\ x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ z^k - z^* \end{bmatrix} \right) + 2\alpha(\tau - 1)D_f(x^*, x_1^{k+1}, u_1^{k+1}).$$

By Lemma 2.1 we get that

$$\begin{aligned} D_f(x_1^k, x^*, u^*) - D_f(x_1^{k+1}, x^*, u^*) - D_f(x_1^k, x_1^{k+1}, u_1^{k+1}) &= \langle u_1^{k+1} - u^*, x_1^k - x_1^{k+1} \rangle \\ &= \frac{1}{\alpha} \langle z^{k+1} - z^* - (x_1^{k+1} - x^*), x_1^k - x_1^{k+1} \rangle \\ &= \frac{1}{\alpha} \langle z^k - z^* + \theta(x_2^k - x_1^k) - (x_1^{k+1} - x^*), x_1^k - x_1^{k+1} \rangle \\ &= \frac{1}{\alpha} \mathcal{Q} \left(\underbrace{\begin{bmatrix} -\theta & \frac{\theta}{2} & \frac{\theta-1}{2} & 0 & \frac{1}{2} \\ * & 0 & -\frac{\theta}{2} & 0 & 0 \\ * & * & 1 & 0 & -\frac{1}{2} \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix}}_{=:Q^{(1)}}, \begin{bmatrix} x_1^k - x^* \\ x_2^k - x^* \\ x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ z^k - z^* \end{bmatrix} \right) \end{aligned}$$

and

$$\begin{aligned}
D_f(x^\star, x_1^k, u_1^k) - D_f(x^\star, x_1^{k+1}, u_1^{k+1}) - D_f(x_1^{k+1}, x_1^k, u_1^k) &= \langle u_1^{k+1} - u_1^k, x^\star - x_1^{k+1} \rangle \\
&= -\frac{1}{\alpha} \langle z^{k+1} - z^k - (x_1^{k+1} - x_1^k), x_1^{k+1} - x^\star \rangle \\
&= -\frac{1}{\alpha} \langle \theta(x_2^k - x_1^k) - (x_1^{k+1} - x_1^k), x_1^{k+1} - x^\star \rangle \\
&= -\frac{1}{\alpha} \mathcal{Q} \left(\underbrace{\begin{bmatrix} 0 & 0 & \frac{1-\theta}{2} & 0 & 0 \\ * & 0 & \frac{\theta}{2} & 0 & 0 \\ * & * & -1 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix}}_{=:Q^{(2)}}, \begin{bmatrix} x_1^k - x^\star \\ x_2^k - x^\star \\ x_1^{k+1} - x^\star \\ x_2^{k+1} - x^\star \\ z^k - z^\star \end{bmatrix} \right).
\end{aligned}$$

Using these equalities we find that

$$\begin{aligned}
V_k^{(1)} - V_{k+1}^{(1)} - R_k^{(1)} &= \mathcal{Q} \left(Q_V^{(1)} - M^\top Q_V^{(1)} M - Q_R^{(1)}, \begin{bmatrix} x_1^k - x^\star \\ x_2^k - x^\star \\ x_1^{k+1} - x^\star \\ x_2^{k+1} - x^\star \\ z^k - z^\star \end{bmatrix} \right) \\
&\quad + 2\alpha(1 - \tau) \left(D_f(x_1^k, x^\star, u^\star) - D_f(x_1^{k+1}, x^\star, u^\star) - D_f(x_1^k, x_1^{k+1}, u_1^{k+1}) \right) \\
&= \mathcal{Q} \left(Q_V^{(1)} - M^\top Q_V^{(1)} M - Q_R^{(1)} + 2(1 - \tau)Q^{(1)}, \begin{bmatrix} x_1^k - x^\star \\ x_2^k - x^\star \\ x_1^{k+1} - x^\star \\ x_2^{k+1} - x^\star \\ z^k - z^\star \end{bmatrix} \right)
\end{aligned}$$

and similarly

$$\begin{aligned}
V_k^{(2)} - V_{k+1}^{(2)} - R_k^{(2)} &= \mathcal{Q} \left(Q_V^{(2)} - M^\top Q_V^{(2)} M - Q_R^{(2)}, \begin{bmatrix} x_1^k - x^\star \\ x_2^k - x^\star \\ x_1^{k+1} - x^\star \\ x_2^{k+1} - x^\star \\ z^k - z^\star \end{bmatrix} \right) \\
&\quad + 2\alpha(\tau - 1) \left(D_f(x^\star, x_1^k, u_1^k) - D_f(x^\star, x_1^{k+1}, u_1^{k+1}) - D_f(x_1^{k+1}, x_1^k, u_1^k) \right) \\
&= \mathcal{Q} \left(Q_V^{(2)} - M^\top Q_V^{(2)} M - Q_R^{(2)} + 2(1 - \tau)Q^{(2)}, \begin{bmatrix} x_1^k - x^\star \\ x_2^k - x^\star \\ x_1^{k+1} - x^\star \\ x_2^{k+1} - x^\star \\ z^k - z^\star \end{bmatrix} \right).
\end{aligned}$$

This gives quadratic expressions for the left hand side of equation **(LE)**. We also note that the

right hand side of **(LE)** can be expressed as

$$\begin{aligned}
\theta \alpha I_k &= \theta \alpha \left(\langle u_1^k - u^*, x_1^k - x^* \rangle + \langle u_2^k + u^*, x_2^k - x^* \rangle + \langle u_1^{k+1} - u^*, x_1^{k+1} - x^* \rangle + \langle u_2^{k+1} + u^*, x_2^{k+1} - x^* \rangle \right) \\
&= \theta \left\langle z^k - z^* - (x_1^k - x^*), x_1^k - x^* \right\rangle \\
&\quad + \theta \left\langle z^k - z^* + \theta(x_2^k - x_1^k) - (x_1^{k+1} - x^*), x_1^{k+1} - x^* \right\rangle \\
&\quad + \theta \left\langle x_1^k - x^* + \tau(x_1^k - x_2^k) - (z^k - z^*), x_2^k - x^* \right\rangle \\
&\quad + \theta \left\langle x_1^{k+1} - x^* + \tau(x_1^{k+1} - x_2^{k+1}) - (z^k - z^*) - \theta(x_2^k - x_1^k), x_2^{k+1} - x^* \right\rangle \\
&= \theta \mathcal{Q} \left(\underbrace{\begin{bmatrix} -1 & \frac{1+\tau}{2} & -\frac{\theta}{2} & \frac{\theta}{2} & \frac{1}{2} \\ * & -\tau & \frac{\theta}{2} & -\frac{\theta}{2} & -\frac{1}{2} \\ * & * & -1 & \frac{1+\tau}{2} & \frac{1}{2} \\ * & * & * & -\tau & -\frac{1}{2} \\ * & * & * & * & 0 \end{bmatrix}}_{=:Q_I}, \begin{bmatrix} x_1^k - x^* \\ x_2^k - x^* \\ x_1^{k+1} - x^* \\ x_2^{k+1} - x^* \\ z^k - z^* \end{bmatrix} \right). \tag{44}
\end{aligned}$$

Lastly we verify that

$$Q_V^{(i)} - M^\top Q_V^{(i)} M - Q_R^{(i)} + 2(1 - \tau)Q^{(i)} = \theta Q_I,$$

where Q_I is given by (44), for $i \in \{1, 2\}$:

$$\begin{aligned} Q_V^{(1)} - M^\top Q_V^{(1)} M - Q_R^{(1)} + 2(1 - \tau)Q^{(1)} &= \begin{bmatrix} \theta - \tau + 1 & -\frac{\theta(1+\tau)}{2} & 0 & 0 & \tau - 1 - \frac{\theta}{2} \\ * & \theta\tau & 0 & 0 & \frac{\theta}{2} \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 1 \end{bmatrix} \\ &- \begin{bmatrix} \theta^2 & -\theta^2 & -\theta(\tau - 1 - \frac{\theta}{2}) & -\frac{\theta^2}{2} & -\theta \\ * & \theta^2 & \theta(\tau - 1 - \frac{\theta}{2}) & \frac{\theta^2}{2} & \theta \\ * & * & 1 - \tau + \theta & -\frac{\theta(1+\tau)}{2} & \tau - 1 - \frac{\theta}{2} \\ * & * & * & \theta\tau & \frac{\theta}{2} \\ * & * & * & * & 1 \end{bmatrix} \\ &- \begin{bmatrix} \theta(2\tau - \theta) - \tau + 1 & -\theta(2\tau - \theta) & -(1 - \tau) & 0 & 0 \\ * & \theta(2\tau - \theta) & 0 & 0 & 0 \\ * & * & 1 - \tau & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix} \\ &+ \begin{bmatrix} -\theta & \frac{\theta}{2} & \frac{\theta-1}{2} & 0 & \frac{1}{2} \\ * & 0 & -\frac{\theta}{2} & 0 & 0 \\ * & * & 1 & 0 & -\frac{1}{2} \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix} 2(1 - \tau) \\ &= \begin{bmatrix} -1 & \frac{1+\tau}{2} & -\frac{\theta}{2} & \frac{\theta}{2} & \frac{1}{2} \\ * & -\tau & \frac{\theta}{2} & -\frac{\theta}{2} & -\frac{1}{2} \\ * & * & -1 & \frac{1+\tau}{2} & \frac{1}{2} \\ * & * & * & -\tau & -\frac{1}{2} \\ * & * & * & * & 0 \end{bmatrix} \theta \end{aligned}$$

and

$$\begin{aligned}
Q_V^{(2)} - M^\top Q_V^{(2)} M - Q_R^{(2)} + 2(1 - \tau)Q^{(2)} = & \begin{bmatrix} \theta + \tau - 1 & -\frac{\theta(1+\tau)}{2} & 0 & 0 & -\frac{\theta}{2} \\ * & \theta\tau & 0 & 0 & \frac{\theta}{2} \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 1 \end{bmatrix} \\
- & \begin{bmatrix} \theta^2 & -\theta^2 & \frac{\theta^2}{2} & -\frac{\theta^2}{2} & -\theta \\ * & \theta^2 & -\frac{\theta^2}{2} & \frac{\theta^2}{2} & \theta \\ * & * & \theta + \tau - 1 & -\frac{\theta(1+\tau)}{2} & -\frac{\theta}{2} \\ * & * & * & \theta\tau & \frac{\theta}{2} \\ * & * & * & * & 1 \end{bmatrix} \\
- & \begin{bmatrix} \theta(2 - \theta) + \tau - 1 & \theta(\theta - 1 - \tau) & (\theta - 1)(\tau - 1) & 0 & 0 \\ * & \theta(2\tau - \theta) & -\theta(\tau - 1) & 0 & 0 \\ * & * & \tau - 1 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix} \\
+ & \begin{bmatrix} 0 & 0 & \frac{1-\theta}{2} & 0 & 0 \\ * & 0 & \frac{\theta}{2} & 0 & 0 \\ * & * & -1 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & 0 \end{bmatrix} 2(1 - \tau) \\
= & \begin{bmatrix} -1 & \frac{1+\tau}{2} & -\frac{\theta}{2} & \frac{\theta}{2} & \frac{1}{2} \\ * & -\tau & \frac{\theta}{2} & -\frac{\theta}{2} & -\frac{1}{2} \\ * & * & -1 & \frac{1+\tau}{2} & \frac{1}{2} \\ * & * & * & -\tau & -\frac{1}{2} \\ * & * & * & * & 0 \end{bmatrix} \theta,
\end{aligned}$$

which shows equality **(LE)** of Proposition 4.1.

Acknowledgments

M. Nilsson, A. Åkerman and P. Giselsson acknowledge support from the ELLIIT Strategic Research Area and the Wallenberg AI, Autonomous Systems, and Software Program (WASP), funded by the Knut and Alice Wallenberg Foundation. Additionally, P. Giselsson acknowledges support from the Swedish Research Council.

References

- [1] J. Douglas and H. H. Rachford. “On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables”. In: *Transactions of the American Mathematical Society* 82.2 (July 1956), pp. 421–439. DOI: 10.1090/S0002-9947-1956-0084194-4.
- [2] P.-L. Lions and B. Mercier. “Splitting algorithms for the sum of two nonlinear operators”. In: *SIAM Journal on Numerical Analysis* 16.6 (1979), pp. 964–979.
- [3] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. 2nd ed. CMS Books in Mathematics. Cham: Springer, 2017. DOI: 10.1007/978-3-319-48311-5.
- [4] J. Eckstein. “Splitting Methods for Monotone Operators with Applications to Parallel Optimization”. MIT LIDS-TH-1877, Department of Civil Engineering. PhD thesis. Cambridge, MA: Massachusetts Institute of Technology, June 1989. URL: <https://dspace.mit.edu/handle/1721.1/14356>.
- [5] J. Eckstein and D. P. Bertsekas. “On the Douglas–Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators”. In: *Mathematical Programming* 55 (Apr. 1992), pp. 293–318. DOI: 10.1007/BF01581204.
- [6] B. Martinet. “Brève communication. Régularisation d’inéquations variationnelles par approximations successives”. In: *Revue française d’informatique et de recherche opérationnelle. Série rouge* 4.R3 (1970), pp. 154–158. DOI: 10.1051/m2an/197004R301541.
- [7] R. T. Rockafellar. “Monotone Operators and the Proximal Point Algorithm”. In: *SIAM Journal on Control and Optimization* 14.5 (Aug. 1976), pp. 877–898. DOI: 10.1137/0314056.
- [8] J. M. Borwein and B. Sims. “The Douglas–Rachford Algorithm in the Absence of Convexity”. In: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Vol. 49. Springer Optimization and Its Applications. New York, NY: Springer, 2011, pp. 93–109. DOI: 10.1007/978-1-4419-9569-8_6.
- [9] R. Hesse and D. R. Luke. “Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems”. In: *SIAM Journal on Optimization* 23.4 (2013), pp. 2397–2419. DOI: 10.1137/120902653.
- [10] H. H. Bauschke, J. Y. B. Cruz, T. T. A. Nghia, H. M. Phan, and X. Wang. “The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle”. In: *Journal of Approximation Theory* 185 (2014), pp. 63–79.
- [11] R. Hesse, D. R. Luke, and P. Neumann. “Alternating Projections and Douglas–Rachford for Sparse Affine Feasibility”. In: *IEEE Transactions on Signal Processing* 62.18 (2014), pp. 4868–4881. DOI: 10.1109/TSP.2014.2339801.
- [12] D. Boley. “Local linear convergence of the alternating direction method of multipliers on quadratic or linear programs”. In: *SIAM Journal on Optimization* 23.4 (2013), pp. 2183–2207.
- [13] L. Demanet and X. Zhang. “Eventual linear convergence of the Douglas–Rachford iteration for basis pursuit”. In: *Mathematics of Computation* 85.297 (2016), pp. 209–238. DOI: 10.1090/mcom/2965.
- [14] J. Liang, J. Fadili, and G. Peyré. “Local Convergence Properties of Douglas–Rachford and ADMM”. In: *Journal of Optimization Theory and Applications* 172.3 (2017), pp. 874–913. DOI: 10.1007/s10957-017-1061-z.

- [15] P. Giselsson and S. Boyd. “Linear Convergence and Metric Selection for Douglas–Rachford Splitting and ADMM”. In: *IEEE Transactions on Automatic Control* 62.2 (2017), pp. 532–544. DOI: 10.1109/TAC.2016.2564160.
- [16] P. Giselsson. “Tight global linear convergence rate bounds for Douglas–Rachford splitting”. In: *Journal of Fixed Point Theory and Applications* 19.4 (2017), pp. 2241–2270. DOI: 10.1007/s11784-017-0417-1.
- [17] W. M. Moursi and L. Vandenbergh. “Douglas–Rachford Splitting for the Sum of a Lipschitz Continuous and a Strongly Monotone Operator”. In: *Journal of Optimization Theory and Applications* 183.1 (2019), pp. 179–198. DOI: 10.1007/s10957-019-01517-8.
- [18] E. K. Ryu, A. B. Taylor, C. Bergeling, and P. Giselsson. “Operator Splitting Performance Estimation: Tight Contraction Factors and Optimal Parameter Selection”. In: *SIAM Journal on Optimization* 30.3 (2020), pp. 2251–2271. DOI: 10.1137/19M1304854.
- [19] G. Li and T. K. Pong. “Douglas–Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems”. In: *Mathematical Programming* 159.1–2 (2016), pp. 371–401. DOI: 10.1007/s10107-015-0963-5.
- [20] A. Themelis and P. Patrinos. “Douglas–Rachford splitting and ADMM for nonconvex optimization: Tight convergence results”. In: *SIAM Journal on Optimization* 30.1 (2020), pp. 149–181. DOI: 10.1137/18M1163993.
- [21] E. K. Ryu. “Uniqueness of DRS as the 2 operator resolvent-splitting and impossibility of 3 operator resolvent-splitting”. In: *Mathematical Programming* 182.1 (2020), pp. 233–273.
- [22] Y. Malitsky and M. K. Tam. “Resolvent splitting for sums of monotone operators with minimal lifting”. In: *Mathematical Programming* 201.1 (2023), pp. 231–262. DOI: 10.1007/s10107-022-01906-4.
- [23] F. J. A. Artacho, R. I. Boţ, and D. Torregrosa-Belén. “A primal-dual splitting algorithm for composite monotone inclusions with minimal lifting”. In: *Numerical Algorithms* 93.1 (2023), pp. 103–130. DOI: 10.1007/s11075-022-01405-9.
- [24] M. Morin, S. Banert, and P. Giselsson. “Frugal Splitting Operators: Representation, Minimal Lifting, and Convergence”. In: *SIAM Journal on Optimization* 34.2 (2024), pp. 1595–1621. DOI: 10.1137/22M1531105.
- [25] A. Åkerman, E. Chenchene, P. Giselsson, and E. Naldi. “Splitting the Forward–Backward Algorithm: A Full Characterization”. In: *arXiv* (2025). DOI: 10.48550/arXiv.2504.10999. arXiv: 2504.10999 [math.OA]. URL: <https://arxiv.org/abs/2504.10999>.
- [26] M. Upadhyaya, S. Banert, A. B. Taylor, and P. Giselsson. “Automated tight Lyapunov analysis for first-order methods”. In: *Mathematical Programming* 209 (2025), pp. 133–170. DOI: 10.1007/s10107-024-02061-8.
- [27] M. Upadhyaya, A. B. Taylor, S. Banert, and P. Giselsson. *AutoLyap: A Python package for computer-assisted Lyapunov analyses for first-order methods*. 2025. DOI: 10.48550/arXiv.2506.24076. arXiv: 2506.24076 [math.OA].
- [28] R. Glowinski and A. Marroco. “Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation–dualité, d’une classe de problèmes de Dirichlet non linéaires”. In: *RAIRO. Analyse numérique* 9.R2 (1975), pp. 41–76. DOI: 10.1051/m2an/197509r200411.
- [29] D. Gabay and B. Mercier. “A dual algorithm for the solution of nonlinear variational problems via finite element approximation”. In: *Computers & Mathematics with Applications* 2 (1976), pp. 17–40. DOI: 10.1016/0898-1221(76)90003-1.

- [30] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. “Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers”. In: *Foundations and Trends in Machine Learning* 3.1 (2011), pp. 1–122. DOI: 10.1561/22000000016.
- [31] D. Gabay. “Applications of the Method of Multipliers to Variational Inequalities”. In: *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*. Ed. by R. Glowinski and A. Marroco. Amsterdam: North-Holland, 1983, pp. 299–331.
- [32] A. Chambolle and T. Pock. “A first-order primal-dual algorithm for convex problems with applications to imaging”. In: *Journal of mathematical imaging and vision* 40.1 (2011), pp. 120–145.
- [33] D. O’Connor and L. Vandenberghe. “On the equivalence of the primal-dual hybrid gradient method and Douglas–Rachford splitting”. In: *Mathematical Programming* 179.1 (2020), pp. 85–108.
- [34] B. He and X. Yuan. “Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective”. In: *SIAM Journal on Imaging Sciences* 5.1 (2012), pp. 119–149.
- [35] S. Banert, M. Upadhyaya, and P. Giselsson. “The Chambolle–Pock method converges weakly with $\theta > 1/2$ and $\tau\sigma\|L\|^2 < 4/(1+2\theta)$ ”. In: *Optimization Letters* (2025), pp. 1–18.
- [36] R. T. Rockafellar. *Convex analysis*. Vol. 18. Princeton, New Jersey, USA: Princeton university press, 1970.
- [37] M. Nilsson, A. Åkerman, and P. Giselsson. *Symbolic verification of Extending Douglas–Rachford Splitting for Convex Optimization*. 2025. URL: https://github.com/MaxNilsson-phd/Symbolic_verification_of_Extending_Douglas--Rachford_Splitting_for_Convex_Optimization.git (visited on 11/23/2025).
- [38] E. K. Ryu and W. Yin. *Large-Scale Convex Optimization: Algorithms & Analyses via Monotone Operators*. Cambridge University Press, 2022.
- [39] L. Chen, D. Sun, and K.-C. Toh. “A note on the convergence of ADMM for linearly constrained convex optimization problems”. In: *Computational Optimization and Applications* 66.2 (Aug. 2016), pp. 327–343. DOI: 10.1007/s10589-016-9864-7. URL: <http://dx.doi.org/10.1007/s10589-016-9864-7>.
- [40] Z. Li and M. Yan. “New convergence analysis of a primal-dual algorithm with large step-sizes”. In: *Advances in Computational Mathematics* 47.1 (2021), p. 9.
- [41] Y. Ma, X. Cai, B. Jiang, and D. Han. “Understanding the convergence of the preconditioned PDHG method: a view of indefinite proximal ADMM”. In: *Journal of Scientific Computing* 94.3 (2023), p. 60.
- [42] M. Yan and Y. Li. “On the improved conditions for some primal-dual algorithms”. In: *Journal of scientific computing* 99.3 (2024), p. 74.
- [43] L. Condat. “A Primal–Dual Splitting Method for Convex Optimization Involving Lipschitzian, Proximable and Linear Composite Terms”. In: *Journal of Optimization Theory and Applications* 158.2 (2013), pp. 460–479. DOI: 10.1007/s10957-012-0245-9. URL: <https://doi.org/10.1007/s10957-012-0245-9>.
- [44] M. N. Dao and H. M. Phan. “Adaptive Douglas–Rachford Splitting Algorithm for the Sum of Two Operators”. In: *SIAM Journal on Optimization* 29.4 (2019), pp. 2697–2724. DOI: 10.1137/18M121160X.

- [45] P. Giselsson and W. M. Moursi. “On compositions of special cases of Lipschitz continuous operators”. In: *Fixed Point Theory and Algorithms for Sciences and Engineering* 2021.1 (2021), p. 25. DOI: 10.1186/s13663-021-00709-0.
- [46] S. Bartz, M. N. Dao, and H. M. Phan. “Conical averagedness and convergence analysis of fixed point algorithms”. In: *Journal of Global Optimization* 82.2 (2022), pp. 351–373. DOI: 10.1007/s10898-021-01057-4.
- [47] D. Davis and W. Yin. “A Three-Operator Splitting Scheme and Its Optimization Applications”. In: *Set-Valued and Variational Analysis* 25.4 (2017), pp. 829–858. DOI: 10.1007/s11228-017-0421-z.
- [48] E. K. Ryu, R. Hannah, and W. Yin. “Scaled relative graphs: nonexpansive operators via 2D Euclidean geometry”. In: *Mathematical Programming* 194 (2022), pp. 569–619. DOI: 10.1007/s10107-021-01639-w.