ASYMMETRIC FORWARD-BACKWARD-ADJOINT SPLITTING FOR SOLVING MONOTONE INCLUSIONS INVOLVING THREE OPERATORS

PUYA LATAFAT AND PANAGIOTIS PATRINOS

Abstract. In this work we propose a new splitting technique, namely Asymmetric Forward-Backward-Adjoint splitting, for solving monotone inclusions involving three terms, a maximally monotone, a cocoercive and a bounded linear operator. Classical operator splitting methods, like Douglas-Rachford and Forward-Backward splitting are special cases of our new algorithm. Asymmetric Forward-Backward-Adjoint splitting unifies, extends and sheds light on the connections between many seemingly unrelated primal-dual algorithms for solving structured convex optimization problems proposed in recent years. More importantly, it greatly extends the scope and applicability of splitting techniques to a wider variety of problems. One important special case leads to a Douglas-Rachford type scheme that includes a third cocoercive operator.

Keywords. convex optimization, monotone inclusion, operator splitting, primal-dual algorithms

1. Introduction

This paper considers two types of general problems. The focus of the first part of the paper is on solving monotone inclusion problems of the form

$$0 \in Ax + Mx + Cx, \tag{1}$$

where A is a maximally monotone operator, M is a bounded linear operator and C is cocoercive[1]. The most well known algorithms for solving monotone inclusion problems are Forward-Backward splitting (FBS), Douglas-Rachford splitting (DRS) and Forward-Backward-Forward splitting (FBFS) [4, 10, 28–30, 35]. The operator splitting schemes FBS and DRS are not well equipped to handle (1) since they are designed for monotone inclusions involving the sum of two operators. The FBFS splitting can solve (1) by considering $M + C$ as one Lipschitz continuous operator. However, being blind to the fact that C is cocoercive, it would require two evaluations of C per iteration. Many other variations of the three main splittings have been proposed over time that can be seen as intelligent applications of these classical methods (see for example [5, 6, 8, 14, 36]).

(P. Latafat) IMT School for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca, Italy
(P. Patrinos) Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven-Heverlee, Belgium
E-mail addresses: puya.latafat@imtlucca.it, panos.patrinos@esat.kuleuven.be.

[1] C is β-cocoercive with respect to the P norm if for some $\beta \in]0, +\infty[$ the following holds

$$(\forall z \in \mathcal{H})(\forall z' \in \mathcal{H}) \quad \langle Cz - Cz', z - z' \rangle \geq \beta \|Cz - Cz'\|_{P^{-1}}^2.$$

The main contribution of the paper is a new algorithm called *Asymmetric-Forward-Backward-Adjoint splitting* (AFBA) to solve the monotone inclusion (1), without resorting to any kind of reformulation of the problem. One important property of AFBA is that it includes asymmetric preconditioning. This gives great flexibility to the algorithm, and indeed it is the key for recovering and unifying existing primal-dual proximal splitting schemes for convex optimization and devising new ones. More importantly, it can deal with problems involving 3 operators, one of which is cocoercive. It is observed that FBS, DRS, the Proximal Point Algorithm (PPA) can be derived as special cases of our method. Another notable special case is the method proposed by Solodov and Tseng for variational inequalities in [33, Algorithm 2.1]. Moreover, when the cocoercive term, C, is absent in (1), in a further special case, it coincides with the FBFS when its Lipschitz operator is skew-adjoint. Recently a new splitting scheme was proposed in [18] for solving monotone inclusions involving the sum of three operators, one of which is cocoercive. This method can be seen as Douglas-Rachford splitting with an extra forward step for the cocoercive operator and at this point it seems that it can not be derived by manipulating one of the main three splitting algorithms. As a special case of our scheme, we propose an algorithm that also bares heavy resemblance to the classic Douglas-Rachford splitting with an extra forward step (see Algorithm 2). The proposed algorithm is different than the one of [18], in that the forward step precedes the two backward updates.

As another contribution of the paper, big-$O(1/(n+1))$ and little-$o(1/(n+1))$ convergence rates are derived for AFBA (see Theorem 3.2). It is observed that in many cases these convergence rates are guaranteed under mild conditions. In addition, under metric subregularity of the underlying operator, linear convergence is guaranteed without restrictions on the parameters (see Theorem 3.3). Given that AFBA generalizes a wide range of algorithms, this analysis provides a systematic way to deduce convergence rates for many algorithms.

The focus of the second half of the paper, in the simpler form, is on solving convex optimization problems of the form

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

where $\mathcal{H}$ is a real Hilbert space and L is a bounded linear operator. The functions f, g and h are proper, closed convex functions and in addition h is Lipschitz differentiable. The equivalent monotone inclusion problem takes the form of finding $x\in\mathcal{H}$ such that

$$0\in Ax+L^*BLx+Cx,$$

where $A=\partial f$, $B=\partial g$ are maximally monotone operators defined on $\mathcal{H}$ and the operator $C=\nabla h$ is cocoercive. Similar to the problem (1), the classical methods FBS, DRS, FBFS are not suitable for solving problems of the form (2) (without any reformulation) because they all deal with problems involving two operators. Furthermore, these methods usually require calculation of the proximal mapping of the composition of a function with a linear operator which is not trivial in general or requires matrix inversion (see [30] for a survey on proximal algorithms). In the recent years in order to solve problem (2), with or without the cocoercive term, many authors have considered the corresponding saddle point problem. This approach yields the primal and dual solutions simultaneously (hence the name primal-dual splittings) and eliminates the need to calculate the proximal mapping of a linearly

composed function. The resulting algorithms only require matrix vector products, gradient and proximal updates (see [3,8,11,14,36] for more discussion). We follow the same approach and notice that it is quite natural to embed the optimality condition of the saddle point problem associated to (2) in the form of the monotone inclusion (1). Subsequently, by appealing to AFBA, we can generate new algorithms and recover many existing methods, such as the ones proposed in [6,8,14,21,24,36], as special cases. In many of the cases, we extend the range of acceptable stepsizes and relaxation parameters under which the methods are convergent. Additionally, the convergence rates for these methods are implied by our results for AFBA.

The paper is organized as follows. Section 2 is devoted to introducing notation and reviewing basic definitions. In Section 3, we present and analyze the convergence and rate of convergence of AFBA. Its relation to classical splitting methods is discussed in Section 4. In Section 5, we consider the saddle point problem associated to a generalization of (2). By applying AFBA and properly choosing the parameters we are able to generate a large class of algorithms. We then consider some important special cases and discuss their relation to existing methods. These connections are summarized in the form of a diagram in Figure 1.

2. Backround and Preliminary Results

In this section we recap the basic definitions and results that will be needed subsequently (see [1] for detailed discussion).

Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces. We denote the scalar product and the induced norm of a Hilbert space by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$ respectively. Id denotes the identity operator. We denote by $\mathscr{B}(\mathcal{H},\mathcal{G})$ the space of bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$ and set $\mathscr{B}(\mathcal{H})=\mathscr{B}(\mathcal{H},\mathcal{H})$. The space of self-adjoint operators is denoted by $\mathcal{S}(\mathcal{H})=\{L\in\mathscr{B}(\mathcal{H})|L=L^*\}$, where L^* denotes the adjoint of L. The *Loewner partial ordering* on $\mathcal{S}(\mathcal{H})$ is denoted by $\succeq$. Let $\tau\in]0,+\infty[$ and define the space of τ-strongly positive self-adjoint operators by $\mathcal{S}_\tau(\mathcal{H})=\{U\in\mathcal{S}(\mathcal{H})|U\succeq\tau\,\mathrm{Id}\}$. For $U\in\mathcal{S}_\tau(\mathcal{H})$, define the scalar product and norm by $\langle x,y\rangle_U=\langle x,Uy\rangle$, and $\|x\|_U=\sqrt{\langle x,Ux\rangle}$. We also define the Hilbert space $\mathcal{H}_U$ by endowing $\mathcal{H}$ with the scalar product $\langle x,y\rangle_U$. One has $|\langle x,y\rangle|\leq\|x\|_U\|y\|_{U^{-1}}$ and $\|x\|_U=\sqrt{\langle x,Ux\rangle}$. The operator norm induced by $\|\cdot\|_U$ is $\|L\|_U=\sup_{x\in\mathcal{H},\|x\|_U=1}\|Lx\|_U=\|U^{1/2}LU^{-1/2}\|$.

Let $A:\mathcal{H}\to 2^{\mathcal{H}}$ be a set-valued operator. The domain of A is denoted by $\mathrm{dom}(A)=\{x\in\mathcal{H}|Ax\neq\emptyset\}$, its graph by $\mathrm{gra}(A)=\{(x,u)\in\mathcal{H}\times\mathcal{H}|u\in Ax\}$ and the set of zeros of A is $\mathrm{zer}(A)=\{x\in\mathcal{H}|0\in Ax\}$. The inverse of A is defined through its graph: $\mathrm{gra}(A^{-1})=\{(u,x)\in\mathcal{H}\times\mathcal{H}|(x,u)\in\mathrm{gra}(A)\}$. The *resolvent* of A is given by $J_A=(\mathrm{Id}+A)^{-1}$. Furthermore, A is *monotone* if $\langle x-y,u-v\rangle\geq 0$ for all $(x,u),(y,v)\in\mathrm{gra}(A)$, and *maximally monotone* if it is monotone and there exists no monotone operator $B:\mathcal{H}\to 2^{\mathcal{H}}$ such that $\mathrm{gra}(A)\subset\mathrm{gra}(B)$ and $A\neq B$.

The set of proper lower semicontinuous convex functions from $\mathcal{H}$ to $]-\infty,+\infty]$ is denoted by $\Gamma_0(\mathcal{H})$. The *Fenchel conjugate* of $f\in\Gamma_0(\mathcal{H})$, denoted $f^*\in\Gamma_0(\mathcal{H})$, is defined by $f^*:\mathcal{H}\to]-\infty,+\infty]:u\mapsto\sup_{z\in\mathcal{H}}(\langle u,z\rangle-f(z))$. The *Fenchel-Young inequality*, $\langle x,u\rangle\leq f(x)+f^*(u)$ for all $x,u\in\mathcal{H}$, holds for $f:\mathcal{H}\to]-\infty,+\infty]$ proper. Throughout this paper we make extensive use of this inequality in the special case when $f=\frac{1}{2}\|\cdot\|_U^2$ for U strongly positive. The *infimal convolution* of $f,g:\mathcal{H}\to]-\infty,+\infty]$ is denoted by $f\,\square\,g:\mathcal{H}\to]\infty,+\infty]:x\mapsto\inf_{y\in\mathcal{H}}(f(y)+g(x-y))$. If $f\in\Gamma_0(\mathcal{H})$ then the *subdifferential* of f, denoted by ∂f, is the maximally monotone operator $\partial f:\mathcal{H}\to 2^{\mathcal{H}}:x\mapsto\{u\in\mathcal{H}|(\forall y\in\mathcal{H})\,\langle y-x,u\rangle+f(x)\leq f(y)\}$,

with inverse $(\partial f)^{-1} = \partial f^*$. The resolvent of ∂f is called *proximal operator* and is uniquely determined by $\operatorname{prox}_f(x) \coloneqq J_{\partial f}(x) = \operatorname{argmin}_{z\in\mathcal{H}} f(z) + \frac{1}{2}\|x - z\|^2$.

Let X be a nonempty closed convex set in $\mathcal{H}$. The *indicator function* of X, denoted $\iota_X : \mathcal{H} \to]-\infty, +\infty]$, is defined by $\iota_X(x) = 0$, if $x \in X$ and $\iota_X(x) = +\infty$, otherwise. The *normal cone* of X is the maximally monotone operator $N_X \coloneqq \partial\iota_X$. The distance to X with respect to $\|\cdot\|_U$ is denoted by $d_U(z, X) = \inf_{z^\star\in X} \|z - z^\star\|_U$, the projection of z onto X with respect to $\|\cdot\|_U$ is denoted by $\Pi_X^U(z)$, and the absence of superscript implies the same definitions with respect to the canonical norm.

3. Asymmetric Forward-Backward-Adjoint Method

Let $\mathcal{H}$ be a real Hilbert space and consider the problem of finding $z \in \mathcal{H}$ such that

$$0 \in Tz \quad \text{where} \quad T \coloneqq A + M + C, \tag{3}$$

where operators A, M, C satisfy the following assumption:

Assumption 3.1. Throughout the paper the following hold:

- (i) Operator $A : \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone and $M \in \mathscr{B}(\mathcal{H})$ is monotone.
- (ii) Operator $C : \mathcal{H} \to \mathcal{H}$ is β-cocoercive with respect to $\|\cdot\|_P$, where $\beta \in]1/4, +\infty[$ and $P \in \mathcal{S}_\rho(\mathcal{H})$ for some $\rho \in]0, \infty[$, *i.e.*

$$(\forall z \in \mathcal{H})(\forall z' \in \mathcal{H}) \quad \langle Cz - Cz', z - z'\rangle \geq \beta\|Cz - Cz'\|_{P^{-1}}^2.$$

It is important to notice that the freedom in choosing P is a crucial part of our method. In Assumption 3.1(ii) we consider cocoercivity with respect to $\|\cdot\|_P$ with $\beta \in]1/4, +\infty[$. However, this is by no means a restriction of our setting; another approach would have been to consider cocoercivity with respect to the canonical norm $\|\cdot\|$ with $\beta \in]0, +\infty[$ but this would lead to statements involving $\|P\|$ and $\|P^{-1}\|$. Indeed convergence with respect to $\|\cdot\|$ and $\|\cdot\|_P$ are equivalent but in using $\|\cdot\|_P$ we simplify the notation substantially.

In addition, let S be a strongly positive, self-adjoint operator, $K \in \mathscr{B}(\mathcal{H})$ a skew-adjoint operator, *i.e.*, $K^* = -K$ and $H = P + K$. Then, the algorithm for solving the monotone inclusion described above is as follows:

Algorithm 1 Asymmetric Forward-Backward-Adjoint Splitting (AFBA)

Inputs: $z_0 \in \mathcal{H}$
for $n = 0, \ldots,$ **do**
$\bar{z}_n = (H + A)^{-1}(H - M - C)z_n$
$\tilde{z}_n = \bar{z}_n - z_n$
$\alpha_n = \dfrac{\lambda_n\|\tilde{z}_n\|_P^2}{\|(H + M^*)\,\tilde{z}_n\|_{S^{-1}}^2}$
$z_{n+1} = z_n + \alpha_n S^{-1}(H + M^*)\tilde{z}_n$

Before proceeding with the convergence analysis let us define

$$D = (H + M^*)^* S^{-1}(H + M^*). \tag{4}$$

Since $P \in \mathcal{S}_\rho(\mathcal{H})$ for some $\rho \in]0, \infty[$, K is skew-adjoint, and $M \in \mathscr{B}(\mathcal{H})$ is monotone, it follows that $\langle (H + M^*)z, z\rangle \geq \rho\|z\|^2$ for all $z \in \mathcal{H}$, and we have

$$(\forall z \in \mathcal{H}) \quad \langle z, Dz\rangle = \|(H + M^*)z\|_{S^{-1}}^2 \geq \rho^2\|S\|^{-1}\|z\|^2. \tag{5}$$

Hence, $D \in \mathcal{S}_\nu(\mathcal{H})$ with $\nu = \rho^2\|S\|^{-1}$. Notice that the denominator of α_n in Algorithm 1 is equal to the left hand side of (5) for $z = \tilde{z}_n$ and thus it is bounded below by $\rho^2\|S\|^{-1}\|\tilde{z}_n\|^2$.

3.1. **Convergence Analysis.** In this section we analyze convergence and rate of convergence of Algorithm 1. We also consider a special case of the algorithm in which it is possible to relax strong positivity of P to positivity. We begin by stating our main convergence result. The proof relies on showing that the sequence $(z_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\operatorname{zer}(A+M+C)$ in the Hilbert space $\mathcal{H}_S$.

Theorem 3.1. *Consider Algorithm 1 under Assumption 3.1 and assume* $\operatorname{zer}(T) \neq \emptyset$ *where* $T = A + M + C$. *Let* $\sigma \in]0,\infty[$, $S \in \mathcal{S}_\sigma(\mathcal{H})$, $K \in \mathscr{B}(\mathcal{H})$ *a skew-adjoint operator, and* $H = P + K$. *Let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence such that*

$$(\lambda_n)_{n\in\mathbb{N}} \subseteq [0,\delta] \quad \text{with} \quad \delta = 2 - \frac{1}{2\beta}, \quad \delta > 0, \quad \liminf_{n\to\infty} \lambda_n(\delta - \lambda_n) > 0. \tag{6}$$

Then the following hold:

(i) $(z_n)_{n\in\mathbb{N}}$ *is Fejér monotone with respect to* $\operatorname{zer}(T)$ *in the Hilbert space* $\mathcal{H}_S$.
(ii) $(\tilde{z}_n)_{n\in\mathbb{N}}$ *converges strongly to zero.*
(iii) $(z_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\operatorname{zer}(T)$.

Furthermore, when $C \equiv 0$ *all of the above statements hold with* $\delta = 2$.

Proof. The operators $\tilde{A} = P^{-1}(A+K)$ and $\tilde{B} = P^{-1}(M+C-K)$ are monotone in the Hilbert space $\mathcal{H}_P$. We observe that

$$\bar{z}_n = (H+A)^{-1}(H-M-C)z_n = (\mathrm{Id}+\tilde{A})^{-1}(\mathrm{Id}-\tilde{B})z_n.$$

Therefore $z_n - \tilde{B}z_n \in \bar{z}_n + \tilde{A}\bar{z}_n$, or $-\tilde{z}_n - \tilde{B}z_n \in \tilde{A}\bar{z}_n$. Since $-\tilde{B}z^\star \in \tilde{A}z^\star$ for $z^\star \in \operatorname{zer}(T)$ by monotonicity of $\tilde{A}$ on $\mathcal{H}_P$ we have

$$\langle \tilde{B}z_n - \tilde{B}z^\star + \tilde{z}_n, z^\star - \bar{z}_n\rangle_P \geq 0.$$

Then,

$$\begin{aligned} 0 &\leq \langle \tilde{B}z_n - \tilde{B}z^\star + \tilde{z}_n, z^\star - \bar{z}_n\rangle_P \\ &= \langle P^{-1}(M+C-K)z_n - P^{-1}(M+C-K)z^\star + \tilde{z}_n, z^\star - \bar{z}_n\rangle_P. \\ &= \langle (M-K)(z_n - z^\star) + Cz_n - Cz^\star + P\tilde{z}_n, z^\star - \bar{z}_n\rangle. \end{aligned} \tag{7}$$

On the other hand

$$\begin{aligned} \langle Cz_n - Cz^\star, z^\star - \bar{z}_n\rangle &= \langle Cz_n - Cz^\star, z_n - \bar{z}_n\rangle + \langle Cz_n - Cz^\star, z^\star - z_n\rangle \\ &\leq \frac{\epsilon}{2}\|\tilde{z}_n\|_P^2 + \frac{1}{2\epsilon}\|Cz_n - Cz^\star\|_{P^{-1}}^2 + \langle Cz_n - Cz^\star, z^\star - z_n\rangle \\ &\leq \frac{\epsilon}{2}\|\tilde{z}_n\|_P^2 + \left(1 - \frac{1}{2\epsilon\beta}\right)\langle Cz_n - Cz^\star, z^\star - z_n\rangle. \end{aligned}$$

The first inequality follows from Fenchel-Young inequality for $\frac{\epsilon}{2}\|\cdot\|_P^2$, while the second from β-cocoercivity of C with respect to $\|\cdot\|_P$. Set $\epsilon := \frac{1}{2\beta}$ so that

$$\langle Cz_n - Cz^\star, z^\star - \bar{z}_n\rangle \leq \frac{1}{4\beta}\|\tilde{z}_n\|_P^2. \tag{8}$$

In turn, (7), (8) and monotonicity of $M-K$, yield

$$\begin{aligned}
0 &\leq \langle (M-K)(z_n - z^\star) + Cz_n - Cz^\star + P\tilde{z}_n, z^\star - \bar{z}_n\rangle \\
&\leq \langle (M-K)(z_n - z^\star), z^\star - z_n\rangle + \langle (M-K)(z_n - z^\star), z_n - \bar{z}_n\rangle \\
&+ \frac{1}{4\beta}\|\tilde{z}_n\|_P^2 + \langle P\tilde{z}_n, z^\star - z_n\rangle + \langle P\tilde{z}_n, z_n - \bar{z}_n\rangle \\
&\leq \langle z_n - z^\star, (M^* + K)(z_n - \bar{z}_n)\rangle + \frac{1}{4\beta}\|\tilde{z}_n\|_P^2 + \langle P\tilde{z}_n, z^\star - z_n\rangle - \|\tilde{z}_n\|_P^2 \\
&= \langle z_n - z^\star, -(M^* + H)\tilde{z}_n\rangle - \left(1 - \frac{1}{4\beta}\right)\|\tilde{z}_n\|_P^2,
\end{aligned}$$

or equivalently

$$\langle z_n - z^\star, (M^* + H)\tilde{z}_n\rangle \leq -\left(1 - \frac{1}{4\beta}\right)\|\tilde{z}_n\|_P^2. \tag{9}$$

For notational convenience define $\delta := 2 - \frac{1}{2\beta}$. We show that $\|z_n - z^\star\|_S^2$ is decreasing using (9) together with step 3 and 4 of Algorithm 1:

$$\begin{aligned}
\|z_{n+1} - z^\star\|_S^2 &= \|z_n - z^\star + \alpha_n S^{-1}(H + M^*)\tilde{z}_n\|_S^2 \\
&= \|z_n - z^\star\|_S^2 + 2\alpha_n\langle z_n - z^\star, (H + M^*)\tilde{z}_n\rangle + \alpha_n^2\|(H + M^*)\tilde{z}_n\|_{S^{-1}}^2 \\
&\leq \|z_n - z^\star\|_S^2 - \alpha_n\delta\|\tilde{z}_n\|_P^2 + \alpha_n^2\|(H + M^*)\tilde{z}_n\|_{S^{-1}}^2 \\
&= \|z_n - z^\star\|_S^2 - \delta\lambda_n\frac{\|\tilde{z}_n\|_P^4}{\|(H + M^*)\tilde{z}_n\|_{S^{-1}}^2} + \lambda_n^2\frac{\|\tilde{z}_n\|_P^4}{\|(H + M^*)\tilde{z}_n\|_{S^{-1}}^2} \\
&= \|z_n - z^\star\|_S^2 - \lambda_n(\delta - \lambda_n)\|\left(H + M^*\right)\tilde{z}_n\|_{S^{-1}}^{-2}\|\tilde{z}_n\|_P^4 \qquad (10) \\
&= \|z_n - z^\star\|_S^2 - \lambda_n(\delta - \lambda_n)\|P^{-1/2}S^{-1/2}\left(H + M^*\right)\tilde{z}_n\|_P^{-2}\|\tilde{z}_n\|_P^4 \\
&\leq \|z_n - z^\star\|_S^2 - \lambda_n(\delta - \lambda_n)\|P^{-1/2}S^{-1/2}\left(H + M^*\right)\|_P^{-2}\|\tilde{z}_n\|_P^2 \\
&= \|z_n - z^\star\|_S^2 - \lambda_n(\delta - \lambda_n)\|S^{-1/2}\left(H + M^*\right)P^{-1/2}\|^{-2}\|\tilde{z}_n\|_P^2. \qquad (11)
\end{aligned}$$

Furthermore, when $C \equiv 0$ all the above analysis holds with $\delta = 2$.

(*i*): Inequality (11) and $(\lambda_n)_{n\in\mathbb{N}} \subseteq [0, \delta]$ show that $(z_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\operatorname{zer}(T)$ in the Hilbert space $\mathcal{H}_S$.

(*ii*): From (11) and $\liminf_{n\to\infty}\lambda_n(\delta - \lambda_n) > 0$, it follows that $\tilde{z}_n \to 0$.

(*iii*): Define

$$w_n := -(H - M)\tilde{z}_n + C\bar{z}_n - Cz_n. \tag{12}$$

It follows from (12), linearity of $H - M$, cocoercivity of C and (*ii*) that

$$w_n \to 0. \tag{13}$$

By step 1 of Algorithm 1 we have $(H - M - C)z_n \in (H + A)\bar{z}_n$, which together with (12) yields

$$w_n \in T\bar{z}_n. \tag{14}$$

Now let z be a weak sequential cluster point of $(z_n)_{n\in\mathbb{N}}$, say $z_{k_n} \rightharpoonup z$. It follows from (*ii*) that $\bar{z}_{k_n} \rightharpoonup z$, and from (13) that $w_{k_n} \to 0$. Altogether, by (14), the members of the sequence $(\bar{z}_{k_n}, w_{k_n})_{n\in\mathbb{N}}$ belong to $\operatorname{gra}(T)$. Additionally, by [1, Example 20.28, 20.29 and Corollary 24.4(i)], T is maximally monotone. Then, an appeal to [1, Proposition 20.33(ii)] yields $(z, 0) \in \operatorname{gra}(T)$. This together with (*i*) and [1, Theorem 5.5] completes the proof. □

Equation (11) implies that the sequence $(\min_{i=1\dots n}\|\tilde{z}_i\|_P^2)_{n\in\mathbb{N}}$, the cumulative minimum of $(\|\tilde{z}_n\|_P^2)_{n\in\mathbb{N}}$, converges sublinearly. Our next goal is to derive big-$O(1/(n+1))$ and little-$o(1/(n+1))$ convergence rates for the sequence itself. This is established below, under further restrictions on $(\lambda_n)_{n\in\mathbb{N}}$, by showing that the sequence $\left(\|\tilde{z}_n\|_D^2\right)_{n\in\mathbb{N}}$ is monotonically nonincreasing and summable.

Theorem 3.2 (Convergence rates)**.** *Consider Algorithm 1 under the assumptions of Theorem 3.1. Let c_1 and c_2 be two positive constants satisfying*

$$c_1 P \preceq D \preceq c_2 P, \tag{15}$$

with D defined in (4). Assume

$$(\lambda_n)_{n\in\mathbb{N}} \subseteq \left]0, c_1\delta/c_2\right], \tag{16}$$

where δ is defined in (6). Suppose that $\tau \coloneqq \inf_{n\in\mathbb{N}} \lambda_n(\delta-\lambda_n) > 0$. Then the following convergence estimates hold:

$$\|\tilde{z}_n\|_D^2 \leq \frac{c_2^2}{\tau(n+1)}\|z_0 - z^\star\|_S^2 \quad \textit{and} \quad \|\tilde{z}_n\|_D^2 = o(1/(n+1)).$$

Proof. Using the monotonicity of A and Step 1 of Algorithm 1

$$0 \leq \langle (H-M)(z_n - z_{n+1}) - H(\bar{z}_n - \bar{z}_{n+1}) + Cz_{n+1} - Cz_n, \bar{z}_n - \bar{z}_{n+1}\rangle. \tag{17}$$

On the other hand we have

$$\begin{aligned}
\langle Cz_{n+1} - Cz_n, \bar{z}_n - \bar{z}_{n+1}\rangle =& \langle Cz_{n+1} - Cz_n, \tilde{z}_n - \tilde{z}_{n+1}\rangle + \langle Cz_{n+1} - Cz_n, z_n - z_{n+1}\rangle \\
\leq& \tfrac{\epsilon}{2}\|\tilde{z}_n - \tilde{z}_{n+1}\|_P^2 + \tfrac{1}{2\epsilon}\|Cz_{n+1} - Cz_n\|_{P^{-1}}^2 \\
&+ \langle Cz_{n+1} - Cz_n, z_n - z_{n+1}\rangle \\
\leq& \tfrac{\epsilon}{2}\|\tilde{z}_n - \tilde{z}_{n+1}\|_P^2 + \left(1 - \tfrac{1}{2\epsilon\beta}\right)\langle Cz_{n+1} - Cz_n, z_n - z_{n+1}\rangle.
\end{aligned}$$

The first inequality follows from the Fenchel-Young inequality for $\frac{\epsilon}{2}\|\cdot\|_P^2$, and the second inequality follows from β-cocoercivity of C with respect to $\|\cdot\|_P$. Set $\epsilon = \frac{1}{2\beta}$ so that

$$\langle Cz_{n+1} - Cz_n, \bar{z}_n - \bar{z}_{n+1}\rangle \leq \tfrac{1}{4\beta}\|\tilde{z}_n - \tilde{z}_{n+1}\|_P^2. \tag{18}$$

Using (17), (18) and monotonicity of M we have

$$\begin{aligned}
0 &\leq \tfrac{1}{4\beta}\|\tilde{z}_n - \tilde{z}_{n+1}\|_P^2 + \langle -M(z_n - z_{n+1}) - H(\tilde{z}_n - \tilde{z}_{n+1}), \bar{z}_n - \bar{z}_{n+1}\rangle \\
&= \tfrac{1}{4\beta}\|\tilde{z}_n - \tilde{z}_{n+1}\|_P^2 + \langle -M(z_n - z_{n+1}) - H(\tilde{z}_n - \tilde{z}_{n+1}), \tilde{z}_n - \tilde{z}_{n+1}\rangle \\
&\quad + \langle -M(z_n - z_{n+1}) - H(\tilde{z}_n - \tilde{z}_{n+1}), z_n - z_{n+1}\rangle \\
&\leq \tfrac{1}{4\beta}\|\tilde{z}_n - \tilde{z}_{n+1}\|_P^2 + \langle -M(z_n - z_{n+1}) - H(\tilde{z}_n - \tilde{z}_{n+1}), \tilde{z}_n - \tilde{z}_{n+1}\rangle \\
&\quad + \langle -H(\tilde{z}_n - \tilde{z}_{n+1}), z_n - z_{n+1}\rangle \qquad (19) \\
&= -\left(1 - \tfrac{1}{4\beta}\right)\|\tilde{z}_n - \tilde{z}_{n+1}\|_P^2 - \langle (M+H^*)(z_n - z_{n+1}), \tilde{z}_n - \tilde{z}_{n+1}\rangle. \qquad (20)
\end{aligned}$$

It follows from (20) and Step 4 of Algorithm 1 that

$$\begin{aligned}
\left(1 - \tfrac{1}{4\beta}\right)\|\tilde{z}_n - \tilde{z}_{n+1}\|_P^2 &\leq \langle -(M+H^*)(z_n - z_{n+1}), \tilde{z}_n - \tilde{z}_{n+1}\rangle \\
&= \langle \alpha_n (H+M^*)^* S^{-1}(H+M^*)\tilde{z}_n, \tilde{z}_n - \tilde{z}_{n+1}\rangle \\
&\leq \langle \alpha_n D\tilde{z}_n, \tilde{z}_n - \tilde{z}_{n+1}\rangle. \qquad (21)
\end{aligned}$$

Let us show that $(\|\tilde{z}_n\|_D^2)_{n\in\mathbb{N}}$ is monotonically nonincreasing. Using the identity

$$\|a\|_D^2 - \|b\|_D^2 = 2\langle Da, a-b\rangle - \|a-b\|_D^2, \tag{22}$$

we have

$$\begin{aligned}
\|\tilde{z}_n\|_D^2 - \|\tilde{z}_{n+1}\|_D^2 &= 2\langle D\tilde{z}_n, \tilde{z}_n - \tilde{z}_{n+1}\rangle - \|\tilde{z}_n - \tilde{z}_{n+1}\|_D^2 \\
&\geq \tfrac{2}{\alpha_n}(1-\tfrac{1}{4\beta})\|\tilde{z}_n - \tilde{z}_{n+1}\|_P^2 - \|\tilde{z}_n - \tilde{z}_{n+1}\|_D^2 \\
&\geq \tfrac{c_1\delta}{\lambda_n}\|\tilde{z}_n - \tilde{z}_{n+1}\|_P^2 - \|\tilde{z}_n - \tilde{z}_{n+1}\|_D^2 \\
&\geq \left(\tfrac{c_1\delta}{c_2\lambda_n} - 1\right)\|\tilde{z}_n - \tilde{z}_{n+1}\|_D^2
\end{aligned}$$

where the inequalities follow from (21), the definition of α_n and (15). Therefore $\|\tilde{z}_n\|_D^2$ is nonincreasing as long as (16) is satisfied. It follows from (10) and (15) that

$$\begin{aligned}
\|z_{n+1} - z^\star\|_S^2 &\leq \|z_n - z^\star\|_S^2 - \lambda_n(\delta-\lambda_n)\|\left(H + M^*\right)\tilde{z}_n\|_{S^{-1}}^{-2}\|\tilde{z}_n\|_P^4 \\
&= \|z_n - z^\star\|_S^2 - \lambda_n(\delta-\lambda_n)\|\tilde{z}_n\|_D^{-2}\|\tilde{z}_n\|_P^4 \\
&\leq \|z_n - z^\star\|_S^2 - c_2^{-2}\lambda_n(\delta-\lambda_n)\|\tilde{z}_n\|_D^2.
\end{aligned}$$

Summing over n yields $\sum_{i=0}^{\infty}\lambda_i(\delta-\lambda_i)\|\tilde{z}_i\|_D^2 \leq c_2^2\|z_0 - z^\star\|_S^2$. Therefore we have

$$\textstyle\sum_{i=0}^{\infty}\|\tilde{z}_i\|_D^2 \leq \frac{c_2^2}{\tau}\|z_0 - z^\star\|_S^2. \tag{23}$$

On the other hand since $\|\tilde{z}_n\|_D^2$ is nonincreasing, we have $\|\tilde{z}_n\|_D^2 \leq \frac{1}{n+1}\sum_{i=0}^{n}\|\tilde{z}_i\|_D^2$. Combining this with (23) establishes the big-O convergence. The little-o convergence follows from [17, Lemma 3-(1a)]. □

We can show stronger convergence results, *i.e.*, linear convergence rate, under *metric subregulariy* assumption for T. We restate the following definition from [19]:

Definition 3.1 (Metric subregularity)**.** A mapping F is metrically subregular at $\bar{x}$ for $\bar{y}$ if $(\bar{x},\bar{y}) \in \operatorname{gra} F$ and there exists $\eta \in [0,\infty[$, a neighborhood $\mathcal{U}$ of $\bar{x}$ and $\mathcal{V}$ of $\bar{y}$ such that

$$d(x, F^{-1}\bar{y}) \leq \eta d(\bar{y}, Fx \cap \mathcal{V}) \;\text{ for all } x \in \mathcal{U}. \tag{24}$$

This is equivalent to *calmness* of the operator F^{-1} at $\bar{y}$ for $\bar{x}$ [19, Theorem 3.2]. The above two properties are weaker versions of metric regularity and Aubin property, respectively. We refer the reader to [32, Chapter 9] and [20, Chapter 3] for an extensive discussion. In Theorem 3.3, we derive linear convergence rates when the operator $T = A + M + C$ is metrically subregular at all $z^\star \in \operatorname{zer}(T)$ for 0. Metric subregularity is used in [26] to show linear convergence of Krasnosel'skiĭ-Mann iterations for finding a fixed point of a nonexpansive mapping.

Theorem 3.3 (Linear convergence)**.** *Consider Algorithm 1 under the assumptions of Theorem 3.1. Suppose that T is metrically subregular at all $z^\star \in \operatorname{zer}(T)$ for 0, cf. (24). If either $\mathcal{H}$ is finite-dimensional or $\mathcal{U} = \mathcal{H}$, then $(d_S(z_n, \operatorname{zer}(T)))_{n\in\mathbb{N}}$ converges Q-linearly to zero, $(z_n)_{n\in\mathbb{N}}$ and $(\|\tilde{z}_n\|_P)_{n\in\mathbb{N}}$ converge R-linearly to some $z^\star \in \operatorname{zer}(T)$ and zero, respectively.*[2]

[2]The sequence $(x_n)_{n\in\mathbb{N}}$ converges to $x^\star$ *R*-linearly if there is a sequence of nonnegative scalars $(v_n)_{n\in\mathbb{N}}$ such that $\|x_n - x^\star\| \leq v_n$ and $(v_n)_{n\in\mathbb{N}}$ converges *Q*-linearly[3] to zero.

[3]The sequence $(x_n)_{n\in\mathbb{N}}$ converges to $x^\star$ *Q*-linearly with *Q*-factor given by $\sigma \in]0,1[$, if for n sufficiently large $\|x_{n+1} - x^\star\| \leq \sigma\|x_n - x^\star\|$ holds.

Proof. It follows from metric subregularity of T at all $z^\star \in \operatorname{zer}(T)$ for 0 that

$$d(x, \operatorname{zer}(T)) \leq \eta \|y\| \quad \forall x \in \mathcal{U} \text{ and } y \in Tx \text{ with } \|y\| \leq \nu, \tag{25}$$

for some $\nu \in]0, \infty[$ and $\eta \in [0, \infty[$ and a neighborhood $\mathcal{U}$ of $\operatorname{zer}(T)$. Consider w_n defined in (12). It was shown in (13) that $w_n \to 0$ and if $\mathcal{H}$ is a finite-dimensional Hilbert space, Theorem 3.1(*ii*)-(*iii*) yield that $\bar{z}_n$ converges to a point in $\operatorname{zer}(T)$. Then there exists $\bar{n} \in \mathbb{N}$ such that for $n > \bar{n}$ we have $\|w_n\| \leq \nu$ and a neighborhood $\mathcal{U}$ of $\operatorname{zer}(T)$ exists with $\bar{z}_n \in \mathcal{U}$ (This holds trivially when $\mathcal{U} = \mathcal{H}$). Consequently (25) yields $d(\bar{z}_n, \operatorname{zer}(T)) \leq \eta \|w_n\|$. In addition, triangle inequality and Lipschitz continuity of C yield

$$\begin{aligned} \|w_n\| &= \|H\tilde{z}_n - M\tilde{z}_n - C\bar{z}_n + Cz_n\| \leq \|(H-M)\tilde{z}_n\| + \|C\bar{z}_n - Cz_n\| \\ &\leq \left(\|H - M\| + \tfrac{1}{\beta}\|P\|\right) \|\tilde{z}_n\|. \end{aligned}$$

Consider the projection of $\bar{z}_n$ onto $\operatorname{zer}(T)$, $\Pi_{\operatorname{zer}(T)}(\bar{z}_n)$. By definition $\|\bar{z}_n - \Pi_{\operatorname{zer}(T)}(\bar{z}_n)\| = d(\bar{z}_n, \operatorname{zer}(T))$ (the minimum is attained since T is maximally monotone [1, Proposition 23.39]), and we have

$$\begin{aligned} \|z_n - \Pi_{\operatorname{zer}(T)}(\bar{z}_n)\| &\leq \|\bar{z}_n - \Pi_{\operatorname{zer}(T)}(\bar{z}_n)\| + \|\tilde{z}_n\| = d(\bar{z}_n, \operatorname{zer}(T)) + \|\tilde{z}_n\| \\ &\leq \xi\eta\|\tilde{z}_n\| + \|\tilde{z}_n\| \leq (\xi\eta + 1)\|P^{-1}\|^{1/2}\|\tilde{z}_n\|_P, \end{aligned} \tag{26}$$

where $\xi = \|H - M\| + \frac{1}{\beta}\|P\|$. It follows from (26) that

$$d_S^2(z_n, \operatorname{zer}(T)) \leq \|z_n - \Pi_{\operatorname{zer}(T)}(\bar{z}_n)\|_S^2 \leq (\xi\eta + 1)^2\|P^{-1}\|\|S\|\|\tilde{z}_n\|_P^2. \tag{27}$$

For $\Pi^S_{\operatorname{zer}(T)}(z_n)$ by its definition we have $\|z_n - \Pi^S_{\operatorname{zer}(T)}(z_n)\|_S = d_S(z_n, \operatorname{zer}(T))$, and since inequality (10) holds for all $z^\star \in \operatorname{zer}(T)$, it follows that

$$\begin{aligned} d_S^2(z_{n+1}, \operatorname{zer}(T)) &\leq \|z_{n+1} - \Pi^S_{\operatorname{zer}(T)}(z_n)\|_S^2 \\ &\leq \|z_n - \Pi^S_{\operatorname{zer}(T)}(z_n)\|_S^2 - \lambda_n(\delta - \lambda_n)\|(H + M^*)\tilde{z}_n\|_{S^{-1}}^{-2}\|\tilde{z}_n\|_P^4 \\ &= d_S^2(z_n, \operatorname{zer}(T)) - \lambda_n(\delta - \lambda_n)\|(H + M^*)\tilde{z}_n\|_{S^{-1}}^{-2}\|\tilde{z}_n\|_P^4 \qquad (28) \\ &\leq d_S^2(z_n, \operatorname{zer}(T)) - \lambda_n(\delta - \lambda_n)\|S^{-1/2}(H + M^*)P^{-1/2}\|^{-2}\|\tilde{z}_n\|_P^2 \qquad (29) \\ &\leq d_S^2(z_n, \operatorname{zer}(T)) - \tfrac{\lambda_n(\delta - \lambda_n)}{(\xi\eta+1)^2\|P^{-1}\|\|S\|\|S^{-1/2}(H+M^*)P^{-1/2}\|^2} d_S^2(z_n, \operatorname{zer}(T)), \end{aligned}$$

where in the last inequality we used (27). It follows from (6) that there exists $\tilde{n} \in \mathbb{N}$ such that $(\lambda_n(\delta - \lambda_n))_{n > \tilde{n}} \subseteq [\bar{\tau}, \infty[$ for some $\bar{\tau} > 0$. Therefore, $(d_S(z_n, \operatorname{zer}(T)))_{n \in \mathbb{N}}$ converges Q-linearly to zero. R-linear convergence of $(\|\tilde{z}_n\|_P)_{n \in \mathbb{N}}$ follows from (29) and Q-linear convergence of $(d_S(z_n, \operatorname{zer}(T)))_{n \in \mathbb{N}}$. Step 4 of Algorithm 1 and (28) yield

$$\|z_{n+1} - z_n\|_S^2 = \lambda_n^2\|(H + M^*)\tilde{z}_n\|_{S^{-1}}^{-2}\|\tilde{z}_n\|_P^4 \leq \frac{\delta^2}{\bar{\tau}}\left(d_S^2(z_n, \operatorname{zer}(T)) - d_S^2(z_{n+1}, \operatorname{zer}(T))\right).$$

Therefore, $(\|z_{n+1} - z_n\|_S)_{n \in \mathbb{N}}$ converges R-linearly to zero. This is equivalent to saying that there exists $c \in]0,1[$, $\kappa \in]0, \infty[$, $\underline{n} \in \mathbb{N}$ such that for all $n \geq \underline{n}$, $\|z_{n+1} - z_n\|_S \leq \kappa c^n$ holds. Thus, for any $j > k \geq \underline{n}$ we have

$$\|z_j - z_k\|_S \leq \textstyle\sum_{i=k}^{j-1}\|z_{i+1} - z_i\|_S \leq \sum_{i=k}^{j-1}\kappa c^i \leq \sum_{i=k}^{\infty}\kappa c^i = \frac{\kappa}{1-c}c^k. \tag{30}$$

Hence, the sequence $(z_n)_{n \in \mathbb{N}}$ is a Cauchy sequence, and therefore converges to some $z \in \mathcal{H}$. From uniqueness of weak limit and Theorem 3.1(*iii*) we have $z \in \operatorname{zer}(T)$. Let $j \to \infty$ in (30) to obtain R-linear convergence of $(z_n)_{n \in \mathbb{N}}$. □

In the special case when $C \equiv 0$, M is skew-adjoint , $K = M$ and $S = P$, the operator $P \in \mathscr{B}(\mathcal{H})$ can be a self-adjoint, positive operator rather than a strongly positive operator. Under these assumptions AFBA simplifies to the following iteration:

$$\bar{z}_n = (H + A)^{-1} P z_n \tag{31a}$$

$$z_{n+1} = z_n + \lambda_n(\bar{z}_n - z_n). \tag{31b}$$

Notice that if P was strongly positive, this could simply be seen as proximal point algorithm in a different metric applied to the operator $A + M$, but we have relaxed this assumption and only require P to be positive. Before providing convergence results for this algorithm we begin with the following lemma, showing that the mapping $(H+A)^{-1}$ has full domain and is continuous when H has a block triangular structure with strongly positive diagonal blocks, even though its symmetric part, P, might not be strongly positive. This lemma motivates the assumption on continuity of $(H + A)^{-1}P$ in Theorem 3.4. As an application of this theorem in Proposition 4.1(*iii*), when P is positive with a two-by-two block structure (see (57a) in the limiting case $\theta = 2$), DRS is recovered.

Lemma 3.1. *Let $\mathcal{H} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_N$, where $\mathcal{H}_1, \cdots, \mathcal{H}_N$ are real Hilbert spaces. Suppose that A is block separable and H has a conformable lower (upper) triangular partitioning,* i.e.*,*

$$\begin{aligned} A &: z \mapsto (A_1 z_1, \cdots, A_N z_N), \\ H &: z \mapsto (H_{11} z_1, H_{21} z_1 + H_{22} z_2, \cdots, \sum_{j=1}^{N} H_{Nj} z_j), \end{aligned} \tag{32}$$

where $z_i \in \mathcal{H}_i$ for $i = 1, \cdots, N$, $z = (z_1, \cdots, z_N) \in \mathcal{H}$, and $H_{ij} \in \mathscr{B}(\mathcal{H}_j, \mathcal{H}_i)$ for $i, j = 1, \cdots, N$. For $i = 1, \cdots, N$, assume that A_i is maximally monotone, and $H_{ii} \in \mathcal{S}_{\tau_i}(\mathcal{H}_i)$ with $\tau_i \in]0, \infty]$. Then, the mapping $(H+A)^{-1}$ is continuous and has full domain, i.e.*, $\operatorname{dom}((H + A)^{-1}) = \mathcal{H}$. Furthermore, the update $\bar{z} = (H + A)^{-1} z$ is carried out using*

$$\bar{z}_i = \begin{cases} (H_{11} + A_1)^{-1} z_1, & i = 1; \\ (H_{ii} + A_i)^{-1} \left(z_i - \sum_{j=1}^{i-1} H_{ij} \bar{z}_j \right), & i = 2, \cdots, N, \end{cases} \tag{33}$$

where $\bar{z} = (\bar{z}_1, \cdots, \bar{z}_N) \in \mathcal{H}$.

Proof. We consider a block lower triangular H as in (32), the analysis for upper triangular case is identical. The goal is to consider A_i's separately. Let $\bar{z} = (H + A)^{-1} z$ with $\bar{z} = (\bar{z}_1, \cdots, \bar{z}_N)$. The block triangular structure of H in (32) yields the equivalent inclusion $z_i \in A_i \bar{z}_i + \sum_{j=1}^{i} H_{ij} \bar{z}_j$, for $i = 1, \cdots, N$. This is equivalent to (33), in which, each $\bar{z}_i$ is evaluated using z_i and $\bar{z}_j$ for $j < i$. For the first block we have $\bar{z}_1 = (H_{11} + A_1)^{-1} z_1$. Since A_1 is monotone and H_{11} is strongly monotone, it follows that $H_{11} + A_1$ is strongly monotone, which in turn implies that $(H_{11} + A_1)^{-1}$ is cocoercive and, as such, at most single-valued and continuous. Since A_1 is maximally monotone and H_{11} is strongly positive we have

$$\operatorname{dom}\left((H_{11} + A_1)^{-1}\right) = \operatorname{ran}(H_{11} + A_1) = \operatorname{ran}(\mathrm{Id} + A_1 H_{11}^{-1}) = \mathcal{H}_1,$$

where the last equality follows from maximal monotonicity of $A_1 H_{11}^{-1}$ in the Hilbert space defined by endowing $\mathcal{H}_1$ with the scalar product $\langle \cdot, \cdot \rangle_{H_{11}^{-1}}$, and Minty's theorem [1, Theorem 21.1]. For the second block in (33) we have $\bar{z}_2 = (H_{22} + A_2)^{-1}(z_2 -$

$H_{21}\bar{z}_1$). Hence, by the same argument used for previous block, $(H_{22}+A_2)^{-1}$ is continuous and has full domain. Since $(z_2 - H_{21}\bar{z}_1)$ and $(H_{22}+A_2)^{-1}$ are continuous, so is their composition $(H_{22}+A_2)^{-1}(z_2 - H_{21}\bar{z}_1)$. Follow the same argument for the remaining blocks in (33) to conclude that $(H+A)^{-1}$ is continuous and has full domain. □

Next theorem provides convergence and rate of convergence results for algorithm (31a)-(31b) in finite-dimensions by employing the same idea used in [14, Theorem 3.3]. The idea is to consider the operator $R = P + \mathrm{Id} - Q$, where Q is the orthogonal projection onto $\mathrm{ran}(P)$. The proof presented here is for a general P and it coincides with the one of Condat [14, Theorem 3.3] for a special choice of P (when P is defined as in (65) with $\theta = 2$).

Theorem 3.4. *Suppose that $\mathcal{H}$ is finite-dimensional. Let $P \in \mathcal{S}(\mathcal{H})$, $P \succeq 0$, $M \in \mathscr{B}(\mathcal{H})$ a skew-adjoint operator and $H = P + M$. Consider the iteration (31a)-(31b) and assume $\mathrm{zer}(T) \neq \emptyset$ where $T = A + M$. Furthermore, assume that $(H+A)^{-1}P$ is continuous. Let $(\lambda_n)_{n\in\mathbb{N}}$ be uniformly bounded in the interval $]0, 2[$. Then,*

(i) $(z_n)_{n\in\mathbb{N}}$ converges to a point in $\mathrm{zer}(T)$.

(ii) Let Q be the orthogonal projection onto $\mathrm{ran}(P)$, and $R = P + \mathrm{Id} - Q$. The following convergence estimates hold:

$$\|Pz_{n+1} - Pz_n\|^2 \leq \tfrac{\|P\|}{\tau(n+1)}\|Qz_0 - Qz^\star\|_R^2, \tag{34}$$

for some constant $\tau > 0$, and $\|Pz_{n+1} - Pz_n\|^2 = o(1/(n+1))$.

Proof. (*i*): Since P is not strongly positive, it does not define a valid inner product. Consider $R \coloneqq P + \mathrm{Id} - Q$, where Q is the orthogonal projection onto $\mathrm{ran}(P)$. We show that by construction R is strongly positive. By the spectral theorem we can write $Pz_1 = U\Lambda U^* z_1$, where U is an orthonormal basis consisting of eigenvectors of A. Consider two sets: $s_1 = \{i | \lambda_i \neq 0\}$ and $s_2 = \{i | \lambda_i = 0\}$. Denote by U_1 the orthonormal basis made up of u_i for $i \in s_1$. Then $\mathrm{ran}\,(P) = \mathrm{ran}\,(U_1)$ and we have $Q = U_1 U_1^*$. For any $z \in \mathcal{H}$, $z = z_1 + z_2$ where $z_1 = Qz$ and $z_2 = (\mathrm{Id} - Q)z$. Then $Rz = Pz + z - Qz = Pz_1 + z_2$ and $\langle Rz, z\rangle = \langle Pz_1 + z_2, z\rangle = \langle Pz_1, z_1\rangle + \|z_2\|^2$. If $z_1 = 0$ then $\langle Rz, z\rangle = \|z_2\|^2 = \|z\|^2$. Suppose that $z_2 \neq 0$ and $z_1 \neq 0$. Denote by $\lambda_{\min}$ the smallest non zero eigenvalue of P. We have

$$\begin{aligned}\langle Rz, z\rangle &= \langle Pz_1, z_1\rangle + \|z_2\|^2 = \langle U\Lambda U^* z_1, z_1\rangle + \|z_2\|^2 = \langle \Lambda U^* z_1, U^* z_1\rangle + \|z_2\|^2 \\ &\geq \lambda_{\min}\|U_1^* z_1\|^2 + \|z_2\|^2 = \lambda_{\min}\langle z_1, Qz_1\rangle + \|z_2\|^2 = \lambda_{\min}\|z_1\|^2 + \|z_2\|^2 \\ &\geq \min\{1, \lambda_{\min}\}\|z\|^2.\end{aligned}$$

If $z_2 = 0$ the above analysis holds with $z = z_1$ and result in strong positivity parameter equal to $\lambda_{\min}$.

We continue by noting that by definition we have $Q \circ P = P$, and symmetry of P yields $P \circ Q = P$. Therefore, $R \circ Q = P$ and for $z \in \mathcal{H}$ we have

$$\langle Pz, z\rangle = \langle QPz, z\rangle = \langle Pz, Qz\rangle = \langle RQz, Qz\rangle, \tag{35}$$

which will be used throughout this proof. Observe now that for $z^\star \in \mathrm{zer}(T) \neq \emptyset$ we have $-Mz^\star \in Az^\star$. By monotonicity of A and (31a) we have $\langle M\bar{z}_n - Mz^\star +$

$P\tilde{z}_n, z^\star - \bar{z}_n\rangle \geq 0$. Then

$$\begin{aligned} 0 &\leq \langle M\bar{z}_n - Mz^\star + P\tilde{z}_n, z^\star - \bar{z}_n\rangle = \langle P\tilde{z}_n, z^\star - \bar{z}_n\rangle \\ &= \langle P\tilde{z}_n, Qz^\star - Q\bar{z}_n\rangle + \langle RQ\tilde{z}_n, Q\tilde{z}_n\rangle - \langle RQ\tilde{z}_n, Q\tilde{z}_n\rangle \\ &= \langle P\tilde{z}_n, Qz^\star - Qz_n\rangle - \|Q\tilde{z}_n\|_R^2, \end{aligned} \tag{36}$$

where the equalities follows from skew-symmetricity of M and (35). We show that $\|Qz_n - Qz^\star\|_R$ is decreasing using (36):

$$\begin{aligned} \|Qz_{n+1} - Qz^\star\|_R^2 &= \|Qz_n - Qz^\star + \lambda_n Q\tilde{z}_n\|_R^2 \\ &= \|Qz_n - Qz^\star\|_R^2 + 2\lambda_n\langle Qz_n - Qz^\star, P\tilde{z}_n\rangle + \lambda_n^2\|Q\tilde{z}_n\|_R^2 \\ &\leq \|Qz_n - Qz^\star\|_R^2 - \lambda_n(2-\lambda_n)\|Q\tilde{z}_n\|_R^2. \end{aligned} \tag{37}$$

Let us define the sequence $x_n = Qz_n$ and $\bar{x}_n = Q\bar{z}_n$ for every $n \in \mathbb{N}$. Then, since $P \circ Q = P$ the iteration for $x_n = Qz_n$ is written as

$$\begin{aligned} \bar{x}_n &= Q(H+A)^{-1}P(x_n) \\ x_{n+1} &= x_n + \lambda_n(\bar{x}_n - x_n). \end{aligned} \tag{38}$$

Let $G = (H+A)^{-1}P$ and $G' = Q \circ G$. It follows from $H = P + M$ and (31a) that

$$0 \in T(G(z)) + PG(z) - Pz. \tag{39}$$

Use (39) and monotonicity of T at $z^\star \in \operatorname{zer}(T)$ and $G(z^\star)$ to derive

$$0 \leq \langle G(z^\star) - z^\star, Pz^\star - PG(z^\star)\rangle. \tag{40}$$

In view of (40) and positivity of P, we have $\langle G(z^\star) - z^\star, PG(z^\star) - Pz^\star\rangle = 0$, and by [1, Corollary 18.17], $PG(z^\star) - Pz^\star = 0$. Hence, since $R \circ Q = P$, we have $RQG(z^\star) - RQz^\star = 0$, and strong positivity of R implies $Qz^\star = QG(z^\star) = QG(Qz^\star)$, where the last equality is due to $G \circ Q = G$. Thus, $Qz^\star$ is a fixed point of $G' = QG$. We showed that if $z^\star \in \operatorname{zer}(T)$ then $Qz^\star \in \operatorname{Fix}(G')$, *i.e.*,

$$Q\operatorname{zer}(T) \subseteq \operatorname{Fix}(G'). \tag{41}$$

Furthermore, for any $x^\star \in \operatorname{Fix}(G')$ we have $Px^\star = PG'(x^\star) = PQG(x^\star) = PG(x^\star)$. Combine this with (39) to derive $G(x^\star) \in \operatorname{zer}(T)$. Therefore, $x^\star = G'(x^\star) = QG(x^\star) \in Q\operatorname{zer}(T)$. This shows that if $x^\star \in \operatorname{Fix}(G')$, then $x^\star \in Q\operatorname{zer}(T)$, *i.e.*, $\operatorname{Fix}(G') \subseteq Q\operatorname{zer}(T)$. Combine this with (41) to conclude that the two sets $\operatorname{Fix}(G')$ and $Q\operatorname{zer}(T)$ are the same. On the other hand, we rewrite (37) for x_n and $\bar{x}_n$:

$$\|x_{n+1} - Qz^\star\|_R^2 \leq \|x_n - Qz^\star\|_R^2 - \lambda_n(2-\lambda_n)\|\bar{x}_n - x_n\|_R^2. \tag{42}$$

Therefore, $(x_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\operatorname{Fix}(G')$ in the Hilbert space $\mathcal{H}_R$. Since $(\lambda_n)_{n\in\mathbb{N}}$ is uniformly bounded in $]0,2[$, it follows that

$$G'(x_n) - x_n = \bar{x}_n - x_n \to 0. \tag{43}$$

Let x be a sequential cluster point of $(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \to x$. G' is continuous since $G' = Q \circ G$ and G is assumed to be continuous. Thus, it follows from (43) that $G'(x) - x = 0$, *i.e.*, $x \in \operatorname{Fix}(G')$. This together with Fejér monotonicity of x_n with respect to $\operatorname{Fix}(G')$ and [1, Theorem 5.5] yields $x_n \to x \in \operatorname{Fix}(G')$.

The proof is completed by first using $G \circ Q = G$ and continuity of G to deduce that $\bar{z}_n = G(z_n) = G(x_n)$ converges to $G(x^\star) \in \operatorname{zer}(T)$, and then arguing for convergence of z_n. We skip the details here because they are identical to the last part of the proof in [14, Theorem 3.3]).

(*ii*): Follow the procedure in the proof of Theorem 3.2 to derive (19), except that in this case the cocoercive term is absent. This yields

$$0 \leq \langle -M(z_n - z_{n+1}) - H(\tilde{z}_n - \tilde{z}_{n+1}), \tilde{z}_n - \tilde{z}_{n+1}\rangle - \langle H(\tilde{z}_n - \tilde{z}_{n+1}), z_n - z_{n+1}\rangle. \tag{44}$$

Since $H = P + M$ and M is skew-symmetric, (44) simplifies to

$$\begin{aligned} 0 &\leq \langle -P(\tilde{z}_n - \tilde{z}_{n+1}), \tilde{z}_n - \tilde{z}_{n+1}\rangle + \langle -P(z_n - z_{n+1}), \tilde{z}_n - \tilde{z}_{n+1}\rangle \\ &= -\|Q\tilde{z}_n - Q\tilde{z}_{n+1}\|_R^2 + \lambda_n \langle P\tilde{z}_n, \tilde{z}_n - \tilde{z}_{n+1}\rangle, \end{aligned} \tag{45}$$

where we used (35) and (31b). Using identity (22), we derive

$$\begin{aligned} \|Q\tilde{z}_n\|_R^2 - \|Q\tilde{z}_{n+1}\|_R^2 &= 2\langle RQ\tilde{z}_n, Q\tilde{z}_n - Q\tilde{z}_{n+1}\rangle - \|Q\tilde{z}_n - Q\tilde{z}_{n+1}\|_R^2 \\ &= 2\langle P\tilde{z}_n, \tilde{z}_n - \tilde{z}_{n+1}\rangle - \|Q\tilde{z}_n - Q\tilde{z}_{n+1}\|_R^2 \\ &\geq \left(\tfrac{2}{\lambda_n} - 1\right) \|Q\tilde{z}_n - Q\tilde{z}_{n+1}\|_R^2, \end{aligned} \tag{46}$$

where we made use of (45). Consider (37) and sum over n to derive

$$\textstyle\sum_{i=0}^{\infty} \lambda_i(2 - \lambda_i)\|Q\tilde{z}_i\|_R^2 \leq \|Qz_0 - Qz^\star\|_R^2. \tag{47}$$

Inequality (46) shows that $\|Q\tilde{z}_n\|_R^2$ is monotonically nonincreasing. Combine this with (47) and uniform boundedness of λ_n, *i.e.*, $(\lambda_n)_{n\in\mathbb{N}} \subseteq [\epsilon, 2-\epsilon]$ for some $\epsilon > 0$, to derive

$$\|Q\tilde{z}_n\|_R^2 \leq \frac{1}{(n+1)\epsilon^2}\|Qz_0 - Qz^\star\|_R^2. \tag{48}$$

Furthermore, it follows from (38) and definition of $x_n, \bar{x}_n$ that

$$\|x_{n+1} - x_n\|_R^2 = \lambda_n^2\|Q\tilde{z}_n\|_R^2 \leq (2-\epsilon)^2\|Q\tilde{z}_n\|_R^2. \tag{49}$$

Combine (49) and (48) to derive

$$\|x_n - x_{n+1}\|_R^2 \leq \frac{(2-\epsilon)^2}{(n+1)\epsilon^2}\|Qz_0 - Qz^\star\|_R^2. \tag{50}$$

This establishes big-O convergence for $(x_n)_{n\in\mathbb{N}}$. The little-o convergence of $\|Q\tilde{z}_n\|_R^2$ and subsequently $\|x_n - x_{n+1}\|_R^2$ follows from (46), (47) and [17, Lemma 3-(1a)]. We derive from (35) that $\|x_n - x_{n+1}\|_R^2 = \langle z_n - z_{n+1}, P(z_n - z_{n+1})\rangle$. Then it follows from [1, Corollary 18.17] that

$$\|Pz_n - Pz_{n+1}\|^2 \leq \|P\|\|x_n - x_{n+1}\|_R^2. \tag{51}$$

Set $\tau = \frac{\epsilon^2}{(2-\epsilon)^2}$, and combine (51) with (50) to yield big-O convergence for the sequence $(Pz_n)_{n\in\mathbb{N}}$. Similarly little-o convergence follows from that property of $\|x_n - x_{n+1}\|_R^2$. □

4. Operator Splitting Schemes as Special Cases

We are ready to consider some important special cases to illustrate the importance of parameters S, P and K. Further discussion on other special choices for the parameters appear in Section 5 in the framework of convex optimization with the understanding that it is straightforward to adapt the same analysis for the corresponding monotone inclusion problem.

4.1. **Forward-Backward Splitting.** When $H = \gamma^{-1}\,\mathrm{Id}$, $S = \mathrm{Id}$ and $M \equiv 0$, Algorithm 1 reduces to FBS. Let β be the cocoercivity constant of C with respect to the canonical norm $\|\cdot\|$, then β/γ is the cocoercivity constant with respect to the P norm and condition (6) of Theorem 3.1 becomes

$$(\lambda_n)_{n\in\mathbb{N}} \subseteq [0,\delta] \quad \text{with} \quad \delta = 2 - \frac{\gamma}{2\beta}, \quad \gamma \in]0, 4\beta[, \quad \liminf_{n\to\infty} \lambda_n(\delta - \lambda_n) > 0. \tag{52}$$

This allows a wider range of parameters than the standard ones found in the literature. The standard convergence results for FBS are based on the theory of averaged operators (see [13] and the references therein) and yield the same conditions as in (52) but with $\gamma \in]0, 2\beta]$ (see also [14, Lemma 4.4] and a slightly more conservative version in [1, Theorem 25.8]). Additionally, if $C \equiv 0$, FBS reduces to the classical PPA.

The convergence rate for FBS follows directly from Theorem 3.2. Since $D = \gamma^{-2}\,\mathrm{Id}$ and $P = \gamma^{-1}\,\mathrm{Id}$, (15) holds with $c_1 = c_2 = \gamma^{-1}$. Consequently, if $(\lambda_n(\delta - \lambda_n))_{n\in\mathbb{N}} \subseteq [\tau, \infty[$ for some $\tau > 0$, we have

$$\|\tilde{z}_n\|^2 \leq \frac{1}{\tau(n+1)}\|z_0 - z^\star\|^2,$$

and $\|\tilde{z}_n\|^2 = o(1/(n+1))$.

4.2. **Solodov and Tseng.** In Algorithm 1, set $C \equiv 0$, $A = N_X$ and $H = \mathrm{Id}$, where N_X is the normal cone operator of X, and X is a nonempty closed, convex set in $\mathcal{H}$. Then we recover the scheme proposed by Solodov and Tseng [33, Algorithm 2.1].

4.3. **Forward-Backward-Forward Splitting.** Consider Algorithm 1 when M is skew-adjoint and set $H = \gamma^{-1}\,\mathrm{Id}$, $S = \mathrm{Id}$. We can enforce $\alpha_n = \gamma$ by choosing $\lambda_n = (\gamma\|(\gamma^{-1}\,\mathrm{Id} + M^*)\tilde{z}_n\|/\|\tilde{z}_n\|)^2$. It remains to show that the sequence $(\lambda_n)_{n\in\mathbb{N}}$ satisfies the conditions of Theorem 3.1. Since M is skew-adjoint, we have $\lambda_n = 1 + (\gamma\|M\tilde{z}_n\|/\|\tilde{z}_n\|)^2$, and if the stepsize satisfies $\gamma \in]0, \|M\|^{-1}\sqrt{1 - 1/(2\beta)}[$, then $(\lambda_n)_{n\in\mathbb{N}}$ is uniformly bounded between 0 and δ (in fact it is larger than 1) and thus satisfies (6). Under these assumptions Algorithm 1 simplifies to

$$\begin{aligned} \bar{z}_n &= (\mathrm{Id} + \gamma A)^{-1}(\mathrm{Id} - \gamma M - \gamma C)z_n \\ z_{n+1} &= \bar{z}_n - \gamma M\,(\bar{z}_n - z_n)\,. \end{aligned}$$

This algorithm resembles the FBFS [35]. Indeed, if $C \equiv 0$, then the range for the stepsize simplifies to $\gamma \in]0, \|M\|^{-1}[$ and yields the FBFS when its Lipschitz operator is the skew-adjoint operator M.

4.4. **Douglas-Rachford Type with a Forward Term.** We now focus our attention on a choice for P, K and S that lead to a new Douglas-Rachford type splitting with a forward term. In Section 3 we consider more general S, P, K and the algorithm presented here can be derived as a special case in Section 5.4 in which we also discuss a 3-block ADMM algorithm. Consider the problem of finding $x \in \mathcal{H}$ such that

$$0 \in Dx + Ex + Fx, \tag{54}$$

together with the dual inclusion problem of finding $y \in \mathcal{H}$ such that there exists $x \in \mathcal{H}$,

$$\begin{cases} 0 \in Dx + Fx + y \\ 0 \in E^{-1}y - x. \end{cases} \tag{55}$$

where $D : \mathcal{H} \to 2^{\mathcal{H}}$, $E : \mathcal{H} \to 2^{\mathcal{H}}$ are maximally monotone and $F : \mathcal{H} \to \mathcal{H}$ is η-cocoercive with respect to the canonical norm. Let $\mathcal{K}$ be the Hilbert direct sum $\mathcal{K} = \mathcal{H} \oplus \mathcal{H}$. The pair $(x^\star, y^\star) \in \mathcal{K}$ is called a primal-dual solution to (54) if it satisfies (55). Let $(x^\star, y^\star) \in \mathcal{K}$ be a primal-dual solution, then $x^\star$ solves the primal problem (54) and $y^\star$ the dual (55). In this section, we assume that there exists $x^\star$ such that $x^\star \in \operatorname{zer}(D + E + F)$. This assumption yields that the set of primal-dual solutions is nonempty (see [3, 11] and the references therein for more discussion). Reformulate (55) in the form of (3) by defining

$$A : \mathcal{K} \to 2^{\mathcal{K}} : (x, y) \mapsto (Dx, E^{-1}y), \tag{56a}$$

$$M \in \mathscr{B}(\mathcal{K}) : (x, y) \mapsto (y, -x), \tag{56b}$$

$$C : \mathcal{K} \to \mathcal{K} : (x, y) \mapsto (Fx, 0). \tag{56c}$$

The operators A and M are maximally monotone [1, Proposition 20.23 and Example 20.30]. It is easy to verify, by definition of cocoercivity, that C is cocoercive. Let $\gamma > 0$, $\theta \in [0, 2]$, (the case of $\theta = 2$ can only be considered in the absence of the cocoercive term and results in classic DR, see Proposition 4.1). Set

$$P \in \mathscr{B}(\mathcal{K}) : (x, y) \mapsto \left(\gamma^{-1}x - \tfrac{1}{2}\theta y, -\tfrac{1}{2}\theta x + \gamma y\right), \tag{57a}$$

$$K \in \mathscr{B}(\mathcal{K}) : (x, y) \mapsto \left(\tfrac{1}{2}\theta y, -\tfrac{1}{2}\theta x\right), \tag{57b}$$

$$S \in \mathscr{B}(\mathcal{K}) : (x, y) \mapsto \left((3 - \theta)\gamma^{-1}x - y, -x + \gamma y\right). \tag{57c}$$

The operators P and S defined in (57) are special cases of (65) and (74b) when $L = \mathrm{Id}$, $\gamma_1 = \gamma_2^{-1} = \gamma$. It follows from Lemmas 5.1 and 5.3 that they are strongly positive for $\theta \in [0, 2[$. Then $H = P + K$ becomes

$$H \in \mathscr{B}(\mathcal{K}) : (x, y) \mapsto (\gamma^{-1}x, -\theta x + \gamma y). \tag{58}$$

Notice that H has the block triangular structure described in Lemma 3.1. By using this structure as in (33) and substituting (57), (58) in Algorithm 1, after some algebraic manipulations involving Moreau's identity as well as a change of variables $s_n \coloneqq x_n - \gamma y_n$ (see proof of Proposition 4.1 for details), we derive the following algorithm:

Algorithm 2 Douglas-Rachford Type with a Forward Term

Inputs: $x_0 \in \mathcal{H}$, $s_0 \in \mathcal{H}$
for $n = 0, \ldots$ **do**
$\bar{x}_n = J_{\gamma D}(s_n - \gamma F x_n)$
$r_n = J_{\gamma E}(\theta \bar{x}_n + (2 - \theta)x_n - s_n)$
$s_{n+1} = s_n + \rho_n(r_n - \bar{x}_n)$
$x_{n+1} = x_n + \rho_n(\bar{x}_n - x_n)$

In the special case when $\rho_n = 1$, the last line in Algorithm 2 becomes obsolete and $\bar{x}_n$ can be replaced with x_{n+1}. The next proposition provides the convergence properties for Algorithm 2.

Proposition 4.1. *Consider the sequences $(x_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$ generated by Algorithm 2. Let $\eta \in]0,+\infty[$ be the cocoercivity constant of F. Suppose that one of the following holds:*

(i) $\theta \in [0,2[$, $\gamma \in]0,\eta(4-\theta^2)[$ and the sequence of relaxation parameters $(\rho_n)_{n\in\mathbb{N}}$ is uniformly bounded in the interval

$$(\rho_n)_{n\in\mathbb{N}} \subseteq \left]0, \frac{4-\theta^2-\gamma/\eta}{(2-\theta)(2+\sqrt{2-\theta})}\right[. \tag{59}$$

(ii) $F \equiv 0$, $\theta \in [0,2[$, $\gamma \in]0,\infty[$, and sequence $(\rho_n)_{n\in\mathbb{N}}$ uniformly bounded in the interval

$$(\rho_n)_{n\in\mathbb{N}} \subseteq \left]0, 2-\sqrt{2-\theta}\right[.$$

(iii) $F \equiv 0$, $\theta = 2$, $\gamma \in]0,\infty[$, $(\rho_n)_{n\in\mathbb{N}}$ uniformly bounded in the interval $]0,2[$, and $\mathcal{K}$ is finite-dimensional.

Then, there exists a pair of solutions $(x^\star, y^\star) \in \mathcal{K}$ to (55) such that the sequences $(x_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$ converge weakly to $x^\star$ and $x^\star - \gamma y^\star$, respectively.

Proof. See Appendix. □

Next proposition provides convergence rate results for Algorithm 2 when $\theta = 2$, based on Theorem 3.4. Similarly, for the case when $\theta \in [0,2[$, convergence rates can be deduced based on Theorem 3.2. However, we omit it in this work. Furthermore, when metric subregulariy assumption in Theorem 3.3 holds, linear convergence follows without any additional assumptions.

Proposition 4.2 (Convergence rate)**.** *Let $\mathcal{K}$ be finite-dimensional. Consider the sequences $(x_n)_{n\in\mathbb{N}}$ and $(s_n)_{n\in\mathbb{N}}$ generated by Algorithm 2. Let $F \equiv 0$, $\theta = 2$, $\gamma \in]0,\infty[$, and $(\rho_n)_{n\in\mathbb{N}}$ be uniformly bounded in the interval $]0,2[$. Then*

$$\|s_{n+1} - s_n\|^2 \leq \tfrac{\gamma}{\tau(n+1)}\|Qz_0 - Qz^\star\|_R^2,$$

and $\|s_{n+1}-s_n\|^2 = o(1/(n+1))$ for some constant $\tau > 0$, where Q is the orthogonal projection onto $\operatorname{ran}(P)$, $R = P + \operatorname{Id} - Q$, and $z_n = \big(x_n, \gamma^{-1}(x_n - s_n)\big)$.

Proof. See Appendix. □

Remark 4.1. Recently, another three-operator splitting algorithm was proposed in [18] which can also be seen as a generalization of Douglas-Rachford method to accommodates a third cocoercive operator. In the aforementioned paper, the forward step takes place after the first backward update, while in Algorithm 2 it precedes the backward update. Whether this is better in practice or not is yet to be seen.

Furthermore, the parameter range prescribed in [18, Theorem 3.1] is simply $\gamma \in]0,2\eta[$ and $(\rho_n)_{n\in\mathbb{N}} \subseteq]0, 2-\frac{\gamma}{2\eta}[$, while for Algorithm 2 it consists of $\gamma \in]0,\eta(4-\theta^2)[$ with relaxation parameter uniformly bounded in the interval (59). For $\theta \in [0,\sqrt{2}[$, Algorithm 2 can have larger step size but it is important to notice that this might not necessarily be advantageous in practice because the upper bound for the relaxation parameter in (59) decreases as we reduce θ. For example if we fix $\rho_n = 1$, conditions of Proposition 4.1 become $\theta \in]1,2[$ and $\gamma/\eta \in]0,(2-\theta)(\theta-\sqrt{2-\theta})[$. This step size is always smaller that the one of [18]. However, if the relaxation parameter ρ_n is selected to be small enough then γ can take values larger than the one allowed in [18].

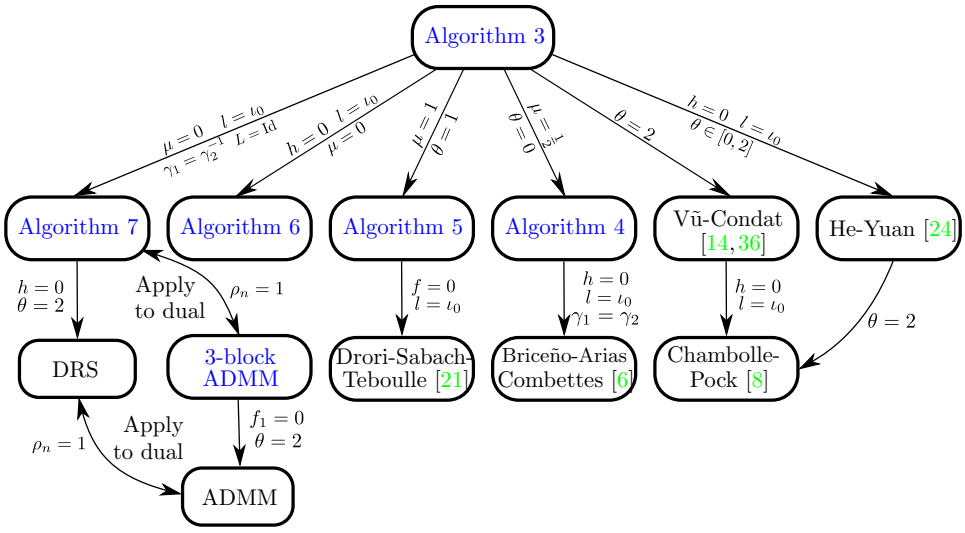

FIGURE 1. Algorithm 3 and its special cases

Remark 4.2. In Algorithm 2 the case $\theta = 2$, $\rho_n = 1$ with $F \equiv 0$ (see Proposition 4.1(*iii*)) yields the classical DRS [28]. This choice of P is precisely the one considered in [14, Section 3.1.1].

5. Structured Nonsmooth Convex Optimization

One of the characteristics of Asymmetric Forward-Backward-Adjoint splitting (AFBA) introduced in Section 3, is availability of the parameters P, K and S which are independent of each other. In the general form, P is a strongly positive operator to be chosen but it directly effects the convergence through the cocoercivity constant in Assumption 3.1(*ii*). S and K are arbitrary strongly positive and skew-adjoint operators, respectively. This introduces a lot of flexibility which proves essential in the development of this section. We will see that by properly choosing these parameters we can recover and generalize several well known schemes proposed in the recent years. Specifically, we will recover the algorithms proposed by Vũ and Condat [14,36], Briceño-Arias and Combettes [6] and the one of Drori and Sabach and Teboulle [21]. These algorithms belong to the class of so called primal-dual algorithms and owe their popularity to their simplicity and special structure. They have been used to solve a wide range of problems arising in image processing, machine learning and control, see for example [3,8,15]. Recently, randomized versions have also been proposed for distributed optimization (refer to [2,12,31]). As the last contribution of the paper we will present a generalization of the classic ADMM to three blocks. Figure 1 summarizes the connections between these algorithms in the form of a diagram. We start with the following convex optimization problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \; f(x) + h(x) + (g \,\square\, l)(Lx), \tag{60}$$

and its dual

$$\underset{y\in\mathcal{G}}{\text{minimize}}\ (f+h)^*(-L^*y)+(g\,\square\, l)^*(y), \tag{61}$$

where $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces, $f\in\Gamma_0(\mathcal{H})$, $g\in\Gamma_0(\mathcal{G})$, $L\in\mathscr{B}(\mathcal{H},\mathcal{G})$ and $h:\mathcal{H}\to\mathbb{R}$ is convex, differentiable on $\mathcal{H}$ with β_h-Lipschitz gradient for some $\beta_h\in]0,+\infty[$. Furthermore, $l\in\Gamma_0(\mathcal{G})$ is β_l^{-1}-strongly convex for some $\beta_l\in]0,+\infty[$ or equivalently ∇l^* is β_l-Lipschitz. The infimal convolution $g\,\square\, l$ can be seen as a regularization of g by l and when $l=\iota_{\{0\}}$, $(g\,\square\, l)$ simply becomes g and $\nabla l^*=0$. In our analysis we will always consider the special cases $l=\iota_{\{0\}}$ and $h\equiv 0$ separately because they result in less conservative conditions.

Throughout this section we assume that there exists $x^\star$ such that

$$x^\star\in\operatorname{zer}\left(\partial f+\nabla h+L^*(\partial g\,\square\,\partial l)L\right). \tag{62}$$

The interested reader is referred to [11, Proposition 4.3] for conditions on existence of such $x^\star$. We consider the saddle point problem corresponding to (60). This allows us to exploit Algorithm 1, and develop a unifying algorithm. We do all this in the context of optimization, with the understanding that it is straightforward to adapt the same analysis for the corresponding monotone inclusion problem. The saddle point problem is

$$\underset{x\in\mathcal{H}}{\text{minimize}}\ \underset{y\in\mathcal{G}}{\text{maximize}}\ f(x)+h(x)+\langle Lx,y\rangle-(g\,\square\, l)^*(y).$$

The optimality conditions are

$$\begin{cases}0\in\partial f(x)+\nabla h(x)+L^*y,\\ 0\in\partial g^*(y)+\nabla l^*(y)-Lx.\end{cases} \tag{63}$$

It follows from (62) that the set of solutions to (63) is nonempty. We say that $(x^\star,y^\star)$ is a primal-dual solution if it satisfies (63). Furthermore, if $(x^\star,y^\star)$ is a solution pair to (63), then $x^\star$ is a solution for the primal problem (60) and $y^\star$ is a solution to the dual problem (61). For further discussion on duality see [11, 14, 16] and the references therein. Let $\mathcal{K}$ be the Hilbert direct sum $\mathcal{K}=\mathcal{H}\oplus\mathcal{G}$ and define the operators

$$A:\mathcal{K}\to 2^{\mathcal{K}}:(x,y)\mapsto(\partial f(x),\partial g^*(y)), \tag{64a}$$
$$M\in\mathscr{B}(\mathcal{K}):(x,y)\mapsto(L^*y,-Lx), \tag{64b}$$
$$C:\mathcal{K}\to\mathcal{K}:(x,y)\mapsto(\nabla h(x),\nabla l^*(y)). \tag{64c}$$

The operator A is maximally monotone [1, Theorem 21.2 and Proposition 20.23], and the operator C is cocoercive (see Lemma 5.2). The monotone inclusion problem (63) can be written in the form of (3):

$$0\in Az+Mz+Cz,$$

where $z=(x,y)$. Let $\theta\in[0,+\infty[$ and set

$$\begin{aligned}P\in\mathscr{B}(\mathcal{K}):(x,y)&\mapsto\left(\gamma_1^{-1}x-\tfrac{1}{2}\theta L^*y,-\tfrac{1}{2}\theta Lx+\gamma_2^{-1}y\right),\\ K\in\mathscr{B}(\mathcal{K}):(x,y)&\mapsto\left(\tfrac{1}{2}\theta L^*y,-\tfrac{1}{2}\theta Lx\right).\end{aligned} \tag{65}$$

Then $H=P+K$ yields

$$H\in\mathscr{B}(\mathcal{K}):(x,y)\mapsto\left(\gamma_1^{-1}x,-\theta Lx+\gamma_2^{-1}y\right). \tag{66}$$

Remark 5.1. We make the above choices for clarity of exposition. It is straightforward to adopt the same analysis when γ_1 and γ_2 are replaced by general strongly positive operators Γ_1 and Γ_2.

Operator H defined in (66) has the block triangular structure described in Lemma 3.1. Therefore, in view of (33), $\bar{z}_n = (H+A)^{-1}(H-M-C)z_n$ becomes

$$\bar{x}_n = \operatorname{prox}_{\gamma_1 f}(x_n - \gamma_1 L^* y_n - \gamma_1 \nabla h(x_n)), \tag{67a}$$

$$\bar{y}_n = \operatorname{prox}_{\gamma_2 g^*}(y_n + \gamma_2 L((1-\theta)x_n + \theta \bar{x}_n) - \gamma_2 \nabla l^*(y_n)). \tag{67b}$$

The following two lemmas will play an important role in development of this section. Lemma 5.1 provides a tight estimate for the strong positivity parameter of P, while in Lemma 5.2 we develop estimates for the cocoercivity parameter of C given by (64c).

Lemma 5.1. *Consider P defined by (65). Let $\gamma_1, \gamma_2 > 0$ and*

$$\gamma_1^{-1} - \frac{\gamma_2}{4}\theta^2 \|L\|^2 > 0. \tag{68}$$

Then $P \in \mathcal{S}_\tau(\mathcal{K})$ where

$$\tau = \tfrac{1}{2}\gamma_1^{-1} + \tfrac{1}{2}\gamma_2^{-1} - \tfrac{1}{2}\sqrt{\theta^2\|L\|^2 + (\gamma_1^{-1} - \gamma_2^{-1})^2}. \tag{69}$$

Proof. Let $z = (x, y)$. First consider the case when $\|L\| > 0$ and $\theta \in]0, \infty[$. We have

$$\begin{aligned}
\langle z, Pz\rangle &= \langle \gamma_1^{-1}x - \tfrac{1}{2}\theta L^* y, x\rangle + \langle -\tfrac{1}{2}\theta L x + \gamma_2^{-1} y, y\rangle \\
&= \gamma_1^{-1}\|x\|^2 + \gamma_2^{-1}\|y\|^2 - \theta\langle Lx, y\rangle \\
&\geq (\gamma_1^{-1} - \tfrac{\theta\|L\|^2}{2\epsilon})\|x\|^2 + (\gamma_2^{-1} - \tfrac{\theta\epsilon}{2})\|y\|^2,
\end{aligned} \tag{70}$$

where we used the Fenchel-Young inequality for $\frac{\epsilon}{2}\|\cdot\|^2$. Select $\epsilon = -\frac{1}{\theta}(\gamma_1^{-1} - \gamma_2^{-1}) + \sqrt{\|L\|^2 + \frac{1}{\theta^2}(\gamma_1^{-1} - \gamma_2^{-1})^2}$, to maximize the strong positivity parameter. It is easy to verify that $\epsilon > 0$. Substitute ϵ in (70) to obtain $\langle z, Pz\rangle \geq \tau\|z\|^2$, where τ is given by (69) and is positive as long as (68) is satisfied. The case when either $\|L\| = 0$ or $\theta = 0$ results in τ-strongly positive P with $\tau = \min\{\gamma_1^{-1}, \gamma_2^{-1}\}$, which is also captured by (69). □

Lemma 5.2. *Let $\beta_h \in]0, +\infty[$ and $\beta_l \in]0, +\infty[$ be the Lipschitz constants of ∇h and ∇l^*, respectively. Let P be defined by (65) and assume that $\gamma_1, \gamma_2 > 0$ are such that (68) is satisfied. Then C given by (64c) is β-cocoercive with respect to the P norm, where*

$$\beta = \tau \min\{\beta_l^{-1}, \beta_h^{-1}\}, \tag{71}$$

with τ defined in (69). If in addition $l = \iota_{\{0\}}$ then

$$\beta = \beta_h^{-1}\left(\gamma_1^{-1} - \frac{\gamma_2}{4}\theta^2\|L\|^2\right). \tag{72}$$

Proof. See Appendix. □

Remark 5.2. It is easy to derive a more conservative strong positivity parameter in Lemma 5.1 similar to [36, Equation (3.20)]:

$$\tau = \min\{\gamma_1^{-1}, \gamma_2^{-1}\}\left(1 - \frac{\theta}{2}\sqrt{\gamma_1\gamma_2\|L\|^2}\right). \tag{73}$$

It must be clear that τ in (71) can be replaced with (73). However, (73) does not result in simplification of our convergence analysis and the strong positivity parameter (69) is larger (less conservative) than (73), as long as P is a strongly positive operator. Notice that according to (71) larger τ results in larger cocoercivity parameter for C and hence is less conservative. A more general version of Lemma 5.2 for several composite functions can be easily derived in a similar way but we will not consider it in this article.

Algorithm 1 gives us an extra degree of freedom in choosing S. Our aim here is to select S so as to derive an easy to implement scheme without sacrificing flexibility and generality of the algorithm. To this end, let us define $S_1 \in \mathscr{B}(\mathcal{K})$, $S_2 \in \mathscr{B}(\mathcal{K})$ as follows:

$$S_1 : (x,y) \mapsto (\gamma_1^{-1}x + (1-\theta)L^*y, (1-\theta)Lx + \gamma_2^{-1}y + \gamma_1(1-\theta)(2-\theta)LL^*y), \quad (74a)$$

$$S_2 : (x,y) \mapsto (\gamma_1^{-1}x + \gamma_2(2-\theta)L^*Lx - L^*y, -Lx + \gamma_2^{-1}y). \quad (74b)$$

Then, let $\mu \in [0,1]$ and define

$$S = \left(\mu S_1^{-1} + (1-\mu)S_2^{-1}\right)^{-1}, \quad (75)$$

so that S^{-1} is a convex combination of S_1^{-1} and S_2^{-1}. For operator D defined in (4) we have

$$D = \mu D_1 + (1-\mu)D_2, \quad (76)$$

for $\mu \in [0,1]$, where $D_1 \in \mathscr{B}(\mathcal{K})$ and $D_2 \in \mathscr{B}(\mathcal{K})$ follow from (4) by substituting S_1 and S_2, respectively:

$$D_1 : (x,y) \mapsto (\gamma_1^{-1}x - L^*y, -Lx + \gamma_2^{-1}y + \gamma_1(2-\theta)LL^*y), \quad (77a)$$

$$D_2 : (x,y) \mapsto (\gamma_1^{-1}x + \gamma_2(1-\theta)(2-\theta)L^*Lx + (1-\theta)L^*y, (1-\theta)Lx + \gamma_2^{-1}y). \quad (77b)$$

This choice is simple enough, yet it allows us to unify and generalize several well known methods. The following lemma establishes conditions for strong positivity of S, D and will be used throughout this chapter.

Lemma 5.3. *Let $\gamma_1, \gamma_2 > 0$ and assume that (68) holds. Then S and D defined in (75) and (76) are strongly positive.*

Proof. See Appendix. □

Lemma 5.3 shows that this choice of S poses no additional constraint on the parameters because it is strongly positive under the same condition required for strong positivity of P. By using (67) and S defined in (75) (or equivalently D defined in (76)), we derive the following algorithm from Algorithm 1.

Algorithm 3

Inputs: $x_0 \in \mathcal{H}$, $y_0 \in \mathcal{G}$
for $n = 0, \dots$ **do**
$\bar{x}_n = \operatorname{prox}_{\gamma_1 f}(x_n - \gamma_1 L^* y_n - \gamma_1 \nabla h(x_n))$
$\bar{y}_n = \operatorname{prox}_{\gamma_2 g^*}(y_n + \gamma_2 L((1-\theta)x_n + \theta\bar{x}_n) - \gamma_2 \nabla l^*(y_n))$
$\tilde{x}_n = \bar{x}_n - x_n, \quad \tilde{y}_n = \bar{y}_n - y_n$
Compute α_n according to (78)
$x_{n+1} = x_n + \alpha_n(\tilde{x}_n - \mu\gamma_1(2-\theta)L^*\tilde{y}_n)$
$y_{n+1} = y_n + \alpha_n(\gamma_2(1-\mu)(2-\theta)L\tilde{x}_n + \tilde{y}_n)$

$$\alpha_n = \lambda_n \frac{\gamma_1^{-1}\|\tilde{x}_n\|^2 + \gamma_2^{-1}\|\tilde{y}_n\|^2 - \theta\langle \tilde{x}_n, L^*\tilde{y}_n\rangle}{V(\tilde{x}_n, \tilde{y}_n)}, \tag{78}$$

where

$$\begin{aligned} V(\tilde{x}_n, \tilde{y}_n) = {} & \gamma_1^{-1}\|\tilde{x}_n\|^2 + \gamma_2^{-1}\|\tilde{y}_n\|^2 + (1-\mu)\gamma_2(1-\theta)(2-\theta)\|L\tilde{x}_n\|^2 \\ & + \mu\gamma_1(2-\theta)\|L^*\tilde{y}_n\|^2 + 2((1-\mu)(1-\theta)-\mu)\langle \tilde{x}_n, L^*\tilde{y}_n\rangle. \end{aligned} \tag{79}$$

The convergence properties of Algorithm 3 are stated in the following theorem. When l is the indicator of $\{0\}$ or $h \equiv 0$, it is possible to derive less conservative conditions. Theorem 5.1 distinguishes three cases: the general case; the case with $l = \iota_{\{0\}}$; and the case when $h \equiv 0$, $l = \iota_{\{0\}}$. The analysis for $h \equiv 0$ and general l is similar.

Theorem 5.1. *Consider the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ generated by Algorithm 3. Let $\beta_h \in]0, +\infty[$ and $\beta_l \in]0, +\infty[$ be the Lipschitz constants of ∇h and ∇l^*, respectively. Let $\theta \in [0, \infty[$ and suppose that one of the following holds:*

(i) $(\lambda_n)_{n\in\mathbb{N}}$ is such that (6) holds with

$$\delta = 2 - \frac{1}{2\tau}\max\{\beta_l, \beta_h\}, \tag{80}$$

where τ is defined in (69) and

$$4\tau \min\{\beta_h^{-1}, \beta_l^{-1}\} > 1. \tag{81}$$

(ii) $l = \iota_{\{0\}}$, $(\lambda_n)_{n\in\mathbb{N}}$ is such that (6) holds with

$$\delta = 2 - \tfrac{\beta_h}{2}\left(\gamma_1^{-1} - \tfrac{\gamma_2}{4}\theta^2\|L\|^2\right)^{-1} \quad \text{and} \quad \gamma_1^{-1} - \tfrac{\gamma_2}{4}\theta^2\|L\|^2 > \tfrac{\beta_h}{4}.$$

(iii) $l = \iota_{\{0\}}$, $h \equiv 0$ and $(\lambda_n)_{n\in\mathbb{N}}$ is such that (6) holds with $\delta = 2$ and

$$\gamma_1^{-1} - \frac{\gamma_2}{4}\theta^2\|L\|^2 > 0. \tag{82}$$

(iv) $l = \iota_{\{0\}}$, $h \equiv 0$, $\theta = 2$ and $\mathcal{K}$ is finite-dimensional. $(\lambda_n)_{n\in\mathbb{N}}$ is uniformly bounded in the interval $]0, 2[$ and $\gamma_1^{-1} - \gamma_2\|L\|^2 \geq 0$.

Then there exists a pair of solutions $(x^\star, y^\star) \in \mathcal{K}$ to (63) such that the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ converge weakly to $x^\star$ and $y^\star$, respectively.

Proof. (*i*): The proof relies on Theorem 3.1. From (81) we have that (68) holds. Therefore, strong positivity of P and S follow from Lemmas 5.1 and 5.3. Use Lemma 5.2 and substitute (71) in (6) to derive (80) and (81). Therefore, weak convergence follows by invoking Theorem 3.1. (*ii*): Mimic the proof of (*i*) but replace the cocoercivity parameter, β, with (72) instead of (71). (*iii*): The proof is similar to previous parts, except that we use Theorem 3.1 with $C \equiv 0$. (*iv*): In this case $P \in \mathcal{S}(\mathcal{K})$, $P \succeq 0$. In view of (64a), (66) and Lemma 3.1 $(H + A)^{-1}P$ is continuous, and thus Theorem 3.4 completes the proof. □

Remark 5.3. Condition (81) of Theorem 5.1 can be simplified for some practical special cases. For example, if $\gamma_1 = \gamma_2 = \gamma$, (81) becomes $\gamma < 4\left(2\theta\|L\| + \max\{\beta_h, \beta_l\}\right)^{-1}$.

Remark 5.4. Algorithm 3 requires 4 matrix vector products in the general case. However, we consider two special cases $\mu = 1$ and $\mu = 0$ in which only 2 matrix vector products are needed. Consider the case where $\mu = 0$. Initially calculate Lx_0 and $L\bar{x}_0$. This yields $L\tilde{x}_0$, α_0 and consequently $Lx_1 = Lx_0 + \alpha_0 L\tilde{x}_0$. Continue in

the same fashion. Therefore at each iteration we only need to calculate $L\bar{x}_n$. A similar trick can be applied when $\mu = 1$.

Remark 5.5. Algorithm 3, is not symmetric with respect to the primal and dual variables. Switching the role of the primal and dual variables we obtain

$$\begin{aligned}
\bar{y}_n &= \operatorname{prox}_{\gamma_2 g^*}(y_n + \gamma_2 L x_n - \gamma_2 \nabla l^*(y_n)) \\
\bar{x}_n &= \operatorname{prox}_{\gamma_1 f}(x_n - \gamma_1 L^*((1-\theta)y_n + \theta \bar{y}_n) - \gamma_1 \nabla h(x_n)) \\
\tilde{x}_n &= \bar{x}_n - x_n, \tilde{y}_n = \bar{y}_n - y_n \\
y_{n+1} &= y_n + \alpha_n(\mu(2-\theta)\gamma_2 L\tilde{x}_n + \tilde{y}_n) \\
x_{n+1} &= x_n + \alpha_n(\tilde{x}_n - (1-\mu)(2-\theta)\gamma_1 L^*\tilde{y}_n),
\end{aligned}$$

with

$$\alpha_n = \lambda_n \frac{\gamma_1^{-1}\|\tilde{x}_n\|^2 + \gamma_2^{-1}\|\tilde{y}_n\|^2 + \theta\langle \tilde{x}_n, L^*\tilde{y}_n\rangle}{V(\tilde{x}_n, \tilde{y}_n)},$$

where

$$\begin{aligned}
V(\tilde{x}_n, \tilde{y}_n) = {} & \gamma_1^{-1}\|\tilde{x}_n\|^2 + \gamma_2^{-1}\|\tilde{y}_n\|^2 + \mu\gamma_2(2-\theta)\|L\tilde{x}_n\|^2 \\
& + (1-\mu)\gamma_1(1-\theta)(2-\theta)\|L^*\tilde{y}_n\|^2 \\
& - 2((1-\mu)(1-\theta) - \mu)\langle \tilde{x}_n, L^*\tilde{y}_n\rangle,
\end{aligned}$$

for $\mu \in [0,1]$ and $\theta \in [0, +\infty[$.

5.1. **Special Case $\theta = 2$.** One notable special case of Algorithm 3 is the algorithm proposed by Vũ and Condat in [14, 36]. Substitute $\theta = 2$ in Algorithm 3 to obtain

$$\begin{aligned}
\bar{x}_n &= \operatorname{prox}_{\gamma_1 f}(x_n - \gamma_1 L^* y_n - \gamma_1 \nabla h(x_n)) \\
\bar{y}_n &= \operatorname{prox}_{\gamma_2 g^*}(y_n + \gamma_2 L(2\bar{x}_n - x_n) - \gamma_2 \nabla l^*(y_n)) \\
\tilde{x}_n &= \bar{x}_n - x_n, \tilde{y}_n = \bar{y}_n - y_n \\
x_{n+1} &= x_n + \lambda_n \tilde{x}_n \\
y_{n+1} &= y_n + \lambda_n \tilde{y}_n.
\end{aligned}$$

The convergence results are summarized in the following proposition and they are direct consequences of Theorems 3.2, 3.4 and 5.1 and for this reason the proof is omitted.

Proposition 5.1. *Consider the sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ generated by Algorithm 3 with $\theta = 2$. Let $\beta_h \in]0, +\infty[$ and $\beta_l \in]0, +\infty[$ be the Lipschitz constants of ∇h and ∇l^*, respectively. The following results hold:*

(1) (Convergence) Suppose one of the following holds:

(i) $(\lambda_n)_{n\in\mathbb{N}}$ is such that (6) holds where δ is defined in (80) with $\theta = 2$.

(ii) $l = \iota_{\{0\}}$, $(\lambda_n)_{n\in\mathbb{N}}$ is such that (6) holds with $\delta = 2 - \frac{\beta_h}{2}\left(\gamma_1^{-1} - \gamma_2\|L\|^2\right)^{-1}$, and

$$\gamma_1^{-1} - \gamma_2\|L\|^2 > \frac{\beta_h}{4}.$$

(iii) $l = \iota_{\{0\}}$, $h \equiv 0$ and $(\lambda_n)_{n\in\mathbb{N}}$ is such that (6) holds with $\delta = 2$ and

$$\gamma_1^{-1} - \gamma_2\|L\|^2 > 0,$$

(iv) $l = \iota_{\{0\}}$, $h \equiv 0$ and $\mathcal{K}$ is finite-dimensional. $(\lambda_n)_{n\in\mathbb{N}}$ is uniformly bounded in the interval $]0, 2[$ and $\gamma_1^{-1} - \gamma_2\|L\|^2 \geq 0$.

Then there exists a pair of solutions $(x^\star, y^\star) \in \mathcal{K}$ *to* (63) *such that the sequences* $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$ *converge weakly to* $x^\star$ *and* $y^\star$*, respectively.*

(2) (Convergence rate)

(i) Let $(\lambda_n(\delta-\lambda_n))_{n\in\mathbb{N}} \subseteq [\underline{\tau}, \infty[$ *for some* $\underline{\tau} > 0$*. For (1-i)-(1-iii) the following convergence rates hold:*

$$\|\tilde{z}_n\|_P^2 \leq \frac{1}{\underline{\tau}(n+1)}\|z_0 - z^\star\|_P^2 \quad \text{and} \quad \|\tilde{z}_n\|_P^2 = o(1/(n+1)).$$

(ii) For (1-iv) the convergence rates in (34) *holds, with* R, Q *defined therein.*

Remark 5.6. The conditions in Proposition 5.1(*1-i*) are less restrictive than the conditions of [36, Theorem 3.1]. In the case when $l = \iota_{\{0\}}$, a less conservative condition was proposed in [14, theorem 3.1], which is also more conservative than Proposition 5.1(*1-ii*). The difference is that in the latter $\gamma_1^{-1} - \gamma_2\|L\|^2$ needs to be larger than $\beta_h/4$ rather than larger or equal to $\beta_h/2$. In the special case when $\mathcal{K}$ is finite-dimensional, Proposition 5.1(*1-iv*) coincides with [14, theorem 3.3]. Proposition 5.1(2) Further provides big-O and little-o convergence rates by only requiring $(\lambda_n(2-\lambda_n))_{n\in\mathbb{N}} \subseteq [\underline{\tau}, \infty[$ for some $\underline{\tau} > 0$.

5.2. **Special Case $\theta = 0$, $\mu = \frac{1}{2}$.** Another special case of Algorithm 3 generalizes the algorithm proposed by Briceño-Arias and Combettes in [6]. Specifically, setting $\theta = 0$ in Algorithm 3, the Gauss-Seidel-type proximal updates are lost but positive definiteness of P boils down to $\gamma_i > 0$, for $i = 1, 2$. In addition, we set $\mu = \frac{1}{2}$ and select λ_n such that the algorithm converges with constant stepsize $\alpha_n = 1$ (see the proof of Proposition 5.2). This leads to Algorithm 4:

Algorithm 4

Inputs: $x_0 \in \mathcal{H}$, $y_0 \in \mathcal{G}$
for $n = 0, \ldots$ **do**
$\bar{x}_n = \operatorname{prox}_{\gamma_1 f}(x_n - \gamma_1 L^* y_n - \gamma_1 \nabla h(x_n))$
$\bar{y}_n = \operatorname{prox}_{\gamma_2 g^*}(y_n + \gamma_2 L x_n - \gamma_2 \nabla l^*(y_n))$
$\tilde{x}_n = \bar{x}_n - x_n, \tilde{y}_n = \bar{y}_n - y_n$
$x_{n+1} = \bar{x}_n - \gamma_1 L^* \tilde{y}_n$
$y_{n+1} = \bar{y}_n + \gamma_2 L \tilde{x}_n$

The convergence results for Algorithm 4 are stated in the following proposition.

Proposition 5.2. *Consider the sequences* $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$ *generated by Algorithm 4. Let* $\beta_h \in]0, +\infty[$ *and* $\beta_l \in]0, +\infty[$ *be the Lipschitz constants of* ∇h *and* ∇l^**, respectively. Suppose that one of the following holds:*

(i) $\beta = \tau \min\{\frac{1}{\beta_l}, \frac{1}{\beta_h}\}$ *with* τ *defined in* (69) *and*

$$\gamma_1^{-1} - \gamma_2\|L\|^2 > \frac{1}{2\beta\gamma_1}. \tag{83}$$

(ii) $l = \iota_{\{0\}}$ *and*

$$\gamma_1^{-1} - \gamma_2\|L\|^2 > \frac{\beta_h}{2}. \tag{84}$$

(iii) $l = \iota_{\{0\}}$, $h \equiv 0$ *and* $\gamma_1^{-1} - \gamma_2\|L\|^2 > 0$.

Then there exists a pair of solutions $(x^\star, y^\star) \in \mathcal{K}$ *to* (63) *such that the sequences* $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$ *converge weakly to* $x^\star$ *and* $y^\star$*, respectively.*

Proof. See Appendix. □

Remark 5.7. We can derive $o(1/(n+1))$ convergence rate for Algorithm 4 using Theorem 3.2. It must be noted that this leads to more restrictive conditions compared to the ones required for convergence in Proposition 5.2. It was shown in the proof of Proposition 5.2 that $(\lambda_n(\delta - \lambda_n))_{n\in\mathbb{N}} \subseteq [\underline{\tau}, \infty[$ for some $\underline{\tau} > 0$. Furthermore, P and D are given by (65) and (76) with $\theta = 0$ and $\mu = \frac{1}{2}$, and it's easy to verify that (15) holds for $c_1 = 1$ and $c_2 = 1 + \gamma_1\gamma_2\|L\|^2$. According to Theorem 3.2, the $o(1/(n+1))$ convergence rate is achieved if $(\lambda_n)_{n\in\mathbb{N}} \subseteq [0, c_1\delta/c_2]$. By substituting c_1 and c_2, we derive the following sufficient condition

$$(1 + \gamma_1\gamma_2\|L\|^2)^2 \leq 2 - \tfrac{1}{2\beta}, \tag{85}$$

where β is defined in (71), and in the special case when $l = \iota_{\{0\}}$, it can be replaced by the simpler term in (72).

Remark 5.8. In Algorithm 4, if $\gamma_1 = \gamma_2 = \gamma$, $h \equiv 0$ and $l = \iota_{\{0\}}$, the error-free version of the algorithm proposed by Briceño-Arias and Combettes [6, Algorithm (4.8)] is recovered. Furthermore, it follows from (85) that $o(1/(n+1))$ convergence rate is guaranteed if

$$\gamma^2\|L\|^2 \leq \sqrt{2} - 1.$$

It is worth mentioning that another splitting was proposed by Combettes and Pesquet in [11, Theorem 4.2] that solves the minimization problem (60). The aforementioned algorithm requires two evaluations ∇h and ∇l^* per iteration.

5.3. **Special Case $\theta = 1$, $\mu = 1$.** As another special case, we derive an algorithm that generalizes the scheme proposed recently by Drori and Sabach and Teboulle in [21]. When $\theta = 1$, Algorithm 3 essentially leads to Arrow-Hurwicz updates. In addition, we set $\mu = 1$, and select λ_n in order to have $\alpha_n = 1$ (see proof of Proposition 5.3). This simplifies the iterations and results in Algorithm 5:

Algorithm 5

Inputs: $x_0 \in \mathcal{H}$, $y_0 \in \mathcal{G}$
for $n = 0, \ldots$ **do**
$\bar{x}_n = \operatorname{prox}_{\gamma_1 f}(x_n - \gamma_1 L^* y_n - \gamma_1 \nabla h(x_n))$
$y_{n+1} = \operatorname{prox}_{\gamma_2 g^*}(y_n + \gamma_2 L\bar{x}_n - \gamma_2 \nabla l^*(y_n))$
$x_{n+1} = \bar{x}_n - \gamma_1 L^*(y_{n+1} - y_n)$

In the next proposition we establish convergence results for Algorithm 5. The proof is similar to that of Proposition 5.2 and is omitted here.

Proposition 5.3. *Consider the sequences* $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$ *generated by Algorithm 5. Let* $\beta_h \in]0, +\infty[$ *and* $\beta_l \in]0, +\infty[$ *be the Lipschitz constants of* ∇h *and* ∇l^**, respectively. Suppose that one of the following holds:*

(i) $\beta = \tau \min\{\frac{1}{\beta_l}, \frac{1}{\beta_h}\}$ *where* τ *is defined in* (69) *and*
  (a) $2\beta > 1$,
  (b) $\gamma_1^{-1} - \gamma_2 \left(1 + \frac{1}{2(2\beta-1)}\right)^2 \|L\|^2 > 0$.

(ii) $l = \iota_{\{0\}}$ *and*

$$\beta_h \gamma_1 < 2 - \gamma_1\gamma_2\|L\|^2 - \sqrt{\gamma_1\gamma_2\|L\|^2}. \tag{86}$$

(iii) $l = \iota_{\{0\}}$, $h \equiv 0$ *and*

$$\gamma_1^{-1} - \gamma_2\|L\|^2 > 0.$$

Then there exists a pair of solutions $(x^\star, y^\star) \in \mathcal{K}$ *to* (63) *such that the sequences* $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$ *converge weakly to* $x^\star$ *and* $y^\star$*, respectively.*

Remark 5.9. In Remark 5.7 we deduced $o(1/(n+1))$ convergence rate for Algorithm 4 using Theorem 3.2. We can similarly deduce the same rate for Algorithm 5. However, also in this case Theorem 3.2 imposes more restrictive conditions compared to the ones required for convergence in Proposition 5.3. It is easy to verify that in this case (15) holds for $c_1 = 2/(2+\sqrt{\gamma_1\gamma_2\|L\|^2})$, $c_2 = 2/(2-\sqrt{\gamma_1\gamma_2\|L\|^2})$. According to Theorem 3.2 we have $o(1/(n+1))$ convergence rate if $(\lambda_n)_{n\in\mathbb{N}} \subseteq [0, c_1\delta/c_2]$. With c_1 and c_2 defined as above we derive the following sufficient condition

$$\frac{1}{\beta} < 4 - \frac{4\left(2+\sqrt{\gamma_1\gamma_2\|L\|^2}\right)}{\left(2-\sqrt{\gamma_1\gamma_2\|L\|^2}\right)^2}, \tag{87}$$

with $\beta = \tau \min\{\frac{1}{\beta_l}, \frac{1}{\beta_h}\}$, where τ is defined in (69). For $l = \iota_{\{0\}}$, (87) becomes

$$\beta_h\gamma_1 < 4 - \gamma_1\gamma_2\|L\|^2 - \frac{\left(2+\sqrt{\gamma_1\gamma_2\|L\|^2}\right)^3}{4-\gamma_1\gamma_2\|L\|^2}.$$

If in addition $h \equiv 0$, it further simplifies to $\gamma_1\gamma_2\|L\|^2 < \frac{21-5\sqrt{17}}{2} = 0.192$.

Remark 5.10. Algorithm 5 with $f \equiv 0$ and $l = \iota_{\{0\}}$ reduces to the algorithm proposed by Drori, Sabach and Teboulle in [21]. The authors require $\beta_h\gamma_1 \leq 1$, $\gamma_1\gamma_2\|L\|^2 \leq 1$, which is less restrictive that (86).

5.4. **Special Case:** $\mu = 0$**.** In this section we discuss three algorithms that are all special cases of Algorithm 3 when $\mu = 0$. Algorithm 7 was already discovered in Section 4. Here we briefly restate it in the optimization framework. Algorithm 8 is an application of Algorithm 7 to a dual problem formulation. Algorithm 6 is derived from Algorithm 3 when $\mu = 0$, $h \equiv 0$ and $l = \iota_{\{0\}}$, by selecting λ_n in order to have constant stepsize $\alpha_n = 1$. A more general algorithm can be considered that includes h and l, but this would complicate the conditions for selecting the parameters.

Algorithm 6

Inputs: $x_0 \in \mathcal{H}$, $y_0 \in \mathcal{G}$
for $n = 0, \ldots$ **do**
$x_{n+1} = \operatorname{prox}_{\gamma_1 f}(x_n - \gamma_1 L^* y_n)$
$\bar{y}_n = \operatorname{prox}_{\gamma_2 g^*}(y_n + \gamma_2 L((1-\theta)x_n + \theta x_{n+1}))$
$y_{n+1} = \bar{y}_n + \gamma_2(2-\theta)L(x_{n+1} - x_n)$

The next proposition provides the convergence results for Algorithm 6 through the same reasoning used in Propositions 5.2 and 5.3.

Proposition 5.4. *Assume that* $h \equiv 0$ *and* $l = \iota_{\{0\}}$*. Consider the sequences* $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$ *generated by Algorithm 6. Let* $\theta \in [0, \infty[$ *and suppose that one of the following holds:*

(i) $\gamma_1^{-1} - \gamma_2(\theta^2 - 3\theta + 3)\|L\|^2 > 0$,
(ii) $\theta = 2$, $\mathcal{K}$ *is finite-dimensional and* $\gamma_1^{-1} - \gamma_2\|L\|^2 \geq 0$.

Then there exists a pair of solutions $(x^\star, y^\star) \in \mathcal{K}$ *to* (63) *such that the sequences* $(x_n)_{n\in\mathbb{N}}$ *and* $(y_n)_{n\in\mathbb{N}}$ *converge weakly to* $x^\star$ *and* $y^\star$*, respectively.*

Remark 5.11. Algorithm 2 was introduced in Section 4 for the monotone inclusion problem (54). It can be rediscovered here by considering Algorithm 3 with $\mu = 0$ and setting $l = \iota_{\{0\}}$, L to be the identity and $\gamma_1 = \gamma_2^{-1} = \gamma$ and following the same algebraic manipulations. For this reason we only state it in the framework of optimization and refer the reader to Proposition 4.1 for convergence results.

Algorithm 7 Douglas-Rachford Type with a Forward Term

Inputs: $x_0 \in \mathcal{H}$, $s_0 \in \mathcal{H}$
for $n = 0, \ldots$ **do**
$\bar{x}_n = \operatorname{prox}_{\gamma f}(s_n - \gamma\nabla h(x_n))$
$r_n = \operatorname{prox}_{\gamma g}(\theta\bar{x}_n + (2-\theta)x_n - s_n)$
$s_{n+1} = s_n + \rho_n(r_n - \bar{x}_n)$
$x_{n+1} = x_n + \rho_n(\bar{x}_n - x_n)$

ADMM Form. Consider the following problem

$$\underset{x_1,x_2,x_3}{\text{minimize}} \quad f_1(x_1) + f_2(x_2) + f_3(x_3) \tag{88a}$$
$$\text{subject to} \quad L_1x_1 + L_2x_2 + L_3x_3 = b, \tag{88b}$$

where $\mathcal{H}_1, \ldots, \mathcal{H}_4$ are real Hilbert spaces, $f_i \in \Gamma_0(\mathcal{H}_i)$, $L_i \in \mathscr{B}(\mathcal{H}_i, \mathcal{H}_4)$ for $i = 1, 2, 3$ and $b \in \mathcal{H}_4$. Additionally, f_1 is ξ-strongly convex for some $\xi \in]0, +\infty[$. It is well known that the classic Alternating Direction Method of Multipliers (ADMM) is equivalent to Douglas-Rachford algorithm applied to the dual problem (see [4] and the references therein). We derive a new 3-block ADMM iteration in a similar way. Consider the dual problem

$$\underset{y}{\text{minimize}}\ d_1(y) + d_2(y) + d_3(y), \tag{89}$$

where $d_1(y) = f_1^*(-L_1^*y)$, $d_2(y) = f_2^*(-L_2^*y)$ and $d_3(y) = f_3^*(-L_3^*y) + \langle y, b\rangle$. By strong convexity of f_1 we have that d_1 has a $\xi^{-1}\|L_1\|^2$-Lipschitz gradient. Form the augmented Lagrangian

$$\mathcal{L}_\gamma(x_1, x_2, x_3, y) = \sum_{i=1}^{3} f_i(x_i) + \left\langle y, \sum_{i=1}^{3} L_ix_i - b \right\rangle + \frac{\gamma}{2}\left\| \sum_{i=1}^{3} L_ix_i - b \right\|^2.$$

We apply Algorithm 7 with $\rho_n = 1$ to (89) and after a change of order and some algebraic manipulations we derive Algorithm 8 (the procedure is similar to the one found in [22, Section 3.5.6] for the classic ADMM). Our 3-block ADMM can be written as

Algorithm 8 3-block ADMM

Inputs: $(x_{1,0}, x_{2,0}, x_{3,0}) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$, $(y_0, y_1) \in \mathcal{H}_4 \times \mathcal{H}_4$
for $n = 0, \ldots$ **do**
$\bar{y}_n = (\theta - 1)y_n + (2 - \theta)y_{n-1}$
$x_{1,n+1} = \operatorname{argmin}_{x_1} \mathcal{L}_0(x_1, x_{2,n}, x_{3,n}, y_n)$
$x_{2,n+1} = \operatorname{argmin}_{x_2} \mathcal{L}_\gamma(x_{1,n}, x_2, x_{3,n}, \bar{y}_n)$
$x_{3,n+1} = \operatorname{argmin}_{x_3} \mathcal{L}_\gamma(x_{1,n+1}, x_{2,n+1}, x_3, \bar{y}_n)$
$y_{n+1} = \bar{y}_n + \gamma(L_1 x_{1,n+1} + L_2 x_{2,n+1} + L_3 x_{3,n+1} - b)$

Proposition 5.5. *Let $\mathcal{H}_1, \ldots, \mathcal{H}_4$ be finite-dimensional, $f_i \in \Gamma_0(\mathcal{H}_i)$, $b \in \mathcal{H}_4$, $L_i \in \mathscr{B}(\mathcal{H}_i, \mathcal{H}_4)$ for $i = 1, 2, 3$ and $\ker(L_2) = \{0\}$, $\ker(L_3) = \{0\}$. Let $\xi \in]0, +\infty[$ be the strong convexity constant of f_1. Assume that the set of saddle points of (88), denoted by Σ, is nonempty. Let $\theta \in]1, 2[$ and*

$$\gamma < \xi(2 - \theta)(\theta - \sqrt{2 - \theta})/\|L_1\|^2. \tag{90}$$

Then the sequence $(x_{1,n}, x_{2,n}, x_{3,n}, y_n)_{n\in\mathbb{N}}$ generated by Algorithm 8, converges to some $(x_1^\star, x_2^\star, x_3^\star, y^\star) \in \Sigma$.

Proof. Let $(x_1^\star, x_2^\star, x_3^\star, y^\star)$ denote a KKT point of (88), *i.e.*,

$$\begin{cases} 0 \in \partial f_i(x_i^\star) + L_i^* y^\star, & \text{for } i = 1, 2, 3 \\ 0 = b - L_1 x_1^\star - L_2 x_2^\star - L_3 x_3^\star. \end{cases} \tag{91}$$

Algorithm 8 is an implementation of Algorithm 7 (or Algorithm 2 in the framework of general monotone operators) when applied to the dual problem (89). Hence, Proposition 4.1 yields $y_n \to y$, where y is a solution to (89). Let x_1 be a point satisfying $-L_1^* y \in \partial f_1(x_1)$. From strong convexity of f_1 at x_1 and $x_{1,n+1}$, we have

$$(\forall u \in \partial f_1(x_{1,n+1}))(\forall v \in \partial f_1(x_1)) \quad \xi\|x_{1,n+1} - x_1\|^2 \leq \langle u - v, x_{1,n+1} - x_1\rangle.$$

It follows from the optimality condition for the $x_{1,n+1}$ update and the definition of x_1 that

$$\xi\|x_{1,n+1} - x_1\|^2 \leq \langle -L_1^* y_n + L_1^* y, x_{1,n+1} - x_1\rangle \leq \|L_1\|\|x_{1,n+1} - x_1\|\|y_n - y\|.$$

Combine this with the convergence of $(y_n)_{n\in\mathbb{N}}$ to derive $x_{1,n} \to x_1$. From the change of variables to derive Algorithm 8 we have

$$s_n - y_n = -\gamma L_1 x_{1,n} - \gamma L_3 x_{3,n}, \tag{92}$$

which together with the convergence of $(s_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}$ (see Proposition 4.1) and $(x_{1,n})_{n\in\mathbb{N}}$ imply that $(L_3 x_{3,n})_{n\in\mathbb{N}}$ converges to a point. Since $\ker(L_3) = \{0\}$, it follows that $(x_{3,n})_{n\in\mathbb{N}}$ converges to some x_3. From the optimality condition for the $x_{3,n+1}$ update and the last step in Algorithm 8, we have

$$-L_3^* y_{n+1} = -L_3^* \bar{y}_n - \gamma L_3^*(L_1 x_{1,n+1} + L_2 x_{2,n+1} + L_3 x_{3,n+1} - b) \in \partial f_3(x_{3,n+1}).$$

Taking the limit and using [1, Proposition 20.33(iii)], we have $-L_3^* y \in \partial f_3(x_3)$. On the other hand, Theorem 3.1(*ii*) and the last line of Algorithm 8 yield

$$y_{n+1} - y_n \to 0, \quad \text{and} \quad L_1 x_{1,n} + L_2 x_{2,n} + L_3 x_{3,n} \to b. \tag{93}$$

It follows from (92), (93), and the convergence of $(s_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}$ that $(L_2 x_{2,n})_{n\in\mathbb{N}}$ converges to a point. We can now argue almost exactly as we did for $(L_3 x_{3,n})_{n\in\mathbb{N}}$. Since $\ker(L_2) = \{0\}$, we deduce that $(x_{2,n})_{n\in\mathbb{N}}$ converges to some x_2. Combine the optimality condition for the $x_{2,n+1}$ update and the last step in Algorithm 8 with

the convergence of y_n, to derive $-L_2^* y \in \partial f_2(x_2)$. Altogether, we showed that the limit points (x_1, x_2, x_3, y), are jointly optimal by the KKT condition (91). □

Remark 5.12. The convergence rate of Algorithm 8 can be deduced similar to Algorithm 2 from Theorems 3.2 to 3.4 with $\rho_n = 1$. However, we do not consider it in this paper due to lack of space.

Remark 5.13. In the case when $f_1 \equiv 0$ we can choose the limiting value $\theta = 2$ and recover the classical ADMM (see Proposition 4.1(*iii*)). On the other hand if f_2 or f_3 vanish, the Alternating Minimization Method (AMM) [34] is recovered. Finally, when both f_2 and f_3 vanish then the dual ascent method is recovered.

Remark 5.14. In the recent work [18, Algorithm 8], another 3-block ADMM formulation is presented by following similar algebraic manipulations (It is derived by applying their Algorithm 7 to the dual). It should be noted that they do not require rank assumptions on L_2, L_3. In contrast to that work in our version $(x_{1,n})_{n\in\mathbb{N}}$ and $(x_{2,n})_{n\in\mathbb{N}}$ are updated in parallel which corresponds to the fact that in Algorithm 2 the forward step precedes the first prox step. Furthermore, in our algorithm, $(x_{2,n})_{n\in\mathbb{N}}$ and $(x_{3,n})_{n\in\mathbb{N}}$ are updated using the augmented Lagrangian at $(\theta-1)y_n + (2-\theta)y_{n-1}$ rather than y_n. Moreover, the stepsize in [18, Theorem 2.1] has to satisfy $\gamma < 2\xi/\|L_1\|^2$. This is always larger than the stepsize in (90). Refer to Remark 4.1 for further discussion, noting that Algorithm 8 is derived by setting the relaxation parameter, ρ_n, equal to one.

Remark 5.15. Some of the other recent attempts to directly generalize ADMM for 3 blocks include [7,9,23,25,27]. In [9], it was shown through a counterexample that a direct extension of ADMM to more that 2 blocks is not convergent in general. In order to ensure convergence, additional assumptions on strong convexity of the functions or rank of L_i's are needed. In [7] the authors require one function to be strongly convex and L_2 and L_3 to have full column rank, while [25] modify the steps with regularization terms and [27] solves a perturbed problem (see [27] and the references therein for further discussion). In contrast to these papers, the first minimization step of Algorithm 8 consists of minimizing a normal Lagrangian rather than an augmented one (therefore it can be trivially executed in a distributed fashion in the case where f_1 is block-separable) and it can be performed in parallel to the second step.

6. Conclusion and Future Work

In this paper we introduced a unifying operator splitting technique called Asymmetric-Forward-Backward-Adjoint splitting (AFBA) for solving monotone inclusions involving three terms. We discussed how it relates to, unifies and extends classical splitting methods as well as several primal-dual algorithms for structured nonsmooth convex optimization. Asymmetric preconditioning is the main feature of AFBA that can lead to several extensions and new algorithmic schemes. Exploring block triangular choices for the operator H in (66) that lead to Gauss-Seidel-type updates for multi-block structured convex optimization problems is certainly a promising research direction. Another research direction involves investigating if the linear operator M can be replaced by a Lipschitz operator. It is also desirable to study randomized variants of our algorithm for problems with block structure in order to further reduce memory and computational requirements. Some natural applications include distributed optimization, control and image processing.

References

[1] Bauschke, H.H., Combettes, P.L.: Convex analysis and monotone operator theory in Hilbert spaces. Springer Science & Business Media (2011)

[2] Bianchi, P., Hachem, W., Iutzeler, F.: A coordinate descent primal-dual algorithm and application to distributed asynchronous optimization. IEEE Transactions on Automatic Control **61**(10), 2947–2957 (2016)

[3] Boţ, R.I., Csetnek, E.R., Hendrich, C.: Recent developments on primal-dual splitting methods with applications to convex minimization. In: Mathematics Without Boundaries, pp. 57–99. Springer (2014)

[4] Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning **3**(1), 1–122 (2011)

[5] Briceño-Arias, L.M.: Forward-Douglas-Rachford splitting and forward-partial inverse method for solving monotone inclusions. Optimization **64**(5), 1239–1261 (2015)

[6] Briceño-Arias, L.M., Combettes, P.L.: A monotone + skew splitting model for composite monotone inclusions in duality. SIAM Journal on Optimization **21**(4), 1230–1250 (2011)

[7] Cai, X., Han, D., Yuan, X.: The direct extension of admm for three-block separable convex minimization models is convergent when one function is strongly convex. Optimization Online **229**, 230 (2014)

[8] Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision **40**(1), 120–145 (2011)

[9] Chen, C., He, B., Ye, Y., Yuan, X.: The direct extension of admm for multi-block convex minimization problems is not necessarily convergent. Mathematical Programming **155**(1-2), 57–79 (2016)

[10] Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In: Fixed-point algorithms for inverse problems in science and engineering, pp. 185–212. Springer New York (2011)

[11] Combettes, P.L., Pesquet, J.C.: Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. Set-Valued and variational analysis **20**(2), 307–330 (2012)

[12] Combettes, P.L., Pesquet, J.C.: Stochastic quasi-fejér block-coordinate fixed point iterations with random sweeping. SIAM Journal on Optimization **25**(2), 1221–1248 (2015)

[13] Combettes, P.L., Yamada, I.: Compositions and convex combinations of averaged nonexpansive operators. Journal of Mathematical Analysis and Applications **425**(1), 55–70 (2015)

[14] Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. Journal of Optimization Theory and Applications **158**(2), 460–479 (2013)

[15] Condat, L.: A generic proximal algorithm for convex optimization-application to total variation minimization. Signal Processing Letters, IEEE **21**(8), 985–989 (2014)

[16] Davis, D.: Convergence rate analysis of primal-dual splitting schemes. SIAM Journal on Optimization **25**(3), 1912–1943 (2015)

[17] Davis, D., Yin, W.: Convergence rate analysis of several splitting schemes. arXiv preprint arXiv:1406.4834 (2014)

[18] Davis, D., Yin, W.: A three-operator splitting scheme and its optimization applications. arXiv preprint arXiv:1504.01032 (2015)

[19] Dontchev, A.L., Rockafellar, R.T.: Regularity and conditioning of solution mappings in variational analysis. Set-Valued Analysis **12**(1-2), 79–109 (2004)

[20] Dontchev, A.L., Rockafellar, R.T.: Implicit functions and solution mappings. Springer Monographs in Mathematics. Springer **208** (2009)

[21] Drori, Y., Sabach, S., Teboulle, M.: A simple algorithm for a class of nonsmooth convex-concave saddle-point problems. Operations Research Letters **43**(2), 209–214 (2015)

[22] Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Mathematical Programming **55**(1-3), 293–318 (1992)

[23] Han, D., Yuan, X.: A note on the alternating direction method of multipliers. Journal of Optimization Theory and Applications **155**(1), 227–238 (2012)

[24] He, B., Yuan, X.: Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIAM Journal on Imaging Sciences **5**(1), 119–149 (2012)
[25] Li, M., Sun, D., Toh, K.C.: A convergent 3-block semi-proximal ADMM for convex minimization problems with one strongly convex block. Asia-Pacific Journal of Operational Research p. 1550024 (2015)
[26] Liang, J., Fadili, J., Peyré, G.: Convergence rates with inexact non-expansive operators. Mathematical Programming pp. 1–32 (2014)
[27] Lin, T., Ma, S., Zhang, S.: Iteration complexity analysis of multi-block admm for a family of convex minimization without strong convexity. Journal of Scientific Computing **69**(1), 52–81 (2016)
[28] Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM Journal on Numerical Analysis **16**(6), 964–979 (1979)
[29] Moreau, J.J.: Proximité et dualité dans un espace Hilbertien. Bulletin de la Société mathématique de France **93**, 273–299 (1965)
[30] Parikh, N., Boyd, S.: Proximal algorithms. Foundations and Trends in optimization **1**, 123–231 (2013)
[31] Pesquet, J.C., Repetti, A.: A class of randomized primal-dual algorithms for distributed optimization. Journal of Nonlinear and Convex Analysis **16**(12), 2453–2490 (2015)
[32] Rockafellar, R.T., Wets, R.J.B.: Variational analysis, vol. 317. Springer Science & Business Media (2009)
[33] Solodov, M.V., Tseng, P.: Modified projection-type methods for monotone variational inequalities. SIAM Journal on Control and Optimization **34**(5), 1814–1830 (1996)
[34] Tseng, P.: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. SIAM Journal on Control and Optimization **29**(1), 119–138 (1991)
[35] Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. SIAM Journal on Control and Optimization **38**(2), 431–446 (2000)
[36] Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. Advances in Computational Mathematics **38**(3), 667–681 (2013)

7. Appendix

The proofs omitted in the main text are listed below.

Proof of Proposition 4.1. (*i*): Let us start by noting that P and S defined in (57) are special cases of (65) and (75) when $L = \mathrm{Id}$, $\gamma_1 = \gamma_2^{-1} = \gamma$ and, as such, are strongly positive if $\theta \in [0, 2[$ (see Lemmas 5.1 and 5.3). Next, consider step 3 of Algorithm 1 and Substitute the parameters defined in (57), to derive

$$\frac{\alpha_n}{\lambda_n} = \frac{\|\tilde{z}_n\|_P^2}{\|\tilde{z}_n\|_D^2} = \frac{\gamma^{-1}\|\tilde{x}_n\|^2 + \gamma\|\tilde{y}_n\|^2 - \theta\langle \tilde{x}_n, \tilde{y}_n\rangle}{\gamma^{-1}(\theta^2 - 3\theta + 3)\|\tilde{x}_n\|^2 + \gamma\|\tilde{y}_n\|^2 + 2(1-\theta)\langle \tilde{x}_n, \tilde{y}_n\rangle}, \tag{94}$$

where D is defined in (4) and by construction is strongly positive. Let λ_n be equal to the inverse of the right hand side in (94) multiplied by ρ_n where $(\rho_n)_{n\in\mathbb{N}} \subseteq]0, 2[$. This would result in $\alpha_n = \rho_n$ and simplify the iterations. Consider C defined in (56c), we have

$$\|Cz - Cz'\|_{P^{-1}}^2 = \gamma\left(1 - \tfrac{1}{4}\theta^2\right)^{-1}\|Cz - Cz'\|^2.$$

From here it is easy to see that C is β-cocoercive with respect to P norm, where $\beta = \eta\gamma^{-1}(1 - \frac{1}{4}\theta^2)$. In order to apply Theorem 3.1, condition (6) must hold. From strong positivity of D, boundedness of P and the fact that ρ_n is uniformly bounded above 0 it follows that $(\lambda_n)_{n\in\mathbb{N}} \subseteq]\nu_1, \infty[$ for some positive ν_1. Let ν_2 be a positive parameter such that

$$\nu_2 < 2 - \frac{1}{2\beta} - \frac{2\rho_n(2 + \sqrt{2-\theta})}{2+\theta}, \tag{95}$$

holds for all $n \in \mathbb{N}$. It's easy to verify that such ν_2 exists as long as ρ_n is uniformly bounded in the interval (59). For brevity, we define β' such that $\frac{1}{2\beta'} = \frac{1}{2\beta} + \nu_2$. Additionally, introduce the notation $\upsilon = \rho_n^{-1}(2 - \frac{1}{2\beta'})$, $\omega_1 = \upsilon\theta + 2(1-\theta)$, $\omega_2 = \theta^2 - 3\theta + 3$. We proceed by showing that λ_n is smaller than $2 - \frac{1}{2\beta} - \nu_2$. A sufficient condition for this to hold is

$$\xi = \gamma^{-1}(\upsilon - \omega_2)\|\tilde{x}_n\|^2 + \gamma(\upsilon - 1)\|\tilde{y}_n\|^2 - \omega_1\langle \tilde{x}_n, \tilde{y}_n\rangle > 0. \tag{96}$$

Apply the Fenchel-Young inequality for $\frac{\epsilon}{2}\|\cdot\|^2$ to lower bound ξ:

$$\xi \geq \left(\gamma^{-1}(\upsilon - \omega_2) - |\omega_1|\tfrac{\epsilon}{2}\right)\|\tilde{x}_n\|^2 + \left(\gamma(\upsilon - 1) - |\omega_1|\tfrac{1}{2\epsilon}\right)\|\tilde{y}_n\|^2, \tag{97}$$

where $|\cdot|$ is the absolute value. It follows from (95) that $\upsilon > 1$ for all $n \in \mathbb{N}$. Let $\epsilon = \frac{|\omega_1|}{2\gamma(\upsilon-1)}$ so that the term involving $\|\tilde{y}_n\|^2$ in (97) disappears. We obtain

$$\gamma^{-1}\left(\upsilon - \omega_2 - \tfrac{\omega_1^2}{4(\upsilon-1)}\right)\|\tilde{x}_n\|^2 > 0, \tag{98}$$

which is sufficient for (96) to hold. By substituting ω_1 and ω_2 and after some algebraic manipulations we find that the condition (95) is sufficient for the right hand side in (98) to be positive. Consequently, Theorem 3.1 completes the proof of convergence. We showed that we can set $\alpha_n = \rho_n$ by choosing λ_n appropriately. Algorithm 2 follows by setting $\alpha_n = \rho_n$, a change of variables $s_n = x_n - \gamma y_n$, substituting x_{n+1} and application of Moreau's identity.

(*ii*): Mimic the proof of (*i*), but use Theorem 3.1 with $C \equiv 0$, by showing that λ_n is uniformly bounded between 0 and 2.

(*iii*): When $\theta = 2$, we have $P \in \mathcal{S}(\mathcal{K})$, $P \succeq 0$. It follows from (56a), (58) and Lemma 3.1 that $(H + A)^{-1}P$ is continuous. Therefore, since $F \equiv 0$, by appealing to Theorem 3.4 and following the same change of variables as in previous parts the assertion is proved. □

Proof of Proposition 4.2. Following the argument in proof of Proposition 4.1(*iii*) and Theorem 3.4(*ii*) yields

$$\|Pz_{n+1} - Pz_n\|^2 \leq \tfrac{\|P\|}{\tau(n+1)}\|Qz_0 - Qz^\star\|_R^2, \tag{99}$$

and $\|Pz_{n+1} - Pz_n\|^2 = o(1/(n+1))$, where $z_n = \big(x_n, \gamma^{-1}(x_n - s_n)\big)$. Combine this with definition of P, (57a) with $\theta = 2$, to derive

$$\|Pz_{n+1} - Pz_n\|^2 = (1 + \gamma^{-2})\|s_{n+1} - s_n\|^2. \tag{100}$$

Furthermore, simple calculation shows that $P^2 = (\gamma + \gamma^{-1})P$. Hence for all $z \in \mathcal{K}$

$$\begin{aligned}\|Pz\|^2 &= (\gamma + \gamma^{-1})\langle z, Pz\rangle \\ &= (\gamma + \gamma^{-1})\left(\gamma^{-1}\|x\|^2 + \gamma\|y\|^2 - 2\langle x, y\rangle\right) \\ &\leq (\gamma + \gamma^{-1})^2\|z\|^2,\end{aligned} \tag{101}$$

where we used Fenchel-Young inequality with $\epsilon = \gamma$. It follows from (101) that $\|P\| \leq (\gamma + \gamma^{-1})$. Combining this with (99) and (100) completes the proof. □

Proof of Lemma 5.2. For every z and z_1 in $\mathcal{K}$ we have

$$\begin{aligned}\langle Cz - Cz', z - z'\rangle &= \langle \nabla h(x) - \nabla h(x'), x - x'\rangle + \langle \nabla l^*(y) - \nabla l^*(y'), y - y'\rangle \\ &\geq \frac{1}{\beta_h}\|\nabla h(x) - \nabla h(x')\|^2 + \frac{1}{\beta_l}\|\nabla l^*(y) - \nabla l^*(y')\|^2 \\ &\geq \min\{\beta_l^{-1}, \beta_h^{-1}\}\|Cz - Cz'\|^2. \end{aligned} \tag{102}$$

It follows from Lemma 5.1 that $P \in \mathcal{S}_\tau(\mathcal{K})$ with τ defined in (69). Therefore, we have $\|P^{-1}\| \leq \tau^{-1}$. Using (102), we have

$$\begin{aligned}\langle Cz - Cz', z - z'\rangle &\geq \min\{\beta_l^{-1}, \beta_h^{-1}\}\|Cz - Cz'\|^2 \\ &\geq \|P^{-1}\|^{-1}\min\{\beta_l^{-1}, \beta_h^{-1}\}\|Cz - Cz'\|_{P^{-1}}^2 \\ &\geq \tau\min\{\beta_l^{-1}, \beta_h^{-1}\}\|Cz - Cz'\|_{P^{-1}}^2.\end{aligned}$$

Hence, C is β-cocoercive with respect to $\|\cdot\|_P$, where $\beta = \tau\min\{\beta_l^{-1}, \beta_h^{-1}\}$. In the case when $l = \iota_{\{0\}}$, we have $\nabla l^* = 0$ and it is possible to derive less conservative parameters. Define the linear operator $Q : (x, y) \to (x, 0)$, then,

$$\begin{aligned}\|Qz\|_{P^{-1}}^2 &= \langle x, (\gamma_1^{-1}\,\mathrm{Id} - \frac{\gamma_2}{4}\theta^2 L^*L)^{-1}x\rangle \leq \|x\|^2\|(\gamma_1^{-1}\,\mathrm{Id} - \frac{\gamma_2}{4}\theta^2 L^*L)^{-1}\| \\ &\leq \|Qz\|^2(\gamma_1^{-1} - \frac{\gamma_2}{4}\theta^2\|L\|^2)^{-1},\end{aligned}$$

where we used (65). For $\beta_h \in]0, +\infty[$, considering the above inequality and the fact that $Cz = (\nabla h(x), 0)$ we obtain

$$\begin{aligned}\|Cz - Cz'\|_{P^{-1}}^2 &\leq (\gamma_1^{-1} - \frac{\gamma_2}{4}\theta^2\|L\|^2)^{-1}\|\nabla h(x) - \nabla h(x')\|^2 \\ &\leq \beta_h(\gamma_1^{-1} - \frac{\gamma_2}{4}\theta^2\|L\|^2)^{-1}\langle \nabla h(x) - \nabla h(x'), x - x'\rangle \\ &= \beta_h(\gamma_1^{-1} - \frac{\gamma_2}{4}\theta^2\|L\|^2)^{-1}\langle Cz - Cz', z - z'\rangle,\end{aligned}$$

where the last inequality follows from the Baillon-Haddad theorem [1, Corollary 18.16]. This shows that C is β-cocoercive with respect to $\|\cdot\|_P$, with β defined in (72). □

Proof of Lemma 5.3. We start by showing that S_2 is strongly positive. Lemma 5.1 asserts that $P \in \mathcal{S}_\tau(\mathcal{K})$ with τ defined in (69). Simple algebra shows that $\tau \leq \min\{\gamma_1^{-1}, \gamma_2^{-1}\}$. If $\theta \in [2, \infty[$ define a constant c, such that $0 < c < \tau$, otherwise if $\theta \in [0, 2[$ set c such that $0 < c < \min\{\tau, \frac{(2-\theta)\gamma_2^{-1}}{4}\}$. Then for all $z = (x, y) \in \mathcal{K}$, $\langle (P - c\,\mathrm{Id})z, z\rangle > 0$ which implies

$$(\gamma_1^{-1} - c)\|x\|^2 - \tfrac{\theta^2}{4(\gamma_2^{-1} - c)}\|Lx\|^2 > 0. \tag{103}$$

It follows from (74b) that

$$\begin{aligned}\langle z, S_2 z\rangle &= \gamma_1^{-1}\|x\|^2 - 2\langle Lx, y\rangle + \gamma_2^{-1}\|y\|^2 + \gamma_2(2-\theta)\|Lx\|^2 \\ &= c\|z\|^2 + (\gamma_1^{-1} - c)\|x\|^2 - 2\langle Lx, y\rangle + (\gamma_2^{-1} - c)\|y\|^2 + \gamma_2(2-\theta)\|Lx\|^2 \\ &\geq c\|z\|^2 + (\gamma_1^{-1} - c)\|x\|^2 + (\gamma_2^{-1} - c - \epsilon)\|y\|^2 + (\gamma_2(2-\theta) - \epsilon^{-1})\|Lx\|^2,\end{aligned}$$

where the inequality follows from Fenchel-Young inequality for $\frac{\epsilon}{2}\|\cdot\|^2$. Let $\epsilon = \gamma_2^{-1} - c$ so that the term involving $\|y\|^2$ disappears, and use (103) to derive

$$\begin{aligned}\langle z, S_2 z\rangle \geq\ & c\|z\|^2 + (\gamma_1^{-1} - c)\|x\|^2 + \left(\gamma_2\,(2-\theta) - \tfrac{1}{\gamma_2^{-1}-c}\right)\|Lx\|^2 \\ >\ & c\|z\|^2 + \left(\gamma_2(2-\theta) + \left(\tfrac{\theta^2}{4} - 1\right)\tfrac{1}{\gamma_2^{-1}-c}\right)\|Lx\|^2. \end{aligned} \tag{104}$$

It follows from the choice of c for both $\theta \in [0, 2[$ and $\theta \in [2, \infty[$ that the last term in (104) is nonnegative and hence, $S_2 \in \mathcal{S}_c(\mathcal{K})$. The analysis for S_1 is similar and have been omitted here. Hence, it follows that S_1^{-1} and S_2^{-1} are strongly positive and so is their convex combination and its inverse, S. Furthermore, It was shown in (5) that if S and P are strongly positive, K skew adjoint and M monotone, then $D \in \mathcal{S}_\nu(\mathcal{K})$ for some $\nu \in]0, \infty[$, where D was defined in (4). We conclude by noting that D in (76) can be derived by substituting S in (4). □

Proof of Proposition 5.2. Algorithm 4 is a special case of Algorithm 3 and consequently of Algorithm 1. Let us establish the proof by directly showing that Theorem 3.1 is applicable with $(\lambda_n)_{n\in\mathbb{N}}$ selected such that $\alpha_n = 1$, while satisfying (6). For all (*i*)-(*iii*), strong positivity of the operators P, S, and D follow from Lemmas 5.1 and 5.3. Set

$$\lambda_n = \frac{\|\tilde{z}_n\|_D^2}{\|\tilde{z}_n\|_P^2} = 1 + \frac{\gamma_2\|L\tilde{x}_n\|^2 + \gamma_1\|L^*\tilde{y}_n\|^2}{\gamma_1^{-1}\|\tilde{x}_n\|^2 + \gamma_2^{-1}\|\tilde{y}_n\|^2}. \tag{105}$$

Such a choice would yield $\alpha_n = 1$. A sufficient condition for (6) to hold is to show that $(\lambda_n)_{n\in\mathbb{N}}$ is uniformly bounded between 0 and δ. From strong positivity of D and boundedness of P it follows that $(\lambda_n)_{n\in\mathbb{N}} \subseteq [\nu_1, \infty[$ for some positive ν_1. Denote by β the cocoercivity constant of C with respect to $\|\cdot\|_P$. We proceed by showing that $(\lambda_n)_{n\in\mathbb{N}} \subseteq [\nu_1, 2 - \frac{1}{2\beta} - \nu_2]$. Assume that

$$0 < 1 - \gamma_1\gamma_2\|L\|^2 - \tfrac{1}{2\beta}. \tag{106}$$

Then, there exists a positive constant ν_2 such that

$$\nu_2 < 1 - \gamma_1\gamma_2\|L\|^2 - \tfrac{1}{2\beta}. \tag{107}$$

It follows from (105) that

$$\lambda_n \leq 1 + \frac{\gamma_2\|L\|^2\|\tilde{x}_n\|^2 + \gamma_1\|L\|^2\|\tilde{y}_n\|^2}{\gamma_1^{-1}\|\tilde{x}_n\|^2 + \gamma_2^{-1}\|\tilde{y}_n\|^2} < 2 - \tfrac{1}{2\beta} - \nu_2,$$

where in the second inequality we used (107) to upper bound both $\gamma_2\|L\|^2$ and $\gamma_1\|L\|^2$. We showed that (106) is sufficient for (6) to hold. Consequently, Theorem 3.1 completes the proof of convergence.

(*i*): By Lemma 5.2, substitute β in (106) with (71) to derive (83). The algorithm follows by substituting $\alpha_n = 1$ in Algorithm 3 when $\theta = 0$ and $\mu = \frac{1}{2}$.

(*ii*): Similar to (*i*), since $l = \iota_{\{0\}}$, substitute β in (106) with (72), i.e. $\beta = \beta_h^{-1}\gamma_1^{-1}$. This choice of β results in (84) and the algorithm follows.

(*iii*): The proof is similar to previous parts, but we show that $(\lambda_n)_{n\in\mathbb{N}}$ is uniformly bounded between 0 and 2, and then use Theorem 3.1 with $C \equiv 0$. □