ORIGINAL RESEARCH

A Direct Proof of Convergence of Davis–Yin Splitting Algorithm Allowing Larger Stepsizes

Francisco J. Aragón-Artacho[1] [ORCID] **· David Torregrosa-Belén[1]**

Abstract

This note is devoted to the splitting algorithm proposed by Davis and Yin (Set-valued Var. Anal. **25**(4), 829–858, 2017) for computing a zero of the sum of three maximally monotone operators, with one of them being cocoercive. We provide a direct proof that guarantees its convergence when the stepsizes are smaller than four times the cocoercivity constant, thus doubling the size of the interval established by Davis and Yin. As a by-product, the same conclusion applies to the forward-backward splitting algorithm. Further, we use the notion of "strengthening" of a set-valued operator to derive a new splitting algorithm for computing the resolvent of the sum. Last but not least, we provide some numerical experiments illustrating the importance of appropriately choosing the stepsize and relaxation parameters of the algorithms.

Keywords Monotone inclusion · Resolvent · Splitting algorithm · Forward-backward · Strengthening

Mathematics Subject Classification (2010) 47H05 · 90C30 · 65K05

1 Introduction

When a problem has certain structure, it is normally useful to take advantage of it. Following the divide-and-conquer paradigm, splitting algorithms iteratively solve simpler problems which are defined by separately using some parts of the original problem. A particular

✉ Francisco J. Aragón-Artacho
   francisco.aragon@ua.es

   David Torregrosa-Belén
   david.torregrosa@ua.es

[1] Department of Mathematics, University of Alicante, Alicante, Spain

subfamily are projection methods (see, e.g., [10, Chapter 5]), which can be used to find a common point in the intersection of sets, based on projections of points defined in the iterations into each of the sets. These methods are usually variations of classical iterative schemes for finding fixed points of certain type of nonexpansive operators. Monotone operator theory [7] permits to generalize these algorithms to tackle the far more general problem of finding a zero of the sum of maximally monotone operators by using their resolvents instead of the projectors (see Definitions 2.2 and 2.4).

There are many different splitting algorithms for computing a zero of the sum of two maximally monotone operators (see, e.g., [7, Chapter 26]). Theoretically, one can always transform any splitting algorithm for computing zeros of the sum of two operators into a splitting algorithm for computing zeros of the sum of finitely many operators (see, e.g., [7, Proposition 26.4]), thanks to Pierra's product space reformulation [26]. Nevertheless, numerical experience shows that this theoretical trick usually slows down the resulting algorithm (see, e.g., [1, Section 6.1]), especially when the number of operators is large (see, e.g., [2, Section 4] and [8, Section 5]). To alleviate this problem, various schemes requiring one space less in the product space have been recently proposed [9, 14, 23].

Only recently, three-operator splitting algorithms have been developed [18, 21, 27–29]. This note is devoted to one of them, which was introduced by Damek Davis and Wotao Yin in [18], and is commonly referred as *Davis–Yin splitting algorithm*. The algorithm is designed for solving the problem

$$\text{find } x \text{ such that } 0 \in (A + B + T)(x), \tag{1}$$

where all three operators involved are maximally monotone and act on a Hilbert space, and T is also cocoercive (see Definition 2.1). Davis and Yin defined the operator

$$DY_\gamma := J_{\gamma B} \circ \left(2J_{\gamma A} - \text{Id} - \gamma T \circ J_{\gamma A}\right) + \text{Id} - J_{\gamma A}, \tag{2}$$

where $J_{\gamma A}$ and $J_{\gamma B}$ denote the corresponding resolvents, and proved that DY_γ is α-*averaged* for $\alpha = \frac{2\beta}{4\beta - \gamma}$ when $\gamma \in]0, 2\beta[$, where $\beta > 0$ is the cocoercivity constant of T. Then, they defined their splitting algorithm through the standard Krasnosel'skiĭ–Mann iteration

$$x_{k+1} = (1 - \lambda_k)x_k + \lambda_k DY_\gamma(x_k), \quad k = 0, 1, 2, \ldots, \tag{3}$$

with $\lambda_k \in]0, 1/\alpha[$ satisfying the assumptions of [7, Proposition 5.16], from which its convergence to a fixed point x of DY_γ follows. Further, the *shadow sequence* $(J_{\gamma A}(x_k))_{k \in \mathbb{N}}$ weakly converges to a solution of (1), and convergence is strong under additional assumptions. Three well-known splitting algorithms can be obtained as a particular instance of Davis–Yin's, namely the Douglas–Rachford [22] (when $T = 0$), the forward-backward [22, 25] (when $A = 0$) and the backward-forward [3] (when $B = 0$).

In this note we provide a direct proof of the convergence of the iterative method (3) without relying on the averagedness of the operator DY_γ (see Theorem 3.3). Our proof has two key advantages: (i) it permits to simplify the assumptions on the relaxation parameters, and (ii) it allows to choose the stepsize γ in $]0, 4\beta[$ instead of $]0, 2\beta[$. Observe that the operator DY_γ does not need to be averaged when $\gamma > 2\beta$ (for instance, take $A = B = 0$, T the identity, and apply DY_γ to the points $x = 1$ and $z = -1$). As a by-product, this shows that the stepsize in the forward-backward and the backward-forward algorithms can be also chosen in $]0, 4\beta[$. In addition, we derive in Theorem 3.7 a *strengthened* version of Davis–Yin splitting algorithm which permits computing the resolvent of $A + B + T$.

Right before submitting this manuscript, we learnt about a recent preprint by Dao-Phan, which now has been published [17]. Using the notion of *conically averaged* operators introduced in [5], the authors prove in [17, Corollary 4.2] that the operator $(1 - \lambda) \, \text{Id} + \lambda D Y_\gamma$ is $2\lambda\beta/(4\beta - \gamma)$-averaged when $\gamma \in \,]0, 4\beta[$, from which the convergence of (3) for a fixed $\lambda_k = \lambda$ follows.

As a simple motivating example of the importance of the algorithm parameters, consider the problem of finding the minimum norm point in the intersection of two balls $\mathbb{A}$ and $\mathbb{B}$ in the Euclidean space whose intersection has nonempty interior. The problem can be solved with Davis–Yin splitting algorithm, taking A and B as the normal cones to the respective balls, and T as the identity mapping. Since the resolvents of the normal cones are the projectors (see Example 2.7), which we denote by $P_\mathbb{A}$ and $P_\mathbb{B}$, the iterative scheme is given by

$$x_{k+1} = x_k - \lambda_k P_\mathbb{A}(x_k) + \lambda_k P_\mathbb{B} \left((2 - \gamma) P_\mathbb{A}(x_k) - x_k \right), \quad k = 0, 1, 2, \ldots,$$

and $(P_\mathbb{A}(x_k))_{k \in \mathbb{N}}$ converges to the minimum norm point in $\mathbb{A} \cap \mathbb{B}$ (the normal cone sum rule holds). Both the *relaxation parameter* λ_k and the *stepsize* γ have a big influence on the behavior of the algorithm, as shown in Fig. 1.

In this example, since the cocoercivity constant β is equal to 1, [18, Theorem 2.1] guarantees the convergence when the parameter γ is taken in $]0, 2[$, while Theorem 3.3 allows to take $\gamma \in \,]0, 4[$. When the Davis–Yin splitting algorithm is applied to the same problem with different starting points x_0, it can behave very differently depending on the parameters, as shown in Figs. 1 and 2.

In general, larger stepsizes are commonly believed to be associated with faster convergence of algorithms, but this is not always the case, particularly when an algorithm has several parameters. It is important to have in mind that the relaxation parameter λ_k of the Davis–Yin splitting algorithm is upper bounded by $2 - \frac{\gamma}{2\beta}$ and that its value has an important effect. If $\gamma \in \,]0, 2\beta[$, *overrelaxed steps* (i.e., $\lambda_k > 1$) are allowed in (3), while only *underrelaxed steps* can be taken when $\gamma \geq 2\beta$. The fact that both the stepsize and the relaxation parameters are important is especially apparent when one considers the particular case of $A = B = 0$ and $T = \nabla f$ for a differentiable function f whose gradient is Lipschitz continuous with constant $L = \frac{1}{\beta}$. In this case, the iteration (3) reduces to the gradient descent
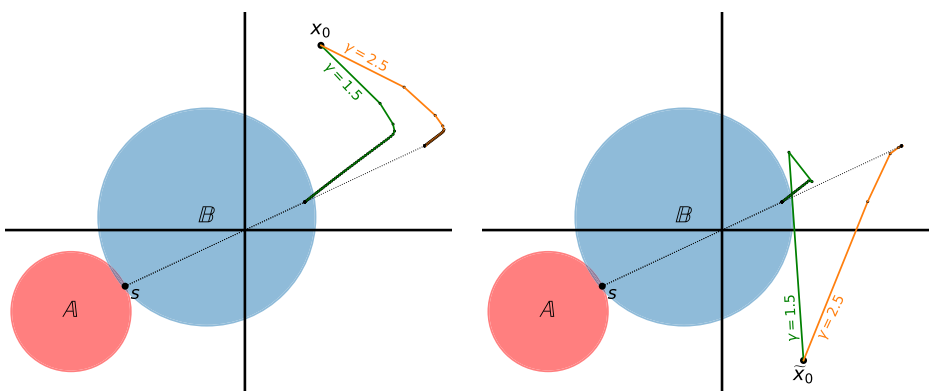
Fig. 1 Behavior of Davis–Yin splitting algorithm for two starting points x_0 and $\widetilde{x}_0$ and two stepsize parameters γ, with $\lambda_k = 0.99(2 - \gamma/2)$. The solution s is obtained after projecting the fixed point onto $\mathbb{A}$
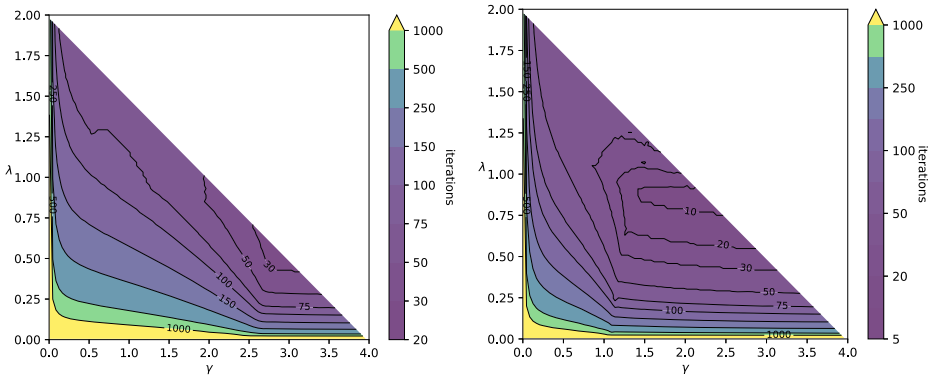
Fig. 2 Number of iterations needed until the shadow sequence gets sufficiently close to the solution s (precisely, $\|P_A(x_k) - s\| < 10^{-10}$) for different values of γ and $\lambda_k = \lambda$, with starting points x_0 (left) and $\widetilde{x}_0$ (right) shown in Fig. 1

scheme:

$$x_{k+1} = x_k - \gamma \lambda_k \nabla f(x_k), \quad k = 0, 1, 2, \ldots. \tag{4}$$

We observe in (4) that the stepsize of the algorithm is actually $\gamma \lambda_k$, so the upper bound $2 - \frac{\gamma}{2\beta}$ on the relaxation parameters λ_k entails $\gamma \lambda_k < 2\beta = \frac{2}{L}$, as expected.

Finally, it is important to recall that in practical applications only a lower bound of the best cocoercivity constant β is usually known, and this can affect the performance of the algorithms. For instance, consider again the application of the Davis–Yin algorithm with starting point $\widetilde{x}_0$ shown on the right in Fig. 2 and imagine that we underestimate β to $\widehat{\beta} = 0.65 < 1 = \beta$. Then, we observe in Fig. 3 how the choice of a stepsize parameter $\gamma \in]0, 2\widehat{\beta}[$ excludes better values like $\widehat{\gamma} \in]2\widehat{\beta}, 4\widehat{\beta}[$. A typical choice for the parameters of the forward-backward algorithm is $\gamma = (2 - \varepsilon)\beta$ and $\lambda_k = 1$, for a small $\varepsilon > 0$ (see, e.g., [13]).
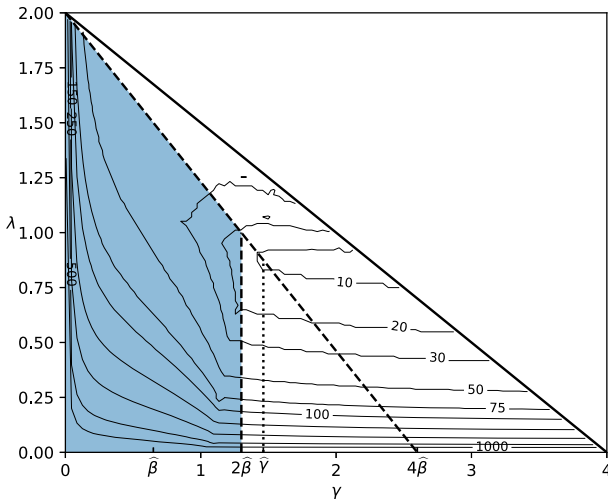
Fig. 3 Repetition of the experiment shown on the right of Fig. 2. When only an approximate value $\widehat{\beta}$ of the cocoercivity constant is known, choosing the stepsize $\gamma \in]0, 2\widehat{\beta}[$ (shaded area) can exclude better choices like $\widehat{\gamma}$

This example shows that, when only an estimate $\widehat{\beta}$ of the best value of β is known, it can be worth testing the performance of the algorithm with parameters $\gamma = (2+\varepsilon)\widehat{\beta}$ and $\lambda_k = 1-\varepsilon$ (i.e., with underrelaxation).

The remainder of this paper is structured as follows. In Section 2 we recall some preliminary notions and results. In Section 3 we provide an alternative proof of convergence of the Davis–Yin splitting algorithm and derive its *strengthened* version for computing the resolvent of the sum. In Section 4 we include some illustrative numerical experiments. We finish with some conclusions in Section 5.

2 Preliminaries

Throughout this paper, $\mathcal{H}$ is a real Hilbert space equipped with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$. We abbreviate *norm convergence* of sequences in $\mathcal{H}$ with $\to$ and we use $\rightharpoonup$ for *weak convergence*.

A *set-valued operator* is a mapping $A : \mathcal{H} \rightrightarrows \mathcal{H}$ that assigns to each point in $\mathcal{H}$ a subset of $\mathcal{H}$, i.e., $A(x) \subseteq \mathcal{H}$ for all $x \in \mathcal{H}$. In the case when A always maps to singletons, i.e., $A(x) = \{u\}$ for all $x \in \mathcal{H}$, A is said to be a *single-valued mapping* and is denoted by $A : \mathcal{H} \to \mathcal{H}$. In an abuse of notation, we may write $A(x) = u$ when $A(x) = \{u\}$. The *domain*, the *range*, the *graph*, the set of *fixed points* and the set of *zeros* of A, are denoted, respectively, by dom A, ran A, gra A, Fix A and zer A; i.e.,

$$\text{dom } A := \{x \in \mathcal{H} : A(x) \neq \emptyset\}, \qquad \text{ran } A := \{u \in \mathcal{H} : \exists x \in \mathcal{H} : u \in A(x)\},$$
$$\text{gra } A := \{(x, u) \in \mathcal{H} \times \mathcal{H} : u \in A(x)\}, \qquad \text{Fix } A := \{x \in \mathcal{H} : x \in A(x)\},$$
$$\text{and} \quad \text{zer } A := \{x \in \mathcal{H} : 0 \in A(x)\}.$$

The *inverse operator* of A, denoted by A^{-1}, is defined through $x \in A^{-1}(u) \iff u \in A(x)$. The *identity operator* is denoted by Id.

Definition 2.1 We say that an operator $T : \mathcal{H} \to \mathcal{H}$ is

(i)   *L-Lipschitz continuous* for $L > 0$ if

$$\|T(x) - T(y)\| \leq L\|x - y\| \quad \forall x, y \in \mathcal{H};$$

(ii)   *β-cocoercive* for $\beta > 0$ if

$$\langle x - y, T(x) - T(y) \rangle \geq \beta\|T(x) - T(y)\|^2 \quad \forall x, y \in \mathcal{H}.$$

Note that, by the Cauchy–Schwarz inequality, any β-cocoercive mapping is $\frac{1}{\beta}$-Lipschitz continuous. When the operator is the gradient of a convex function, the Baillon–Haddad theorem states that both notions are equivalent, see [4, Corolaire 10].

Definition 2.2 Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be a set-valued operator.

(i)   A is said to be *η-monotone* for $\eta \in \mathbb{R}$ if

$$\langle x - y, u - v \rangle \geq \eta\|x - y\|^2 \quad \forall (x, u), (y, v) \in \text{gra } A.$$

Furthermore, an η-monotone operator A is said to be *maximally η-monotone* if there exists no η-monotone operator $B : \mathcal{H} \rightrightarrows \mathcal{H}$ such that gra B properly contains gra A.

(ii)   *A* is said to be *uniformly monotone* with modulus $\phi : \mathbb{R}_+ \to [0, +\infty[$ if ϕ is increasing, vanishes only at 0, and

$$\langle x - y, u - v \rangle \geq \phi(\|x - y\|) \quad \forall (x, u), (y, v) \in \text{gra } A.$$

An operator is monotone (in the classical sense) if it is 0-monotone and it is η-strongly monotone (in the classical sense) if it is η-monotone for $\eta > 0$, in which case it is uniformly monotone with modulus $\phi(t) = \eta t^2$, for $t \in \mathbb{R}_+$.

Definition 2.3 We say that an operator $T : \mathcal{H} \to \mathcal{H}$ is *demiregular* at $x \in \mathcal{H}$ if for all sequences $(x_k)_{k \in \mathbb{N}}$ with $x_k \rightharpoonup x$ and $T(x_k) \to T(x)$, we have $x_k \to x$.

The resolvent operator, whose definition is given next, is one of the main building blocks of splitting algorithms.

Definition 2.4 Given an operator $A \colon \mathcal{H} \rightrightarrows \mathcal{H}$, the *resolvent* of A with parameter $\gamma > 0$ is the operator $J_{\gamma A} \colon \mathcal{H} \rightrightarrows \mathcal{H}$ defined by $J_{\gamma A} := (\text{Id} + \gamma A)^{-1}$.

The following result is a consequence of Minty's theorem [24].

Proposition 2.5 (Resolvents of η-monotone operators) *Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ be η-monotone and let $\gamma > 0$ such that $1 + \gamma\eta > 0$. Then*

(i)   *$J_{\gamma A}$ is single-valued,*
(ii) *$\text{dom } J_{\gamma A} = \mathcal{H}$ if and only if A is maximally η-monotone.*

Proof See [15, Proposition 3.4]. □

Example 2.6 Let $f : \mathcal{H} \to]-\infty, +\infty]$ be a proper, lower semicontinuous (lsc) and convex function. Then, the *subdifferential* of f, which is the operator $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$ defined as

$$\partial f(x) = \{u \in \mathcal{H} : f(x) + \langle u, y - x \rangle \leq f(y), \ \forall y \in \mathcal{H}\},$$

is a maximally monotone operator. Furthermore, it holds that $J_{\gamma \partial f} = \text{prox}_{\gamma f} : \mathcal{H} \to \mathcal{H}$, where $\text{prox}_{\gamma f}$ is the *proximity operator* of f (with parameter γ) defined at $x \in \mathcal{H}$ by

$$\text{prox}_{\gamma f}(x) := \underset{u \in \mathcal{H}}{\text{argmin}} \left(f(u) + \frac{1}{2\gamma} \|x - u\|^2 \right),$$

see, e.g., [7, Theorem 20.25 & Example 23.3]. Some functions are *prox-friendly*, which means that their proximity operator is easy to compute, see [11] for various examples. This is the case for the ℓ_1-norm, whose proximity operator is the result of applying the *soft thresholding* function:

$$\text{prox}_{\gamma \|\cdot\|_1}(x) = \text{sign}(x) \odot [|x| - \gamma]_+,$$

where $\odot$ denotes element-wise product and $[\,\cdot\,]_+$ and $|\cdot|$ are applied element-wise. That is, its i-th component is given by

$$\text{prox}_{\gamma \|\cdot\|_1}(x)_i = \begin{cases} x_i + \gamma, & \text{if } x_i < -\gamma, \\ 0, & \text{if } |x_i| \leq \gamma, \\ x_i - \gamma, & \text{if } x_i > \gamma, \end{cases}$$

for $i = 1, 2, \ldots, n$.

Example 2.7 Given a nonempty set $C \subseteq \mathcal{H}$, the *indicator function* of C, $\iota_C : \mathcal{H} \to]-\infty, \infty]$, is defined as

$$\iota_C(x) := \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{if } x \notin C. \end{cases}$$

When C is a convex set, ι_C is a convex function whose subdifferential becomes the *normal cone* to C, $N_C : \mathcal{H} \rightrightarrows \mathcal{H}$, given by

$$\partial \iota_C(x) = N_C(x) := \begin{cases} \{u \in \mathcal{H} : \langle u, c - x \rangle \leq 0, \ \forall c \in C\}, & \text{if } x \in C, \\ \emptyset, & \text{otherwise.} \end{cases}$$

When C is nonempty, closed and convex, the normal cone N_C is maximally monotone. Furthermore, $J_{N_C} = P_C$, where $P_C : \mathcal{H} \to \mathcal{H}$ denotes the *projector* onto C, which is defined at $x \in \mathcal{H}$ by

$$P_C(x) := \operatorname*{argmin}_{c \in C} \|x - c\|,$$

see, e.g., [7, Example 20.26 & Example 23.4].

Fejér monotonicity is a key property in fixed point theory (see, e.g, [7, Chapter 5]). It will allow us to derive weak convergence of the sequence generated by the Davis–Yin splitting algorithm.

Definition 2.8 Let C be a nonempty subset of $\mathcal{H}$ and let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$. Then $(x_n)_{n \in \mathbb{N}}$ is *Fejér monotone* with respect to C if for all $x \in C$

$$\|x_{n+1} - x\| \leq \|x_n - x\| \quad \forall n \in \mathbb{N}.$$

Proposition 2.9 *Let C be a nonempty subset of $\mathcal{H}$ and let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$. Suppose that $(x_n)_{n \in \mathbb{N}}$ is Fejér monotone with respect to C and that every weak sequential cluster point of $(x_n)_{n \in \mathbb{N}}$ belongs to C. Then $(x_n)_{n \in \mathbb{N}}$ converges weakly to a point in C.*

Proof See [7, Theorem 5.5]. □

3 Davis–Yin Splitting Algorithm

Let $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ be two maximally monotone operators and let $T : \mathcal{H} \to \mathcal{H}$ be cocoercive. Consider the problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in (A + B + T)(x). \tag{5}$$

The following lemma characterizes the set of zeros of the latter sum of operators in terms of the set

$$\Omega_\gamma := \left\{ x \in \mathcal{H} : J_{\gamma A}(x) = J_{\gamma B}\big(2J_{\gamma A}(x) - x - \gamma T(J_{\gamma A}(x))\big) \right\}, \tag{6}$$

with $\gamma > 0$, and shows that $\Omega_\gamma = \text{Fix } DY_\gamma$, where

$$\text{Fix } DY_\gamma = \{u + \gamma y : u \in \text{zer}(A + B + T), y \in (-B(u) - T(u)) \cap A(u)\}, \tag{7}$$

as proved in [18, Lemma 2.2].

Lemma 3.1 *For every $\gamma > 0$, it holds*

$$\text{zer}(A + B + T) = J_{\gamma A}(\Omega_\gamma).$$

In particular, $\text{zer}(A + B + T) \neq \emptyset \iff \Omega_\gamma \neq \emptyset$. *Further,* $\Omega_\gamma = \text{Fix } DY_\gamma$.

Proof Observe that

$$u \in \text{zer}\,(A + B + T) \Leftrightarrow -\gamma T(u) \in (\gamma A + \gamma B)(u)$$

$$\Leftrightarrow (\exists x \in \mathcal{H}) \quad x - u \in \gamma A(u), \quad u - x - \gamma T(u) \in \gamma B(u)$$

$$\Leftrightarrow (\exists x \in \mathcal{H}) \quad u = J_{\gamma A}(x), \quad 2u - x - \gamma T(u) \in (\text{Id} + \gamma B)(u)$$

$$\Leftrightarrow (\exists x \in \mathcal{H}) \quad u = J_{\gamma A}(x), \quad u = J_{\gamma B}(2u - x - \gamma T(u)),$$

from where the first claim follows. Further, we have

$$x \in \Omega_\gamma \Leftrightarrow (\exists u \in \text{zer}\,(A + B + T)) \quad u = J_{\gamma A}(x), \quad u = J_{\gamma B}(2u - x - \gamma T(u))$$

$$\Leftrightarrow (\exists u \in \text{zer}\,(A + B + T)) \quad x - u \in \gamma A(u), \quad x - u \in (-\gamma B(u) - \gamma T(u))$$

$$\Leftrightarrow (\exists u \in \text{zer}\,(A + B + T)\,, \exists y \in (-B(u) - T(u)) \cap A(u)), \quad x = u + \gamma y,$$

and thus, $\Omega_\gamma = \text{Fix}\,DY_\gamma$, by (7). $\qquad \square$

Using a technique similar to the one employed in [1, Theorem 8], we can provide a direct proof of the convergence of Davis–Yin splitting algorithm with the additional advantages of both allowing a larger stepsize and having a simpler condition on the relaxation parameters than [18, Theorem 2.1]. The proof makes use of the following technical lemma.

Lemma 3.2 *Let A, $B : \mathcal{H} \rightrightarrows \mathcal{H}$ be two maximally monotone operators and $T : \mathcal{H} \to \mathcal{H}$. Let $x, \hat{x} \in \mathcal{H}$ and $\gamma > 0$, and set $u := J_{\gamma A}(x)$, $\hat{u} := J_{\gamma_A}(\hat{x})$, $v := J_{\gamma B}(2u - x - \gamma T(u))$ and $\hat{v} := J_{\gamma B}(2\hat{u} - \hat{x} - \gamma T(\hat{u}))$. Then, it holds*

$$0 \le \langle x - \hat{x}, (u - v) - (\hat{u} - \hat{v}) \rangle - \|(u - v) - (\hat{u} - \hat{v})\|^2 - \gamma \langle T(u) - T(\hat{u}), v - \hat{v} \rangle. \quad (8)$$

Further, if A (respectively B) is uniformly monotone with modulus ϕ, then (8) holds with 0 replaced by $\gamma\phi(\|u - \hat{u}\|)$ (respectively $\gamma\phi(\|v - \hat{v}\|)$).

Proof Since $x - u \in \gamma A(u)$ and $\hat{x} - \hat{u} \in \gamma A(\hat{u})$, monotonicity of γA yields

$$0 \le \langle (x - u) - (\hat{x} - \hat{u}), u - \hat{u} \rangle. \quad (9)$$

Likewise, since $2u - x - \gamma T(u) - v \in \gamma B(v)$ and $2\hat{u} - \hat{x} - \gamma T(\hat{u}) - \hat{v} \in \gamma B(\hat{v})$, monotonicity of γB implies

$$0 \le \langle (2u - x - \gamma T(u) - v) - (2\hat{u} - \hat{x} - \gamma T(\hat{u}) - \hat{v}), v - \hat{v} \rangle$$

$$= \langle (\hat{v} - \hat{u}) - (v - u), v - \hat{v} \rangle - \langle (x - u) - (\hat{x} - \hat{u}), v - \hat{v} \rangle - \gamma \langle T(u) - T(\hat{u}), v - \hat{v} \rangle. \quad (10)$$

Summing together (9) and (10), we obtain

$$0 \le \langle (x - u) - (\hat{x} - \hat{u}), (u - v) - (\hat{u} - \hat{v}) \rangle$$

$$+ \langle (\hat{v} - \hat{u}) - (v - u), v - \hat{v} \rangle - \gamma \langle T(u) - T(\hat{u}), v - \hat{v} \rangle$$

$$= \langle x - \hat{x}, (u - v) - (\hat{u} - \hat{v}) \rangle - \|(u - v) - (\hat{u} - \hat{v})\|^2 - \gamma \langle T(u) - T(\hat{u}), v - \hat{v} \rangle,$$

which proves (8). The last assertion easily follows from the definition of uniform monotonicity. $\qquad \square$

Theorem 3.3 (Davis–Yin splitting) *Let A, $B : \mathcal{H} \rightrightarrows \mathcal{H}$ be two maximally monotone operators and $T : \mathcal{H} \to \mathcal{H}$ be a β-cocoercive operator, with $\beta > 0$, such that $\text{zer}\,(A + B + T) \ne \emptyset$. Set a stepsize $\gamma \in\,]0, 4\beta[$ and consider a sequence of relaxation parameters $(\lambda_k)_{k \in \mathbb{N}}$*

in $]0, 2 - \gamma/(2\beta)]$ such that $\sum_{k\in\mathbb{N}} \lambda_k \left(2 - \frac{\gamma}{2\beta} - \lambda_k\right) = +\infty$. Given some initial point $x_0 \in \mathcal{H}$, consider the sequences defined by

$$
\begin{cases}
u_k = J_{\gamma A}(x_k) \\
v_k = J_{\gamma B}(2u_k - x_k - \gamma T(u_k)) \\
x_{k+1} = x_k + \lambda_k(v_k - u_k).
\end{cases}
\tag{11}
$$

Then, the sequence $(x_k)_{k\in\mathbb{N}}$ is Fejér monotone with respect to the set Ω_γ given in (6). Moreover, the following assertions hold:

(i) $x_k \rightharpoonup \bar{x} \in \Omega_\gamma$, $u_k \rightharpoonup \bar{u}$, $v_k \rightharpoonup \bar{u}$, $v_k - u_k \to 0$ and $T(u_k) \to T(\bar{u})$ with

$$
\bar{u} = J_{\gamma A}(\bar{x}) = J_{\gamma B}(2\bar{u} - \bar{x} - \gamma T(\bar{u})) \in \mathrm{zer}\,(A + B + T).
\tag{12}
$$

Further, $T(\mathrm{zer}(A + B + T)) = \{T(\bar{u})\}$.

(ii) If either A or B is uniformly monotone on every bounded subset of its domain, or T is demiregular at every point in $\mathrm{zer}\,(A + B + T)$, then $(u_k)_{k\in\mathbb{N}}$ and $(v_k)_{k\in\mathbb{N}}$ converge strongly to $\bar{u} \in \mathrm{zer}\,(A + B + T)$.

Proof Define the sequences

$$
(\forall k \in \mathbb{N}) \qquad z_k := \gamma T(u_k) \qquad \text{and} \qquad w_k := v_k - u_k
$$

and note the following relations that (11) yields

$$
(u_k, x_k - u_k) \in \mathrm{gra}\,\gamma A \quad \text{and} \quad (v_k, 2u_k - x_k - z_k - v_k) \in \mathrm{gra}\,\gamma B.
\tag{13}
$$

Pick any $x \in \Omega_\gamma$ and denote $u := J_{\gamma A}(x)$. By definition of Ω_γ, we have $u = J_{\gamma B}(2u - x - \gamma T(u))$. Applying Lemma 3.2 to x and $\hat{x} := x_k$, observing that $\hat{u} = u_k$, $v = u$ and $\hat{v} = v_k$, yields

$$
0 \leq \langle x - x_k, w_k \rangle - \|w_k\|^2 - \gamma \langle T(u) - T(u_k), u - v_k \rangle.
\tag{14}
$$

The first two terms in (14) multiplied by $2\lambda_k$ can be expressed as

$$
\begin{aligned}
2\lambda_k \left(\langle x - x_k, w_k \rangle - \|w_k\|^2\right) &= 2\langle x - x_k, x_{k+1} - x_k \rangle - 2\lambda_k\|w_k\|^2 \\
&= \|x_k - x\|^2 - \|x_{k+1} - x\|^2 + \lambda_k(\lambda_k - 2)\|w_k\|^2.
\end{aligned}
\tag{15}
$$

Now, using the β-cocoercivity of T, the last term in (14) can be expressed as

$$
\begin{aligned}
-\gamma \langle T(u) - T(u_k), u - v_k \rangle &= -\gamma \langle T(u) - T(u_k), u - u_k \rangle + \gamma \langle T(u) - T(u_k), w_k \rangle \\
&\leq -\beta\gamma\|T(u) - T(u_k)\|^2 + \gamma \langle T(u) - T(u_k), w_k \rangle.
\end{aligned}
\tag{16}
$$

Using Cauchy–Schwarz and Young's inequalities, the last term in (16) can be estimated as

$$
\gamma \langle T(u) - T(u_k), w_k \rangle \leq \beta\gamma\|T(u) - T(u_k)\|^2 + \frac{\gamma}{4\beta}\|w_k\|^2.
\tag{17}
$$

Combining (14)–(17), we have

$$
\|x_{k+1} - x\|^2 + \lambda_k(2 - \lambda_k)\|w_k\|^2 \leq \|x_k - x\|^2 + \frac{2\gamma\lambda_k}{4\beta}\|w_k\|^2.
$$

As a result, we reach the expression

$$
\|x_{k+1} - x\|^2 + \lambda_k \left(2 - \frac{\gamma}{2\beta} - \lambda_k\right) \|w_k\|^2 \leq \|x_k - x\|^2.
\tag{18}
$$

Since $\lambda_k \leq 2 - \gamma/(2\beta)$, (18) implies that $(x_k)_{k\in\mathbb{N}}$ is Fejér monotone with respect to Ω_γ and thus, bounded. Since resolvents are nonexpansive and T is $\frac{1}{\beta}$-Lipschitz continuous (by Cauchy–Schwarz), it follows that $(u_k)_{k\in\mathbb{N}}$, $(z_k)_{k\in\mathbb{N}}$ and $(v_k)_{k\in\mathbb{N}}$ are bounded.

(i) The Fejér monotonicity of $(x_k)_{k\in\mathbb{N}}$ implies that the sequence $(\|x_k - x\|)_{k\in\mathbb{N}}$ is nonincreasing and convergent. Telescoping (18), we obtain

$$\sum_{k\in\mathbb{N}} \lambda_k \left(2 - \frac{\gamma}{2\beta} - \lambda_k\right) \|w_k\|^2 \leq \|x_0 - x\|^2,$$

which implies $\liminf_{k\to\infty} \|w_k\| = 0$, since $\sum_{k\in\mathbb{N}} \lambda_k \left(2 - \frac{\gamma}{2\beta} - \lambda_k\right) = +\infty$. To prove that $w_k \to 0$, it suffices to show that the sequence $(\|w_k\|)_{k\in\mathbb{N}}$ is nonincreasing. Applying Lemma 3.2 with $x := x_{k+1}$ and $\hat{x} := x_k$ yields

$$0 \leq \langle x_{k+1} - x_k, w_k - w_{k+1}\rangle - \|w_{k+1} - w_k\|^2 - \gamma\langle T(u_{k+1}) - T(u_k), v_{k+1} - v_k\rangle.$$

The first two terms multiplied by 2 can be expressed as

$$2\langle \lambda_k w_k, w_k - w_{k+1}\rangle - 2\|w_{k+1} - w_k\|^2 = \lambda_k^2\|w_k\|^2 - \|w_{k+1} - w_k\|^2 - \|w_{k+1} - w_k + \lambda_k w_k\|^2,$$

while the third term is equal to

$$\begin{aligned}
-\gamma\langle T(u_{k+1}) &- T(u_k), v_{k+1} - v_k\rangle \\
&= -\gamma\langle T(u_{k+1}) - T(u_k), w_{k+1} - w_k\rangle - \gamma\langle T(u_{k+1}) - T(u_k), u_{k+1} - u_k\rangle \\
&\leq \gamma\beta\|T(u_{k+1}) - T(u_k)\|^2 + \frac{\gamma}{4\beta}\|w_{k+1} - w_k\|^2 - \gamma\beta\|T(u_{k+1}) - T(u_k)\|^2 \\
&= \frac{\gamma}{4\beta}\|w_{k+1} - w_k\|^2,
\end{aligned}$$

where we have used again Young's inequality and the cocoercivity of T. Therefore, we deduce

$$\begin{aligned}
0 &\leq \lambda_k^2\|w_k\|^2 - \|w_{k+1} - w_k + \lambda_k w_k\|^2 + \left(\frac{\gamma}{2\beta} - 1\right)\|w_{k+1} - w_k\|^2 \\
&= \lambda_k^2\|w_k\|^2 - \lambda_k^2\|w_k\|^2 + 2\lambda_k\langle w_{k+1} - w_k, -w_k\rangle + \left(\frac{\gamma}{2\beta} - 2\right)\|w_{k+1} - w_k\|^2 \\
&= \lambda_k\|w_k\|^2 - \lambda_k\|w_{k+1}\|^2 + \left(\frac{\gamma}{2\beta} - 2 + \lambda_k\right)\|w_{k+1} - w_k\|^2,
\end{aligned}$$

that is,

$$\lambda_k\|w_{k+1}\|^2 \leq \lambda_k\|w_k\|^2 - \left(2 - \frac{\gamma}{2\beta} - \lambda_k\right)\|w_{k+1} - w_k\|^2 \leq \lambda_k\|w_k\|^2,$$

so $(\|w_k\|)_{k\in\mathbb{N}}$ is nonincreasing, since $\lambda_k > 0$. Hence, we have proved that $w_k \to 0$.

Let $(\bar{x}, \bar{u}, \bar{z})$ be a weak sequential cluster point of the bounded sequence $(x_k, u_k, z_k)_{k\in\mathbb{N}}$. Hence, there is a subsequence of $(x_{k_n}, u_{k_n}, z_{k_n})_{n\in\mathbb{N}}$ which is weakly convergent to $(\bar{x}, \bar{u}, \bar{z})$. Now, consider the operator $S : \mathcal{H}^3 \rightrightarrows \mathcal{H}^3$ given by

$$S := \begin{pmatrix} (\gamma A)^{-1} \\ (\gamma T)^{-1} \\ \gamma B \end{pmatrix} + \begin{pmatrix} 0 & 0 & -\mathrm{Id} \\ 0 & 0 & -\mathrm{Id} \\ \mathrm{Id} & \mathrm{Id} & 0 \end{pmatrix},$$

which is maximally monotone, because it is the sum of a maximally monotone operator and a skew-symmetric matrix (see, e.g., [7, Example 20.35 & Corollary 25.5(i)]). From (13), it follows that

$$\begin{pmatrix} u_{k_n} - v_{k_n} \\ u_{k_n} - v_{k_n} \\ u_{k_n} - v_{k_n} \end{pmatrix} \in S \begin{pmatrix} x_{k_n} - u_{k_n} \\ z_{k_n} \\ v_{k_n} \end{pmatrix}.$$

As the graph of a maximally monotone operator is sequentially closed in the weak-strong topology (see, e.g., [7, Proposition 20.38]), taking the limit as $n \to \infty$ and observing that $x_{k_n} - u_{k_n} \rightharpoonup \bar{x} - \bar{u}$ and $v_{k_n} \rightharpoonup \bar{u}$ (since $w_{k_n} = v_{k_n} - u_{k_n} \to 0$), we deduce that

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \in \left(\begin{pmatrix} (\gamma A)^{-1} \\ (\gamma T)^{-1} \\ \gamma B \end{pmatrix} + \begin{pmatrix} 0 & 0 & -\mathrm{Id} \\ 0 & 0 & -\mathrm{Id} \\ \mathrm{Id} & \mathrm{Id} & 0 \end{pmatrix} \right) \begin{pmatrix} \bar{x} - \bar{u} \\ \bar{z} \\ \bar{u} \end{pmatrix}.$$

The latter inclusion is equivalent to

$$\bar{u} = J_{\gamma A}(\bar{x}), \quad \bar{z} = \gamma T(\bar{u}) \quad \text{and} \quad \bar{u} = J_{\gamma B}(2\bar{u} - \bar{x} - \bar{z}), \qquad (19)$$

which implies $\bar{x} \in \Omega_\gamma$. Therefore, every weak sequential cluster point of $(x_k)_{k\in\mathbb{N}}$ is contained in Ω_γ, and Proposition 2.9 implies that $(x_k)_{k\in\mathbb{N}}$ is weakly convergent to a point $\bar{x} \in \Omega_\gamma$. Then (19) shows that $\bar{u} = J_{\gamma A}(\bar{x})$ and $\bar{z} = \gamma T(\bar{u})$ are the unique cluster points of $(u_k)_{k\in\mathbb{N}}$ and $(z_k)_{k\in\mathbb{N}}$, respectively, and hence $u_k \rightharpoonup \bar{u}$, $v_k \rightharpoonup \bar{u}$ and $z_k \rightharpoonup \bar{z}$.

Moreover, since x was arbitrarily chosen in Ω_γ, (14) and (16) also hold with u replaced by $\bar{u}$ and x replaced by $\bar{x}$. From the resulting inequalities, we obtain

$$\begin{aligned} \beta\gamma \|T(\bar{u}) - T(u_k)\|^2 &\leq \langle \bar{x} - x_k, w_k \rangle + \langle u_k - \bar{u}, w_k \rangle \\ &\quad + \langle \bar{u} - v_k, w_k \rangle + \gamma \langle T(\bar{u}) - T(u_k), w_k \rangle, \end{aligned} \qquad (20)$$

and thus $T(u_k) \to T(\bar{u})$. Now, by Lemma 3.1, we know that $\bar{u} \in \mathrm{zer}(A + B + T)$.

Finally, pick any $\tilde{u} \in \mathrm{zer}(A + B + T)$. By Lemma 3.1, there is $\tilde{x} \in \Omega_\gamma$ such that $\tilde{u} = J_{\gamma A}(\tilde{x})$. Setting $x = \tilde{x}$ at the beginning of the proof, (20) becomes

$$\beta\gamma \|T(\tilde{u}) - T(u_k)\|^2 \leq \langle \tilde{x} - x_k, w_k \rangle + \langle u_k - \tilde{u}, w_k \rangle + \langle \tilde{u} - v_k, w_k \rangle + \gamma \langle T(\tilde{u}) - T(u_k), w_k \rangle.$$

Since $x_k \rightharpoonup \bar{x}$, $u_k \rightharpoonup \bar{u}$, $v_k \rightharpoonup \bar{u}$, $w_k \to 0$ and $T(u_k) \to T(\bar{u})$, the inequality above implies $T(\bar{u}) = T(\tilde{u})$. This proves that $T(\mathrm{zer}(A + B + T)) = \{T(\bar{u})\}$.

(ii) Assume first that A is uniformly monotone. Since the sequence $(u_k)_{k\in\mathbb{N}}$ is bounded, the set $\{\bar{u}\} \cup \{u_k, k \geq 0\} \subset \mathrm{dom}\, A$ is bounded. Thus, using uniform monotonicity in Lemma 3.2 with $x := \bar{x}$ and $\hat{x} := x_k$, we obtain the stronger inequality

$$\gamma\phi(\|\bar{u} - u_k\|) \leq \langle \bar{x} - x_k, w_k \rangle - \|w_k\|^2 - \gamma\langle T(\bar{u}) - T(u_k), \bar{u} - v_k \rangle, \qquad (21)$$

which entails $\gamma\phi(\|\bar{u} - u_k\|) \to 0$. Since ϕ is increasing, we deduce that $u_k \to \bar{u}$, which implies $v_k \to \bar{u}$. When B is uniformly monotone, the result similarly follows.

Finally, suppose that the demiregularity assumption holds. By (i), we know that $u_k \rightharpoonup \bar{u}$ and $T(u_k) \to T(\bar{u})$, so the demiregularity of T at $\bar{u}$ implies that $u_k \to \bar{u}$. Since $v_k - u_k \to 0$, we also obtain that $v_k \to \bar{u}$. □

Remark 3.4 (i) The stepsize γ in [18, Theorem 2.1] is assumed to be in $]0, 2\beta\varepsilon[$, with $\varepsilon \in]0, 1[$, while Theorem 3.3 allows to take stepsizes in the interval $]0, 4\beta[$, which is twice larger. Note that our assumption is required to guarantee that $2 - \gamma/(2\beta) > 0$. The relaxation parameters $(\lambda_k)_{k\in\mathbb{N}}$ in [18, Theorem 2.1] must be taken in $]0, 2 - \varepsilon[$, while the interval given in Theorem 3.3 is $]0, 2 - \gamma/(2\beta)]$. If $\gamma \in]0, 2\beta\varepsilon[$, we have $2 - \varepsilon < 2 - \gamma/(2\beta)$. Thus, Theorem 3.3 additionally allows to take some of the relaxation parameters equal to

$2 - \gamma/(2\beta)$ (but not all of them, as we need $\sum_{k\in\mathbb{N}} \lambda_k \left(2 - \frac{\gamma}{2\beta} - \lambda_k\right) = +\infty$, unless either A or B is uniformly monotone). Finally, unlike [18, Theorem 2.1], we do not require the assumption $\inf_{k\in\mathbb{N}} \lambda_k > 0$.

(ii) In Theorem 3.3(ii), even when $\sum_{k\in\mathbb{N}} \lambda_k \left(2 - \frac{\gamma}{2\beta} - \lambda_k\right) < +\infty$, we have proved that the sequence $(u_k)_{k\in\mathbb{N}}$ (respectively $(v_k)_{k\in\mathbb{N}}$) is strongly convergent to $\bar{u}$ when A (respectively B) is uniformly monotone.

(iii) Observe that it is also possible to prove $x_k \rightharpoonup \bar{x} \in \Omega_\gamma$ using the notion of *conically averaged operators* recently introduced in [5], not only for a fixed relaxation parameter $\lambda_k = \lambda$, as it was done in [17, Corollary 4.2]. Indeed, by [17, Theorem 4.1], the operator DY_γ in (2) is conically $(2 - \gamma/(2\beta))^{-1}$-averaged, so [5, Proposition 2.9] can be applied to deduce the convergence of the Krasnosel'skiĭ–Mann iteration (3) to a fixed point of DY_γ, which belongs to Ω_γ by Lemma 3.1.

As a corollary, we obtain the following convergence result for the forward-backward splitting algorithm that allows doubling the range of the stepsizes assumed in [7, Theorem 26.14] (which is a particular case of [12, Proposition 4.4]). Although this wider range of the stepsizes has been shown before in [19–21], it has not yet become widely known in the literature.

Corollary 3.5 *Let $B : \mathcal{H} \rightrightarrows \mathcal{H}$ be a maximally monotone operator and $T : \mathcal{H} \to \mathcal{H}$ be a β-cocoercive operator, with $\beta > 0$, such that $\operatorname{zer}(B + T) \neq \emptyset$. Set a stepsize $\gamma \in\ ]0, 4\beta[$ and consider a sequence of relaxation parameters $(\lambda_k)_{k\in\mathbb{N}}$ in $]0, 2 - \gamma/(2\beta)]$ such that $\sum_{k\in\mathbb{N}} \lambda_k \left(2 - \frac{\gamma}{2\beta} - \lambda_k\right) = +\infty$. Given some initial point $x_0 \in \mathcal{H}$, consider the sequences defined by*

$$\begin{cases} y_k = x_k - \gamma T(x_k) \\ x_{k+1} = x_k + \lambda_k (J_{\gamma B}(y_k) - x_k). \end{cases}$$

Then, the following assertions hold:

(i) *$(x_k)_{k\in\mathbb{N}}$ converges weakly to a point $\bar{x} \in \operatorname{zer}(B + T)$ and $(T(x_k))_{k\in\mathbb{N}}$ converges strongly to the unique dual solution $T(\bar{x})$.*

(ii) *If either B is uniformly monotone on every bounded subset of its domain, or T is demiregular at every point in $\operatorname{zer}(B + T)$, then $(x_k)_{k\in\mathbb{N}}$ converges strongly to $\bar{x} \in \operatorname{zer}(B + T)$.*

Proof Apply Theorem 3.3 with $A = 0$. By Theorem 3.3(i), $T(x) \to T(\bar{x})$ and $T(\operatorname{zer}(B + T)) = \{T(\bar{x})\}$, which is the solution to the dual problem, see [7, Proposition 26.1(iv)]. $\square$

We conclude this section by deriving a splitting algorithm for computing the resolvent of $A + B + T$. To this aim, we use the systematic framework developed in [1], based on the notion of strengthening of an operator.

Definition 3.6 Let $\theta > 0$, $\sigma \in \mathbb{R}$ and let $w \in \mathcal{H}$. Given $A : \mathcal{H} \rightrightarrows \mathcal{H}$, the (θ, σ)-*strengthening with inner perturbation w* of A is the operator $A_w^{(\theta,\sigma)} : \mathcal{H} \rightrightarrows \mathcal{H}$ defined by

$$A_w^{(\theta,\sigma)} := A \circ (\theta \operatorname{Id} - w) + \sigma \operatorname{Id}.$$

Theorem 3.7 (Strengthened-Davis–Yin splitting) *Let A, $B : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally α_A-monotone and α_B-monotone operators, respectively, and let $T : \mathcal{H} \to \mathcal{H}$ be a β-cocoercive and maximally α_T-monotone operator, with $\beta > 0$. Let $\theta > 0$, $\sigma_A, \sigma_B \in \mathbb{R}$ and $\sigma_T \geq 0$ be such that*

$$\sigma_A + \sigma_B + \sigma_T > 0 \quad and \quad (\theta \alpha_A + \sigma_A, \theta \alpha_B + \sigma_B, \theta \alpha_T + \sigma_T) \in \mathbb{R}_+^3 \setminus \{0_3\}. \tag{22}$$

Let $\mu := (\theta/\beta + \sigma_T)^{-1}$ and $\gamma \in {]0, 4\mu[}$. Consider a sequence of relaxation parameters $(\lambda_k)_{k \in \mathbb{N}}$ in ${]0, 2 - \gamma/(2\mu)]}$ verifying $\sum_{k \in \mathbb{N}} \lambda_k \left(2 - \frac{\gamma}{2\mu} - \lambda_k\right) = +\infty$. Suppose $q \in \operatorname{ran}\left(\operatorname{Id} + \frac{\theta}{\sigma_A + \sigma_B + \sigma_T}(A + B + T)\right)$. Given any $x_0 \in \mathcal{H}$, consider the sequences

$$\begin{cases} u_k = J_{\frac{\gamma \theta}{1 + \gamma \sigma_A} A}\left(\dfrac{1}{1 + \gamma \sigma_A}(x_k + \gamma \sigma_A q)\right) \\[2ex] v_k = J_{\frac{\gamma \theta}{1 + \gamma \sigma_B} B}\left(\dfrac{1}{1 + \gamma \sigma_B}((2 - \gamma \sigma_T)u_k - x_k - \theta \gamma T(u_k) + \gamma(\sigma_B + \sigma_T)q)\right) \\[2ex] x_{k+1} = x_k + \lambda_k(v_k - u_k). \end{cases} \tag{23}$$

Then $(u_k)_{k \in \mathbb{N}}$ and $(v_k)_{k \in \mathbb{N}}$ are weakly convergent to $J_{\frac{\theta}{\sigma_A + \sigma_B + \sigma_T}(A+B+T)}(q)$, and $(x_k)_{k \in \mathbb{N}}$ is weakly convergent to $\bar{x}$, with

$$J_{\frac{\gamma \theta}{1 + \gamma \sigma_A} A}\left(\frac{1}{1 + \gamma \sigma_A}(\bar{x} + \gamma \sigma_A q)\right) = J_{\frac{\theta}{\sigma_A + \sigma_B + \sigma_T}(A+B+T)}(q).$$

Further, if $\theta \alpha_A + \sigma_A > 0$ (respectively $\theta_B \alpha + \sigma_B > 0$) then the convergence of $(u_k)_{k \in \mathbb{N}}$ (respectively $(v_k)_{k \in \mathbb{N}}$) is strong, even when $\sum_{k \in \mathbb{N}} \lambda_k \left(2 - \frac{\gamma}{2\mu} - \lambda_k\right) < +\infty$.

Proof Set $\hat{x}_0 := \frac{1}{\theta}(x_0 - q)$ and consider the sequences

$$\begin{cases} \hat{u}_k = J_{\gamma A_{-q}^{(\theta, \sigma_A)}}(\hat{x}_k) \\[1.5ex] \hat{v}_k = J_{\gamma B_{-q}^{(\theta, \sigma_B)}}\left(2\hat{u}_k - \hat{x}_k - \gamma T_{-q}^{(\theta, \sigma_T)}(\hat{u}_k)\right) \\[1.5ex] \hat{x}_{k+1} = \hat{x}_k + \lambda_k(\hat{v}_k - \hat{u}_k). \end{cases} \tag{24}$$

By (22) and [16, Proposition 2.1], the operators $A_{-q}^{(\theta, \sigma_A)}$, $B_{-q}^{(\theta, \sigma_B)}$ and $T_{-q}^{(\theta, \sigma_T)}$ are maximally monotone, and by [1, Theorem 1(iii)], $T_{-q}^{(\theta, \sigma_T)}$ is μ-cocoercive. By assumption, $q \in \operatorname{ran}\left(\operatorname{Id} + \frac{\theta}{\sigma_A + \sigma_B + \sigma_T}(A + B + T)\right)$, and thus (22) and [1, Proposition 3] imply that

$$\operatorname{zer}\left(A_{-q}^{(\theta, \sigma_A)} + B_{-q}^{(\theta, \sigma_B)} + T_{-q}^{(\theta, \sigma_T)}\right) = \left\{\frac{1}{\theta}\left(J_{\frac{\theta}{\sigma_A + \sigma_B + \sigma_T}(A+B+T)}(q) - q\right)\right\}. \tag{25}$$

By Theorem 3.3(i), $\hat{u}_k \rightharpoonup \hat{u}$ and $\hat{v}_k \rightharpoonup \hat{u}$, with

$$\hat{u} \in \operatorname{zer}\left(A_{-q}^{(\theta, \sigma_A)} + B_{-q}^{(\theta, \sigma_B)} + T_{-q}^{(\theta, \sigma_T)}\right),$$

and $\hat{x}_k \rightharpoonup \hat{x}$, where $\hat{x}$ satisfies

$$\hat{u} = J_{\gamma A_{-q}^{(\theta, \sigma_A)}}(\hat{x}) \in \operatorname{zer}\left(A_{-q}^{(\theta, \sigma_A)} + B_{-q}^{(\theta, \sigma_B)} + T_{-q}^{(\theta, \sigma_T)}\right). \tag{26}$$

If $\theta\alpha_A + \sigma_A > 0$ (respectively $\theta\alpha_B + \sigma_B > 0$), then $\hat{u}_k \to \hat{u}$ (respectively $\hat{v}_k \to \hat{u}$) by Theorem 3.3(ii), even if $\sum_{k\in\mathbb{N}} \lambda_k \left(2 - \frac{\gamma}{2\mu} - \lambda_k \right) < +\infty$. Thanks to [16, Proposition 2.1], we may rewrite (24) as

$$
\begin{cases}
\theta\hat{u}_k + q = J_{\frac{\gamma\theta}{1+\gamma\sigma_A} A} \left(\frac{\theta}{1+\gamma\sigma_A} \hat{x}_k + q \right) \\
\theta\hat{v}_k + q = J_{\frac{\gamma\theta}{1+\gamma\sigma_B} B} \left(\frac{\theta}{1+\gamma\sigma_B} \left(2\hat{u}_k - \hat{x}_k - \gamma \left(T(\theta\hat{u}_k + q) + \sigma_T \hat{u}_k \right) \right) + q \right)
\end{cases}
$$

Further, by (26), (25) and [16, Proposition 2.1],

$$
J_{\frac{\theta}{\sigma_A+\sigma_B+\sigma_T}(A+B+T)}(q) = \theta J_{\gamma A^{(\theta,\sigma_A)}_{-q}} (\hat{x}) + q = J_{\frac{\gamma\theta}{1+\gamma\sigma_A} A} \left(\frac{\theta}{1+\gamma\sigma_A} \hat{x} + q \right).
$$

The result follows by making the change of variables $(x_k, u_k, v_k) := (\theta\hat{x}_k + q, \theta\hat{u}_k + q, \theta\hat{v}_k + q)$ for all $k \in \mathbb{N}$ and $\bar{x} := \theta\hat{x} + q$. The final assertion is a consequence of Remark 3.4(ii). □

Remark 3.8 Another way of computing the resolvent with parameter $\mu > 0$ of $A + B + T$ at $q \in \mathcal{H}$ is applying the Davis–Yin splitting algorithm to A, B and $\widetilde{T} := \frac{1}{\mu}(\mathrm{Id} - q) + T$, where $\widetilde{T}$ is $\left(\beta^{-1} + \mu^{-1} \right)^{-1}$-cocoercive, by [1, Theorem 1(iii)], and β is the cocoercivity constant of T. Note that this is a particular instance covered by Theorem 3.7, taking $\sigma_T = \frac{1}{\mu}$, $\sigma_A = \sigma_B = 0$ and $\theta = 1$.

4 Numerical Experiments

In this section we provide some numerical examples of the algorithms developed in the previous section. These experiments aim not to be exhaustive and only intend to show the importance of appropriately choosing the stepsize and the relaxation parameters of the algorithms.

4.1 A Feasibility Problem with Hard and Soft Constraints

Let $\mathbb{A}, \mathbb{B}, \mathbb{C} \subseteq \mathbb{R}^n$ be three closed and convex sets with nonempty intersection of the relative interiors of $\mathbb{A}$ and $\mathbb{B}$. Suppose $\mathbb{A}$ and $\mathbb{B}$ are *hard constraints*, which need to be satisfied, and $\mathbb{C}$ is a third *soft constraint*, which does not necessarily need to be fulfilled, but whose violation we want to reduce as much as possible. Imagine that, at the same time, we would like to find a point in $\mathbb{A} \cap \mathbb{B}$ as close as possible to a point $q \in \mathbb{R}^n$. This problem can be written as

$$
\underset{x\in\mathbb{A}\cap\mathbb{B}}{\arg\min} \ \frac{1}{2} d^2(x, \mathbb{C}) + \frac{\rho}{2} \|x - q\|^2, \tag{27}
$$

where $d^2(x, \mathbb{C}) := \|x - P_{\mathbb{C}}(x)\|^2$ and $\rho > 0$ is a regularization parameter specifying the importance of remaining close to the point q. Problem (27) can be reformulated as

$$
\underset{x\in\mathbb{R}^n}{\arg\min} \ \iota_{\mathbb{A}}(x) + \iota_{\mathbb{B}}(x) + \frac{1}{2}\|x - q\|^2 + \frac{1}{2\rho} d^2(x, \mathbb{C}),
$$

whose solution is given by $\text{prox}_{\left(\iota_{\mathbb{A}}+\iota_{\mathbb{B}}+\frac{1}{2\rho}d^2(\cdot,\mathbb{C})\right)}(q)$. The subdifferential sum rule (see, e.g., [7, Corollary 16.50(v)]) guarantees the equality

$$\text{prox}_{\left(\iota_{\mathbb{A}}+\iota_{\mathbb{B}}+\frac{1}{2\rho}d^2(\cdot,\mathbb{C})\right)}(q) = J_{\left(\partial\iota_{\mathbb{A}}+\partial\iota_{\mathbb{B}}+\nabla\left(\frac{1}{2\rho}d^2(\cdot,\mathbb{C})\right)\right)}(q) = J_{\left(N_{\mathbb{A}}+N_{\mathbb{B}}+\frac{1}{\rho}(\text{Id}-P_{\mathbb{C}})\right)}(q),$$

and thus, solving (27) boils down to computing the resolvent at q of the sum of the three maximally monotone operators $A := N_{\mathbb{A}}$, $B := N_{\mathbb{B}}$ and $T := \frac{1}{\rho}(\text{Id}-P_{\mathbb{C}})$, with T being $\frac{1}{\rho}$-cocoercive (see, e.g., [7, Corollary 12.31]).

To illustrate on the problem (27) the behavior of the Davis–Yin algorithm and its strengthened version derived in Theorem 3.7, we retake our simple introductory example of two balls $\mathbb{A}$ and $\mathbb{B}$ centered at $(-1.6, -0.75)$ and $(-0.35, 0.12)$, with radii 0.55 and 1, respectively. We chose these values to make the problem slightly challenging. We now add a new third ball $\mathbb{C}$ with center $(1, -1)$ and radius 0.5, the point $q := (-1.75, 1.5)$ and take $\rho := 1$. Observe that any combination of $\sigma_A \geq 0$, $\sigma_B \geq 0$ and $\sigma_T \geq 0$ such that $\theta := \sigma_A + \sigma_B + \sigma_T > 0$ satisfies the hypotheses of Theorem 3.7. Although finding the best values is beyond the scope of this work, for comparison, we tested the result of running the algorithm (23) with $(\sigma_A, \sigma_B, \sigma_T) = (0, 0, 1/\mu)$ (which corresponds to Davis–Yin splitting, see Remark 3.8) and $(\sigma_A, \sigma_B, \sigma_T) = (0, 1, 1)$, using as starting point $x_0 := (0.7, 1.7)$. In accordance with Theorem 3.7, the stepsize γ must be chosen so that $\frac{\gamma}{\mu} \in \,]0, 4[$, for $\mu = ((\sigma_A + \sigma_B + \sigma_T)\rho + \sigma_T)^{-1}$. In Fig. 4 we have represented the iterates for $\lambda_k = 0.99(2 - \frac{\gamma}{2\mu})$ and for two values of $\frac{\gamma}{\mu}$, namely 1.5 (overrelaxation) and 2.5 (underrelaxation).

In order to obtain the best combination of the stepsize and relaxation parameters, we run the algorithms for every possible value of $(\frac{\gamma}{\mu}, \lambda)$ on a grid with 4950 points in $]0, 4[\times]0, 2[$. The algorithms were stopped when the norm of the difference between the shadow sequence $P_A(x_k)$ and the solution to the problem was smaller than 10^{-8}. The solution, which is approximately equal to $(-1.227559, -0.3452923)$, was computed in Maple by numerically solving the KKT conditions with high precision. A contour plot representing the number of iterations is shown in Fig. 5. The minimum number of iterations for Davis–Yin was 17 and it was attained at $(\frac{\gamma}{\mu}, \lambda) = (3.11, 0.43)$, and for the strengthened-Davis–Yin was 16 and
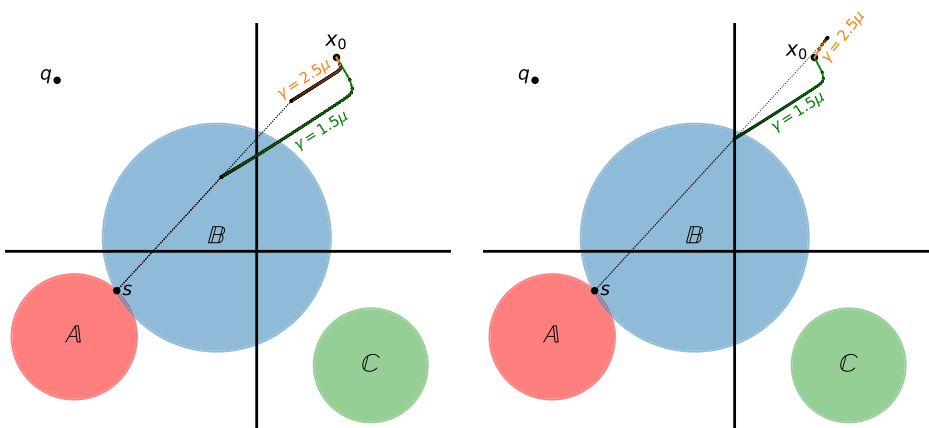
Fig. 4 Behavior of the iterates of the Davis–Yin (left) and the strengthened-Davis–Yin (right) splitting algorithms for the problem (27) for two stepsize parameters γ and $\lambda_k = 0.99(2 - \gamma/(2\mu))$. Since $\sigma_A = 0$, the solution is obtained after projecting the fixed point onto the set $\mathbb{A}$
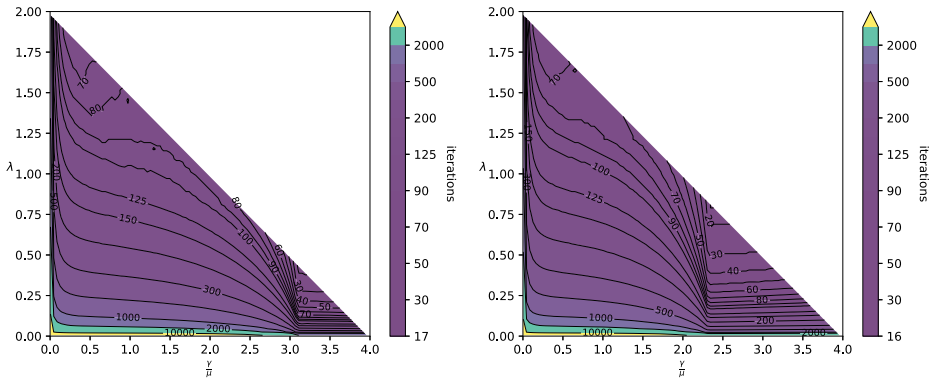
Fig. 5 Number of iterations needed until the shadow sequence is sufficiently close to the solution s when the Davis–Yin (left) and the strengthened-Davis–Yin (right) splitting algorithms are applied for different values of γ and $\lambda_k = \lambda$, with the experiment setting shown in Fig. 4

it was reached at three pair of values of $\frac{\gamma}{\mu}$ and λ, namely $\frac{\gamma}{\mu} = 2.34$, $\lambda \in \{0.79, 0.81\}$ and $\frac{\gamma}{\mu} = 2.39$, $\lambda = 0.79$.

4.2 Image Recovery Via ℓ_1 Regularization

The restoration of blurred images using ℓ_1 regularization has become a standard application in the literature to test the performance of forward-backward algorithms, see [6]. This consists in solving a minimization problem of the form

$$\underset{x \in \mathbb{R}^n}{\arg\min} \ \mu \|x\|_1 + \frac{1}{2} \|Mx - b\|_2^2, \tag{28}$$

where $M \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ is the observed blurred image (the vectorization of the two-dimensional matrix) and $\mu > 0$ is a regularization parameter. Setting $B = \partial(\mu\|\cdot\|_1)$ and $T = M^T(Mx - b)$, this problem can be reformulated as finding a zero of the sum $B + T$ of two maximally monotone operators. Since T is Lipschitz continuous, we can employ the forward-backward algorithm (i.e., Davis–Yin with $A = 0$), to solve (28). Note that the proximity operator of the ℓ_1-norm is the well-known soft thresholding function from Example 2.6. As pixel values must be in [0, 1], it is more realistic to solve instead the problem

$$\underset{x \in [0,1]^n}{\arg\min} \ \mu \|x\|_1 + \frac{1}{2} \|Mx - b\|_2^2,$$

Setting $A = N_{[0,1]^n}$ and B and T as above, this problem can be solved without much additional effort using the Davis–Yin splitting algorithm.

For our tests we replicated the wavelet-based restoration method in [6, Section 5.1.], including the additional constraint $x \in [0, 1]^n$. We also ran our experiments without this constraint (applying thus forward-backward) and the results were basically the same, so we do not include them for brevity. We employed as observed images the widely-used 256×256 pixels cameraman image and a picture of a symbol from the University of Alicante: the sculpture "Dibuixar l'espai" (by Pepe Azorín), with a resolution of 600×800 pixels. The images, shown in Fig. 6, were subjected to a Gaussian 9×9 blur with standard deviation 4, followed by an additive zero-mean Gaussian noise with standard deviation 10^{-3}. We chose $M = RW$, where R is the matrix representing the blur operator and W is the inverse of the

Fig. 6 Original (left), observed blurred (middle) and restored (right) images, showing the cameraman at the top and the sculpture "Dibuixar l'espai" at the bottom. The Davis–Yin algorithm was applied for 200 iterations with $\gamma = 1.98$ and $\lambda = 0.99$, using as starting point the observed blurred image

three stage Haar wavelet transform. The regularization parameter was taken as $\mu = 2 \cdot 10^{-5}$. The Lipschitz constant of T is the spectral radius of $M^T M$, which is equal to 1. Thus, T is 1-cocoercive and the stepsize in the Davis–Yin algorithm can be chosen in the interval $]0, 4[$. For values of (γ, λ) on a grid with 4950 points in $]0, 4[\times]0, 2[$, we performed 200 iterations of the algorithm taking as initial image the observed blurred image. Figure 7 shows the value of the objective function in the final iteration. We observe a symmetry with respect to the diagonal. The lowest values of the objective function were 0.349 for the cameraman and 2.684 for the sculpture, and they were both attained at $(\gamma, \lambda) = (1.98, 0.99)$.
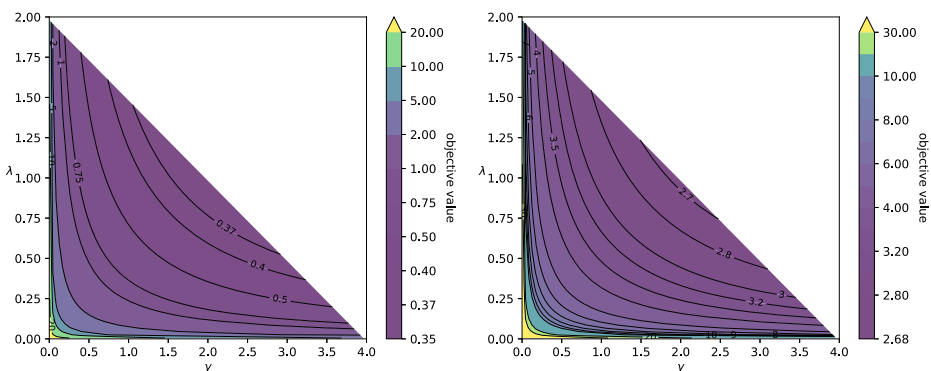
Fig. 7 Objective function value after 200 iterations of the forward-backward algorithm applied to the cameraman (left) and the sculpture "Dibuixar l'espai" (right), for different values of γ and λ, and taking as starting point the observed blurred image

5 Conclusions

We have presented an alternative proof of convergence for the Davis–Yin splitting algorithm without requiring the Davis–Yin operator (2) to be averaged. The proof was solely based on monotone operator theory and has the additional advantage of allowing larger stepsizes, up to four times the cocoercivity constant of the single-valued operator, doubling thus the range of values allowed in [18]. As a consequence, the same conclusion applies to the forward-backward splitting algorithm. We have also derived a *strengthened* version of the algorithm for computing the resolvent of the sum, based on the framework developed in [1]. The numerical experiments included show the importance of appropriately selecting the stepsize and relaxation parameters. In most of our tests, the behavior of the algorithm with respect to the parameters was symmetric, as the one shown in Fig. 7. Selecting the best parameters is not a simple task, but even so, it is clear that having more freedom in the choice of the stepsize parameter can be advantageous.

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

References

1. Aragón Artacho, F.J., Campoy, R., Tam, M.K.: Strengthened splitting methods for computing resolvents. Comput. Optim. Appl. **80**, 549–585 (2021)
2. Aragón Artacho, F.J., Censor, Y., Gibali, A.: The cyclic Douglas–Rachford algorithm with r-sets-Douglas–Rachford operators. Optim. Methods Softw. **34**(4), 875–889 (2019)
3. Attouch, H., Peypouquet, J., Redont, P.: Backward-forward algorithms for structured monotone inclusions in Hilbert spaces. J. Math. Anal. Appl. **457**, 1095–1117 (2018)
4. Baillon, J.-B., Haddad, G.: Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones. Israel. J. Math. **26**, 137–150 (1997)
5. Bartz, S., Dao, M.N., Phan, H.M.: Conical averagedness and convergence analysis of fixed point algorithms. J. Glob Optim. https://doi.org/10.1007/s10898-021-01057-4 (2021)
6. Beck, A., Teboulle, M.: A Fast Iterative Shrinkage-Tresholding Algorithm for linear inverse problems. SIAM J. Imaging Sci. **2**(1), 183–202 (2009)

7. Bauschke, H.H., Combettes, P.L. Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd edn. Springer, Berlin (2017)
8. Borwein, J.M., Tam, M.T.: A cyclic Douglas–Rachford iteration scheme. J. Optim. Theory Appl. **160**, 1–29 (2014)
9. Campoy, R.: A product space reformulation with reduced dimension for splitting algorithms. arXiv:1910.14185 (2021)
10. Cegielski, A.: Iterative Methods for Fixed Point Problems in Hilbert Spaces Lecture Notes in Mathematics, vol. 2057. Springer, Heidelberg (2012)
11. Combettes, P.L., Pesquet, J.-C.: Proximal Splitting Methods in Signal Processing. In: Bauschke, H.H., Burachik, R.S., Combettes, P.L., Elser, V., Luke, D.R., Wolkowicz, H. (eds.) Fixed-Point Algorithms for Inverse Problems in Science and Engineering, pp. 185-212. Springer, New York (2011)
12. Combettes, P.L., Yamada, I.: Compositions and convex combinations of averaged nonexpansive operators. J. Math. Anal. Appl. **425**, 55–70 (2015)
13. Condat, L., Kitahara, D., Contreras, A., Hirabayashi, A.: Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists. arXiv:1912.00137 (2021)
14. Dao, M.N., Dizon, N., Hogan, J.A., Tam, M.K.: Constraint reduction reformulations for projection algorithms with applications to wavelet construction. J. Optim. Theory Appl. **190**, 201–233 (2021)
15. Dao, M.N., Phan, H.M.: Adaptive Douglas–Rachford splitting algorithm for the sum of two operators. SIAM J. Optim. **29**(4), 2697–2724 (2019)
16. Dao, M.N., Phan, H.M.: Computing the resolvent of the sum of operators with application to best approximation problems. Optim. Lett. **14**, 1193–1205 (2020)
17. Dao, M.N., Phan, H.M.: An adaptive splitting algorithm for the sum of two generalized monotone operators and one cocoercive operator. Fixed Point Theory Algorithms Sci. Eng. **2021**, 16 (2021). https://doi.org/10.1186/s13663-021-00701-8
18. Davis, D., Yin, W.: A three-operator splitting scheme and its optimization applications. Set-valued Var. Anal. **25**(4), 829–858 (2017)
19. Giselsson, P.: Nonlinear forward-backward splitting with projection correction. SIAM J. Optim. **31**(3), 2199–2226 (2021)
20. Giselsson, P., Moursi, W.M.: On compositions of special cases of Lipschitz continuous operators. Fixed Point Theory Algorithms Sci. Eng. **2021**, 25 (2021). https://doi.org/10.1186/s13663-021-00709-0
21. Latafat, P., Patrinos, P.: Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators. Comput. Optim. Appl. **68**, 57–93 (2017)
22. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)
23. Malitsky, Y., Tam, M.K.: Resolvent splitting for sums of monotone operators with minimal lifting. arXiv:2108.02897 (2021)
24. Minty, G.: Monotone (nonlinear) operators in a Hilbert space. Duke Math. J. **29**, 341–34 (1962)
25. Passty, G.B.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. J. Math. Anal. Appl. **72**, 383–390 (1979)
26. Pierra, G.: Decomposition through formalization in a product space. Math. Program. **28**, 96–115 (1984)
27. Rieger, J., Tam, M.K.: Backward-forward-reflected-backward splitting for three operator monotone inclusions. Appl. Math. Comput. **381**, 125248 (2020)
28. Ryu, E.K.: Uniqueness of DRS as the 2 operator resolvent-splitting and impossibility of 3 operator resolvent-splitting. Math. Program. **182**, 233–273 (2020)
29. Ryu, E.K., Vũ, B.C.: Finding the forward-Douglas–Rachford-forward method. J. Optim. Theory Appl. **184**, 858–876 (2020)