Projective Splitting Methods for Pairs of Monotone Operators

Jonathan Eckstein* B. F. Svaiter†

August 13, 2003

Abstract

By embedding the notion of splitting within a general separator projection algorithmic framework, we develop a new class of splitting algorithms for the sum of two general maximal monotone operators in Hilbert space. Our algorithms are essentially standard projection methods, using splitting decomposition to construct separators. These projective algorithms converge under more general conditions than prior splitting methods, allowing the proximal parameter to vary from iteration to iteration, and even from operator to operator, while retaining convergence for essentially arbitrary pairs of operators. The new projective splitting class also contains noteworthy preexisting methods either as conventional special cases or excluded boundary cases.

1 Introduction

This paper considers *splitting* methods for solving the inclusion

$$0 \in A(x) + B(x), \tag{1}$$

where A and B are set-valued maximal monotone operators on some real Hilbert space $\mathcal{H}$. Splitting methods for this problem are algorithms that do not attempt to evaluate the resolvent mapping $(I + \lambda(A + B))^{-1}$ of the combined operator $A + B$ (where $\lambda > 0$ is some scalar) but instead only evaluate resolvent mappings $(I + \lambda A)^{-1}$ and $(I + \lambda B)^{-1}$ of the individual operators A and B. Such methods have numerous applications in constructing decomposition methods for convex optimization and monotone variational inequalities.

*Business School and RUTCOR, 640 Bartholomew Road, Busch Campus, Rutgers University, Piscataway NJ 08854 USA, jeckstei@rutcor.rutgers.edu.

†IMPA, Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina, 110. Rio de Janeiro, RJ, CEP 22460-320, Brazil, benar@impa.br. Partially supported by CNPq Grant 302748/2002-4 and by PRONEX–Optimization.

At present, the *Douglas-Rachford* class of algorithms constitute the predominant methods of this form. Given a fixed scalar $\eta > 0$ and a sequence $\{\rho_k\} \subset (0, 2)$, this class of methods may be expressed via the recursion

$$x^{k+1} = \left[\left(1 - \frac{\rho_k}{2}\right) I + \frac{\rho_k}{2}(2(I + \eta B)^{-1} - I)(2(I + \eta B)^{-1} - I)\right](x^k). \tag{2}$$

This approach is known to converge [4] for arbitrary maximal monotone A and B, under the conditions that $0 < \underline{\rho} \leq \rho_k \leq \overline{\rho} < 2$ for all k, and (1) has a solution. Classical Douglas-Rachford splitting is the special case $\rho_k \equiv 1$, whose convergence proof in the set-valued monotone context dates back to [10]. Also present in [10] is the case $\rho_k \equiv 2$, which is known as *Peaceman-Rachford splitting*; however, this case requires additional assumptions on either A or B to guarantee convergence. A concise and elegant approach to understanding this class of methods may be found in [8, Section 1], and is elaborated in [5, Sections 1.1–1.2]. Note also that there exist variants where the recursion (2) may be evaluated approximately, which we omit here in the interest of brevity.

A very broad range of decomposition algorithms for monotone inclusions, monotone variational inequalities, and convex optimization are in fact special cases of the Douglas-Rachford class of methods. This observation includes methods derived from Spingarn's principle of *partial inverses* [18].

Practical computational experience with Douglas-Rachford methods has proved somewhat mixed. In addition, they have the drawback that all known convergence proofs require the proximal parameter η to remain fixed throughout the algorithm; it cannot — at least in theory — be varied from iteration to iteration.

Splitting methods outside the class of encompassed by (2) and its approximate relatives have historically been burdened with a variety of restrictive assumptions. *Double-backward methods* [9, 13] use the simple recursion

$$x^{k+1} = (I + \lambda_k B)^{-1}(I + \lambda_k A)^{-1}(x^k).$$

These methods are suitable for situations where A and B share a common root, that is, $A^{-1}(0) \cap B^{-1}(0) \neq \emptyset$. Otherwise, convergence requires that $\lambda_k \to 0$ in a particular way, and even then $\{x^k\}$ may not converge, but only a certain ergodic sequence of its weighted averages.

Forward-backward splitting [6] methods offer a more popular alternative, and use the fundamental recursion

$$x^{k+1} \in (I + \lambda_k B)^{-1}(I - \lambda_k A)(x^k);$$

the next iterate x^{k+1} is no longer determined uniquely by this recursion unless A is single-valued. These methods are essentially generalizations of the classical gradient projection method for constrained convex optimization and monotone variational inequalities, and inherit restrictions similar to those methods. Typically, one must assume that A is Lipschitz continuous, and the stepsizes λ_k must fall in a range dictated by A's modulus of continuity. These restrictions are weakened somewhat in [19], but convergence still fails for general maximal monotone A and B.

This paper develops a new approach to splitting by embedding it within the framework of *separator projection* algorithms. We discuss the basic theoretical properties of this framework in Section 2. Fundamentally, given an iterate $x^k \in \mathcal{H}$ and a closed convex set $S \subset \mathcal{H}$, we suppose that we can construct a separating hyperplane φ_k between x^k and S, and then obtain x^{k+1} by projecting onto this hyperplane, possibly with a relaxation factor $\rho_k \in (0, 2)$. This recursion produces a sequence $\{x^k\}$ Féjer monotone to S, and under the proper conditions (weakly) convergent to a point in S.

Our main idea is to apply splitting-style decomposition techniques to the construction of the separators φ_k within this framework. Furthermore, we do not use $S = (A + B)^{-1}(0) \subset \mathcal{H}$, but instead employ an expanded version of the solution set we denote $S_{\mathrm{e}}(A, B)$, which is a closed convex subset of $\mathcal{H} \times \mathcal{H}$. Thus, the separator projection method operates in $\mathcal{H} \times \mathcal{H}$. We develop the basic properties of $S_{\mathrm{e}}(A, B)$, along with related separators and projection algorithms, in Section 3.

In Section 4, we present a new family of splitting algorithms, based on a particular decomposition approach to constructing the separator φ_k: to compute a separator, one need only evaluate resolvents of the form $(I + \mu_k A)^{-1}$ and $(I + \lambda_k B)^{-1}$. Performing the appropriate projections, which are simple closed-form calculations, one obtains a sequence weakly convergent to a solution of (1). Our family of splitting methods converges for essentially arbitrary maximal monotone A and B, the only restrictions being that $(A + B)^{-1}(0) \neq \emptyset$ and, if $\mathcal{H}$ is infinite-dimensional, that $A+B$ must be maximal. Nevertheless, the proximal parameter may vary from iteration to iteration, a property not shared by prior splitting methods applicable to general maximal monotone A and B. A further, completely novel feature is that the proximal parameter can vary from operator to operator: one can in the general case construct a valid separator by evaluating $(I + \mu_k A)^{-1}$ and $(I + \lambda_k B)^{-1}$, with $\mu_k \neq \lambda_k$.

Section 5 next considers some reformulations and special cases of our algorithmic family. In Section 5.1, we include an additional scaling factor $\eta > 0$ that imparts additional flexibility to the new algorithm family. With the aid of this extra factor, we then consider some special cases. In Section 5.2, we develop a special case that resembles the $n = 2$ case of Spingarn's splitting method [18], and is in fact identical if one makes some further restrictions to its parameters.

Section 5.3 then considers some other special cases which do not appear to have an analog in the prior literature. Finally, Section 5.4 considers Douglas-Rachford methods. This class of methods also turns out to be a special case of our projective framework, but with "boundary" parameter settings excluded by our convergence theory.

Although we do not pursue the implications here, we note that our new family of splitting methods, because of its generality, can clearly be used wherever Douglas-Rachford and other techniques have been applied in the past. Applications include the alternating direction method of multipliers for decomposing convex programs in “Fenchel” form [6, 4, 3], and complementarity problems [5]. Replacing Douglas-Rachford splitting with our projective approach would produce new versions of the algorithms in these references. Naturally, computational work will be needed to ascertain if anything is gained in practice.

Finally, we note that our approach to splitting can be further generalized, for example, to sums of more than two operators, or to allow approximate computation of resolvents. We will discuss such generalizations in forthcoming follow-up work.

2 A generic separator projection framework

Suppose that $\mathcal{H}$ is a real Hilbert space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$, and we wish to find a point p^* in some closed convex set $S \subset \mathcal{H}$. We do not know S explicitly, but we are able to detect if a given point $\overline{p} \in \mathcal{H}$ is in S, and to construct a separating hyperplane between S and any point $\overline{p} \in \mathcal{H}\backslash S$. We call a function $\varphi : \mathcal{H} \to \mathbb{R}$ a *separator* for $S \subset \mathcal{H}$ and $\overline{p} \in \mathcal{H}\backslash S$ if it is both affine and continuous, and has the properties

$$\varphi(\overline{p}) > 0 \qquad\qquad \varphi(p^*) \leq 0 \quad \forall\, p^* \in S. \tag{3}$$

Note that if a separator exists for S and every $\overline{p} \in \mathcal{H}\backslash S$, then S must be closed and convex. If φ is a separator for S and $\bar{p}$, one must have $\|\nabla\varphi\| > 0$, since otherwise the two conditions in (3) cannot both hold.

If we are able to generate such separators, a natural way to use them is through successive projection. Defining the halfspace $H(\varphi)$ via $H(\varphi) = \{p \in \mathcal{H} \mid \varphi(p) \leq 0\}$, projection onto $H(\varphi)$ may be accomplished via

$$P_{H(\varphi)}(q) = q - \frac{\max\{0, \varphi(q)\}}{\|\nabla\varphi\|^2}\nabla\varphi = \begin{cases} q & \text{if } \varphi(q) \leq 0 \\ q - \frac{\varphi(q)}{\|\nabla\varphi\|^2}\nabla\varphi & \text{if } \varphi(q) > 0. \end{cases}$$

In practice, it is often helpful to introduce a relaxation factor $\rho \in (0, 2)$: if the current iterate is $p^k \notin S$ and φ_k is a separator for S and p^k, we compute next iterate via

$$p^{k+1} = (1 - \rho_k)p^k + \rho_k P_{H(\varphi_k)}(p^k) = p^k - \rho_k \frac{\varphi_k(p^k)}{\|\nabla\varphi_k\|^2}\nabla\varphi_k,$$

where $\rho_k \in (0, 2)$. Bounding $\{\rho_k\}$ strictly away from 0 and 2, we arrive at the following “skeleton” algorithm:

Algorithm 1 (Generic projection framework) *Given:*

- *A real Hilbert space* $\mathcal{H}$ *and initial point* $p^0 \in \mathcal{H}$
- *Scalar constants* $\underline{\rho}, \overline{\rho}$ *with* $0 < \underline{\rho} \leq \overline{\rho} < 2$.

Starting with $k = 0$, *execute:*

1. *If $p^k \in S$, halt.*
2. *Find some separator φ_k for S and p^k.*
3. *Choose any ρ_k with $\underline{\rho} \le \rho_k \le \overline{\rho}$, and set*

$$p^{k+1} = p^k - \rho_k \frac{\varphi_k(p^k)}{\|\nabla\varphi_k\|^2}\nabla\varphi_k,$$

and repeat with $k \leftarrow k+1$.

The basic properties of this algorithmic form may be derived by an analysis much like that of classical projection algorithms, dating back to Cimmino [2] and Kaczmarz [7] in the late 1930's. A comprehensive review of projection algorithms may be found in [1].

Proposition 1 *Suppose Algorithm 1 does not halt. For all $k \ge 0$, define the positive scalar δ_k by*

$$\delta_k \stackrel{\text{def}}{=} \frac{\varphi_k(p^k)}{\|\nabla\varphi_k\|}. \tag{4}$$

Then, for all $k \ge 0$ and $p^* \in S$,

$$\left\|p^{k+1} - p^*\right\|^2 \le \left\|p^k - p^*\right\|^2 - \underline{\rho}(2-\overline{\rho})\delta_k^2. \tag{5}$$

The sequence $\{p^k\} \subset \mathcal{H}$ is thus Féjer monotone to S, that is, $\left\|p^{k+1} - p^*\right\| \le \left\|p^k - p^*\right\|$ for all $k \ge 0$ and $p^* \in S$. Furthermore, if $S \neq \emptyset$

$$\sum_{k=0}^{\infty}\delta_k^2 < \infty \qquad\qquad \lim_{k\to\infty}\delta_k = 0. \tag{6}$$

Proof. For all $k \ge 0$, define the vector

$$d^k \stackrel{\text{def}}{=} \frac{\varphi_k(p^k)}{\|\nabla\varphi_k\|^2}\nabla\varphi_k,$$

so $p^{k+1} = p^k - \rho_k d^k$. Picking any $p^* \in S$, we therefore have

$$\left\|p^{k+1} - p^*\right\|^2 = \left\|p^k - p^*\right\|^2 - 2\rho_k\langle p^k - p^*, d^k\rangle + \rho_k^2\left\|d^k\right\|^2. \tag{7}$$

¿From the optimality conditions for the projection operator $P_{H(\varphi_k)}$, along with $p^* \in S \subseteq H(\varphi_k)$, we have

$$\langle p^* - P_{H(\varphi_k)}(p^k), p^k - P_{H(\varphi_k)}(p^k)\rangle \le 0.$$

Substituting $P_{H(\varphi_k)}(p^k) = p^k - d^k$ into this relation, we obtain

$$\langle p^* - (p^k - d^k), d^k \rangle \leq 0 \qquad \Leftrightarrow \qquad \langle p^k - p^*, d^k \rangle \geq \left\| d^k \right\|^2$$

(this relation is just the well-established fact that $P_{H(\varphi_k)}$ must be *firmly nonexpansive*). Substituting this relation into (7),

$$\begin{aligned} \left\| p^{k+1} - p^* \right\|^2 &\leq \left\| p^k - p^* \right\|^2 - \rho_k(2 - \rho_k) \left\| d^k \right\|^2 \\ &\leq \left\| p^k - p^* \right\|^2 - \underline{\rho}(2 - \overline{\rho}) \left\| d^k \right\|^2, \end{aligned}$$

which is equivalent to (5). The remaining assertions follow immediately from (5). □

Since $\{p^k\}$ is Féjer monotone to S, we immediately have:

Corollary 2 *In Algorithm 1, if $\{p^k\}$ has a limit point $p^\infty \in S$, then it converges to $p^\infty \in S$.*

While this generic projection framework always yields sequences Féjer monotone to S, it may not achieve convergence in general, even in the weak topology. The reason is that one might choose separators that are asymptotically "shallow", causing the method to stall. However, one can assert that $\{p^k\}$ cannot have multiple weak limit points in S:

Proposition 3 *In Algorithm 1, $\{p^k\}$ has at most one weak limit point in S.*

Proof. Let $\mathcal{K}_1, \mathcal{K}_2 \subseteq \{0, 1, 2, \ldots\}$ be two infinite sets such that $p^k \xrightarrow{\mathsf{w}}_{\mathcal{K}_i} p_i^\infty \in S$, $i = 1, 2$. By Opial's Lemma [12],

$$\begin{aligned} \liminf_{\substack{k\to\infty \\ k \in \mathcal{K}_1}} \left\| p^k - p_2^\infty \right\| &\geq \liminf_{\substack{k\to\infty \\ k \in \mathcal{K}_1}} \left\| p^k - p_1^\infty \right\| \\ \liminf_{\substack{k\to\infty \\ k \in \mathcal{K}_2}} \left\| p^k - p_1^\infty \right\| &\geq \liminf_{\substack{k\to\infty \\ k \in \mathcal{K}_2}} \left\| p^k - p_2^\infty \right\| \end{aligned}$$

with equality holding in both relations if and only if $p_1^\infty = p_2^\infty$. Since $p_1^\infty, p_2^\infty \in S$, $\left\| p^k - p_1^\infty \right\|$ and $\left\| p^k - p_2^\infty \right\|$ are nonincreasing over all $k \geq 0$, and so

$$\begin{aligned} \liminf_{\substack{k\to\infty \\ k \in \mathcal{K}_1}} \left\| p^k - p_1^\infty \right\| &= \liminf_{\substack{k\to\infty \\ k \in \mathcal{K}_2}} \left\| p^k - p_1^\infty \right\| = \liminf_{k\to\infty} \left\| p^k - p_1^\infty \right\| \\ \liminf_{\substack{k\to\infty \\ k \in \mathcal{K}_2}} \left\| p^k - p_2^\infty \right\| &= \liminf_{\substack{k\to\infty \\ k \in \mathcal{K}_1}} \left\| p^k - p_2^\infty \right\| = \liminf_{k\to\infty} \left\| p^k - p_2^\infty \right\| \end{aligned}$$

Therefore, $p_1^\infty = p_2^\infty$. □

Note that the classical proximal point algorithm [14] for finding a zero of a general maximal monotone operator T (including the possibility of over- or under-relaxation; see for example [4]) is an instance of Algorithm 1, where $S = T^{-1}(0)$. To see this, let

$$r^k = (I + \lambda_k T)^{-1}(p^k).$$

Therefore, r^k and some $v^k \in \mathcal{H}$ satisfy the the conditions

$$v^k \in T(r^k) \qquad\qquad \lambda_k v^k + r^k - p^k = 0. \tag{8}$$

Since $0 \in T(p^*)$ for all $p^* \in S$, the monotonicity of T gives

$$\langle v^k, p^* - r^k\rangle \le 0 \qquad \forall\, p^* \in S,$$

while

$$\langle v^k, p^k - r^k\rangle = \lambda_k \left\|v^k\right\|^2 = \left\|p^k - r^k\right\|^2 / \lambda_k \ge 0,$$

with equality holding if and only if $p^k = r^k \in T^{-1}(0) = S$. Thus, if $p^k \notin T^{-1}(0)$, the affine function

$$\phi_k(p) = \langle v^k, p - r^k\rangle$$

is a separator for $T^{-1}(0)$ and p^k. So, if we implement Algorithm 1 with this choice of separator, that is, taking $\varphi_k = \phi_k$ in step 2, we get

$$p^{k+1} = p^k - \rho_k \frac{\varphi_k(p^k)}{\|\nabla\varphi_k\|^2}\nabla\varphi_k = (1-\rho_k)p^k + \rho_k r^k.$$

Thus, there are commonalities between the analyses of projection methods and the proximal point family of algorithms.

The hybrid projection-proximal point method of [16] and its generalizations in [15, 17] are based on this equivalence between proximal point algorithms and projection methods; in these works, an *approximate* solution of the “proximal system” (8), using a relative error criterion, is shown to provide a sufficiently deep separator to assure convergence. In [16], p^{k+1} is obtained by projecting onto this separator without a relaxation factor (that is, with $\rho_k = 1$). In [17], under- and over-relaxation appears in a similar context.

3 The set $S_{\mathrm{e}}(A, B)$ and decomposable separators

We now turn to the problem of solving $0 \in A(x) + B(x)$, where A and B are arbitrary maximal monotone operators on $\mathcal{H}$. We will use the generic projection framework of Section 2; however, the separators will not be for the set $(A+B)^{-1}(0)$, but for a higher-dimensional set $S_{\mathrm{e}}(A, B) \subseteq \mathcal{H} \times \mathcal{H}$ defined via

$$S_{\mathrm{e}}(A, B) \stackrel{\text{def}}{=} \left\{(x, b) \in \mathcal{H} \times \mathcal{H} \;\middle|\; b \in B(x), -b \in A(x)\right\}.$$

Clearly, a point $x \in \mathcal{H}$ satisfies $0 \in A(x) + B(x)$ if and only if there exists $b \in B(x)$ such that $-b \in A(x)$, so $S_e(A, B)$ is nonempty if and only if $(A+B)^{-1}(0)$ is nonempty, and applying the trivial projection $(x, y) \mapsto x$ to $S_e(A, B)$ yields $(A + B)^{-1}(0)$.

We endow $\mathcal{H} \times \mathcal{H}$ with the canonical inner product

$$\langle (x, y), (z, w) \rangle = \langle x, z \rangle + \langle y, w \rangle$$

induced by the inner product $\langle \cdot, \cdot \rangle$ of $\mathcal{H}$, and the corresponding norm

$$\|(x, y)\| = \sqrt{\|x\|^2 + \|y\|^2}\,.$$

Our basic strategy will be to find separators between $S_e(A, B)$ and an arbitrary point $(z^k, w^k) \in \mathcal{H} \times \mathcal{H}$ by calculations involving A and B individually, but not the joint operator $A + B$.

Lemma 4 *Suppose $b \in B(x)$ and $a \in A(y)$, where $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ are monotone operators. Define the function $\varphi : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ via*

$$\varphi(z, w) = \langle z - x, b - w \rangle + \langle z - y, a + w \rangle. \tag{9}$$

Then φ is affine, continuous, and $\varphi(z, w) \le 0$ for all $(z, w) \in S_e(A, B)$. Furthermore, one has

$$\nabla\varphi = \begin{pmatrix} a + b \\ x - y \end{pmatrix} \qquad\qquad \|\nabla\varphi\|^2 = \|a + b\|^2 + \|x - y\|^2\,, \tag{10}$$

and $\nabla\varphi = 0$ if and only if $(x, b) \in S_e(A, B)$, $x = y$, and $a = -b$.

Proof. Direct calculation produces

$$\varphi(z, w) = \langle z, a + b \rangle + \langle x - y, w \rangle - \langle x, b \rangle - \langle y, a \rangle, \tag{11}$$

so φ is affine and continuous. If $(z, w) \in S_e(A, B)$, then one must have $w \in B(z)$, and $\langle z - x, b - w \rangle \le 0$ by the monotonicity of B. Similarly, $-w \in A(z)$, and the monotonicity of A yields $\langle z - y, a + w \rangle \le 0$, so $\varphi(z, w) \le 0$. Next, (10) follows immediately from (11), and from (11) one has that $\nabla\varphi = 0$ if and only if $x = y$ and $a = -b$. In that case, since $b \in B(x)$ and $a \in A(y)$, one also has $(x, b) \in S_e(A, B)$. $\square$

Lemma 5 *Given any $(\overline{z}, \overline{w}) \in \mathcal{H} \times \mathcal{H}$, $(\overline{z}, \overline{w}) \notin S_e(A, B)$, there exists a separator $\varphi_{\overline{z},\overline{w}}$ for $S_e(A, B)$ and $(\overline{z}, \overline{w})$. Consequently, $S_e(A, B)$ must be a closed convex set.*

Proof. Take any $(\overline{z}, \overline{w}) \in \mathcal{H} \times \mathcal{H}$, $(\overline{z}, \overline{w}) \notin S_e(A, B)$. Let (x, b) and (y, a) be the unique elements of $\mathcal{H} \times \mathcal{H}$ such that

$$x + b = \overline{z} + \overline{w} \qquad\qquad b \in B(x) \tag{12}$$

$$y + a = \overline{z} - \overline{w} \qquad\qquad a \in A(y). \tag{13}$$

To see that (x, b) exists and is unique, note that (12) is equivalent to the conditions

$$\begin{aligned} x &= (I+B)^{-1}(\overline{z}+\overline{w}) \\ b &= \overline{z}+\overline{w}-x. \end{aligned}$$

The operator $(I+B)^{-1}$ is everywhere-defined and single-valued [11], so these conditions determine x and b uniquely. The argument for (y, a) is similar. Now define

$$\varphi_{\overline{z},\overline{w}}(z,w) \overset{\text{def}}{=} \langle z-x, b-w\rangle + \langle z-y, a+w\rangle.$$

By Lemma 4, $\varphi_{\overline{z},\overline{w}}(z,w) \leq 0$ for all $(z,w) \in S_{\mathrm{e}}(A,B)$. To establish that $\varphi_{\overline{z},\overline{w}}$ is a separator, it thus suffices to show that $\varphi_{\overline{z},\overline{w}}(\overline{z},\overline{w}) > 0$. From (12) and (13), respectively, we have $b-\overline{w} = \overline{z}-x$ and $a+\overline{w} = \overline{z}-y$. Substituting these expressions into the definition of $\varphi_{\overline{z},\overline{w}}$, we obtain

$$\varphi_{\overline{z},\overline{w}}(\overline{z},\overline{w}) = \|\overline{z}-x\|^2 + \|\overline{z}-y\|^2 \geq 0.$$

In the case $\varphi_{\overline{z},\overline{w}}(\overline{z},\overline{w}) = 0$, we would immediately have $x = y = \overline{z}$. Then, substituting $x = \overline{z}$ and $y = \overline{z}$ into (12) and (13) , we would deduce $b = -a = \overline{w}$. Since $b \in B(x)$ and $a \in A(y)$, we would then reach the contradictory conclusion that $(\overline{z},\overline{w}) \in S_{\mathrm{e}}(A,B)$. Therefore, $\varphi_{\overline{z},\overline{w}}(\overline{z},\overline{w}) > 0$, and $\varphi_{\overline{z},\overline{w}}$ is a separator.

Then closedness and convexity of $S_{\mathrm{e}}(A,B)$ now follows by a standard construction which we sketch briefly for completeness. Define

$$S' \overset{\text{def}}{=} \bigcap_{\substack{(\overline{z},\overline{w}) \in \mathcal{H}\times\mathcal{H} \\ (\overline{z},\overline{w}) \notin S_{\mathrm{e}}(A,B)}} \left\{(z,w) \in \mathcal{H}\times\mathcal{H} \;\middle|\; \varphi_{\overline{z},\overline{w}}(z,w) \leq 0\right\}.$$

S' is closed and convex, since it is the intersection of a collection of closed half-spaces. It is straightforward to show that S' contains every member of $S_{\mathrm{e}}(A,B)$ and excludes all $(\overline{z},\overline{w}) \notin S_{\mathrm{e}}(A,B)$, so $S_{\mathrm{e}}(A,B) = S'$, implying $S_{\mathrm{e}}(A,B)$ is closed and convex. □

We now specialize the framework of Algorithm 1 to the case $S = S_{\mathrm{e}}(A,B)$ and φ_k constructed as in Lemma 4:

Algorithm 2 (Generic projection/splitting framework for $S_{\mathrm{e}}(A,B)$) *Given:*

- *A general Hilbert space* $\mathcal{H}$
- *Maximal monotone operators* $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ *with* $(A+B)^{-1}(0)$ *nonempty*
- *Arbitrary vectors* $z^0, w^0 \in \mathcal{H}$
- *Scalar constants* $\underline{\rho}, \overline{\rho}$ *with* $0 < \underline{\rho} \leq \overline{\rho} < 2$.

Starting with $k = 0$*, execute:*

1. *Choose* $(x^k, b^k) \in \text{Gph}(B)$ *and* $(y^k, a^k) \in \text{Gph}(A)$ *such that either* $(x^k, b^k) = (y^k, -a^k)$ *or*
$$\langle z^k - x^k, b^k - w^k\rangle + \langle z^k - y^k, a^k + w^k\rangle > 0.$$
If $(x^k, b^k) = (y^k, -a^k)$*, set* $(z^{k+1}, w^{k+1}) = (x^k, b^k)$*, and halt; otherwise, continue.*

2. *Choose some* ρ_k *with* $\underline{\rho} \le \rho_k \le \overline{\rho}$*, and let*
$$\begin{aligned}\sigma_k &= \frac{\langle z^k - x^k, b^k - w^k\rangle + \langle z^k - y^k, a^k + w^k\rangle}{\|a^k + b^k\|^2 + \|x^k - y^k\|^2}\\ z^{k+1} &= z^k - \rho_k\sigma_k(a^k + b^k)\\ w^{k+1} &= w^k - \rho_k\sigma_k(x^k - y^k).\end{aligned}$$
Let $k \leftarrow k+1$*, and repeat.*

If $(x^k, b^k) = (y^k, -a^k)$ and the algorithm halts in step 1, then $(z^{k+1}, w^{k+1}) = (x^k, b^k) \in S_{\text{e}}(A, B)$, and an element of $S_{\text{e}}(A, B)$ has been found. In the context of this algorithm, let $\varphi_k(\cdot, \cdot)$ be defined by
$$\varphi_k(z, w) = \langle z - x^k, b^k - w\rangle + \langle z - y^k, a^k + w\rangle. \tag{14}$$
By step 1 and Lemma 4, each time step 2 is executed, φ_k is a separator for $S_{\text{e}}(A, B)$ and (z^k, w^k). Moreover, step 2 of this algorithm is exactly Step 3 of Algorithm 1, implemented with the separator φ_k defined above; note also that the denominator $\|a^k + b^k\|^2 + \|x^k - y^k\|^2$ must be positive, so σ_k is well-defined. Therefore, Algorithm 2 is essentially a special case of Algorithm 1, where $S = S_{\text{e}}(A, B)$. Therefore, the following result follows directly from Propositions 1-3 and (10).

Proposition 6 *In Algorithm 2, suppose that* $S_{\text{e}}(A, B) \neq \emptyset$*. Then the algorithm either halts with* $(z^{k+1}, w^{k+1}) \in S_{\text{e}}(A, B)$*, or* $\{(z^k, w^k)\}$ *is an infinite sequence Féjer monotone to* $S_{\text{e}}(A, B)$*, and one also has* $\sum_{k=0}^{\infty} \delta_k^2 < \infty$ *and* $\lim_{k\to\infty} \delta_k = 0$*, where*
$$\delta_k = \frac{\langle z^k - x^k, b^k - w^k\rangle + \langle z^k - y^k, a^k + w^k\rangle}{\sqrt{\|a^k + b^k\|^2 + \|x^k - y^k\|^2}}.$$
Furthermore, if $\{(z^k, w^k)\}$ *has a limit point in* $S_{\text{e}}(A, B)$*, it converges to a point in* $S_{\text{e}}(A, B)$*. If all weak limit points of* $\{(z^k, w^k)\}$ *are in* $S_{\text{e}}(A, B)$*, it converges weakly to a point in* $S_{\text{e}}(A, B)$*.*

4 A family of projective splitting methods

To transform Algorithm 2 into a workable method, one must specify how to obtain $(x^k, b^k) \in \text{Gph}(B)$ and $(y^k, a^k) \in \text{Gph}(A)$ meeting the requirements of step 1. This section describes a parameterized family of procedures for doing so. For each k, the calculation of (x^k, b^k) and (y^k, a^k) is described by three scalar parameters $\lambda_k, \mu_k > 0$ and α_k. The construction resembles the proof of Lemma 5, but is more general.

Algorithm 3 (A projective splitting family) *Suppose we are given the same objects as in Algorithm 2, as well as scalar constants* $\overline{\lambda} \geq \underline{\lambda} > 0$*. Starting with* $k = 0$*, execute:*

1. *Choose some* $\alpha_k \in \mathbb{R}$ *and* $\lambda_k, \mu_k \in [\underline{\lambda}, \overline{\lambda}]$ *satisfying the condition*

$$\mu_k/\lambda_k - (\alpha_k/2)^2 > 0. \tag{15}$$

Let (x^k, b^k) *and* (y^k, a^k) *be the unique points in* $\mathrm{Gph}(B)$ *and* $\mathrm{Gph}(A)$*, respectively, such that*

$$x^k + \lambda_k b^k = z^k + \lambda_k w^k \tag{16}$$

$$y^k + \mu_k a^k = (1 - \alpha_k) z^k + \alpha_k x^k - \mu_k w^k. \tag{17}$$

If $x^k - y^k = 0$ *and* $a^k + b^k = 0$*, set* $(z^{k+1}, w^{k+1}) = (x^k, b^k)$ *and halt.*

2. *Otherwise, as in Algorithm 2, choose some* ρ_k *with* $\underline{\rho} \leq \rho_k \leq \overline{\rho}$*, and let*

$$\begin{aligned} \sigma_k &= \frac{\langle z^k - x^k, b^k - w^k \rangle + \langle z^k - y^k, a^k + w^k \rangle}{\|a^k + b^k\|^2 + \|x^k - y^k\|^2} \\ z^{k+1} &= z^k - \rho_k \sigma_k (a^k + b^k) \\ w^{k+1} &= w^k - \rho_k \sigma_k (x^k - y^k). \end{aligned}$$

Let $k \leftarrow k + 1$*, and repeat.*

Note that (16) implies $x^k = (I + \lambda_k B)^{-1}(z^k + \lambda_k w^k)$, and the operator $(I + \lambda_k B)^{-1}$ is everywhere-defined and single-valued by the maximal monotonicity of B [11]. Rearranging (16), one has $b^k = w^k + (1/\lambda_k)(z^k - x^k)$, so (x^k, b^k) exists and is unique. Similarly,

$$y^k = (I + \mu_k A)^{-1} \left((1 - \alpha_k) z^k + \alpha_k x^k - \mu_k w^k \right),$$

so the maximal monotonicity of A guarantees the existence and uniqueness of y^k and $a^k = (1/\mu_k) \left((1 - \alpha_k) z^k + \alpha_k x^k - y^k \right) - w^k$.

Algorithm 3 is a true splitting method for the problem $0 \in A(z) + B(z)$, in that it uses only the individual resolvent mappings $(I + \mu_k A)^{-1}$ and $(I + \lambda_k B)^{-1}$, and never works directly with the operator $A + B$.

We will now show that if $\inf_{k \geq 0} \{\mu_k/\lambda_k - (\alpha_k/2)^2\} > 0$, the algorithm converges in the weak topology to a solution of $0 \in A(z) + B(z)$. As the proof is somewhat lengthy, we divide it into two main parts: Propositions 7 and 9 below. The first proposition establishes the basic properties of the separator φ_k computed via (16)-(17), and that step 1 is valid implementation of step 1 of Algorithm 2. The second proposition completes the proof of convergence.

Proposition 7 *In Algorithm 3, the function*

$$\varphi_k(z,w) = \langle z - x^k, b^k - w\rangle + \langle z - y^k, a^k + w\rangle$$

has the property

$$\varphi_k(z^k,w^k) \geq \xi_k \left[\left\|z^k - x^k\right\|^2 + \left\|z^k - y^k\right\|^2\right], \tag{18}$$

where

$$\begin{aligned}
\xi_k &\stackrel{\text{def}}{=} \frac{(\mu_k/\lambda_k + 1) - \sqrt{(\mu_k/\lambda_k+1)^2 - 4(\mu_k/\lambda_k - \alpha_k^2/4)}}{2\mu_k} && (19)\\
&\geq \frac{\mu_k/\lambda_k - \alpha^2/4}{\mu_k(\mu_k/\lambda_k+1)} \;>\; 0. && (20)
\end{aligned}$$

In particular, $\varphi_k(z^k,w^k) \geq 0$*. Moreover,* $\varphi_k(z^k,w^k) = 0$ *can only occur if* $(x^k,b^k) = (y^k,-a^k) = (z^k,w^k) \in S_{\mathrm{e}}(A,B)$*.*

Proof. Rearranging (16) and (17) yields

$$\begin{aligned}
b^k - w^k &= (1/\lambda_k)(z^k - x^k) && (21)\\
a^k + w^k &= (1/\mu_k)\left[(z^k - y^k) - \alpha_k(z^k - x^k)\right] && (22)
\end{aligned}$$

Substituting these identities into the formula (14) for $\varphi_k(z^k,w^k)$ gives

$$\varphi_k(z^k,w^k) = \frac{1}{\lambda_k}\left\|z^k - x^k\right\|^2 + \frac{1}{\mu_k}\left\|z^k - y^k\right\|^2 - \frac{\alpha_k}{\mu_k}\langle z^k - x^k, z^k - y^k\rangle. \tag{23}$$

Applying the Cauchy-Schwarz inequality and interpreting the resulting expression as a quadratic form applied to $\left(\|z^k - x^k\|, \|z^k - y^k\|\right) \in \mathbb{R}^2$, we obtain

$$\begin{aligned}
\varphi_k(z^k,w^k) &\geq \frac{1}{\lambda_k}\left\|z^k - x^k\right\|^2 + \frac{1}{\mu_k}\left\|z^k - y^k\right\|^2 - \frac{|\alpha_k|}{\mu_k}\left\|z^k - y^k\right\|\left\|z^k - x^k\right\| && (24)\\
&= \begin{bmatrix} \|z^k - x^k\| \\ \|z^k - y^k\| \end{bmatrix}^\top \begin{bmatrix} \frac{1}{\lambda_k} & -\frac{|\alpha_k|}{2\mu_k} \\ -\frac{|\alpha_k|}{2\mu_k} & \frac{1}{\mu_k} \end{bmatrix} \begin{bmatrix} \|z^k - x^k\| \\ \|z^k - y^k\| \end{bmatrix}. && (25)
\end{aligned}$$

Calculating the smaller eigenvalue of the matrix in (25), we obtain ξ_k as defined in (19), and therefore (18) holds. To prove the inequality passing from (19) to (20), we write

$$\xi_k = \frac{\mu_k/\lambda_k + 1}{2\mu_k}\left[1 - \sqrt{1 - 4\,\frac{\mu_k/\lambda_k - \alpha_k^2/4}{(\mu_k/\lambda_k + 1)^2}}\,\right], \tag{26}$$

and use the concavity inequality $\sqrt{1-h} \leq 1 - h/2$ for $h \leq 1$. The positivity of the resulting expression in (20) follows from the assumptions in step 1 of the algorithm.

It immediately follows from the positivity of ξ_k that $\varphi_k(z^k, w^k) \geq 0$. Finally, when $\varphi_k(z^k, w^k) = 0$, (18) implies $x^k = y^k = z^k$. From (21) and (22), one then has $b^k = -a^k = w^k$, and so $(x^k, b^k) = (y^k, -a^k) = (z^k, w^k) \in S_{\mathrm{e}}(A, B)$. □

Note that it is evident from (26) that the sign of the eigenvalue ξ_k is identical to that of $\mu_k/\lambda_k - (\alpha_k/2)^2$. In view of (20), we can thus use $\mu_k/\lambda_k - (\alpha_k/2)^2$ to guide our convergence analysis, instead of the more cumbersome expression (19).

We now proceed to complete our proof of convergence. First, we state a simple lemma that we will employ twice in the main proof.

Lemma 8 *Suppose $\{r^k\}, \{u^k\}, \{v^k\} \subset \mathcal{H}$ and $\{\lambda_k\} \subset (0, \infty)$ are sequences such that $\{r^k\}$ is bounded and*

$$u^k + \lambda_k v_k = r_k \qquad\qquad v^k \in T(u^k) \tag{27}$$

for all k, where T is a monotone operator. If $\{\lambda_k\}$ is bounded, then $\{u^k\}$ is bounded. If $\inf_k\{\lambda_k\} > 0$, then v_k is bounded.

Proof. In order to prove the first claim, suppose $\{\lambda_k\}$ is bounded. Pick any point $(\overline{u}, \overline{v}) \in \mathrm{Gph}(T)$ (the hypotheses imply T's graph is nonempty). Since $\{\lambda_k\}$ is bounded, so is the sequence $\{\overline{u} + \lambda_k \overline{v}\}$. Since $\{r^k\}$ is also bounded, there exists $\beta > 0$ such that $\|r^k - (\overline{u} + \lambda_k \overline{v})\| \leq \beta$ for all k. We have $u^k = (I + \lambda_k T)^{-1}(r^k)$ for all k, and using the nonexpansiveness of $(I + \lambda_k T)^{-1}$ [11], we obtain

$$\begin{aligned} \left\|u^k - \overline{u}\right\| &= \left\|(I + \lambda_k T)^{-1}(r^k) - (I + \lambda_k T)^{-1}(\overline{u} + \lambda_k \overline{v})\right\| \\ &\leq \left\|r^k - (\overline{u} + \lambda_k \overline{v})\right\| \\ &\leq \beta, \end{aligned}$$

and so $\{u^k\}$ is also bounded.

To prove the second claim, we apply some simple transformations to (27), obtaining the equivalent conditions

$$v^k + (1/\lambda_k)u^k = r_k/\lambda_k \qquad\qquad u^k \in T^{-1}(v^k).$$

If $\inf_k\{\lambda_k\} > 0$, then $\{1/\lambda_k\} \subset (0, \infty)$ and $\{r^k/\lambda_k\} \subset \mathcal{H}$ are bounded. Since T is monotone, so is T^{-1}. Thus, we may apply the first claim with a suitable redefinition of variables to conclude that $\{v^k\}$ is bounded. □

Proposition 9 *In Algorithm 3, assume that $(A + B)^{-1}(0) \neq \emptyset$, and that either $\mathcal{H}$ is finite-dimensional or $A + B$ is maximal monotone. If*

$$\inf_{k \geq 0} \left\{\mu_k/\lambda_k - (\alpha_k/2)^2\right\} > 0, \tag{28}$$

then either the algorithm halts with $(z^{k+1}, w^{k+1}) \in S_{\mathrm{e}}(A,B)$, or $\{z^k\}$, $\{x^k\}$, and $\{y^k\}$ all converge weakly to some z^∞ such that $0 \in A(z^\infty) + B(z^\infty)$. In this case, $\{w^k\}$ and $\{b^k\}$ converge weakly to some w^∞ such that $w^\infty \in B(z^\infty)$ and $-w^\infty \in A(z^\infty)$, while $\{a^k\}$ converges to $-w^\infty$.

Proof. Proposition 7 shows that step 1 of Algorithm 3 produces pairs $(x^k, b^k) \in \mathrm{Gph}(B)$, and $(y^k, a^k) \in \mathrm{Gph}(A)$ meeting the assumptions of Algorithm 2. Therefore, we may apply Proposition 6 to conclude that either we halt with $(z^{k+1}, w^{k+1}) \in S_{\mathrm{e}}(A,B)$, or $\{(z^k, w^k)\}$ is an infinite sequence Féjer monotone to $S_{\mathrm{e}}(A,B)$, with

$$\delta_k \;=\; \frac{\varphi_k(z^k, w^k)}{\sqrt{\|a^k + b^k\|^2 + \|x^k - y^k\|^2}} \;\to\; 0. \tag{29}$$

We now show that the denominator in (29) is bounded. First, since $\{(z^k, w^k)\}$ is bounded by Féjer monotonicity and $\{\lambda_k\}$ is bounded, the sequence $\{z^k + \lambda_k w^k\}$ appearing on the right-hand side of (16) is bounded. Since $\{\lambda_k\} \subset [\underline{\lambda}, \overline{\lambda}]$, Lemma 8 implies that $\{x^k\}$ and $\{b^k\}$ are bounded.

We next remark that from (15),

$$|\alpha_k| \le 2\sqrt{\mu_k/\lambda_k} \le 2\sqrt{\overline{\lambda}/\underline{\lambda}},$$

so $\{\alpha_k\}$ must also be a bounded sequence. It follows, using the boundedness of $\{(z^k, w^k)\}$ and $\{x^k\}$, that the right-hand side of (17) is bounded. Since $\{\mu_k\} \subset [\underline{\lambda}, \overline{\lambda}]$, Lemma 8 also implies that $\{y^k\}$ and $\{a^k\}$ are bounded.

Thus, all the variables appearing in the denominator in (29) are bounded, and consequently $\varphi_k(z^k, w^k) \to 0$. Using (19)-(20), we have

$$\xi_k \;\ge\; \frac{\inf_{k\ge 0}\{\mu_k/\lambda_k - (\alpha_k/2)^2\}}{\overline{\lambda}(\overline{\lambda}/\underline{\lambda} + 1)} \;>\; 0.$$

for all $k \ge 0$, so $\{\xi_k\}$ is bounded away from zero. From (18) and $\varphi_k(z^k, w^k) \to 0$, we then obtain $z^k - x^k \to 0$ and $z^k - y^k \to 0$, which lead immediately to $x^k - y^k \to 0$. Then, using (21)-(22), $\lambda_k, \mu_k \ge \underline{\lambda} > 0$, and the boundedness of $\{\alpha_k\}$, we deduce $b^k - w^k \to 0$ and $a^k + w^k \to 0$, which in turn yield $a^k + b^k \to 0$.

Since $\{(z^k, w^k)\}$ is bounded, it must have at least one weak limit point. Let (z^∞, w^∞) be any of its weak limit points, with $\mathcal{K} \subseteq \mathbb{N}$ an infinite set such that $(z^k, w^k) \xrightarrow{\mathsf{w}}_{\mathcal{K}} (z^\infty, w^\infty)$. Using $(z^k - x^k), (z^k - y^k), (b^k - w^k), (a^k + w^k) \to 0$, we also have $x^k \xrightarrow{\mathsf{w}}_{\mathcal{K}} z^\infty$, $y^k \xrightarrow{\mathsf{w}}_{\mathcal{K}} z^\infty$, $b^k \xrightarrow{\mathsf{w}}_{\mathcal{K}} w^\infty$ and $a^k \xrightarrow{\mathsf{w}}_{\mathcal{K}} -w^\infty$.

Now, suppose $\mathcal{H}$ is finite-dimensional. Then, since we have $(x^k, b^k) \in \mathrm{Gph}(B)$ and $(y^k, a^k) \in \mathrm{Gph}\, A$ for all k, we may take limits over $k \in \mathcal{K}$ and use the maximality of A and B to obtain $(z^\infty, w^\infty) \in \mathrm{Gph}(B)$ and $(z^\infty, -w^\infty) \in \mathrm{Gph}(A)$. It follows that $(z^\infty, -w^\infty) \in S_{\mathrm{e}}(A,B)$. Corollary 2 then gives $(z^k, w^k) \to (z^\infty, w^\infty)$.

On the other hand, suppose $\mathcal{H}$ is infinite-dimensional. In this, case we assume $A+B$ is maximal. Since $(x^k, b^k) \in \operatorname{Gph}(B)$ and $(y^k, a^k) \in \operatorname{Gph} A$ for all k, we may use $x^k - y^k \to 0$ and $a^k + b^k \to 0$ in the context of Proposition 10 in Appendix A to conclude that $(z^\infty, -w^\infty) \in S_{\mathrm{e}}(A, B)$. As the choice of the weak limit point (z^∞, w^∞) above was arbitrary, all weak limit points of $\{(z^k, w^k)\}$ are in $S_{\mathrm{e}}(A, B)$. Therefore, Proposition 6 guarantees $(z^k, w^k) \xrightarrow{\mathsf{w}} (z^\infty, w^\infty) \in S_{\mathrm{e}}(A, B)$.

Finally, we may again use $(z^k - x^k), (z^k - y^k), (b^k - w^k), (a^k + w^k) \to 0$ to obtain $x^k, y^k \xrightarrow{\mathsf{w}} z^\infty$, $b^k \xrightarrow{\mathsf{w}} w^\infty$, and $a^k \xrightarrow{\mathsf{w}} -w^\infty$, where one may dispense with the "w" in the finite-dimensional case. □

It might appear that one way to generalize the Algorithm 3 would be to substitute an affine combination $(1-\beta_k)w^k + \beta_k b^k$ for w^k in (17). Letting the first affine combination factor be α'_k instead of α_k, this modification yields

$$y^k + \mu_k a^k = (1-\alpha'_k)z_k + \alpha'_k x_k - \mu_k \left[(1-\beta_k)w^k + \beta_k b^k\right]. \tag{30}$$

However, substituting $b^k = w^k + (1/\lambda_k)(z^k - x^k)$ into this equation yields

$$\begin{aligned} y^k + \mu_k a^k &= (1-\alpha'_k)z_k + \alpha'_k x_k - \mu_k \left[w^k + \frac{\beta_k}{\lambda_k}(z^k - x^k)\right] && (31) \\ &= \left(1 - \left(\alpha'_k + \frac{\beta_k \mu_k}{\lambda_k}\right)\right) z^k + \left(\alpha'_k + \frac{\beta_k \mu_k}{\lambda_k}\right) x^k - \mu_k w^k, && (32) \end{aligned}$$

so introducing such an extra parameter β_k is equivalent to simply taking $\alpha_k = \alpha'_k + \beta_k \mu_k / \lambda_k$ in (17).

For reference, the full set of recursions for Algorithm 3 is

$$\begin{aligned} x^k + \lambda_k b^k &= z^k + \lambda_k w^k, && \text{where } b^k \in B(x^k) && (33) \\ y^k + \mu_k a^k &= (1-\alpha_k)z_k + \alpha_k x_k - \mu_k w^k, && \text{where } a^k \in A(y^k) && (34) \\ \sigma_k &= \frac{\langle z^k - x^k, b^k - w^k\rangle + \langle z^k - y^k, a^k + w^k\rangle}{\|a^k + b^k\|^2 + \|x^k - y^k\|^2} && && (35) \\ z^{k+1} &= z^k - \rho_k \sigma_k (a^k + b^k) && && (36) \\ w^{k+1} &= w^k - \rho_k \sigma_k (x^k - y^k). && && (37) \end{aligned}$$

The algorithm (33)-(37) is very general: one may use a different proximal parameter λ_k in each iteration, and even different proximal parameters μ_k and λ_k for the operators A and B. The affine combination factors α_k may also be varied, along with the relaxation factors ρ_k. In classical Douglas-Rachford splitting, only ρ_k may be varied; one must effectively have $\lambda_k = \mu_k = \lambda > 0$ for all k; see Section 5.4 for further discussion.

Note also that, with some care, it is also possible to exchange the roles of the operators A and B during the execution of the algorithm, either periodically or as often as every iteration. However, we do not explicitly include this possibility in (33)-(37), since it would make the notation more complicated.

5 Reformulations and special cases

5.1 Including a scale factor

In some splitting methods, it is helpful to reformulate the problem $0 \in A(x) + B(x)$ via multiplying through by some scalar $\eta > 0$, yielding the problem

$$0 \in \eta A(x) + \eta B(x).$$

This trivial reformulation leaves the solution set unchanged, but can change the form of Douglas-Rachford-based splitting algorithms. A similar effect occurs in the method (33)-(37). If we apply Algorithm 3 to ηA and ηB, instead of to A and B, the set $S_{\mathrm{e}}(A, B)$ is transformed in the simple manner

$$S_{\mathrm{e}}(\eta A, \eta B) = \{(z, \eta w) \mid (z, w) \in S_{\mathrm{e}}(A, B)\}.$$

Applying (33)-(37) with the substitutions $A \to \eta A$, $B \to \eta B$, and also the change of variables

$$w^k \to \eta w^k \qquad a^k \to \eta a^k \qquad b^k \to \eta b^k, \tag{38}$$

one obtains

$$\begin{aligned}
x^k + \lambda_k \eta b^k &= z^k + \lambda_k \eta w^k, & \eta b^k \in \eta B(x^k) \quad (39)\\
y^k + \mu_k \eta a^k &= (1-\alpha_k) z_k + \alpha_k x_k - \mu_k \eta w^k, & \eta a^k \in \eta A(y^k) \quad (40)\\
\sigma_k &= \frac{\langle z^k - x^k, \eta b^k - \eta w^k\rangle + \langle z^k - y^k, \eta a^k + \eta w^k\rangle}{\|\eta a^k + \eta b^k\|^2 + \|x^k - y^k\|^2} & (41)\\
z^{k+1} &= z^k - \rho_k \sigma_k (\eta a^k + \eta b^k) & (42)\\
\eta w^{k+1} &= \eta w^k - \rho_k \sigma_k (x^k - y^k). & (43)
\end{aligned}$$

Simplifying and dividing both the numerator and denominator of (41) by η, one arrives at

$$\begin{aligned}
x^k + \lambda_k \eta b^k &= z^k + \lambda_k \eta w^k, & \text{where } b^k \in B(x^k) \quad (44)\\
y^k + \mu_k \eta a^k &= (1-\alpha_k) z_k + \alpha_k x_k - \mu_k \eta w^k, & \text{where } a^k \in A(y^k) \quad (45)\\
\sigma_k &= \frac{\langle z^k - x^k, b^k - w^k\rangle + \langle z^k - y^k, a^k + w^k\rangle}{\eta\,\|a^k + b^k\|^2 + \frac{1}{\eta}\|x^k - y^k\|^2} & (46)\\
z^{k+1} &= z^k - \rho_k \sigma_k \eta (a^k + b^k) & (47)\\
w^{k+1} &= w^k - \frac{\rho_k \sigma_k}{\eta}(x^k - y^k). & (48)
\end{aligned}$$

The η factors could then be eliminated from (44) and (45) by a simple redefinition of $\{\lambda_k\}$ and $\{\mu_k\}$. In either case, if the hypotheses of Proposition 9 hold, $\{(z^k, w^k)\}$ will be weakly convergent to a point in $S_{\mathrm{e}}(A, B)$.

Note also that under if we change the definitions of $\{w^k\}$, $\{a^k\}$, and $\{b^k\}$ as in (38), the identity (23) becomes

$$\varphi_k(z^k, w^k) = \frac{1}{\eta\lambda_k}\left\|z^k - x^k\right\|^2 + \frac{1}{\eta\mu_k}\left\|z^k - y^k\right\|^2 - \frac{\alpha_k}{\eta\mu_k}\langle z^k - x^k, z^k - y^k\rangle. \tag{49}$$

In theory, the scaling factor η must remain fixed throughout the algorithm. In practice, however, it might be periodically adjusted to assure that one is giving appropriate relative weight to respective "dual" and "primal" optimality conditions $x^k - y^k = 0$ and $a^k + b^k = 0$. For example, one could periodically adjust η so that the relative step lengths obtained from (47) and (48) after canceling the identical factors $\rho_k\sigma_k$, namely

$$\frac{\eta\left\|a^k + b^k\right\|}{\|z^k\|} \qquad \frac{\left\|x^k - y^k\right\|}{\eta\left\|w^k\right\|},$$

have approximately the same magnitude. This heuristic yields

$$\eta \approx \sqrt{\frac{\|x^k - y^k\|\,\|z^k\|}{\|a^k + b^k\|\,\|w^k\|}}.$$

To retain the theoretical convergence results of Proposition 9, such an adjustment could only be done a finite number of times.

We now consider some illustrative special cases of the method.

5.2 A "parallel" special case: Spingarn's splitting method

Consider specializing Algorithm (44)-(48) by setting $\alpha_k = 0$ for all k. Then,

$$\inf_{k\geq 0}\left\{\mu_k/\lambda_k - (\alpha_k/2)^2\right\} \;=\; \inf_{k\geq 0}\{\mu_k/\lambda_k\} \;\geq\; \underline{\lambda}/\overline{\lambda} \;>\; 0,$$

so (28) is satisfied and Proposition 9 applies, implying weak convergence. With a few simplifications and substitutions, one obtains the method

$$x^k + \lambda_k\eta b^k = z^k + \lambda_k\eta w^k, \qquad \text{where } b^k \in B(x^k) \tag{50}$$

$$y^k + \mu_k\eta a^k = z^k - \mu_k\eta w^k, \qquad \text{where } a^k \in A(y^k) \tag{51}$$

$$\sigma_k = \frac{\frac{1}{\lambda_k}\left\|z^k - x^k\right\|^2 + \frac{1}{\mu_k}\left\|z^k - y^k\right\|^2}{\eta^2\left\|a^k + b^k\right\|^2 + \left\|x^k - y^k\right\|^2} \tag{52}$$

$$z^{k+1} = z^k - \rho_k\sigma_k\eta(a^k + b^k) \tag{53}$$

$$w^{k+1} = w^k - \frac{\rho_k\sigma_k}{\eta}(x^k - y^k), \tag{54}$$

weakly convergent under the assumptions $0 < \underline{\lambda} \leq \lambda_k, \mu_k \leq \overline{\lambda}$ and $0 < \underline{\rho} \leq \rho_k \leq \overline{\rho} < 2$ for all $k \geq 0$.

Setting $\alpha_k = 0$ makes steps (50) and (51) independent, possibly allowing them to be performed in parallel.

This algorithm bears a close resemblance to two-operator case of Spingarn's splitting method [18, Section 5]. In our notation, that method takes the following form:

$$x^k + \eta b^k = z^k + \eta w^k, \qquad \text{where } b^k \in B(x^k) \tag{55}$$
$$y^k + \eta a^k = z^k - \eta w^k, \qquad \text{where } a^k \in A(y^k) \tag{56}$$
$$z^{k+1} = (1-\rho_k)z^k + \rho_k \tfrac{1}{2}(x^k + y^k) \tag{57}$$
$$w^{k+1} = (1-\rho_k)w^k + \rho_k \tfrac{1}{2}(b^k - a^k). \tag{58}$$

(Note that [18] presents only the case $\rho_k \equiv 1$, but a general ρ_k obeying $0 < \underline{\rho} \le \rho_k \le \overline{\rho} < 2$ is easily added to the algorithm in the manner of [4].)

The first two steps (50)-(51) and (55)-(56) of these methods are identical if we fix $\lambda_k = \mu_k = 1$ for all $k \ge 0$. The remaining calculations appear superficially different, but in fact Spingarn's method is a special case of (50)-(54), as we shall now establish.

In (50)-(54), fix $\lambda_k = \mu_k = 1$ for all $k \ge 0$. Then we may rearrange (50) and (51) respectively into

$$b^k = w^k + \tfrac{1}{\eta}(z^k - x^k) \qquad a^k = -w^k + \tfrac{1}{\eta}(z^k - y^k).$$

Adding these two equations, one obtains

$$a^k + b^k = \tfrac{1}{\eta}\left[(z^k - y^k) + (z^k - x^k)\right] \tag{59}$$
$$= \tfrac{1}{\eta}\left[2z^k - (x^k + y^k)\right]. \tag{60}$$

Substituting (59) and the identity $x^k + y^k = (z^k - y^k) - (z^k - x^k)$ into the denominator of (52), we have

$$\eta^2 \left\|a^k + b^k\right\|^2 + \left\|x^k - y^k\right\|^2 = 2\left\|z^k - x^k\right\|^2 + 2\left\|z^k - y^k\right\|^2,$$

and therefore, in view of $\lambda_k = \mu_k = 1$, we have $\sigma_k = 1/2$ for all $k \ge 0$. Substituting $\sigma_k = 1/2$ and (60) into (53), one has

$$z^{k+1} = z^k - \rho_k \tfrac{1}{2}\big(2z^k - (x^k + y^k)\big) = (1-\rho_k)z^k + \rho_k \tfrac{1}{2}(x^k + y^k),$$

which is identical to (57). Again rearranging (50)-(51) with $\lambda_k = \mu_k = 1$, we also have

$$x^k = z^k + \eta(w^k - b^k) \qquad y^k = z^k - \eta(w^k + a^k),$$

and thus

$$x^k - y^k = \eta\left(2w^k - a^k + b^k\right).$$

Substituting this expression, along with $\sigma_k = 1/2$, into (54) yields

$$w^{k+1} = w^k - \rho_k \tfrac{1}{2}\left(2w^k - a^k + b^k\right) = (1-\rho_k)w^k + \rho_k \tfrac{1}{2}(b^k - a^k),$$

which is identical to (58). Therefore, we conclude that algorithm (55)-(58) is equivalent to (50)-(54) under the restriction $\lambda_k = \mu_k = 1$ for all k. Thus, the Spingarn splitting algorithm is a special case of (50)-(54).

The extra generality provided by the form (50)-(54) should be useful, since the proximal parameters may be varied from iteration to iteration, and even from operator to operator, while the overall computational effort per iteration should very similar to (55)-(58).

Incidentally, the original Spingarn splitting method is actually an instance of Douglas-Rachford splitting, but in $\mathcal{H} \times \mathcal{H}$. Define the operators $T_1, T_2 : \mathcal{H} \times \mathcal{H} \rightrightarrows \mathcal{H} \times \mathcal{H}$ as follows:

$$T_1(x,y) = B(x) \times A(y) \tag{61}$$

$$T_2(x,y) = \begin{cases} \{(w,-w) \mid w \in \mathcal{H}\} & x = y \\ \emptyset & x \neq y. \end{cases} \tag{62}$$

T_1 is maximal monotone by the maximal monotonicity of A and B. It is also easily verified that T_2 is maximal monotone, either directly, or because it is the normal cone mapping of the linear subspace $\{(x,y) \in \mathcal{H} \times \mathcal{H} \mid x = y\}$. A zero of $T_1 + T_2$ is a point of the form $(z,z) \in \mathcal{H} \times \mathcal{H}$, where $A(z) + B(z) \ni 0$. Applying Douglas-Rachford splitting to T_1 and T_2 yields the method (55)-(58).

5.3 Some "sequential" special cases

Instead of setting $\alpha_k = 0$ for all k, we could set $\alpha_k = 1$ for all k. Then (28) reduces to

$$\inf_{k \geq 0} \{\mu_k / \lambda_k\} > 1/4.$$

Under this condition, Proposition 9 applies. Including a scaling factor η as in (44)-(48), we then obtain the method

$$x^k + \lambda_k \eta b^k = z^k + \lambda_k \eta w^k, \qquad \text{where } b^k \in B(x^k) \tag{63}$$

$$y^k + \mu_k \eta a^k = x^k - \mu_k \eta w^k, \qquad \text{where } a^k \in A(y^k) \tag{64}$$

$$\sigma_k = \frac{\langle z^k - w^k, b^k - w^k \rangle + \langle z^k - y^k, a^k + w^k \rangle}{\eta \left\| a^k + b^k \right\|^2 + \frac{1}{\eta} \left\| x^k - y^k \right\|^2} \tag{65}$$

$$z^{k+1} = z^k - \rho_k \sigma_k \eta (a^k + b^k) \tag{66}$$

$$w^{k+1} = w^k - \frac{\rho_k \sigma_k}{\eta}(x^k - y^k), \tag{67}$$

weakly convergent under the assumptions $0 < \underline{\lambda} \leq \lambda_k, \mu_k \leq \overline{\lambda}$, $\mu_k \geq (1/4 + \epsilon)\lambda_k$, and $0 < \underline{\rho} \leq \rho_k \leq \overline{\rho} < 2$ for all $k \geq 0$, where $\epsilon > 0$ is an arbitrary small scalar. The method is similar to (50)-(54), except that x^k replaces z^k in (64). Thus, one must execute (63) before (64); these steps cannot be performed simultaneously like (50)-(51). A similar observation can be made whenever $\alpha_k \neq 0$. A possible advantage of (64) is that it makes use of more recent information than (51), namely x^k instead of z^k.

Another possibility is to set $\alpha_k = \mu_k/\lambda_k$ for all $k \geq 0$. In this case, we have, as in (30)-(32),

$$\begin{aligned} y^k + \mu_k a^k &= \left(1 - \frac{\mu_k}{\lambda_k}\right) z^k + \frac{\mu_k}{\lambda_k} x^k - \mu_k w^k \\ &= z^k - \mu_k \left(w^k + \frac{1}{\lambda_k}\left(z^k - x^k\right)\right) \\ &= z^k - \mu_k b^k. \end{aligned}$$

In this case, condition (28) reduces to

$$\inf_{k\geq 0} \left\{ \frac{\mu_k}{\lambda_k} - \frac{1}{4}\left(\frac{\mu_k}{\lambda_k}\right)^2 \right\} > 0,$$

which, since μ_k/λ_k must be positive, is equivalent to $\sup_{k\geq 0}\{\mu_k/\lambda_k\} < 4$. We thus obtain from Proposition 9 the (weak) convergence of the method

$$x^k + \lambda_k \eta b^k = z^k + \lambda_k \eta w^k, \qquad \text{where } b^k \in B(x^k) \tag{68}$$

$$y^k + \mu_k \eta a^k = z^k - \mu_k \eta b^k, \qquad \text{where } a^k \in A(y^k) \tag{69}$$

$$\sigma_k = \frac{\langle z^k - w^k, b^k - w^k\rangle + \langle z^k - y^k, a^k + w^k\rangle}{\eta \left\|a^k + b^k\right\|^2 + \frac{1}{\eta}\left\|x^k - y^k\right\|^2} \tag{70}$$

$$z^{k+1} = z^k - \rho_k \sigma_k \eta (a^k + b^k) \tag{71}$$

$$w^{k+1} = w^k - \frac{\rho_k \sigma_k}{\eta}(x^k - y^k), \tag{72}$$

under the assumptions $0 < \underline{\lambda} \leq \lambda_k, \mu_k \leq \overline{\lambda}$, $\mu_k \leq (4-\epsilon)\lambda_k$, and $0 < \underline{\rho} \leq \rho_k \leq \overline{\rho} < 2$ for all $k \geq 0$, where $\epsilon > 0$ is once again some small scalar. This method resembles (63)-(67), except that rather than replacing z^k in (50) with x^k, recursion (69) replaces w^k with b^k.

5.4 Douglas-Rachford: a forbidden boundary case

Suppose we wish to use as much recent information as possible by replacing *both* z^k and w^k in (51) with x^k and b^k, respectively. Following (30)-(32), these substitutions are equivalent to taking $\alpha_k = 1 + \mu_k/\lambda_k$ for all $k \geq 0$. However, in this case conditions (15) and (28) require

$$\frac{\mu_k}{\lambda_k} - \frac{1}{4}\left(1 + \frac{\mu_k}{\lambda_k}\right)^2 > 0 \quad \Leftrightarrow \quad \left(1 + \frac{\mu_k}{\lambda_k}\right)^2 - \frac{4\mu_k}{\lambda_k} < 0 \quad \Leftrightarrow \quad \left(1 - \frac{\mu_k}{\lambda_k}\right)^2 < 0,$$

which clearly cannot be satisfied. Then closest one can come to meeting this condition is to set $\mu_k = \lambda_k$ and thus $\alpha_k = 2$ for every k, in which case $\mu_k/\lambda_k - (\alpha_k/2)^2 = 0$, rather than $\mu_k/\lambda_k - (\alpha_k/2)^2 > 0$ as required. The hypotheses of Proposition 9 are therefore not satisfied. We may consider this situation a "boundary" case, since for arbitrarily small $\epsilon > 0$, the choice $\lambda_k = \mu_k$ and $\alpha_k = 2 - \epsilon$ for all k *would* satisfy the conditions of Proposition 9.

However, it is informative to continue the analysis by fixing $\lambda_k = \mu_k = 1$ and $\alpha_k = 2$ for all k. Then, (44)-(45) reduce to

$$\begin{aligned} x^k + \eta b^k &= z^k + \eta w^k && (73) \\ y^k + \eta a^k &= x_k - \eta b^k. && (74) \end{aligned}$$

From (74), one has $x^k - y^k = \eta(a^k + b^k)$, and hence

$$\eta \left\|a^k + b^k\right\|^2 + \tfrac{1}{\eta} \left\|x^k - y^k\right\|^2 = 2\eta \left\|a^k + b^k\right\|^2 .$$

Next, from (49) with that $\lambda_k = \mu_k = 1$ and $\alpha_k = 2$, we have

$$\begin{aligned} \varphi_k(z^k, w^k) &= \tfrac{1}{\eta} \left\|z^k - x^k\right\|^2 + \tfrac{1}{\eta} \left\|z^k - y^k\right\|^2 - \tfrac{2}{\eta}\langle z^k - x^k, z^k - y^k\rangle \\ &= \tfrac{1}{\eta} \left\|x^k - y^k\right\|^2 \\ &= \eta \left\|a^k + b^k\right\|^2 , \end{aligned}$$

where the last equality follows from (74). Therefore, we have

$$\sigma_k = \frac{\eta \left\|a^k + b^k\right\|^2}{2\eta \left\|a^k + b^k\right\|^2} = 1/2 \qquad \forall\, k \geq 0.$$

Steps (47)-(48) then become

$$\begin{aligned} z^{k+1} &= z_k - \frac{\rho_k \eta}{2}(a^k + b^k), \\ w^{k+1} &= w_k - \frac{\rho_k}{2\eta}(x^k - y^k) = w_k - \frac{\rho_k}{2}(a^k + b^k). \end{aligned}$$

At the first step of the next iteration, one then has $b^{k+1} \in B(x^{k+1})$ and

$$\begin{aligned} x^{k+1} + \eta b^{k+1} &= z^{k+1} + \eta w^{k+1} \\ &= z_k - \frac{\rho_k \eta}{2}(a^k + b^k) + \eta \left(w_k - \frac{\rho_k}{2}(a^k + b^k)\right) \\ &= (z^k + \eta w^k) - \eta\rho_k(a^k + b^k) \\ &= (x^k + \eta b^k) - \eta\rho_k(a^k + b^k) \\ &= \rho_k(x^k - \eta a^k) + (1 - \rho_k)(x^k + \eta b^k) \\ &= \rho_k(y^k + \eta b^k) + (1 - \rho_k)(x^k + \eta b^k) \\ &= \rho_k y^k + (1 - \rho_k)x^k + \eta b^k, \end{aligned}$$

where the we use the substitutions $x^k + \eta b^k = z^k + \eta w^k$ from (73) and $x^k - \eta a^k = y^k + \eta b^k$ from (74). Thus, we conclude that the sequence $\{(x^k, b^k, y^k, a^k)\}$ produced by the method has the properties

$$\begin{aligned} y^k + \eta a^k &= x^k - \eta b^k & a^k &\in A(y^k), \\ x^{k+1} + \eta b^{k+1} &= \rho_k y^k + (1-\rho_k)x^k + \eta b^k & b^{k+1} &\in B(x^{k+1}) \end{aligned}$$

for all $k \geq 0$. These are exactly the recursions used in generalized Douglas-Rachford splitting, as described in [4]. In particular, if we fix $\rho_k \equiv 1$, we obtain classical Douglas-Rachford splitting. We thus conclude that Douglas-Rachford splitting is an excluded boundary special case of Algorithm 3, where one has $\mu_k/\lambda_k - (\alpha_k/2)^2 = 0$.

All known analyses of Douglas-Rachford splitting are essentially equivalent to showing that the sequence $\{x^k + \eta b^k\}$ is Féjer monotone to the set

$$\begin{aligned} S_\eta(A,B) &\stackrel{\text{def}}{=} \{z + \eta w \mid w \in B(z), -w \in A(z)\} \\ &= [I \;\; \eta I]\, S_{\mathrm{e}}(A,B). \end{aligned}$$

Thus, the convergence mechanism operates in a different, lower-dimensional space than Proposition 9. In theory, the proximal factor η cannot be varied as the algorithm progresses, because the set $S_\eta(A,B)$ will change, and Féjer monotonicity lost.

By using a slightly smaller value of α_k, say $\alpha_k = 2 - \epsilon$, and performing the appropriate projection operation, one can use Proposition 9 to introduce varying proximal factors λ_k and μ_k; for α_k near 2, these factors would have to be almost equal, but they could vary virtually arbitrarily with k. Thus, the projective approach to splitting affords far more flexibility than the classical Douglas-Rachford framework.

References

[1] H. H. Bauschke and J. M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Rev.*, 38(3):367–426, 1996.

[2] G. Cimmino. Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari. *Ric. Sci. Progr. Tecn. Econom. Naz.*, 1:326–333, 1938.

[3] J. Eckstein. Some saddle-function splitting methods for convex programming. *Optim. Meth. Software*, 4(1):75–83, 1994.

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

[5] J. Eckstein and M. C. Ferris. Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control. *INFORMS J. Comput.*, 10(2):218–235, 1998.

[6] D. Gabay. Applications of the method of multipliers to variational inequalities. In M. Fortin and R. Glowinski, editors, *Augmented Lagrangian Methods: Applications to the Solution of Boundary Value Problems*, chapter IX, pages 299–340. North-Holland, Amsterdam, 1983.

[7] S. Kaczmarz. Angenherte Auflsung von Systemen linearer Gleichungen. *Bull. Int. Acad. Polon. Sci. A*, 1937:355–357, 1937.

[8] J. Lawrence and J. E. Spingarn. On fixed points of nonexpansive piecewise isometric mappings. *Proc. London Math. Soc. (3)*, 55(3):605–624, 1987.

[9] P.-L. Lions. Une méthode itérative de résolution d'une inéquation variationnelle. *Israel J. Math.*, 31(2):204–208, 1978.

[10] P.-L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.*, 16(6):964–979, 1979.

[11] G. J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Math. J.*, 29:341–346, 1962.

[12] Z. Opial. Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Amer. Math. Soc.*, 73:591–597, 1967.

[13] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl.*, 72(2):383–390, 1979.

[14] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optimization*, 14(5):877–898, 1976.

[15] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient-proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Anal.*, 7(4):323–345, 1999.

[16] M. V. Solodov and B. F. Svaiter. A hybrid projection-proximal point algorithm. *J. Convex Anal.*, 6(1):59–70, 1999.

[17] M. V. Solodov and B. F. Svaiter. A unified framework for some inexact proximal point algorithms. *Numer. Funct. Anal. Optim.*, 22(7-8):1013–1035, 2001.

[18] J. E. Spingarn. Partial inverse of a monotone operator. *Appl. Math. Optim.*, 10(3):247–265, 1983.

[19] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.*, 38(2):431–446, 2000.

A A technical result for infinite dimension

Let $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximal monotone. Suppose that

$$\begin{aligned} \{(x^k, b^k)\} &\subset \operatorname{Gph}(B) & \qquad (x^k, b^k) &\to (z, w) \\ \{(y^k, a^k)\} &\subset \operatorname{Gph}(A) & \qquad (y^k, a^k) &\to (z, -w). \end{aligned}$$

Then one may immediately conclude, using that $\operatorname{Gph}(A)$ and $\operatorname{Gph}(B)$ are closed, that

$$\|x^k - y^k\| \to 0 \qquad \|a^k + b^k\| \to 0 \qquad (z, w) \in S_{\mathrm{e}}(A, B).$$

If $\mathcal{H}$ is finite-dimensional, this situation occurs near the end of the proof of Proposition 9, after passing to a subsequence. If $\mathcal{H}$ has infinite dimension, however, we have at that point only *weak* convergence of (x^k, b^k) and (y^k, a^k) (again, after passing to a subsequence). Since $\operatorname{Gph}(A)$ and $\operatorname{Gph}(B)$ may not be *weakly* closed, we cannot immediately conclude in this instance that $(z, w) \in S_{\mathrm{e}}(A, B)$.

Fortunately, we can still prove $(z, w) \in S_{\mathrm{e}}(A, B)$ if we make the mild regularity assumption that the operator $A + B$ is maximal, and know that $\|x^k - y^k\| \to 0$ and $\|a^k + b^k\| \to 0$ (strongly). The last two assumptions are already established in the proof of Proposition 9. Formally:

Proposition 10 *Let $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximal monotone, and suppose all of the following hold:*

$$\begin{aligned} \{(x^k, b^k)\} &\subset \operatorname{Gph}(B) & \qquad (x^k, b^k) &\xrightarrow{\mathsf{w}} (z, w) & \qquad \|x^k - y^k\| &\to 0 \\ \{(y^k, a^k)\} &\subset \operatorname{Gph}(A) & \qquad (y^k, a^k) &\xrightarrow{\mathsf{w}} (z, -w) & \qquad \|a^k + b^k\| &\to 0. \end{aligned}$$

Then, if $A + B$ is maximal monotone, one has $(z, w) \in S_{\mathrm{e}}(A, B)$.

Proof. First we claim that

$$0 \in (A + B)(z). \tag{75}$$

To prove this claim, take an arbitrary $(z', w') \in \operatorname{Gph}(A + B)$. Then, there exist $b' \in B(z')$, $a' \in A(z')$ such that $a' + b' = w'$. Therefore,

$$\langle x^k - z', b^k - b' \rangle \geq 0 \qquad \langle y^k - z', a^k - a' \rangle \geq 0.$$

The second of these inequalities can be rewritten as

$$\langle x^k - z', a^k - a' \rangle \geq \langle x^k - y^k, a^k - a' \rangle.$$

Adding $\langle x^k - z', b^k - b' \rangle \geq 0$, we find

$$\langle x^k - z', a^k + b^k - (a' + b') \rangle \geq \langle x^k - y^k, a^k - a' \rangle,$$

or, since $a' + b' = w'$,

$$\langle x^k - z', -w' \rangle \geq -\langle x^k - z', a^k + b^k \rangle + \langle x^k - y^k, a^k - a' \rangle. \tag{76}$$

Since $\|x^k - y^k\| \to 0$ and $\{a^k\}$ is bounded, we have $\langle x^k - y^k, a^k - a' \rangle \to 0$. We also have $\|a^k + b^k\| \to 0$ with $\{x^k\}$ bounded, so $\langle x^k - z', a^k + b^k \rangle \to 0$. Using these two facts and taking limits in (76),

$$\langle z - z', 0 - w' \rangle = \lim_{k\to\infty} \langle x^k - z', 0 - w' \rangle \geq 0.$$

As (z', w') was an arbitrary point in $\mathrm{Gph}(A + B)$, and $A + B$ is maximal monotone, we conclude that $(z, 0) \in \mathrm{Gph}(A + B)$, and our claim (75) holds.

Next, we claim that

$$\lim_{k\to\infty} \langle x^k, b^k \rangle = \langle z, w \rangle \qquad \lim_{k\to\infty} \langle y^k, a^k \rangle = \langle z, -w \rangle. \tag{77}$$

In view of (75), there exists $b \in \mathcal{H}$ such that $b \in B(z)$ and $-b \in A(z)$. Since A and B are monotone,

$$\langle x^k - z, b^k - b \rangle \geq 0 \qquad \langle y^k - z, a^k + b \rangle \geq 0.$$

Equivalently,

$$\langle x^k, b^k \rangle \geq \langle z, b^k - b \rangle + \langle x^k, b \rangle \qquad \langle y^k, a^k \rangle \geq \langle z, a^k + b \rangle - \langle y^k, b \rangle.$$

The right-hand sides of the above inequalities converge to $\langle z, w \rangle$ and $\langle z, -w \rangle$, respectively. Hence,

$$\liminf_{k\to\infty} \langle x^k, b^k \rangle \geq \langle z, w \rangle \qquad \liminf_{k\to\infty} \langle y^k, a^k \rangle \geq -\langle z, w \rangle. \tag{78}$$

Direct manipulation yields

$$\langle y^k, a^k \rangle = -\langle x^k, b^k \rangle + \langle x^k, a^k + b^k \rangle + \langle y^k - x^k, a^k \rangle. \tag{79}$$

As the two last terms on the right hand side of this equation converge to 0, we conclude that

$$\liminf_{k\to\infty} \langle y^k, a^k \rangle = \liminf_{k\to\infty} -\langle x^k, b^k \rangle = -\limsup_{k\to\infty} \langle x^k, b^k \rangle.$$

Combining this equation with the second inequality in (78) we conclude that

$$\limsup_{k\to\infty} \langle x^k, b^k \rangle \leq \langle z, w \rangle.$$

In view of the first inequality in (78), we conclude that $\langle x^k, b^k \rangle$ converges to $\langle z, w \rangle$. Taking limits in (79), we then also conclude that $\langle y^k, a^k \rangle$ converges to $-\langle z, w \rangle$, and the second claim (77) is established.

Now, take any $(x', b') \in \mathrm{Gph}(B)$. Then

$$\langle x^k - x', b^k - b' \rangle \geq 0,$$

or equivalently,

$$\langle x^k, b^k \rangle - \langle x^k, b' \rangle - \langle x', b^k \rangle + \langle x', b' \rangle \geq 0.$$

Taking limits and using (77),

$$\langle z, w \rangle - \langle z, b' \rangle - \langle x', w \rangle + \langle x', b' \rangle \geq 0,$$

which is equivalent to $\langle z - x', w - b' \rangle \geq 0$. As B is maximal monotone and $(x', b') \in \mathrm{Gph}(B)$ was arbitrary, we conclude that $(z, w) \in \mathrm{Gph}(B)$. By similar reasoning, $(z, -w) \in \mathrm{Gph}(A)$, and so $(z, w) \in S_{\mathrm{e}}(A, B)$. □