



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

Nonlinear Analysis 60 (2005) 283–301

**Nonlinear  
Analysis**

[www.elsevier.com/locate/na](http://www.elsevier.com/locate/na)

# The asymptotic behavior of the composition of two resolvents

Heinz H. Bauschke<sup>a</sup>, Patrick L. Combettes<sup>b</sup>, Simeon Reich<sup>c,\*</sup>

<sup>a</sup>*Department of Mathematics and Statistics, University of Guelph, Guelph, Ont., Canada N1G 2W1*

<sup>b</sup>*Laboratoire Jacques-Louis Lions, Université Pierre et Marie Curie – Paris 6, 75005 Paris, France*

<sup>c</sup>*Department of Mathematics, The Technion—Israel Institute of Technology, 32000 Haifa, Israel*

Received 7 July 2004; accepted 23 July 2004

## Abstract

The asymptotic behavior of the composition of two resolvents in a Hilbert space is investigated. Connections are made between the solutions of associated monotone inclusion problems and their dual versions. The applications provided include a study of an alternating minimization procedure and a new proof of von Neumann's classical result on the method of alternating projections.

© 2004 Elsevier Ltd. All rights reserved.

*MSC:* 47H05; 47H09; 47J25; 90C25

*Keywords:* Duality; Firmly nonexpansive operator; Gradient projection method; Hilbert space; Monotone inclusion; Monotone operator; Proximal iteration; Resolvent

## 1. Introduction

Throughout,  $\mathcal{H}$  is a real Hilbert space with inner product  $\langle \cdot | \cdot \rangle$  and induced norm  $\| \cdot \|$ . Let  $A$  and  $B$  be two maximal monotone operators from  $\mathcal{H}$  to  $2^{\mathcal{H}}$  with resolvents  $J_A$  and  $J_B$ , respectively, and let  $\gamma \in ]0, +\infty[$ . Our paper is concerned with the inclusion problem

$$\text{find } (x, y) \in \mathcal{H}^2 \quad \text{such that} \quad (0, 0) \in (\text{Id} - R + \gamma(A \times B))(x, y), \quad (1)$$

\* Corresponding author. Tel.: +972-4-829-4172; fax: +972-4-829-3388.

*E-mail addresses:* [hbauschk@uoguelph.ca](mailto:hbauschk@uoguelph.ca) (H.H. Bauschke), [plc@math.jussieu.fr](mailto:plc@math.jussieu.fr) (P.L. Combettes), [sreich@technion.technion.ac.il](mailto:sreich@technion.technion.ac.il) (S. Reich).

where  $R: (x, y) \mapsto (y, x)$ . This abstract formulation subsumes a wide spectrum of problems in nonlinear analysis and its applications. We thus recover problems arising in variational inequalities [30], best approximation [12], image processing [4,44], mechanics [34], and optimization [1,31]. The dual inclusion problem associated with (1) is

$$\begin{aligned} &\text{find } (x^*, y^*) \in \mathcal{H}^2 \text{ such that} \\ &(0, 0) \in ((\text{Id} - R)^{-1} + (A^{-1} \times B^{-1}) \circ (\text{Id}/\gamma))(x^*, y^*). \end{aligned} \tag{2}$$

Now consider the alternating resolvent method

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad y_n = J_{\gamma B} x_n, \quad x_{n+1} = J_{\gamma A} y_n, \tag{3}$$

where  $\mathbb{N} = \{0, 1, 2, \dots\}$ . The objective of the present paper is to provide a systematic investigation of the asymptotic behavior of the sequences  $(x_n)_{n \in \mathbb{N}}$ ,  $(y_n)_{n \in \mathbb{N}}$ ,  $(y_n - x_n)_{n \in \mathbb{N}}$ , and  $(x_{n+1} - y_n)_{n \in \mathbb{N}}$  generated by this algorithm in connection with the solutions of (1) and (2). When specialized to the case when  $A$  and  $B$  are subdifferentials, our results will be significantly refined and will yield new insights into an alternating minimization procedure.

The remainder of the paper is organized as follows. Section 2 contains basic notation and auxiliary results on nonexpansive and monotone operators. In Section 3, we provide a detailed investigation of the asymptotic behavior of (3). The applications discussed in that section include variational inequalities as well as the problem of finding cycles for inconsistent feasibility problems. In Section 4, the results of Section 3 are sharpened in the context of proximity operators and we obtain new results on the primal and dual behavior of an alternating minimization procedure. Among the applications presented is a new proof of von Neumann’s classical result on the convergence of alternating projections.

## 2. Auxiliary results

We recall some useful results on monotone operators and resolvents. Let  $A: \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be a set-valued operator. The sets  $\text{dom } A = \{x \in \mathcal{H} \mid Ax \neq \emptyset\}$ ,  $\text{ran } A = \{u \in \mathcal{H} \mid (\exists x \in \mathcal{H}) u \in Ax\}$ , and  $\text{gr } A = \{(x, u) \in \mathcal{H}^2 \mid u \in Ax\}$  are the domain, the range, and the graph of  $A$ , respectively. The inverse of  $A$  is the set-valued operator  $A^{-1}$  with graph  $\{(u, x) \in \mathcal{H}^2 \mid u \in Ax\}$ , the resolvent of  $A$  is  $J_A = (\text{Id} + A)^{-1}$ , and the Yosida approximation of  $A$  of index  $\gamma \in ]0, +\infty[$  is

$$J_{\gamma} A = (\text{Id} - J_{\gamma A})/\gamma = (\text{Id} + A^{-1}/\gamma)^{-1} \circ (\text{Id}/\gamma). \tag{4}$$

The operator  $A$  is monotone if  $\langle x - y \mid u - v \rangle \geq 0$ , for all  $(x, u)$  and  $(y, v)$  in  $\text{gr } A$ . If  $A$  is monotone and  $\text{gr } A$  cannot be enlarged without destroying monotonicity, then  $A$  is maximal monotone. A classical result due to Minty [35] implies that  $A$  is maximal monotone if and only if  $J_A$  is firmly nonexpansive with domain  $\mathcal{H}$ . We now provide basic properties of firmly nonexpansive operators (see [24, Sections 1.9 and 1.11] for proofs and additional properties).

**Fact 2.1.** Let  $T: \mathcal{H} \rightarrow \mathcal{H}$  be an operator with domain  $D$ . Then  $T$  is firmly nonexpansive on  $D$  if any of the following equivalent properties holds:

- (i)  $(\forall x \in D)(\forall y \in D) \|Tx - Ty\|^2 \leq \|x - y\|^2 - \|(\text{Id} - T)x - (\text{Id} - T)y\|^2$ .
- (ii)  $2T - \text{Id}$  is nonexpansive, i.e.,  $(\forall x \in D)(\forall y \in D) \|(2T - \text{Id})x - (2T - \text{Id})y\| \leq \|x - y\|$ .
- (iii)  $T = J_A$  for some monotone operator  $A: \mathcal{H} \rightarrow 2^{\mathcal{H}}$ .

The class of firmly nonexpansive operators is not closed under composition. In [15], Bruck and Reich introduced the class of strongly nonexpansive operators: a nonexpansive operator  $T: \mathcal{H} \rightarrow \mathcal{H}$  is strongly nonexpansive on  $D = \text{dom } T$  if  $(x_n - y_n) - (Tx_n - Ty_n) \rightarrow 0$ , whenever  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  are sequences in  $D$  such that  $(x_n - y_n)_{n \in \mathbb{N}}$  is bounded and  $\|x_n - y_n\| - \|Tx_n - Ty_n\| \rightarrow 0$ . We shall utilize the following properties of strongly nonexpansive operators (which hold true in considerably more general settings, see [15,40]). As usual,

$$\text{Fix } T = \{x \in \mathcal{H} \mid Tx = x\} \tag{5}$$

denotes the set of fixed points of an operator  $T: \mathcal{H} \rightarrow \mathcal{H}$ .

**Fact 2.2.** Let  $T$  and  $(T_i)_{1 \leq i \leq m}$  be operators from  $\mathcal{H}$  to  $\mathcal{H}$  with domain  $\mathcal{H}$ . Then the following properties hold:

- (i) If  $T$  is firmly nonexpansive, then it is strongly nonexpansive.
- (ii) If the operators  $(T_i)_{1 \leq i \leq m}$  are strongly nonexpansive, then the composition  $T_1 \cdots T_m$  is also strongly nonexpansive.
- (iii) Suppose that  $T$  is strongly nonexpansive and let  $x_0 \in \mathcal{H}$ . If  $\text{Fix } T \neq \emptyset$ , then the sequence  $(T^n x_0)_{n \in \mathbb{N}}$  converges weakly to some point in  $\text{Fix } T$ ; otherwise,  $\|T^n x_0\| \rightarrow +\infty$ .

**Proof.** (i): See [15, Proposition 2.1]. (ii): See [15, Proposition 1.1]. (iii): See [15, Corollaries 1.3 and 1.4].  $\square$

Now let  $A: \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be a set-valued operator. We use the notation

$$A^\vee = A \circ (-\text{Id}) \tag{6}$$

and

$$\tilde{A} = (-\text{Id}) \circ A^{-1} \circ (-\text{Id}) = -((A^{-1})^\vee). \tag{7}$$

It is straightforward to verify that  $\tilde{A}$  is (maximal) monotone if and only if  $A$  is, that  $y \in \tilde{A}x$  if and only if  $-x \in A(-y)$ , and that

$$\tilde{\gamma A} = \gamma \text{Id} + \tilde{A}. \tag{8}$$

These concepts admit finer descriptions when specialized to subdifferentials. Let  $\Gamma_0(\mathcal{H})$  be the class of all proper lower semicontinuous convex functions from  $\mathcal{H}$  to  $] -\infty, +\infty]$ . Now take  $f \in \Gamma_0(\mathcal{H})$ . Then the subdifferential operator  $\partial f$  of  $f$  is maximal monotone and

the proximity operator  $\text{prox}_f = J_{\partial f}$  has full domain  $\mathcal{H}$  [36]. For  $\gamma \in ]0, +\infty[$ , the function

$$\gamma f = f \square \left( \frac{1}{2\gamma} \|\cdot\|^2 \right) \tag{9}$$

is the Moreau envelope of  $f$  of index  $\gamma$ , where  $\square$  denotes the inf-convolution operator. We have [36]

$$\nabla(\gamma f) = \gamma(\partial f). \tag{10}$$

Now let  $f^*$  denote the conjugate of  $f$ . Then Moreau’s decomposition [36] states that  $\frac{1}{2} \|\cdot\|^2 = (\frac{1}{2} \|\cdot\|^2 \square f) + (\frac{1}{2} \|\cdot\|^2 \square f^*)$ . Differentiation then yields  $\text{Id} = \text{prox}_f + \text{prox}_{f^*}$ . More generally, we deduce from (4) that

$$(\forall x \in \mathcal{H}) \quad x = \text{prox}_{\gamma f} x + \gamma \text{prox}_{f^*/\gamma}(x/\gamma). \tag{11}$$

If  $f$  is the indicator function of a nonempty closed convex set  $C$ , denoted by  $\iota_C$ , then  $\text{prox}_f$  is the projector onto  $C$ , denoted by  $P_C$ .

We conclude this section with a resolvent identity.

**Fact 2.3.** Let  $A, B: \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be maximal monotone operators and let  $\rho \in ]0, +\infty[$ . Then

$$\text{Fix } J_{\rho A}((1 - \rho)\text{Id} + \rho J_B) = \text{Fix } J_A J_B. \tag{12}$$

**Proof.** Take  $x \in \mathcal{H}$ . Then the following equivalences hold:  $x \in \text{Fix}(J_A J_B) \Leftrightarrow J_B x - x \in Ax \Leftrightarrow (1 - \rho)x + \rho J_B x \in x + \rho Ax \Leftrightarrow ((1 - \rho)\text{Id} + \rho J_B)x \in (\text{Id} + \rho A)x \Leftrightarrow x = (\text{Id} + \rho A)^{-1}((1 - \rho)\text{Id} + \rho J_B)x \Leftrightarrow x \in \text{Fix } J_{\rho A}((1 - \rho)\text{Id} + \rho J_B)$ .  $\square$

The following facts appear implicitly in [3,34].

**Fact 2.4.** Let  $H$  be a real Hilbert space and let  $A, B: H \rightarrow 2^H$  be maximal monotone operators. Let  $S$  be the set of solutions to the primal inclusion problem

$$\text{find } x \in H \quad \text{such that} \quad 0 \in Ax + Bx \tag{13}$$

and let  $S^*$  be the set of solutions to the associated dual problem

$$\text{find } x^* \in H \quad \text{such that} \quad 0 \in A^{-1}x^* + \tilde{B}x^*. \tag{14}$$

Then:

- (i)  $S = \{x \in H \mid (\exists x^* \in S^*) \quad x^* \in Ax \quad \text{and} \quad -x^* \in Bx\}$ .
- (ii)  $S^* = \{x^* \in H \mid (\exists x \in S) \quad x \in A^{-1}x^* \quad \text{and} \quad -x \in \tilde{B}x^*\}$ .

### 3. The composition of two resolvents

Throughout this section,  $A, B: \mathcal{H} \rightarrow 2^{\mathcal{H}}$  are maximal monotone operators and  $\gamma \in ]0, +\infty[$ .

3.1. Primal and dual inclusions

A regularization of the problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx \tag{15}$$

is [31,34]

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + \gamma Bx. \tag{16}$$

As seen in Fact 2.4, the dual problem of (16) is

$$\text{find } x^* \in \mathcal{H} \text{ such that } 0 \in A^{-1}x^* + \gamma \tilde{B}x^*. \tag{17}$$

An alternative regularization of (15) is

$$\text{find } y \in \mathcal{H} \text{ such that } 0 \in \gamma Ay + By, \tag{18}$$

the dual of which is

$$\text{find } y^* \in \mathcal{H} \text{ such that } 0 \in \gamma \tilde{A}y^* + B^{-1}y^*. \tag{19}$$

The connection between the composition of the two resolvents  $J_{\gamma A}$  and  $J_{\gamma B}$  and the inclusion problems (16)–(19) will be explored after the following definition.

**Definition 3.1.** Set

- (i)  $S = (\text{Id} - R + \gamma(A \times B))^{-1}(0, 0)$ ;
- (ii)  $S^* = ((\text{Id} - R)^{-1} + (A^{-1} \times B^{-1}) \circ (\text{Id}/\gamma))^{-1}(0, 0)$ ;
- (iii)  $E = (A + \gamma B)^{-1}(0)$  and  $F = (\gamma A + B)^{-1}(0)$ ;
- (iv)  $u^* = J_{(A^{-1} + \tilde{B})/\gamma}(0)$  and  $v^* = J_{(\tilde{A} + B^{-1})/\gamma}(0)$ ; note that, depending on the domains of  $J_{(A^{-1} + \tilde{B})/\gamma}$  and  $J_{(\tilde{A} + B^{-1})/\gamma}$ , the vectors  $u^*$  and  $v^*$  may not exist.

In other words, the solution sets of (1), (2), (16), and (18) are denoted by  $S, S^*, E,$  and  $F,$  respectively.

**Proposition 3.2.** The following identities hold:

- (i)  $E = \text{Fix } J_{\gamma A} J_{\gamma B} = J_{\gamma A}(F)$  and  $F = \text{Fix } J_{\gamma B} J_{\gamma A} = J_{\gamma B}(E)$ .
- (ii)  $S = \text{Fix } J_{\gamma(A \times B)} R = (E \times F) \cap \text{gr } J_{\gamma B}$ .
- (iii)  $S^* = \{(\gamma u^*, \gamma v^*)\}$  and  $v^* = -u^*$ .
- (iv)  $S^* = (R - \text{Id})(S)$ .
- (v)  $J_{\gamma B}|_E: E \rightarrow F: x \mapsto x + \gamma u^*$  is a bijection with inverse  $J_{\gamma A}|_F: F \rightarrow E: y \mapsto y + \gamma v^*$ .
- (vi)  $E = A^{-1}(u^*) \cap (\gamma B)^{-1}(v^*)$  and  $F = (\gamma A)^{-1}(u^*) \cap B^{-1}(v^*)$ .
- (vii)  $S = (E \times F) \cap (R - \text{Id})^{-1}(S^*)$ .

**Proof.** (i) The equivalences

$$\begin{aligned} x \in E &\Leftrightarrow 0 \in Ax + \gamma Bx \Leftrightarrow J_{\gamma B}x \in x + \gamma Ax \\ &\Leftrightarrow x = J_{\gamma A} J_{\gamma B}x \Leftrightarrow x \in \text{Fix } J_{\gamma A} J_{\gamma B} \end{aligned} \tag{20}$$

show that  $E = \text{Fix } J_{\gamma A} J_{\gamma B}$ ; the proof of  $F = \text{Fix } J_{\gamma B} J_{\gamma A}$  is analogous. If  $x \in \text{Fix } J_{\gamma A} J_{\gamma B}$ , then  $J_{\gamma B} x \in \text{Fix } J_{\gamma B} J_{\gamma A}$ . Hence  $J_{\gamma B}(E) \subset F$  and similarly  $J_{\gamma A}(F) \subset E$ . Now apply  $J_{\gamma A}$  and  $J_{\gamma B}$  to the last two inclusions and obtain the remaining inclusions  $E \subset J_{\gamma A}(F)$  and  $F \subset J_{\gamma B}(E)$ , respectively.

(ii) These two identities result from the equivalences

$$\begin{aligned}
 (x, y) \in S &\Leftrightarrow R(x, y) \in (\text{Id} + \gamma(A \times B))(x, y) \\
 &\Leftrightarrow (x, y) \in \text{Fix } J_{\gamma(A \times B)} R \\
 &\Leftrightarrow x = J_{\gamma A} y \quad \text{and} \quad y = J_{\gamma B} x \\
 &\Leftrightarrow x \in \text{Fix } J_{\gamma A} J_{\gamma B}, \quad y \in \text{Fix } J_{\gamma B} J_{\gamma A}, \quad \text{and} \quad y = J_{\gamma B} x \\
 &\Leftrightarrow (x, y) \in E \times F \quad \text{and} \quad (x, y) \in \text{gr } J_{\gamma B},
 \end{aligned} \tag{21}$$

where the last equivalence follows from (i).

(iii) Take  $(x^*, y^*) \in S^*$ , i.e.,

$$(0, 0) \in (A^{-1}(x^*/\gamma) \times B^{-1}(y^*/\gamma)) + (\text{Id} - R)^{-1}(x^*, y^*).$$
 \tag{22}

Then, by Fact 2.4(ii), there exists  $(x, y) \in S$  such that

$$\begin{cases} (x, y) \in A^{-1}(x^*/\gamma) \times B^{-1}(y^*/\gamma), \\ (-x, -y) \in (\text{Id} - R)^{-1}(x^*, y^*). \end{cases} \tag{23}$$

These two inclusions can be rewritten as

$$\begin{cases} (x^*, y^*) \in \gamma(A \times B)(x, y), \\ (-x^*, -y^*) = (\text{Id} - R)(x, y). \end{cases} \tag{24}$$

In turn, (24) is equivalent to

$$\begin{cases} x \in A^{-1}(x^*/\gamma), \\ -y \in \tilde{B}(x^*/\gamma), \\ x^* = y - x = -y^*, \end{cases} \tag{25}$$

or, alternatively, to

$$\begin{cases} -x \in \tilde{A}(y^*/\gamma), \\ y \in B^{-1}(y^*/\gamma), \\ x^* = y - x = -y^*. \end{cases} \tag{26}$$

Adding the two inclusions in both (25) and (26), we arrive at

$$\begin{cases} -x^* = x - y \in (A^{-1} + \tilde{B})(x^*/\gamma), \\ -y^* = y - x \in (\tilde{A} + B^{-1})(y^*/\gamma). \end{cases} \tag{27}$$

In view of Definition 3.1(iv), we conclude that

$$\begin{cases} x^*/\gamma = J_{(A^{-1} + \tilde{B})/\gamma}(0) = u^*, \\ y^*/\gamma = J_{(\tilde{A} + B^{-1})/\gamma}(0) = v^*. \end{cases} \tag{28}$$

We have thus proven that

$$S^* \subset \{(\gamma u^*, \gamma v^*)\} \text{ and, if } S^* \neq \emptyset, \text{ then } v^* = -u^*.$$
 \tag{29}

Now suppose that  $u^* = J_{(A^{-1} + \tilde{B})/\gamma}(0)$  exists. Then  $0 \in \gamma u^* + A^{-1}u^* + \tilde{B}u^* = A^{-1}u^* + \tilde{\gamma}B u^*$  by (8). In other words,  $u^*$  solves the dual problem (17) and, by Fact 2.4(ii), the primal problem (16) also has a solution. Accordingly, by Definition 3.1(iii),  $E \neq \emptyset$ . In turn, it follows from (i) and (ii) that  $S \neq \emptyset$ . Consequently, invoking Fact 2.4(i), we obtain  $S^* \neq \emptyset$  and we therefore deduce from (29) that (iii) holds.

(iv): Since  $S = \emptyset \Leftrightarrow S^* \neq \emptyset$ , it is enough to assume that  $S^* \neq \emptyset$ . Since  $S^* = \{(\gamma u^*, \gamma v^*)\}$ , it follows from (25) that there exists  $(x, y) \in S$  such that  $\gamma u^* = y - x = \gamma v^*$ . Hence  $S^* \subset (\text{Id} - R)(S)$ . Conversely, take  $(x, y) \in S$ . Then, by Fact 2.4(i), (24) holds and therefore  $\gamma u^* = y - x = \gamma v^*$ . Thus  $(\text{Id} - R)(S) \subset S^*$ .

(v): It is clear from (i) that  $J_{\gamma B}|_E: E \rightarrow F$  is surjective. Now take  $x \in E$ . Then (ii) implies that  $(x, J_{\gamma B}x) \in S$ . By (iii) and (iv),  $\gamma(u^*, v^*) = (R - \text{Id})(x, J_{\gamma B}x)$  and hence  $J_{\gamma B}|_E$  is a translation by  $\gamma u^*$ , the inverse of which is a translation by  $-\gamma u^* = \gamma v^*$ , namely,  $J_{\gamma A}|_F$ .

(vi): Fix  $x \in E$  and set  $y = J_{\gamma B}x$ . Then it follows from (i) and (ii) that  $(x, y) \in S$ . In turn, we deduce from (25) and (28) that

$$\begin{cases} x \in A^{-1}u^*, \\ v^* = (x - y)/\gamma = (x - J_{\gamma B}x)/\gamma = \gamma Bx, \end{cases} \tag{30}$$

whence  $x \in A^{-1}u^* \cap (\gamma B)^{-1}(v^*)$ . Accordingly,  $E \subset A^{-1}u^* \cap (\gamma B)^{-1}(v^*)$ . To verify the reverse inclusion, take  $x \in A^{-1}u^* \cap (\gamma B)^{-1}(v^*)$  and set  $y = x + \gamma u^*$ . Then  $x = J_{\gamma A}y$  and  $y = J_{\gamma B}x$ . Thus, (i) yields  $x \in \text{Fix } J_{\gamma A}J_{\gamma B} = E$  and we obtain  $A^{-1}u^* \cap (\gamma B)^{-1}(v^*) \subset E$ . The corresponding identity for  $F$  is derived in the same fashion.

(vii): On the one hand, (ii) yields  $S \subset E \times F$  and, on the other hand, (iv) yields  $S \subset (R - \text{Id})^{-1}(S^*)$ . Therefore  $S \subset (E \times F) \cap (R - \text{Id})^{-1}(S^*)$ . Conversely, let  $(x, y) \in (E \times F) \cap (R - \text{Id})^{-1}(S^*)$ . Then  $(x, y) \in (R - \text{Id})^{-1}(S^*)$  and it follows from (iii) that  $y = x + \gamma u^*$ . Hence, since  $(x, y) \in (E \times F)$ , (v) yields  $y = J_{\gamma B}x$ . In view of (ii), we conclude that  $(x, y) \in S$ .  $\square$

### 3.2. Asymptotic behavior

The objective of this section is to study the asymptotic behavior of the alternating resolvent method

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad y_n = J_{\gamma B}x_n, \quad x_{n+1} = J_{\gamma A}y_n. \tag{31}$$

**Theorem 3.3.** *Let  $\gamma \in ]0, +\infty[$ , let  $A, B: \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be maximal monotone operators such that  $S \neq \emptyset$ , and let  $((x_n, y_n))_{n \in \mathbb{N}}$  be the sequence generated by (31). Then:*

- (i) *The sequence  $((x_n, y_n))_{n \in \mathbb{N}}$  converges weakly to a point in  $S$ .*
- (ii) *For every  $(\bar{x}, \bar{y}) \in S$ ,*

$$\begin{aligned} \sum_{n \in \mathbb{N}} \|(x_n - y_n) - (\bar{x} - \bar{y})\|^2 < +\infty \quad \text{and} \\ \sum_{n \in \mathbb{N}} \|(x_{n+1} - y_n) - (\bar{x} - \bar{y})\|^2 < +\infty. \end{aligned} \tag{32}$$

(iii) The sequence  $((y_n - x_n, x_{n+1} - y_n))_{n \in \mathbb{N}}$  converges strongly to  $\gamma(u^*, v^*)$ .

**Proof.** Let  $(\bar{x}, \bar{y}) \in S$ . Then we derive from Proposition 3.2(i) and (v) the equalities

$$\bar{y} = J_{\gamma B} \bar{x} = \bar{x} + \gamma u^* \quad \text{and} \quad \bar{x} = J_{\gamma A} \bar{y} = \bar{y} + \gamma v^* = \bar{y} - \gamma u^*. \tag{33}$$

Hence, it follows from Fact 2.1(i) and (31) that

$$\begin{aligned} (\forall n \in \mathbb{N}) \quad \|x_{n+1} - \bar{x}\|^2 &= \|J_{\gamma A} J_{\gamma B} x_n - J_{\gamma A} J_{\gamma B} \bar{x}\|^2 \\ &\leq \|J_{\gamma B} x_n - J_{\gamma B} \bar{x}\|^2 - \|(\text{Id} - J_{\gamma A}) J_{\gamma B} x_n \\ &\quad - (\text{Id} - J_{\gamma A}) J_{\gamma B} \bar{x}\|^2 \\ &\leq \|x_n - \bar{x}\|^2 - \|(\text{Id} - J_{\gamma B}) x_n - (\text{Id} - J_{\gamma B}) \bar{x}\|^2 \\ &\quad - \|(\text{Id} - J_{\gamma A}) J_{\gamma B} x_n - (\text{Id} - J_{\gamma A}) J_{\gamma B} \bar{x}\|^2 \\ &= \|x_n - \bar{x}\|^2 - \|(x_n - y_n) - (\bar{x} - \bar{y})\|^2 \\ &\quad - \|(y_n - x_{n+1}) - (\bar{y} - \bar{x})\|^2. \end{aligned} \tag{34}$$

Therefore,

$$\sum_{n \in \mathbb{N}} (\|(x_n - y_n) - (\bar{x} - \bar{y})\|^2 + \|(x_{n+1} - y_n) - (\bar{x} - \bar{y})\|^2) \leq \|x_0 - \bar{x}\|^2 \tag{35}$$

and (ii) is established. In view of (33), we deduce from (ii) that  $y_n - x_n \rightarrow \bar{y} - \bar{x} = \gamma u^*$  and  $x_{n+1} - y_n \rightarrow \bar{x} - \bar{y} = \gamma v^*$ . Hence (iii) holds and, moreover,  $x_{n+1} - x_n \rightarrow 0$ . Since  $J_{\gamma A} J_{\gamma B}$  is nonexpansive, it follows from [38, Theorem 1] that  $(x_n)_{n \in \mathbb{N}}$  converges weakly to a fixed point  $x$  of  $J_{\gamma A} J_{\gamma B}$ . Now let  $y = J_{\gamma B} x$ . Then Proposition 3.2(i) and (ii) yield  $(x, y) \in S$ . Hence, as seen above,  $y_n - x_n \rightarrow y - x$  and therefore  $y_n \rightarrow y$ . Thus,  $((x_n, y_n))_{n \in \mathbb{N}} \rightarrow (x, y) \in S$ .  $\square$

**Remark 3.4.** Some comments on Theorem 3.3 are in order.

- (i) Item (i) in Theorem 3.3 was obtained via different techniques in [34, Section 3.4]. It can also be deduced from results on iterations of firmly nonexpansive operators [32, Théorème 5.5.2], as well as from results on iterations of strongly nonexpansive operators [15]. Items (i) and (ii) can also be derived from results on the asymptotic behavior of averaged operators [20]. Recall that an averaged operator is a strict convex combination of a nonexpansive operator and the identity. This notion was first introduced and analyzed in [6].
- (ii) The sequences  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  in Theorem 3.3 converge strongly when both  $A$  and  $B$  are odd. Indeed, the resolvents and their compositions are odd and averaged and [6, Corollary 2.1] therefore applies. The counterexample of Genel and Lindenstrauss [23] shows that strong convergence of  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  may fail outside Euclidean spaces, even when  $A = 0$  (hence  $J_{\gamma A} = \text{Id}$  has no effect and only the resolvent  $J_{\gamma B}$  is iterated). The situation does not improve for proximity operators or even projectors; see Remark 4.7(ii) below.
- (iii) It would be interesting to find out in what form Theorem 3.3 can be extended to more general Banach space settings, in which firmly nonexpansive operators are defined as



resolvents of accretive operators [24, Section 1.11]. See also [15, Theorem 2.4]; [41] for weak and strong convergence results on the iteration of a single resolvent in Banach spaces.

- (iv) Theorem 3.3 may have a Hilbert ball counterpart based upon resolvents of co-accretive operators (which are firmly nonexpansive of the first kind and therefore (para-) strongly nonexpansive); see [42, p. 540]; [13, Section 9].

We now turn to the inconsistent case.

**Theorem 3.5.** *Let  $\gamma \in ]0, +\infty[$ , let  $A, B: \mathcal{H} \rightarrow 2^{\mathcal{H}}$  be maximal monotone operators such that  $S = \emptyset$ , and let  $((x_n, y_n))_{n \in \mathbb{N}}$  be the sequence generated by (31). Then  $\|x_n\| \rightarrow +\infty$  and  $\|y_n\| \rightarrow +\infty$ .*

**Proof.** We have, for every  $n \in \mathbb{N}$ ,  $x_n = T^n x_0$ , where  $T = J_{\gamma A} J_{\gamma B}$  is strongly nonexpansive (see Fact 2.2(i) and (ii)) with  $\text{Fix } T = \emptyset$  by Proposition 3.2(i) and (vii). Hence  $\|x_n\| \rightarrow +\infty$  by Fact 2.2 (iii). The sequence  $(y_n)_{n \in \mathbb{N}}$  is treated similarly.  $\square$

### 3.3. Applications

In this section, we discuss applications of Theorem 3.3 that feature at least one resolvent which is not a proximity operator.

#### 3.3.1. Variational inequalities

Take  $\gamma \in ]0, +\infty[$ ,  $f \in \Gamma_0(\mathcal{H})$ , and set  $A = \partial f$  so that  $A^{-1} = \partial f^*$ . Then  $E$  is the set of solutions to the variational inequality problem

$$\text{find } x \in \mathcal{H} \quad \text{such that} \quad (\forall z \in \mathcal{H}) \quad \langle x - z | \gamma Bx \rangle + f(x) \leq f(z), \tag{36}$$

which itself reduces to the classical variational inequality problem

$$\text{find } x \in C \quad \text{such that} \quad (\forall z \in C) \quad \langle x - z | \gamma Bx \rangle \leq 0 \tag{37}$$

when  $f = \iota_C$  for some nonempty closed convex set  $C$ . Moreover, algorithm (31) becomes

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad y_n = J_{\gamma B} x_n, \quad x_{n+1} = \text{prox}_{\gamma f} y_n. \tag{38}$$

Let us assume that (36) possesses at least one solution, i.e.,  $E \neq \emptyset$ . Then Proposition 3.2(i) and (ii) imply that  $S \neq \emptyset$ . Consequently, Theorem 3.3 states that  $(x_n)_{n \in \mathbb{N}}$  converges weakly to a point in  $E$  and that  $(y_n - x_n)_{n \in \mathbb{N}}$  converges strongly to the solution  $J_{(\partial f^* + \tilde{B})/\gamma}(0)$  of the associated dual variational inequality, which is derived from (17) to be

$$\text{find } x^* \in \mathcal{H} \quad \text{such that} \\ (\exists y^* \in B^{-1}(-x^*)) (\forall z^* \in \mathcal{H}) \quad \langle x^* - z^* | \gamma x^* - y^* \rangle + f^*(x^*) \leq f^*(z^*). \tag{39}$$

#### 3.3.2. Cycles for compositions of projection operators

Consider a finite family of nonempty closed convex sets  $(S_i)_{1 \leq i \leq m}$  in a real Hilbert space  $H$  with associated projectors  $(P_i)_{1 \leq i \leq m}$ , where  $m \geq 3$  (the case  $m = 2$  will be discussed in

Section 4.3). The convex feasibility problem is to

$$\text{find } x \in \bigcap_{i=1}^m S_i. \tag{40}$$

This formulation has found numerous applications in engineering and in the physical sciences; see, for instance, [10,16,19]. If the intersection of the sets  $(S_i)_{1 \leq i \leq m}$  is nonempty, then one can equivalently look for a fixed point of the composition  $P_m \cdots P_1$  [26]. This reformulation remains useful even when the intersection of the sets  $(S_i)_{1 \leq i \leq m}$  is empty; see [7,11,18] and references therein.

As in [7], it will be convenient to work in the product space  $H^m$ , where we define the product set  $\mathbf{S} = S_1 \times \cdots \times S_m$  and the circular right-shift operator  $R: H^m \rightarrow H^m: (x_1, x_2, \dots, x_m) \mapsto (x_m, x_1, \dots, x_{m-1})$ . The problem of finding a cycle, i.e., a point  $\mathbf{x} = (x_1, x_2, \dots, x_m) \in H^m$  satisfying  $x_1 = P_1 x_m, x_2 = P_2 x_1, \dots, x_m = P_m x_{m-1}$ , is equivalent to finding a fixed point of  $P_{\mathbf{S}} R$ . Although the projector  $P_{\mathbf{S}}$  is a resolvent (actually, a proximity operator; see Section 2), the operator  $R$  is not since it is not firmly nonexpansive. Theorem 3.3 is therefore not directly applicable. However, since  $R$  is an isometry, it is nonexpansive on the entire space  $H^m$ . Hence, Fact 2.1 implies that  $Q = (\text{Id} + R)/2$  is defined everywhere and firmly nonexpansive, and that it is the resolvent of the maximal monotone operator  $B = Q^{-1} - \text{Id}$ . Letting, furthermore,  $A = N_{\mathbf{S}} = \partial I_{\mathbf{S}}$  (the normal cone operator of  $\mathbf{S}$ ) and  $\rho = 2$ , we derive from Fact 2.3 that

$$\text{Fix } P_{\mathbf{S}} R = \text{Fix } P_{\mathbf{S}} Q. \tag{41}$$

Now, let us assume that cycles exist, i.e.,  $\text{Fix } P_{\mathbf{S}} R \neq \emptyset$ . Then we first deduce from Proposition 3.2(v) that every cycle  $\mathbf{x}$  satisfies

$$R\mathbf{x} - \mathbf{x} = 2J_{N_{\mathbf{S}}^{-1} + \tilde{B}}(\mathbf{0}). \tag{42}$$

Furthermore, if we take  $\mathbf{x}_0 \in H^m$  and set

$$(\forall n \in \mathbb{N}) \quad \mathbf{x}_{n+1} = P_{\mathbf{S}} Q \mathbf{x}_n, \tag{43}$$

then Theorem 3.3, Proposition 3.2(i), and (41) imply that  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  converges weakly to some cycle  $\mathbf{x}$ , that  $Q\mathbf{x}_n \rightarrow Q\mathbf{x}$ , and that  $R\mathbf{x}_n - \mathbf{x}_n = 2(Q\mathbf{x}_n - \mathbf{x}_n) \rightarrow 2(Q\mathbf{x} - \mathbf{x}) = R\mathbf{x} - \mathbf{x}$ . We observe that, since  $Q \neq Q^*$  for  $m \geq 3$ ,  $Q$  cannot be a proximity operator [36, Section 3].

### 3.3.3. Cycles for compositions of resolvents

The arguments just presented extend with minor modifications as follows. Suppose that  $(A_i)_{1 \leq i \leq m}$  are finitely many maximal monotone operators on a real Hilbert space  $H$  which admit at least one cycle, i.e., a point  $\mathbf{x} = (x_1, x_2, \dots, x_m) \in H^m$  satisfying  $x_1 = J_{A_1} x_m, x_2 = J_{A_2} x_1, \dots, x_m = J_{A_m} x_{m-1}$ . Now define  $A = A_1 \times \cdots \times A_m$ , let  $R$  be the circular right-shift operator, and set  $Q = (\text{Id} + R)/2$  and  $B = Q^{-1} - \text{Id}$ . Then a cycle is a fixed point of  $J_A R$  and Fact 2.3 (applied to  $A/2$  and  $B$ , with  $\rho = 2$ ) yields  $\text{Fix } J_A R = \text{Fix } J_{A/2} Q$ . Now take  $\mathbf{x}_0 \in H^m$  and generate a sequence  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  via  $\mathbf{x}_{n+1} = J_{A/2} Q \mathbf{x}_n$ , for all  $n \in \mathbb{N}$ . Then it follows from Theorem 3.3 that  $(\mathbf{x}_n)_{n \in \mathbb{N}}$  converges weakly to some cycle  $\mathbf{x}$ ,  $Q\mathbf{x}_n \rightarrow Q\mathbf{x}$ , and  $R\mathbf{x}_n - \mathbf{x}_n \rightarrow 2J_{A^{-1}(2 \cdot) + \tilde{B}}(\mathbf{0})$ .

#### 4. The composition of two proximity operators

We specialize the results of Section 3 on resolvents to proximity operators by setting

$$A = \partial f \quad \text{and} \quad B = \partial g, \quad \text{where} \quad \{f, g\} \subset \Gamma_0(\mathcal{H}). \tag{44}$$

This additional structure makes significant refinements of the previous results possible. Let us introduce the function

$$\Phi: \mathcal{H} \times \mathcal{H} \rightarrow ]-\infty, +\infty]: (x, y) \mapsto f(x) + g(y) + \frac{1}{2\gamma} \|x - y\|^2, \tag{45}$$

where  $\gamma \in ]0, +\infty[$ . It is assumed throughout that

$$\varphi = \inf \Phi(\mathcal{H} \times \mathcal{H}) \in \mathbb{R}. \tag{46}$$

##### 4.1. Primal and dual optimization problems

We first identify the sets  $S$ ,  $E$ , and  $F$  of Definition 3.1 in the present setting.

**Proposition 4.1.** *The following identities hold:*

- (i)  $S = \text{Argmin } \Phi$ .
- (ii)  $E = \text{Fix}(\text{prox}_{\gamma f} \text{prox}_{\gamma g}) = \text{Argmin}(f + \gamma g)$ .
- (iii)  $F = \text{Fix}(\text{prox}_{\gamma g} \text{prox}_{\gamma f}) = \text{Argmin}(\gamma f + g)$ .

**Proof.** Item (i) follows from the equivalences

$$\begin{aligned} (x, y) \text{ minimizes } \Phi \text{ over } \mathcal{H} \times \mathcal{H} &\Leftrightarrow (0, 0) \in \gamma \partial \Phi(x, y) \\ &= (\gamma \partial f(x) + x - y, \gamma \partial g(y) + y - x) \\ &\Leftrightarrow (0, 0) \in (\text{Id} - R + \gamma(\partial f \times \partial g))(x, y). \end{aligned} \tag{47}$$

Next, using (44) and (10), we observe that

$$A + \gamma B = \partial f + \gamma(\partial g) = \partial f + \partial(\gamma g) = \partial(f + \gamma g). \tag{48}$$

This identity and Definition 3.1(iii) imply that

$$x \in E \Leftrightarrow 0 \in (A + \gamma B)(x) \Leftrightarrow 0 \in \partial(f + \gamma g)(x) \Leftrightarrow x \text{ minimizes } f + \gamma g. \tag{49}$$

Hence (ii) is verified. The proof of (iii) is similar.  $\square$

The following definition is justified by the Fenchel–Rockafellar duality theorem (see, for instance, [43, Theorem 1] or [45, Corollary 2.8.5]).

**Definition 4.2.** Consider the problem of minimizing  $f + \gamma g$  over  $\mathcal{H}$ , and let  $w^* \in \mathcal{H}$  be the unique solution of its Fenchel dual problem, i.e.,

$$-\varphi = f^*(w^*) + g^*(-w^*) + \frac{\gamma}{2} \|w^*\|^2 = \min \left( f^* + g^{*\vee} + \frac{\gamma}{2} \|\cdot\|^2 \right) (\mathcal{H}). \tag{50}$$

**Proposition 4.3.** *The following statements hold true:*

- (i)  $w^* = \text{prox}_{(f^*+g^{*\vee})/\gamma}(0) = -\frac{1}{\gamma}\text{prox}_{\gamma(f^*+g^{*\vee})^*}(0) = -\frac{1}{\gamma}\text{prox}_{\gamma(f \square g^\vee)^{**}}(0)$ .
- (ii) *If  $f \square g^\vee$  is lower semicontinuous, then  $w^* = -\frac{1}{\gamma}\text{prox}_{\gamma(f \square g^\vee)}(0)$ . In particular, this happens when the cone generated by  $\text{dom } f^* + \text{dom } g^*$  is a closed vector subspace.*
- (iii) *The vector  $u^*$  in Definition 3.1(iv) exists if and only if  $f + \gamma g$  has minimizers, in which case  $u^* = w^*$ .*

**Proof.** Note that (50) implies that  $f^* + g^{*\vee} \in \Gamma_0(\mathcal{H})$  and, in turn, that  $(f^* + g^{*\vee})^* \in \Gamma_0(\mathcal{H})$ .

(i): The first equality is clear by Definition 4.2, and the second one follows from (11). Since  $(f \square g^\vee)^* = f^* + g^{\vee*} = f^* + g^{*\vee}$ , we have  $(f \square g^\vee)^{**} = (f^* + g^{*\vee})^*$  and hence the third equation is also verified.

(ii): If  $f \square g^\vee$  is lower semicontinuous, then  $f \square g^\vee \in \Gamma_0(\mathcal{H})$  and therefore  $f \square g^\vee = (f \square g^\vee)^{**}$ . In turn, the formula for  $w^*$  follows from (i). Now, since  $\text{dom } f^* + \text{dom } g^* = \text{dom } f^* - \text{dom } g^{*\vee}$ , the stated assumption is simply the Attouch–Brézis qualification condition for  $f^*$  and  $g^{*\vee}$  and it follows from [2, Theorem 1.1] or [45, Theorem 2.8.7] that  $(f^* + g^{*\vee})^* = f \square g^\vee$ , whence the lower semicontinuity of  $f \square g^\vee$ .

(iii): Proposition 3.2 implies that  $u^*$  exists  $\Leftrightarrow S^* \neq \emptyset \Leftrightarrow S \neq \emptyset \Leftrightarrow E \neq \emptyset$ . The equivalence thus follows from Proposition 4.1. Now assume that  $u^*$  exists. Then Definition 3.1(iv) and (44) yield

$$u^* = J_{((\partial f)^{-1} + \tilde{\partial} g)/\gamma}(0) = J_{(\partial f^* + \tilde{\partial}(g^{*\vee}))/\gamma}(0). \tag{51}$$

We deduce that  $0 \in (\gamma \text{Id} + \partial f^* + \tilde{\partial}(g^{*\vee}))(u^*) \subset (\gamma \text{Id} + \partial(f^* + g^{*\vee}))(u^*)$  and furthermore, using (i), that  $u^* = \text{prox}_{(f^*+g^{*\vee})/\gamma}(0) = w^*$ .  $\square$

#### 4.2. Asymptotic behavior

We begin with two preliminary results.

**Lemma 4.4.** *Let  $(x, y, u, v, w) \in \mathcal{H}^5$ . Then*

$$\begin{aligned} \|x - u\|^2 &= \|x - w\|^2 + \|w - v\|^2 - \|x - y\|^2 + \|(x - y) - (u - v)\|^2 \\ &\quad + 2\langle x - w \mid w - v \rangle + 2\langle u - v \mid v - y \rangle. \end{aligned} \tag{52}$$

**Proof.** This is verified by a straightforward calculation.  $\square$

**Proposition 4.5.** *Let  $\Phi$ ,  $\varphi$ , and  $w^*$  be as in (45), (46), and (50), respectively. Let  $((u_n, v_n))_{n \in \mathbb{N}}$  be a sequence in  $\mathcal{H} \times \mathcal{H}$  such that  $\Phi(u_n, v_n) \rightarrow \varphi$ . Then  $(v_n - u_n)/\gamma \rightarrow w^*$ .*

**Proof.** Eq. (50) and the Fenchel–Young inequality imply

$$\begin{aligned}
 0 &\leftarrow \Phi(u_n, v_n) - \varphi \\
 &= f(u_n) + g(v_n) + \frac{1}{2\gamma} \|u_n - v_n\|^2 + f^*(w^*) + g^*(-w^*) + \frac{\gamma}{2} \|w^*\|^2 \\
 &= (f(u_n) + f^*(w^*)) + (g(v_n) + g^*(-w^*)) + \frac{1}{2\gamma} \|u_n - v_n\|^2 + \frac{\gamma}{2} \|w^*\|^2 \\
 &\geq \langle u_n | w^* \rangle + \langle v_n | -w^* \rangle + \frac{1}{2\gamma} \|u_n - v_n\|^2 + \frac{\gamma}{2} \|w^*\|^2 \\
 &= \frac{1}{2\gamma} \|(v_n - u_n) - \gamma w^*\|^2 \\
 &\geq 0.
 \end{aligned}
 \tag{53}$$

Therefore,  $v_n - u_n \rightarrow \gamma w^*$ .  $\square$

The main result of this section is a refinement of Theorem 3.3 for the case of proximity operators.

**Theorem 4.6.** *Let  $\gamma \in ]0, +\infty[$ , let  $f$  and  $g$  be two functions in  $\Gamma_0(\mathcal{H})$ , and let*

$$\Phi: \mathcal{H} \times \mathcal{H} \rightarrow ]-\infty, +\infty]: (x, y) \mapsto f(x) + g(y) + \frac{1}{2\gamma} \|x - y\|^2.
 \tag{54}$$

Assume that  $\varphi = \inf \Phi(\mathcal{H} \times \mathcal{H}) \in \mathbb{R}$  and consider the sequences  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  generated by the method of alternating proximity operators

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad y_n = \text{prox}_{\gamma g}(x_n), \quad x_{n+1} = \text{prox}_{\gamma f}(y_n).
 \tag{55}$$

Then:

- (i)  $(\forall n \in \mathbb{N}) \quad \Phi(x_{n+1}, y_{n+1}) \leq \Phi(x_{n+1}, y_n) \leq \Phi(x_n, y_n)$ .
- (ii)  $\lim_{n \rightarrow +\infty} \Phi(x_{n+1}, y_n) = \lim_{n \rightarrow +\infty} \Phi(x_n, y_n) = \varphi$ .
- (iii) *The sequences  $((y_n - x_n)/\gamma)_{n \in \mathbb{N}}$  and  $((y_n - x_{n+1})/\gamma)_{n \in \mathbb{N}}$  converge strongly to the unique minimizer*

$$w^* = \text{prox}_{(f^* + g^{*\vee})/\gamma}(0) = -\frac{1}{\gamma} \text{prox}_{\gamma(f \square g^\vee)^{**}}(0)
 \tag{56}$$

of the function  $z^* \mapsto f^*(z^*) + g^*(-z^*) + (\gamma/2)\|z^*\|^2$ .

- (iv) *If  $\text{Argmin } \Phi \neq \emptyset$ , then  $(x_n)_{n \in \mathbb{N}}$  converges weakly to a minimizer  $\bar{x}$  of  $f + \gamma g$  and  $(y_n)_{n \in \mathbb{N}}$  converges weakly to a minimizer  $\bar{y}$  of  $\gamma f + g$ . Moreover,*

- (a)  $\bar{y} = \text{prox}_{\gamma g}(\bar{x})$ .
- (b)  $\sum_{n \in \mathbb{N}} \|(x_n - y_n) - (\bar{x} - \bar{y})\|^2 < +\infty$  and  $\sum_{n \in \mathbb{N}} \|(x_{n+1} - y_n) - (\bar{x} - \bar{y})\|^2 < +\infty$ .
- (c)  $\Phi(\bar{x}, \bar{y}) = \varphi$ .
- (d)  $\bar{y} - \bar{x} = \gamma w^*$ .
- (e)  $\sum_{n \in \mathbb{N}} (\Phi(x_{n+1}, y_n) - \varphi) < +\infty$  and  $\sum_{n \in \mathbb{N}} (\Phi(x_n, y_n) - \varphi) < +\infty$ .

(v) If  $\text{Argmin } \Phi = \emptyset$ , then  $\|x_n\| \rightarrow +\infty$  and  $\|y_n\| \rightarrow +\infty$ .

**Proof.** (i): For every  $n \in \mathbb{N}$ , we have

$$\begin{aligned}
 \varphi &\leq \Phi(x_{n+1}, y_{n+1}) \\
 &= f(x_{n+1}) + g(y_{n+1}) + \frac{1}{2\gamma} \|x_{n+1} - y_{n+1}\|^2 \\
 &= f(x_{n+1}) + g(\text{prox}_{\gamma g}(x_{n+1})) + \frac{1}{2\gamma} \|x_{n+1} - \text{prox}_{\gamma g}(x_{n+1})\|^2 \\
 &\leq f(x_{n+1}) + g(y_n) + \frac{1}{2\gamma} \|x_{n+1} - y_n\|^2 \\
 &= \Phi(x_{n+1}, y_n) \\
 &= f(\text{prox}_{\gamma f}(y_n)) + g(y_n) + \frac{1}{2\gamma} \|\text{prox}_{\gamma f}(y_n) - y_n\|^2 \\
 &\leq f(x_n) + g(y_n) + \frac{1}{2\gamma} \|x_n - y_n\|^2 \\
 &= \Phi(x_n, y_n).
 \end{aligned} \tag{57}$$

(ii): In view of (57), the two limits are well defined and coincide, say with  $\psi$ . Thus,

$$\varphi \leq \psi = \lim_{n \rightarrow +\infty} \Phi(x_{n+1}, y_n) = \lim_{n \rightarrow +\infty} \Phi(x_n, y_n). \tag{58}$$

Our next step is to prove that  $\varphi = \psi$ , i.e., by (58),  $\varphi \geq \psi$ . To this end, let  $n \in \mathbb{N}$  and  $(x, y) \in \text{dom } f \times \text{dom } g$ . Lemma 4.4 (with  $(u, v, w)$  replaced by  $(x_n, y_n, x_{n+1})$ ) yields

$$\begin{aligned}
 \|x - x_n\|^2 &= \|x - x_{n+1}\|^2 + \|x_{n+1} - y_n\|^2 - \|x - y\|^2 + \|(x - y) - (x_n - y_n)\|^2 \\
 &\quad + 2\langle x - x_{n+1} | x_{n+1} - y_n \rangle + 2\langle x_n - y_n | y_n - y \rangle.
 \end{aligned} \tag{59}$$

On the other hand, since  $y_n = \text{prox}_{\gamma g}(x_n)$  and  $x_{n+1} = \text{prox}_{\gamma f}(y_n)$ , we have  $x_n - y_n \in \gamma \partial g(y_n)$  and  $y_n - x_{n+1} \in \gamma \partial f(x_{n+1})$ , whence

$$\gamma g(y) - \gamma g(y_n) + \langle y_n - y | x_n - y_n \rangle \geq 0 \tag{60}$$

and

$$\gamma f(x) - \gamma f(x_{n+1}) + \langle x_{n+1} - x | y_n - x_{n+1} \rangle \geq 0. \tag{61}$$

Thus (59), (60), (61), (i), and (58) result in

$$\begin{aligned}
 \|x - x_n\|^2 - \|x - x_{n+1}\|^2 &= \|x_{n+1} - y_n\|^2 - \|x - y\|^2 + \|(x - y) - (x_n - y_n)\|^2 \\
 &\quad + 2\gamma(f(x_{n+1}) + g(y_n) - f(x) - g(y)) \\
 &\quad + 2(\gamma f(x) - \gamma f(x_{n+1}) + \langle x - x_{n+1} | x_{n+1} - y_n \rangle) \\
 &\quad + 2(\gamma g(y) - \gamma g(y_n) + \langle x_n - y_n | y_n - y \rangle) \\
 &\geq \|x_{n+1} - y_n\|^2 - \|x - y\|^2 + \|(x - y) - (x_n - y_n)\|^2 \\
 &\quad + 2\gamma(f(x_{n+1}) + g(y_n) - f(x) - g(y)) \\
 &= 2\gamma(\Phi(x_{n+1}, y_n) - \Phi(x, y)) + \|(y_n - x_n) - (y - x)\|^2 \\
 &\geq 2\gamma(\psi - \Phi(x, y)).
 \end{aligned} \tag{62}$$

Consequently,

$$2\gamma(\Phi(x_{n+1}, y_n) - \Phi(x, y)) + \|(y_n - x_n) - (y - x)\|^2 \leq \|x_n - x\|^2 - \|x_{n+1} - x\|^2. \tag{63}$$

Now suppose that  $\varphi < \psi$ . Then we can pick  $(x, y) \in \text{dom } f \times \text{dom } g$  so that

$$\varphi = \inf \Phi(\mathcal{H} \times \mathcal{H}) \leq \Phi(x, y) < \psi. \tag{64}$$

However, summing (62) over  $n \in \mathbb{N}$ , we arrive at the absurdity  $+\infty > 2\sum_{n \in \mathbb{N}}(\psi - \Phi(x, y)) = +\infty$ . Therefore,  $\varphi \geq \psi$ .

(iii): This claim follows from (ii), Proposition 4.5, and Proposition 4.3(i).

(iv): Suppose that  $\text{Argmin } \Phi \neq \emptyset$ . Then it follows from Proposition 4.1(i) that  $S \neq \emptyset$ . Therefore, (iv) and items (iv)(a) and (b) follow from Theorem 3.3(i) and (ii), Proposition 3.2(ii), and Proposition 4.1(ii) and (iii). (iv)(c): Since  $(\bar{x}, \bar{y}) \in S$  by Theorem 3.3(i), this claim follows at once from Proposition 4.1(i) and (46). (iv)(d): We deduce from (iv)(b) that  $y_n - x_n \rightarrow \bar{y} - \bar{x}$ . On the other hand, (iii) yields  $y_n - x_n \rightarrow \gamma w^*$ . (iv)(e): Let  $(x, y)$  be a minimizer of  $\Phi$ . Then summing (63) over  $n \in \mathbb{N}$ , we obtain the first summability property and, using (i), the second follows.

(v): Suppose that  $\text{Argmin } \Phi = \emptyset$ . Then  $S = \emptyset$  by Proposition 4.1(i) and the claim follows from Theorem 3.5. Alternatively, if the desired conclusion were false, then by (iii) the sequence  $((x_n, y_n))_{n \in \mathbb{N}}$  would have a weakly convergent subsequence, say  $(x_{k_n}, y_{k_n}) \rightharpoonup (\bar{x}, \bar{y})$ . However, the weak lower semicontinuity of  $\Phi$  on  $\mathcal{H} \times \mathcal{H}$  and (ii) would then imply that  $(\bar{x}, \bar{y})$  is a minimizer of  $\Phi$ , which is a contradiction.  $\square$

**Remark 4.7.** Some comments on Theorem 4.6 are in order.

- (i) In [1], Acker and Prestel obtained items (i), (ii), and (v) of Theorem 4.6 by different means. They also established the weak convergence of  $((x_n, y_n))_{n \in \mathbb{N}}$  to a minimizer of  $\Phi$ , as well as the strong convergence of  $(y_n - x_n)_{n \in \mathbb{N}}$ . The identification of the strong limit in (iii) as the solution of a dual optimization problem, as well as the other items in Theorem 4.6, is new.
- (ii) Strong convergence of  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  in Theorem 4.6(iv) is guaranteed when both  $f$  and  $g$  are even (which implies that  $(0, 0)$  is a minimizer of  $\Phi$  and that the subdifferential operators  $\partial f$  and  $\partial g$  are odd so that Remark 3.4(ii) is applicable). On the other hand, it is known that strong convergence may fail to hold, even when both proximity operators are projectors (see [28,33]) or even when only a single proximity operator is iterated (see [13,27]). See also Remark 4.8 below.
- (iii) Along the same lines as in Remark 3.4(iii), it would be interesting to know how these results generalize to Banach space settings. A starting point could be [39, Theorem 3.7(b)], which yields strong convergence of the sequence  $(x_{n+1} - x_n)_{n \in \mathbb{N}}$ .

**Remark 4.8** (*weak-but-not-strong convergence of the gradient projection method*). Let  $C$  and  $D$  be two closed convex subsets of  $\mathcal{H}$  with associated distance functions  $d_C$  and  $d_D$ , respectively, and assume that  $C \cap D \neq \emptyset$ .

(i) Set  $f = \iota_C$  and  $g = \iota_D$ . Then (55) yields the method of alternating projections

$$x_{n+1} = P_C P_D x_n. \tag{65}$$

This iteration converges weakly [14, Theorem 1] to a point in  $C \cap D$  (see also Theorem 4.6(iv)(a)). Now set  $h = d_D^2/2$ , which implies that  $\nabla h = \text{Id} - P_D$  is 1-Lipschitz. Observe that (65) assumes the form of the gradient projection method [25,29]

$$x_{n+1} = P_C(x_n - \nabla h(x_n)). \tag{66}$$

As shown in [28], (65) does not converge strongly in general and, therefore, neither does the gradient projection method.

(ii) Set  $f = \iota_{\mathcal{H}}$  and  $g: \mathcal{H} \rightarrow \mathbb{R}: x \mapsto \frac{1}{2} \min \|(C - x) \cap (x - D)\|^2$ . Then  $\text{prox}_f = \text{Id}$ ,  $\text{prox}_g = (P_C + P_D)/2$  [13, Theorem 6.1], and (55) becomes the method of barycentric projections

$$x_{n+1} = \frac{P_C x_n + P_D x_n}{2}, \tag{67}$$

which is known to converge weakly to a point in  $C \cap D$  [5]. Now set  $h = (d_C^2 + d_D^2)/2$ . Then (67) assumes the form of the (unconstrained) gradient method [22]

$$x_{n+1} = x_n - \nabla h(x_n). \tag{68}$$

As shown in [13, Corollary 7.1], (67) does not converge strongly in general and, therefore, neither does the gradient method.

### 4.3. Applications

Our first application concerns the method of alternating projections. Set  $f = \iota_C$  and  $g = \iota_D$  in (45), where  $C$  and  $D$  are nonempty closed convex sets in  $\mathcal{H}$ . Then minimizing  $\Phi$  reduces to finding a best approximation pair, i.e.,  $(x, y) \in C \times D$  such that  $\|x - y\|$  is equal to the gap  $\inf \|C - D\|$  between  $C$  and  $D$ . It is possible that  $\Phi$  has no minimizers in which case  $u^*$  is not defined: consider, for instance, the case when  $C$  is the horizontal axis and  $D$  is the epigraph of the exponential function in the Euclidean plane. Returning to the general setting, note that iteration (55) becomes the method of alternating projections

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad y_n = P_D(x_n), \quad x_{n+1} = P_C(y_n). \tag{69}$$

Now observe that  $f \square g^\vee = \iota_C \square \iota_{-D} = \iota_{C-D}$ . Hence

$$(f \square g^\vee)^{**} = \iota_{C-D}^{**} = \iota_{\overline{C-D}} \tag{70}$$

is the indicator function of the closure of the Minkowski difference  $C - D$ . Using Proposition 4.3(i) with  $\gamma = 1$ , we determine the dual solution to be (see also [8, Section 2])

$$w^* = -\text{prox}_{(f \square g^\vee)^{**}}(0) = -\text{prox}_{\iota_{\overline{C-D}}}(0) = -P_{\overline{C-D}}(0) = P_{\overline{D-C}}(0). \tag{71}$$

In this setting, Theorem 4.6 recovers most of the results in [9, Section 4], which—when viewed in the product space setting utilized in Section 3.3.2—yield in turn some of the



results on parallel projection methods discussed in [8, Section 6; 9, Section 6; 18]. Alternating projections for two possibly nonintersecting sets were first considered by Cheney in Goldstein [17] in 1959 and then by Gubin et al. [26]; see also [12,33, Corollary 4.6] for related recent results.

As a second application, we propose a new derivation of von Neumann’s method of alternating projections [37]. Let  $C$  and  $D$  be two closed vector subspaces of  $\mathcal{H}$ , and let  $z \in \mathcal{H}$ . Consider the problem of finding  $P_{C \cap D}(z)$ , i.e., the best approximation to  $z$  from  $C \cap D$ . J. von Neumann proved that  $P_{C \cap D}(z)$  can be constructed by alternating projections. We now provide a novel proof of his result (see [37,21, Chapter 9] for further information, and [13, Section 3] for a recent elementary proof).

**Corollary 4.9** (von Neumann). *Suppose that  $C$  and  $D$  are closed vector subspaces of  $\mathcal{H}$  and that  $z \in \mathcal{H}$ . Then*

$$\lim_{n \rightarrow +\infty} (P_D P_C)^n(z) = P_{C \cap D}(z). \tag{72}$$

**Proof.** Denote the orthogonal complements of  $C$  and  $D$  by  $C^\perp$  and  $D^\perp$ , respectively. We define

$$f: \mathcal{H} \rightarrow ]-\infty, +\infty]: x \mapsto \iota_{C^\perp}(x+z) - \frac{1}{2}\|z\|^2 \quad \text{and} \quad g = \iota_{D^\perp}. \tag{73}$$

Then, for every  $z^* \in \mathcal{H}$ ,  $f^*(z^*) = \frac{1}{2}\|z\|^2 - \langle z | z^* \rangle + \iota_C(z^*)$  and  $g^*(z^*) = \iota_D(z^*)$ . Recalling Definition 4.2 with  $\gamma = 1$ , the (Fenchel dual) minimizer  $w^*$  of the function  $z^* \mapsto f^*(z^*) + g^*(-z^*) + \frac{1}{2}\|z^*\|^2 = \frac{1}{2}\|z\|^2 - \langle z | z^* \rangle + \iota_C(z^*) + \iota_D(z^*) + \frac{1}{2}\|z^*\|^2 = \frac{1}{2}\|z^* - z\|^2 + \iota_{C \cap D}(z^*)$  is precisely  $P_{C \cap D}(z)$ . Furthermore, for any  $w \in \mathcal{H}$ , we compute

$$\text{prox}_f(w) = w - P_C w - P_C z \quad \text{and} \quad \text{prox}_g(w) = w - P_D w. \tag{74}$$

Now consider the alternating proximity operator iteration

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad y_n = x_n - P_D x_n, \quad x_{n+1} = y_n - P_C y_n - P_C z. \tag{75}$$

Theorem 4.6(iii) with  $\gamma = 1$  implies that

$$\lim_{n \rightarrow +\infty} y_n - x_n = w^* = P_{C \cap D}(z). \tag{76}$$

Now let  $x_0 = 0$  and observe that (75) yields  $y_0 = 0, x_1 = -P_C z, y_1 = -P_C z + P_D P_C z, x_2 = -P_C z + P_D P_C z - P_C P_D P_C z, y_2 = -P_C z + P_D P_C z - P_C P_D P_C z + P_D P_C P_D P_C z$ , and so forth. Therefore,

$$(\forall n \in \mathbb{N} \setminus \{0\}) \quad y_n - x_n = (P_D P_C)^n(z), \tag{77}$$

and the result follows from (76).  $\square$

**Acknowledgements**

H.H. Bauschke’s work was partially supported by the Natural Sciences and Engineering Research Council of Canada and by a NATO collaborative linkage grant. P.L. Combettes’

work was partially supported by a NATO collaborative linkage grant. S. Reich's work was partially supported by the Israel Science Foundation founded by the Israel Academy of Sciences and Humanities (Grant 592/00), by the Fund for the Promotion of Research at the Technion (Grant 100-104), and by the Technion VPR Fund—Argentinian Research Fund (Grant 100-217).

## References

- [1] F. Acker, M.A. Prestel, Convergence d'un schéma de minimisation alternée, *Ann. Fac. Sci. Toulouse V. Sér. Math.* 2 (1980) 1–9.
- [2] H. Attouch, H. Brézis, Duality for the sum of convex functions in general Banach spaces, in: J.A. Barroso (Ed.), *Aspects of Mathematics and its Applications*, North-Holland Math. Library, vol. 34, North-Holland, Amsterdam, The Netherlands, 1986, pp. 125–133.
- [3] H. Attouch, M. Théra, A general duality principle for the sum of two operators, *J. Convex Anal.* 3 (1996) 1–24.
- [4] J.F. Aujol, G. Aubert, L. Blanc-Féraud, A. Chambolle, Image decomposition into a bounded variation component and an oscillating component, *J. Math. Imaging Vis.*, to appear.
- [5] A. Auslender, *Méthodes Numériques pour la Résolution des Problèmes d'Optimisation avec Contraintes*, Thèse, Université de Grenoble, France, 1969.
- [6] J.-B. Baillon, R.E. Bruck, S. Reich, On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces, *Houston J. Math.* 4 (1978) 1–9.
- [7] H.H. Bauschke, The composition of finitely many projections onto closed convex sets in Hilbert space is asymptotically regular, *Proc. Amer. Math. Soc.* 131 (2003) 141–146.
- [8] H.H. Bauschke, J.M. Borwein, On the convergence of von Neumann's alternating projection algorithm for two sets, *Set-Valued Anal.* 1 (1993) 185–212.
- [9] H.H. Bauschke, J.M. Borwein, Dykstra's alternating projection algorithm for two sets, *J. Approx. Theory* 79 (1994) 418–443.
- [10] H.H. Bauschke, J.M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Rev.* 38 (1996) 367–426.
- [11] H.H. Bauschke, J.M. Borwein, A.S. Lewis, The method of cyclic projections for closed convex sets in Hilbert space, in: Y. Censor, S. Reich (Eds.), *Recent Developments in Optimization Theory and Nonlinear Analysis*, *Contemp. Mathematics*, vol. 204, American Mathematical Society, Providence, RI, 1997, pp. 1–38.
- [12] H.H. Bauschke, P.L. Combettes, D.R. Luke, Finding best approximation pairs relative to two closed convex sets in Hilbert spaces, *J. Approx. Theory* 127 (2004) 178–192.
- [13] H.H. Bauschke, E. Matoušková, S. Reich, Projection and proximal point methods: convergence results and counterexamples, *Nonlinear Anal.* 56 (2004) 715–738.
- [14] L.M. Brègman, The method of successive projection for finding a common point of convex sets, *Soviet Math. Doklady* 6 (1965) 688–692.
- [15] R.E. Bruck, S. Reich, Nonexpansive projections and resolvents of accretive operators in Banach spaces, *Houston J. Math.* 3 (1977) 459–470.
- [16] Y. Censor, S.A. Zenios, *Parallel Optimization—Theory, Algorithms and Applications*, Oxford University Press, New York, 1997.
- [17] W. Cheney, A.A. Goldstein, Proximity maps for convex sets, *Proc. Amer. Math. Soc.* 10 (1959) 448–450.
- [18] P.L. Combettes, Inconsistent signal feasibility problems: least-squares solutions in a product space, *IEEE Trans. Signal Process.* 42 (1994) 2955–2966.
- [19] P.L. Combettes, The convex feasibility problem in image recovery, in: P. Hawkes (Ed.), *Advances in Imaging and Electron Physics*, vol. 95, Academic Press, New York, 1996, pp. 155–270.
- [20] P.L. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, *Optimization*, to appear.
- [21] F. Deutsch, *Best Approximation in Inner Product Spaces*, Springer, New York, 2001.
- [22] J.C. Dunn, Convexity, monotonicity, and gradient processes in Hilbert space, *J. Math. Anal. Appl.* 53 (1976) 145–158.

- [23] A. Genel, J. Lindenstrauss, An example concerning fixed points, *Israel J. Math.* 22 (1975) 81–86.
- [24] K. Goebel, S. Reich, *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings*, Marcel Dekker, New York and Basel, 1984.
- [25] A.A. Goldstein, Convex programming in Hilbert space, *Bull. Amer. Math. Soc.* 70 (1964) 709–710.
- [26] L.G. Gubin, B.T. Polyak, E.V. Raik, The method of projections for finding the common point of convex sets, *USSR Comput. Math. Math. Phys.* 7 (1967) 1–24.
- [27] O. Güler, On the convergence of the proximal point algorithm for convex minimization, *SIAM J. Control Optim.* 29 (1991) 403–419.
- [28] H.S. Hundal, An alternating projection that does not converge in norm, *Nonlinear Anal.* 57 (2004) 35–61.
- [29] E.S. Levitin, B.T. Polyak, Constrained minimization methods, *USSR Comput. Math. Math. Phys.* 6 (1966) 1–50.
- [30] J.L. Lions, *Quelques Méthodes de Résolution des Problèmes aux Limites Non Linéaires*, Dunod, Paris, 1969.
- [31] P. Mahey, D.T. Pham, Partial regularization of the sum of two maximal monotone operators, *RAIRO Modél. Math. Anal. Numér.* 27 (1993) 375–392.
- [32] B. Martinet, *Algorithmes pour la Résolution de Problèmes d’Optimisation et de Minimax*, Thèse, Université de Grenoble, France, 1972.
- [33] E. Matoušková, S. Reich, The Hundal example revisited, *J. Nonlinear Convex Anal.* 4 (2003) 411–427.
- [34] B. Mercier, *Inéquations Variationnelles de la Mécanique* (Publications Mathématiques d’Orsay, no. 80.01). Orsay, France, Université Paris 11, 1980.
- [35] G.J. Minty, Monotone (nonlinear) operators in Hilbert space, *Duke Math. J.* 29 (1962) 341–346.
- [36] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. France* 93 (1965) 273–299.
- [37] J. von Neumann, On rings of operators. Reduction theory, *Ann. of Math.* 50 (1949) 401–485.
- [38] Z. Opial, Weak convergence of the sequence of successive approximations for nonexpansive mappings, *Bull. Amer. Math. Soc.* 73 (1967) 591–597.
- [39] S. Reich, On the asymptotic behavior of nonlinear semigroups and the range of accretive operators, *J. Math. Anal. Appl.* 79 (1981) 113–126.
- [40] S. Reich, A limit theorem for projections, *Linear and Multilinear Algebra* 13 (1983) 281–290.
- [41] S. Reich, I. Shafir, The asymptotic behavior of firmly nonexpansive mappings, *Proc. Amer. Math. Soc.* 101 (1987) 246–250.
- [42] S. Reich, I. Shafir, Nonexpansive iterations in hyperbolic spaces, *Nonlinear Anal.* 15 (1990) 537–558.
- [43] R.T. Rockafellar, Extension of Fenchel’s duality theorem for convex functions, *Duke Math. J.* 33 (1966) 81–89.
- [44] L.A. Vese, S.J. Osher, Modeling textures with total variation minimization and oscillating patterns in image processing, *J. Sci. Comput.* 19 (2003) 553–572.
- [45] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific Publishing, River Edge, NJ, 2002.