

# Chapter 13

## A Note on the Equivalence of Operator Splitting Methods



Walaa M. Moursi and Yuriy Zinchenko

**Abstract** This note provides a comprehensive discussion of the equivalences between some splitting methods. We survey known results concerning these equivalences which have been studied over the past few decades. In particular, we provide simplified proofs of the equivalence of the ADMM and the Douglas–Rachford method and the equivalence of the ADMM with intermediate update of multipliers and the Peaceman–Rachford method.

**Keywords** Alternating Direction Method of Multipliers (ADMM) · Chambolle–Pock method · Douglas–Rachford algorithm · Dykstra method · Equivalence of splitting methods · Fenchel–Rockafellar duality · Peaceman–Rachford algorithm

**AMS 2010 Subject Classification** 47H05, 47H09, 49M27, 49M29, 49N15, 90C25

### 13.1 Introduction

Splitting methods have become popular in solving convex optimization problems that involve finding a minimizer of the sum of two proper lower semicontinuous convex functions. Among these methods are the Douglas–Rachford and

---

W. M. Moursi (✉)  
Electrical Engineering, Stanford University, Stanford, CA, USA

Faculty of Science, Mathematics Department, Mansoura University, Mansoura, Egypt  
e-mail: [wmoursi@stanford.edu](mailto:wmoursi@stanford.edu)

Y. Zinchenko  
University of Calgary, Department of Mathematics and Statistics, Calgary, AB, Canada  
e-mail: [yzinchenko@ucalgary.ca](mailto:yzinchenko@ucalgary.ca)

the Peaceman–Rachford methods introduced in the seminal work of Lions and Mercier [24], the forward-backward method (see, e.g., [12, 17] and [29]), Dykstra’s method (see, e.g., [3] and [10]), and the Method of Alternating Projections (MAP) (see, e.g., [19]).

When the optimization problem features the composition of one of the functions with a bounded linear operator, a popular technique is the Alternating-Direction Method of Multipliers (ADMM) (see [22, Section 4], [16, Section 10.6.4] and also [7, Chapter 15]). The method has a wide range of applications including large-scale optimization, machine learning, image processing and portfolio optimization, see, e.g., [9, 15] and [20]. A powerful framework to use ADMM in the more general setting of *monotone operators* is developed in the work of Briceño-Arias and Combettes [13] (see also [8] and [14]). Another relatively recent method is the Chambolle–Pock method introduced in [11].

Equivalences between splitting methods have been studied over the past four decades. For instance, it is known that ADMM is equivalent to the Douglas–Rachford method [24] (see, also [21]) in the sense that with a careful choice of the starting point, one can prove that the sequences generated by both algorithms coincide. (See, e.g., [22, Section 5.1] and [6, Remark 3.14].) A similar equivalence holds between ADMM (with intermediate update of multiplier) and Peaceman–Rachford method [24] (see [22, Section 5.2]). In [25], the authors proved the correspondence of Douglas–Rachford and Chambolle–Pock methods.

The rest of this paper is organized as follows: Section 13.2 provides a brief literature review of ADMM, Douglas–Rachford and Peaceman–Rachford methods. In Sections 13.3 and 13.4 we explicitly describe the equivalence of ADMM (respectively ADMM with intermediate update of multipliers) and Douglas–Rachford (respectively Peaceman–Rachford) method introduced by Gabay in [22, Sections 5.1&5.2]. We provide simplified proofs of these equivalences. Section 13.5 focuses on the recent work of O’Connor and Vandenberghe concerning the equivalence of Douglas–Rachford and Chambolle–Pock methods (see [25]). Our notation is standard and follows largely, e.g., [5].

## 13.2 Three Techniques

In this paper, we assume that

$X$  and  $Y$  are real Hilbert spaces,

and that

$f: X \rightarrow ]-\infty, +\infty]$ ,  $g: Y \rightarrow ]-\infty, +\infty]$  are convex lower semicontinuous and proper.

**Alternating-Direction Method of Multipliers (ADMM)** In the following we assume that<sup>1</sup>

$L: Y \rightarrow X$  is linear such that  $L^*L$  is invertible,

that

$$\operatorname{argmin}(f \circ L + g) \neq \emptyset, \quad (13.1)$$

and that

$$0 \in \operatorname{sri}(\operatorname{dom} f - L(\operatorname{dom} g)), \quad (13.2)$$

where  $\operatorname{sri} S$  denotes the *strong relative interior* of a subset  $S$  of  $X$  with respect to the closed affine hull of  $S$ . When  $X$  is finite-dimensional we have  $\operatorname{sri} S = \operatorname{ri} S$ , where  $\operatorname{ri} S$  is the *relative interior* of  $S$  defined as the interior of  $S$  with respect to the affine hull of  $S$ .

Consider the problem

$$\underset{y \in Y}{\operatorname{minimize}} \quad f(Ly) + g(y). \quad (13.3)$$

Note that (13.1) and (13.2) imply that (see, e.g., [5, Proposition 27.5(iii)(a)1])

$$\operatorname{argmin}(f \circ L + g) = \operatorname{zer}(\partial(f \circ L) + \partial g) = \operatorname{zer}(L^* \circ (\partial f) \circ L + \partial g) \neq \emptyset. \quad (13.4)$$

In view of (13.4), solving (13.3) is equivalent to solving the inclusion:

$$\text{Find } y \in Y \text{ such that } 0 \in L^*(\partial f(Ly)) + \partial g(y). \quad (13.5)$$

The augmented Lagrangian associated with (13.3) is the function

<sup>1</sup>The adjoint of  $L$  is the unique operator  $L^*: X \rightarrow Y$  that satisfies  $\langle Ly, x \rangle = \langle y, L^*x \rangle$  ( $\forall (x, y) \in X \times Y$ ).

$$\mathcal{L}: X \times Y \times X \rightarrow ]-\infty, +\infty]: (a, b, u) \mapsto f(a) + g(b) + \langle u, Lb - a \rangle + \frac{1}{2} \|Lb - a\|^2. \quad (13.6)$$

The ADMM (see [22, Section 4] and also [16, Section 10.6.4]) applied to solve (13.3) consists in minimizing  $\mathcal{L}$  over  $b$  then over  $a$  and then applying a proximal minimization step with respect to the Lagrange multiplier  $u$ . The method applied with a starting point  $(a_0, u_0) \in X \times X$  generates three sequences  $(a_n)_{n \in \mathbb{N}}$ ;  $(b_n)_{n \geq 1}$  and  $(u_n)_{n \in \mathbb{N}}$  via  $(\forall n \in \mathbb{N})$ :

$$b_{n+1} := (L^*L + \partial g)^{-1}(L^*a_n - L^*u_n), \quad (13.7a)$$

$$a_{n+1} := \text{Prox}_f(Lb_{n+1} + u_n), \quad (13.7b)$$

$$u_{n+1} := u_n + Lb_{n+1} - a_{n+1}, \quad (13.7c)$$

where  $\text{Prox}_f: X \rightarrow X: x \mapsto \text{argmin}_{y \in X} \left( f(y) + \frac{1}{2} \|x - y\|^2 \right)$ .

Let  $(x_n)_{n \in \mathbb{N}}$  be a sequence in  $X$  and let  $\bar{x} \in X$ . In the following we shall use  $x_n \rightarrow \bar{x}$  (respectively  $x_n \rightharpoonup \bar{x}$ ) to indicate that  $(x_n)_{n \in \mathbb{N}}$  converges strongly (respectively weakly) to  $\bar{x}$ .

**Fact 13.1 (Convergence of ADMM (See [22, Theorem 4.1]))** *Let  $(a_0, u_0) \in X \times X$ , and let  $(a_n)_{n \in \mathbb{N}}$ ,  $(b_n)_{n \geq 1}$  and  $(u_n)_{n \in \mathbb{N}}$  be defined as in (13.7). Then, there exists  $\bar{b} \in Y$  such that  $b_n \rightharpoonup \bar{b} \in \text{argmin}(f \circ L + g)$ .*

**The Douglas–Rachford Method** Suppose that  $Y = X$  and that  $L = \text{Id}$ . In this case Problem (13.3) becomes

$$\text{minimize}_{x \in X} f(x) + g(x). \quad (13.8)$$

The Douglas–Rachford (DR) method, introduced in [24], applied to the ordered pair  $(f, g)$  with a starting point  $x_0 \in X$  to solve (13.8) generates two sequences  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  via:

$$y_n := \text{Prox}_f x_n, \quad (13.9a)$$

$$x_{n+1} := T_{\text{DR}} x_n, \quad (13.9b)$$

where

$$T_{\text{DR}} := T_{\text{DR}}(f, g) = \frac{1}{2}(\text{Id} + R_g R_f) = \text{Id} - \text{Prox}_f + \text{Prox}_g(2\text{Prox}_f - \text{Id}), \quad (13.10)$$

and where  $R_f := 2\text{Prox}_f - \text{Id}$ .

To lighten the notation, in the sequel we shall use  $T_{\text{DR}}$  to denote  $T_{\text{DR}}(f, g)$ . Let  $T: X \rightarrow X$ . Recall that the set of fixed points of  $T$ , denoted by  $\text{Fix } T$ , is defined as  $\text{Fix } T := \{x \in X \mid x = Tx\}$ .

**Fact 13.2 (Convergence of Douglas–Rachford Method (See, e.g., [24, Theorem 1] or [5, Corollary 28.3]))** *Let  $x_0 \in X$  and let  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  be defined as in (13.9). Then, there exists  $\bar{x} \in \text{Fix } T_{DR}$  such that  $x_n \rightarrow \bar{x}$  and  $y_n \rightarrow \text{Prox}_f \bar{x} \in \text{argmin}(f + g)$ .*

**The Peaceman–Rachford Method** Let  $h: X \rightarrow ]-\infty, +\infty]$  be proper and let  $\beta > 0$ . We say that  $h$  is *strongly convex* if  $f - \frac{\beta}{2} \|\cdot\|^2$  is convex, i.e.,  $(\forall(x, y) \in \text{dom } f \times \text{dom } f) (\forall\alpha \in ]0, 1[)$  we have  $f(\alpha x + (1 - \alpha)y) + \alpha(1 - \alpha)\phi(\|x - y\|) + \frac{\beta}{2} \|x - y\|^2 \leq \alpha f(x) + (1 - \alpha)f(y)$ .

When  $g$  is strongly convex, the Peaceman–Rachford (PR) method, introduced in [24], can be used to solve (13.8). In this case, given  $x_0 \in X$ , PR method generates the sequences  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  via:

$$y_n := \text{Prox}_f x_n, \tag{13.11a}$$

$$x_{n+1} := T_{PR} x_n, \tag{13.11b}$$

where

$$T_{PR} = T_{PR}(f, g) = R_g R_f = (2 \text{Prox}_g - \text{Id})(2 \text{Prox}_f - \text{Id}). \tag{13.12}$$

To lighten the notation, in the sequel we shall use  $T_{PR}$  to denote  $T_{PR}(f, g)$ .

**Fact 13.3 (Convergence of Peaceman–Rachford Method (See, e.g., [24, Proposition 1] or [5, Proposition 28.8]))** *Suppose that  $g$  is strongly convex. Let  $\bar{y}$  be the unique minimizer of  $f + g$ , let  $x_0 \in X$  and let  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  be defined as in (13.11). Then  $y_n \rightarrow \bar{y}$ .*

In the sequel we use the notation

$$g^\vee: X \rightarrow ]-\infty, +\infty]: x \mapsto g(-x). \tag{13.13}$$

Recall that the Fenchel–Rockafellar dual of (13.3) is

$$\underset{x \in X}{\text{minimize}} \quad f^*(x) + g^*(-L^*x). \tag{13.14}$$

*Remark 13.1*

- (i) One can readily verify that  $\partial g^\vee = (-\text{Id}) \circ \partial g \circ (-\text{Id})$ . Therefore, in view of [27, Theorem A] and [20, Lemma 3.5 on page 125 and Lemma 3.6 on page 133] (see also [4, Corollaries 4.2 and 4.3]) we have<sup>2</sup>

$$T_{DR}(f, g) = T_{DR}(f^*, g^{\vee*}), \tag{13.15}$$

---

<sup>2</sup>It is straightforward to verify that  $g^{\vee*} = (g^*)^\vee$  (see, e.g., [5, Proposition 13.23(v)]).

and

$$T_{\text{PR}}(f, g) = T_{\text{PR}}(f^*, g^{*\vee}). \quad (13.16)$$

- (ii) When  $(L, Y) = (\text{Id}, X)$ , inclusion (13.5) reduces to: Find  $y \in X$  such that  $0 \in \partial f(y) + \partial g(y)$  and the *dual* inclusion (corresponding to the Fenchel–Rockafellar dual (13.14)) is: Find  $y \in X$  such that  $0 \in \partial f^*(y) - \partial g^*(-y) = (\partial f)^{-1}y - (\partial g)^{-1}(-y)$ , which in this case coincide with the Attouch–Thera dual of (13.5) (see [2]).

One can use DR method to solve (13.14) where  $(f, g)$  in Fact 13.2 is replaced by  $(f^*, g^* \circ (-L^*))$ . Recalling (13.15) we learn that  $T_{\text{DR}} = T_{\text{DR}}(f^*, g^* \circ (-L^*)) = T_{\text{DR}}(f^{**}, (g^* \circ (-L^*))^{\vee*}) = T_{\text{DR}}(f, (g^* \circ L^*)^*)$ , where the last identity follows from [5, Proposition 13.44].

In view of (13.10) (13.15), and [5, Proposition 15.23(v)] we have

$$\begin{aligned} T_{\text{DR}} &= T_{\text{DR}}(f^*, g^* \circ (-L^*)) = T_{\text{DR}}(f, (g^* \circ L^*)^*) \\ &= \text{Id} - \text{Prox}_f + \text{Prox}_{(g^* \circ L^*)^*} (2 \text{Prox}_f - \text{Id}). \end{aligned} \quad (13.17)$$

Similarly, under additional assumptions (see Fact 13.3), one can use PR method to solve (13.14) where  $(f, g)$  in Theorem 13.3 is replaced by  $(f^*, g^* \circ (-L^*))$ . In this case (13.12), (13.16) and [5, Proposition 15.23(v)] imply that

$$\begin{aligned} T_{\text{PR}} &= T_{\text{PR}}(f^*, g^* \circ (-L^*)) = T_{\text{PR}}(f, (g^* \circ L^*)^*) \\ &= (2 \text{Prox}_{(g^* \circ L^*)^*} - \text{Id})(2 \text{Prox}_f - \text{Id}). \end{aligned} \quad (13.18)$$

For completeness, we provide a concrete proof of the formula for  $\text{Prox}_{(g^* \circ L^*)^*}$  in Appendix 1 (see Proposition 13.3(viii) below). We point out that the formula for  $\text{Prox}_{(g^* \circ L^*)^*}$  in a more general setting is given in [22, Proposition 4.1] (see also [16, Section 10.6.4]).

### 13.3 ADMM and Douglas–Rachford Method

In this section we discuss the equivalence of ADMM and DR method. This equivalence was first established by Gabay in [22, Section 5.1] (see also [6, Remark 3.14]). Let  $(x_0, a_0, u_0) \in X^3$ . Throughout the rest of this section, we assume that the sequences  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  are as defined in (13.9), i.e.,

$$(x_{n+1}, y_n)_{n \in \mathbb{N}} = (T_{\text{DR}}x_n, \text{Prox}_f x_n), \quad (13.19)$$

$$T_{\text{DR}} := T_{\text{DR}}(f, (g^* \circ L^*)^*) = \text{Id} - \text{Prox}_f + L(L^*L + \partial g)^{-1}L^*(2\text{Prox}_f - \text{Id}). \quad (13.20)$$

Note that the second identity in (13.20) follows from (13.17) and Proposition 13.3(viii). We also assume that

$$(a_n, u_n, b_{n+1})_{n \in \mathbb{N}} \text{ is defined as in (13.7).}$$

The following lemma will be used later to clarify the equivalence of DR and ADMM.

**Lemma 13.1** *Let  $(b_-, a_-, u_-) \in Y \times X \times X$  and set*

$$(b, a, u) := ((L^*L + \partial g)^{-1}(L^*a_- - L^*u_-), \text{Prox}_f(Lb + u_-), u_- + Lb - a), \quad (13.21a)$$

$$(b_+, a_+, u_+) := ((L^*L + \partial g)^{-1}(L^*a - L^*u), \text{Prox}_f(Lb_+ + u), u + Lb_+ - a_+). \quad (13.21b)$$

Then

$$T_{\text{DR}}(Lb + u_-) = Lb_+ + u, \quad (13.22a)$$

$$\text{Prox}_f T_{\text{DR}}(Lb + u_-) = a_+. \quad (13.22b)$$

*Proof* Indeed, it follows from (13.20), (13.21), (13.22a) and (13.22b) that

$$\begin{aligned} T_{\text{DR}}(Lb + u_-) &= (Lb + u_-) - \text{Prox}_f(Lb + u_-) \\ &\quad + L(L^*L + \partial g)^{-1}L^*(2\text{Prox}_f(Lb + u_-) - (Lb + u_-)) \end{aligned} \quad (13.23a)$$

$$= (Lb + u_-) - a + L(L^*L + \partial g)^{-1}L^*(2a - (Lb + u_-)) \quad (13.23b)$$

$$= (Lb + u_-) - a + L(L^*L + \partial g)^{-1}L^*(a - (Lb + u_- - a)) \quad (13.23c)$$

$$= u + L(L^*L + \partial g)^{-1}L^*(L^*a - L^*u) \quad (13.23d)$$

$$= Lb_+ + u, \quad (13.23e)$$

where (13.23a) follows from (13.20), (13.23b) and (13.23d) follow from (13.21a), and (13.23e) follow from (13.21b). This proves (13.22a). Now (13.22b) follows from combining (13.22a) and (13.21b).  $\square$

We now prove the main result in this section by induction.

**Theorem 13.4** *The following hold:*

- (i) **(DR as ADMM Iteration)** Using DR method with a starting point  $x_0 \in X$  to solve (13.14) is equivalent to using ADMM with a starting point  $(a_0, u_0) := (\text{Prox}_f x_0, x_0 - \text{Prox}_f x_0)$  to solve (13.3), in the sense that  $(x_n)_{n \geq 1} = (Lb_n + u_{n-1})_{n \geq 1}$  and  $(y_n)_{n \in \mathbb{N}} = (a_n)_{n \in \mathbb{N}}$ .
- (ii) **(ADMM as DR Iteration)** Using ADMM with a starting point  $(a_0, u_0) \in X \times X$  to solve (13.3) is equivalent to using DR method with a starting point  $x_0 = Lb_1 + u_0$  to solve (13.14), in the sense that  $(x_n)_{n \in \mathbb{N}} = (Lb_{n+1} + u_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}} = (a_{n+1})_{n \in \mathbb{N}}$ .

*Proof* For simplicity, set  $T = T_{\text{DR}}$ . (i): Note that (13.19) implies that  $y_0 = a_0$ . Now, when  $n = 1$  we have

$$\begin{aligned} x_1 &= Tx_0 = x_0 - \text{Prox}_f x_0 + L(L^*L + \partial g)^{-1}L^*(2\text{Prox}_f x_0 - x_0) && \text{(by (13.17))} \\ &= x_0 - a_0 + L(L^*L + \partial g)^{-1}L^*(2a_0 - x_0) && \text{(by (13.19))} \\ &= (x_0 - a_0) + L(L^*L + \partial g)^{-1}L^*(a_0 - (x_0 - a_0)) \\ &= u_0 + L(L^*L + \partial g)^{-1}L^*(a_0 - u_0) \\ &= u_0 + Lb_1. && \text{(by (13.7a))} \end{aligned}$$

Combining with (13.19) and (13.7b) we get  $y_1 = \text{Prox}_f Tx_0 = \text{Prox}_f x_1 = \text{Prox}_f(u_0 + Lb_1) = a_1$ , which verifies the base case. Now suppose for some  $n \geq 1$  we have  $x_n = Lb_n + u_{n-1}$  and  $y_n = a_n$  and use Lemma 13.1 with  $(b_-, a_-, u_-)$  replaced by  $(b_{n-1}, a_{n-1}, u_{n-1})$  to learn that  $x_{n+1} = Lb_{n+1} + u_n$  and  $y_{n+1} = a_{n+1}$ . Consequently,  $(x_n)_{n \geq 1} = (Lb_n + u_{n-1})_{n \geq 1}$  and  $(y_n)_{n \in \mathbb{N}} = (a_n)_{n \in \mathbb{N}}$  as claimed.

(ii): At  $n = 0$ ,  $x_0 = Lb_1 + u_0 = Lb_{0+1} + u_0$ , and therefore (13.9a) implies that  $y_0 = \text{Prox}_f x_0 = \text{Prox}_f(Lb_1 + u_0) = a_1$  by (13.7b). Now suppose that for some  $n \geq 0$  we have  $x_n = Lb_{n+1} + u_n$  and  $y_n = a_{n+1}$ . The conclusion follows by applying Lemma 13.1 with  $(b_-, a_-, u_-)$  replaced by  $(b_n, a_n, u_n)$ .  $\square$

## 13.4 ADMM and Peaceman–Rachford Method

We now turn to the equivalence of ADMM with intermediate update of multiplier and PR method. This equivalence was established in [22, Section 5.2]. Given  $(a_0, u_0) \in X \times X$ , the ADMM with an *intermediate* update of multiplier applied to solve (13.3) generates four sequences  $(a_n)_{n \in \mathbb{N}}$ ,  $(u_n)_{n \in \mathbb{N}}$ ,  $(b_n)_{n \geq 1}$  and  $(w_n)_{n \geq 1}$  via  $(\forall n \in \mathbb{N})$ :

$$b_{n+1} := (L^*L + \partial g)^{-1}(L^*a_n - L^*u_n), \quad (13.24a)$$

$$w_{n+1} := u_n + Lb_{n+1} - a_n, \quad (13.24b)$$



$$a_{n+1} := \text{Prox}_f(Lb_{n+1} + w_{n+1}), \quad (13.24c)$$

$$u_{n+1} := w_{n+1} + Lb_{n+1} - a_{n+1}. \quad (13.24d)$$

**Fact 13.5 (Convergence of ADMM with Intermediate Update of Multipliers (See [22, Theorem 5.4]))** *Suppose that  $g$  is strongly convex. Let  $(a_0, u_0) \in X \times X$ , and let  $(b_n)_{n \geq 1}$ ,  $(w_n)_{n \geq 1}$ ,  $(a_n)_{n \in \mathbb{N}}$  and  $(u_n)_{n \in \mathbb{N}}$  be defined as in (13.24). Then, there exists  $\bar{b} \in Y$  such that  $b_n \rightarrow \bar{b} \in \text{argmin}(f \circ L + g)$ .*

In this section we work under the additional assumption that

$g$  is strongly convex.

Let  $(x_0, a_0, u_0) \in X^3$ . Throughout the rest of this section we assume that the sequences  $(x_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}}$  are as defined in (13.11), i.e.,

$$(x_{n+1}, y_n)_{n \in \mathbb{N}} = (T_{\text{PR}}x_n, \text{Prox}_f x_n)_{n \in \mathbb{N}} \quad (13.25)$$

where

$$T_{\text{PR}} := T_{\text{PR}}(f, (g^* \circ L^*)^*) = 2L(L^*L + \partial g)^{-1}L^*(2\text{Prox}_f - \text{Id}) - 2\text{Prox}_f + \text{Id}. \quad (13.26)$$

Note that the second identity in (13.26) follows from (13.18) and Proposition 13.3(viii). We also assume that

$(a_n, u_n, b_{n+1}, w_{n+1})_{n \in \mathbb{N}}$  is defined as in (13.24).

Before we proceed further, we prove the following useful lemma.

**Lemma 13.2** *Let  $(b_-, w_-, a_-, u_-) \in Y \times X \times X \times X$  and set*

$$(b, w, a, u) = ((L^*L + \partial g)^{-1}(L^*a_- - L^*u_-), u_- + Lb - a_-, \text{Prox}_f(Lb + w), w + Lb - a), \quad (13.27a)$$

$$(b_+, w_+, a_+, u_+) = ((L^*L + \partial g)^{-1}(L^*a - L^*u), u + Lb_+ - a, \text{Prox}_f(Lb_+ + w_+), w_+ + Lb_+ - a_+). \quad (13.27b)$$

Then

$$T_{PR}(Lb + w) = Lb_+ + w_+, \quad (13.28a)$$

$$\text{Prox}_f T_{PR}(Lb + w) = a_+. \quad (13.28b)$$

*Proof* Indeed, by (13.26), (13.27a) and (13.27b) we have

$$\begin{aligned} T_{PR}(Lb + w) &= Lb + w - 2\text{Prox}_f(Lb + w) \\ &\quad + 2L(L^*L + \partial g)^{-1}L^*(2\text{Prox}_f(Lb + w) - (Lb + w)) \\ &\hspace{15em} \text{(by (13.26))} \\ &= Lb + w - a - a + 2L(L^*L + \partial g)^{-1}L^*(a - (Lb + w - a)) \\ &\hspace{15em} \text{(by (13.27a))} \\ &= u - a + 2L(L^*L + \partial g)^{-1}L^*(a - u) = u - a + 2Lb_+ \\ &\hspace{15em} \text{(by (13.27b))} \\ &= Lb_+ + w_+, \hspace{15em} \text{(by (13.27b))} \end{aligned}$$

which proves (13.28a). Now (13.28b) is a direct consequence of (13.28a) in view of (13.27b).  $\square$

We are now ready for the main result in this section.

**Theorem 13.6** *Suppose that  $g$  is strongly convex. Then the following hold:*

- (i) **(PR as ADMM Iteration)** *Using PR method with a starting point  $x_0 \in X$  to solve (13.14) is equivalent to using ADMM with intermediate update of multipliers with starting points  $(a_0, u_0) := (\text{Prox}_f x_0, x_0 - \text{Prox}_f x_0)$  to solve (13.3), in the sense that  $(x_n)_{n \geq 1} = (Lb_n + w_n)_{n \geq 1}$  and  $(y_n)_{n \in \mathbb{N}} = (a_n)_{n \in \mathbb{N}}$ .*
- (ii) **(ADMM as PR Iteration)** *Using ADMM with intermediate update of multipliers with a starting point  $(a_0, u_0) \in X \times X$  to solve (13.3) is equivalent to using PR method with starting point  $x_0 = Lb_1 + w_1$  to solve (13.14), in the sense that  $(x_n)_{n \in \mathbb{N}} = (Lb_{n+1} + w_{n+1})_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}} = (a_{n+1})_{n \in \mathbb{N}}$ .*

*Proof* We proceed by induction. (i): We have

$$\begin{aligned} x_1 &= T_{PR}x_0 = x_0 - 2\text{Prox}_f x_0 + L(L^*L + \partial g)^{-1}L^*(2\text{Prox}_f x_0 - x_0) \\ &\hspace{15em} \text{(by (13.26))} \\ &= x_0 - 2a_0 + 2L(L^*L + \partial g)^{-1}L^*(2a_0 - x_0) \\ &= (x_0 - a_0) - a_0 + 2L(L^*L + \partial g)^{-1}L^*(a_0 - (x_0 - a_0)) \\ &= u_0 - a_0 + 2Lb_1 = u_0 - a_0 + Lb_1 + Lb_1 \\ &= Lb_1 + w_1, \hspace{15em} \text{(by (13.24b))} \end{aligned}$$

which verifies the base case. Now suppose for some  $n \geq 1$  we have  $x_n = Lb_n + w_n$ . The conclusion follows from applying Lemma 13.2 with  $(b_-, w_-, a_-, u_-)$  replaced by  $(b_{n-1}, w_{n-1}, a_{n-1}, u_{n-1})$  in view of (13.24).

(ii): At  $n = 0$ , the base case clearly holds in view of (13.25) and (13.24c). Now suppose that for some  $n \geq 0$  we have  $x_n = Lb_{n+1} + w_{n+1}$  and  $y_n = a_{n+1}$  and use Lemma 13.2 with  $(b_-, w_-, a_-, u_-)$  replaced by  $(b_n, w_n, a_n, u_n)$  in view of (13.24).  $\square$

## 13.5 Chambolle–Pock and Douglas–Rachford Methods

In this section we survey the recent work by O’Connor and Vandenberghe [25] concerning the equivalence of Douglas–Rachford method and Chambolle–Pock method. (For a detailed study of this correspondence in the more general framework of the primal–dual hybrid gradient method and DR method with relaxation, as well as connection to linearized ADMM, we refer the reader to [25].) We work under the assumption that<sup>3</sup>

$$A: X \rightarrow Y \text{ is linear, } \sigma > 0, \tau > 0, \text{ and } \sigma\tau\|A\|^2 < 1. \quad (13.31)$$

Consider the problem

$$\underset{x \in X}{\text{minimize}} \quad f(x) + g(Ax) \quad (13.32)$$

and its Fenchel–Rockafellar dual given by

$$\underset{x \in X}{\text{minimize}} \quad f^*(-Ax) + g^*(x). \quad (13.33)$$

To proceed further, in the following we assume that

$$\text{argmin}(f + g \circ A) \neq \emptyset \text{ and } 0 \in \text{sri}(\text{dom } g - A(\text{dom } f)). \quad (13.34)$$

Note that (13.34) implies that (see, e.g., [5, Proposition 27.5(iii)(a)1])

$$\text{argmin}(f + g \circ A) = \text{zer}(\partial f + \partial(g \circ A)) = \text{zer}(\partial f + A^* \circ (\partial g) \circ A) \neq \emptyset. \quad (13.35)$$

<sup>3</sup>In passing, we point out that, when  $X$  is a finite-dimensional Hilbert space, the condition  $\tau\sigma\|A\|^2 < 1$  can be relaxed to  $\tau\sigma\|A\|^2 \leq 1$ . The convergence in this case is proved in [18, Theorem 3.3].

In view of (13.35), solving (13.32) is equivalent to solving the inclusion:

$$\text{Find } x \in X \text{ such that } 0 \in \partial f(x) + A^*(\partial g(Ax)). \quad (13.36)$$

The Chambolle–Pock (CP) method applied with a starting point  $(u_0, v_0) \in X \times Y$  to solve (13.32) generates the sequences  $(u_n)_{n \in \mathbb{N}}$ , and  $(v_n)_{n \in \mathbb{N}}$  via:

$$u_n = \text{Prox}_{\tau f}(u_{n-1} - \tau A^* v_{n-1}), \quad (13.37a)$$

$$v_n = \text{Prox}_{\sigma g^*}(v_{n-1} + \sigma A(2u_n - u_{n-1})), \quad (13.37b)$$

where  $\tau$  and  $\sigma$  are as defined in (13.31).

**Fact 13.7 (Convergence of Chambolle–Pock Method (See [11, Theorem 1], [30, Theorem 3.1] and Also [18, Theorem 3.1]))** *Let  $(u_0, v_0) \in X \times Y$  and let  $(u_n)_{n \in \mathbb{N}}$  and  $(v_n)_{n \in \mathbb{N}}$  be defined as in (13.37). Then, there exists  $(\bar{u}, \bar{v}) \in X \times Y$  such that  $(u_n, v_n)_{n \in \mathbb{N}} \rightarrow (\bar{u}, \bar{v})$ ,  $\bar{u} \in \text{argmin}(f + g \circ A)$  and  $\bar{v} \in \text{argmin}(f^* \circ (-A^*) + g^*)$ .*

It is known that the method in (13.37) reduces to DR method (see, e.g., [11, Section 4.2]) when  $A = \text{Id}$ . We state this equivalence in Proposition 13.1 below.

**Proposition 13.1 (DR as a CP Iteration)** *Suppose that  $X = Y$ , and that  $A = \text{Id}$ . Then, using DR method, defined as in (13.9), with a starting point  $x_0 \in X$  to solve (13.32) is equivalent to using CP method with a starting point  $(u_0, v_0) \in \{(u, v) \mid u - v = x_0\} \subseteq X \times X$  to solve (13.32) in the sense that  $(x_n)_{n \in \mathbb{N}} = (u_n - v_n)_{n \in \mathbb{N}}$  and  $(y_n)_{n \in \mathbb{N}} = (u_n)_{n \in \mathbb{N}}$ .*

*Proof* We use induction. When  $n = 0$ , the base case is obviously true. Now suppose that for some  $n \geq 0$  we have  $x_n = u_n - v_n$  and  $y_n = u_n$ . Then,

$$x_{n+1} = \text{Prox}_f x_n - \text{Prox}_{g^*}(2 \text{Prox}_f x_n - x_n) \quad (13.38a)$$

$$= \text{Prox}_f(u_n - v_n) - \text{Prox}_{g^*}(2 \text{Prox}_f(u_n - v_n) - (u_n - v_n)) \quad (13.38b)$$

$$= u_{n+1} - \text{Prox}_{g^*}(v_n + 2u_{n+1} - u_n) = u_{n+1} - v_{n+1}. \quad (13.38c)$$

Here (13.38a) follows from Lemma 13.5 below (applied with  $\gamma = 1$ ), (13.38b) follows from the inductive hypothesis, and (13.38c) follows from (13.37) applied with  $(\tau, \sigma, A)$  replaced by  $(1, 1, \text{Id})$ . The claim about  $y_{n+1}$  follows directly and the proof is complete.  $\square$

**Chambolle–Pock as a DR Iteration: The O’Connor–Vandenberghe Technique**

Let  $Z$  be a real Hilbert space. In the following, we assume that  $C : Z \rightarrow Y$  is linear and that

$$B : X \times Z \rightarrow Y : (x, z) \mapsto Ax + Cz \text{ satisfies that } \sigma \tau B B^* = \text{Id}. \quad (13.39)$$

Note that one possible choice of  $C$  is to set  $C^2 := \text{Id} - \sigma \tau A A^*$ , where the existence of  $C$  follows from, e.g., [26, Theorem on page 265]. Now consider the problem

$$\underset{(x,z) \in X \times Z}{\text{minimize}} \quad \tilde{f}(x, z) + g(B(x, z)), \quad (13.40)$$

where

$$\tilde{f}: X \times Z \rightarrow ]-\infty, +\infty] : (x, z) \mapsto f(x) + \iota_{\{0\}}(z). \quad (13.41)$$

The following result, proved in [25, Section 4] in the more general framework of primal-dual hybrid gradient method, provides an elegant way to construct the correspondence between the DR sequence when applied to solve (13.40) and the CP sequence when applied to solve (13.32). We restate the proof for the sake of completeness.

**Proposition 13.2 (CP Corresponds to a DR Iteration)** *Using CP method with starting point  $(u_0, v_0) \in X \times Z$  to solve (13.32) corresponds to using DR with starting point  $\mathbf{x}_0 := (u_0, 0) - \tau B^* v_0 \in X \times Z$  to solve (13.40), in the sense that  $(\mathbf{x}_n)_{n \in \mathbb{N}} = ((u_n, 0) - \tau B^* v_n)_{n \in \mathbb{N}}$  and  $(\mathbf{y}_n)_{n \in \mathbb{N}} = (u_{n+1}, 0)_{n \in \mathbb{N}}$ .*

*Proof* We apply DR to solve (13.40) with  $(\tilde{f}, g)$  replaced by  $(\tau \tilde{f}, \tau g)$ . The proof proceeds by induction. When  $n = 0$ , by assumption we have  $\mathbf{x}_0 = (u_0, 0) - \tau B^* v_0$ . It follows from Proposition 13.4(i)&(vii) below applied with  $\tilde{f}$  replaced by  $\tau \tilde{f}$  that  $\mathbf{y}_0 = \text{Prox}_{\tau \tilde{f}} \mathbf{x}_0 = \text{Prox}_{\tau \tilde{f}}((u_0, 0) - \tau(A^* v_0, C^* v_0)) = \text{Prox}_{\tau \tilde{f}}(u_0 - \tau A^* v_0, -\tau C^* v_0) = (\text{Prox}_{\tau f}(u_0 - \tau A^* v_0), 0)$ . Now suppose that for some  $n \geq 0$  we have

$$\mathbf{x}_n = (u_n, 0) - \tau B^* v_n, \quad (13.42a)$$

$$\mathbf{y}_n = (u_{n+1}, 0). \quad (13.42b)$$

Then

$$(u_{n+1}, 0) - \tau B^* v_{n+1} = (u_{n+1}, 0) - \tau B^* (\text{Prox}_{\sigma g^*}(v_n + \sigma A(2u_{n+1} - u_n))) \quad (13.43a)$$

$$= \mathbf{y}_n - \tau B^* (\text{Prox}_{\sigma g^*}(v_n + \sigma B(2(u_{n+1}, 0) - (u_n, 0)))) \quad (13.43b)$$

$$= \mathbf{y}_n - \tau B^* (\text{Prox}_{\sigma g^*}(\sigma \tau B B^* v_n + \sigma B(2(u_{n+1}, 0) - (u_n, 0)))) \quad (13.43c)$$

$$= \mathbf{y}_n - \tau B^* \text{Prox}_{\sigma g^*}(\sigma B(2(u_{n+1}, 0) - ((u_n, 0) - \tau B^* v_n))) \quad (13.43d)$$

$$= \mathbf{y}_n - \text{Prox}_{(\tau g \circ B)^*}(2\mathbf{y}_n - \mathbf{x}_n) = \mathbf{x}_{n+1}, \quad (13.43e)$$

where (13.43a) follows from (13.37b), (13.43b) follows from (13.42b) and Proposition 13.4(iii) below, (13.43c) follows from (13.39), and (13.43e) follows from (13.42a), Proposition 13.4(viii) and (13.49b) below applied with  $(\gamma, g)$  replaced by  $(\tau, g \circ B)$ .

Now by (13.37a) and Proposition 13.4(vii) below we have

$$(u_{n+2}, 0) = (\text{Prox}_{\tau f}(u_{n+1} - \tau A^* v_{n+1}), 0) \quad (13.44a)$$

$$= \text{Prox}_{\tau \tilde{f}}(u_{n+1} - \tau A^* v_{n+1}, -\tau C^* v_{n+1}) \quad (13.44b)$$

$$= \text{Prox}_{\tau \tilde{f}}((u_{n+1}, 0) - \tau(A^* v_{n+1}, C^* v_{n+1})) \quad (13.44c)$$

$$= \text{Prox}_{\tau \tilde{f}}((u_{n+1}, 0) - \tau B^* v_{n+1}) \quad (13.44d)$$

$$= \text{Prox}_{\tau \tilde{f}} \mathbf{x}_{n+1} = \mathbf{y}_{n+1}. \quad (13.44e)$$

□

**Acknowledgements** WMM was supported by the Pacific Institute of Mathematics Postdoctoral Fellowship and the DIMACS/Simons Collaboration on Bridging Continuous and Discrete Optimization through NSF grant # CCF-1740425.

## Appendices

### Appendix 1

Let  $A: X \rightarrow X$  be linear. Define

$$q_A: X \rightarrow \mathbb{R}: x \mapsto \frac{1}{2} \langle x, Ax \rangle. \quad (13.45)$$

Recall that a linear operator  $A: X \rightarrow X$  is *monotone* if  $(\forall x \in X) \langle x, Ax \rangle \geq 0$ , and is *strictly monotone* if  $(\forall x \in X \setminus \{0\}) \langle x, Ax \rangle > 0$ . Let  $h: X \rightarrow \mathbb{R}$  and let  $x \in X$ . We say that  $h$  is *Fréchet differentiable at  $x$*  if there exists a linear operator  $Dh(x): X \rightarrow \mathbb{R}$ , called the *Fréchet derivative* of  $h$  at  $x$ , such that  $\lim_{0 \neq \|y\| \rightarrow 0} \frac{h(x+y) - h(x) - Dh(x)y}{\|y\|} = 0$ ; and  $h$  is *Fréchet differentiable on  $X$*  if it is Fréchet differentiable at every point in  $X$ .

The following lemma is a special case of [5, Proposition 17.36].

**Lemma 13.3** *Let  $A: X \rightarrow X$  be linear, strictly monotone, self-adjoint and invertible. Then the following hold:*

- (i)  $q_A$  and  $q_{A^{-1}}$  are strictly convex, continuous, Fréchet differentiable. Moreover,  $(\nabla q_A, \nabla q_{A^{-1}}) = (A, A^{-1})$ .
- (ii)  $q_A^* = q_{A^{-1}}$ .

*Proof* Note that, likewise  $A$ ,  $A^{-1}$  is linear, strictly monotone, self-adjoint (since  $(A^{-1})^* = (A^*)^{-1} = A^{-1}$ ) and invertible. Moreover,  $\text{ran } A = \text{ran } A^{-1} = X$ . (i): This follows from [5, Example 17.11 and Proposition 17.36(i)] applied to  $A$  and  $A^{-1}$  respectively. (ii): It follows from [5, Proposition 17.36(iii)], [28, Theorem 4.8.5.4] and the invertibility of  $A$  that  $q_A^* = q_{A^{-1}} + \iota_{\text{ran } A} = q_{A^{-1}} + \iota_X = q_{A^{-1}}$ .  $\square$

**Proposition 13.3** *Let  $L: Y \rightarrow X$  be linear. Suppose that  $L^*L$  is invertible.  $g: Y \rightarrow ]-\infty, +\infty]$  be convex, lower semicontinuous, and proper. Then the following hold:*

- (i)  $\ker L = \{0\}$ .
- (ii)  $L^*L$  is strictly monotone.
- (iii)  $\text{dom}(q_{L^*L} + g)^* = X$ .
- (iv)  $\partial(q_{L^*L} + g) = \nabla q_{L^*L} + \partial g = L^*L + \partial g$ .
- (v)  $(q_{L^*L} + g)^*$  is Fréchet differentiable on  $X$ .
- (vi)  $(L^*L + \partial g)^{-1}$  is single-valued and  $\text{dom}(L^*L + \partial g)^{-1} = X$ .
- (vii)  $\text{Prox}_{g^* \circ L^*} = \text{Id} - L(L^*L + \partial g)^{-1}L^*$ .
- (viii)  $\text{Prox}_{(g^* \circ L^*)^*} = L(L^*L + \partial g)^{-1}L^*$ .

*Proof* (i): Using [5, Fact 2.25(vi)] and the assumption that  $L^*L$  is invertible we have  $\ker L = \ker L^*L = \{0\}$ . (ii): Using (i) we have  $(\forall x \in X \setminus \{0\}) \langle L^*Lx, x \rangle = \langle Lx, Lx \rangle = \|Lx\|^2 > 0$ , hence  $L^*L$  is strictly monotone. (iii): By (ii) and Lemma 13.3(i) applied with  $A$  replaced by  $L^*L$  we have  $\text{dom } q_{L^*L} = \text{dom } q_{L^*L}^* = X$ , hence

$$\text{dom } q_{L^*L} - \text{dom } g = X - \text{dom } g = X. \quad (13.46)$$

It follows from (13.46), [1, Corollary 2.1] and Lemma 13.3(ii)&(i) that  $\text{dom}(q_{L^*L} + g)^* = \text{dom } q_{L^*L}^* + \text{dom } g^* = \text{dom } q_{(L^*L)^{-1}} + \text{dom } g^* = X + \text{dom } g^* = X$ . (iv): Combine (13.46), [1, Corollary 2.1] and Lemma 13.3(i). (v): Since  $q_{L^*L}$  is strictly convex, so is  $q_{L^*L} + g$ , which in view of [5, Proposition 18.9] and (iii) implies that  $(q_{L^*L} + g)^*$  is Fréchet differentiable on  $X = \text{int } \text{dom}(q_{L^*L} + g)^*$ . (vi): Using (iv), Fact 13.8(i) applied with  $f$  replaced by  $q_{L^*L} + g$ , (v) and [5, Proposition 17.31(i)] we have  $(L^*L + \partial g)^{-1} = (\partial(q_{L^*L} + g))^{-1} = \partial(q_{L^*L} + g)^* = \{\nabla(q_{L^*L} + g)^*\}$  is single-valued with  $\text{dom}(L^*L + \partial g)^{-1} = X$ .

(vii): Let  $x \in X = \text{dom}(L^*L + \partial g)^{-1}$  and let  $y \in X$  such that  $y = x - L(L^*L + \partial g)^{-1}L^*x$ . Then using (vi) we have

$$x = y + Lu \quad \text{where} \quad u = (L^*L + \partial g)^{-1}L^*x. \quad (13.47)$$

Consequently,  $L^*y + L^*Lu = L^*x \in L^*Lu + \partial g(u)$ , hence  $L^*y \in \partial g(u)$ , equivalently, in view of Fact 13.8(i) applied with  $f$  replaced by  $g$ ,  $u \in (\partial g)^{-1}(L^*y) = \partial g^*(L^*y)$ . Combining with (13.47) we learn that

$$x \in y + L \circ (\partial g^*) \circ L^*(y). \quad (13.48)$$

Note that [5, Fact 2.25(vi) and Fact 2.26] implies that  $\text{ran } L^* = \text{ran } L^*L = X$ , hence  $0 \in \text{sri}(\text{dom } g^* - \text{ran } L^*)$ . Therefore one can apply [5, Corollary 16.53(i)] to re-write (13.48) as  $x \in (\text{Id} + \partial(g^* \circ L^*))y$ . Therefore,  $y = \text{Prox}_{g^* \circ L^*} x$  by [5, Proposition 16.44]. (viii): Apply Fact 13.8(ii) with  $f$  replaced by  $g^* \circ L^*$ .  $\square$

## Appendix 2

**Lemma 13.4** *Let  $g: Y \rightarrow ]-\infty, +\infty]$  be convex, lower semicontinuous, and proper. Consider the following statements:*

- (i)  $g$  is strongly convex.
- (ii)  $g^*$  is Fréchet differentiable and  $\nabla g^*$  is Lipschitz continuous.
- (iii)  $g^* \circ L^*$  is Fréchet differentiable and  $\nabla(g^* \circ L^*) = L \circ (\nabla g^*) \circ L^*$  is Lipschitz continuous.
- (iv)  $(g^* \circ L^*)^*$  is strongly convex.

Then (i)  $\Leftrightarrow$  (ii)  $\Rightarrow$  (iii)  $\Leftrightarrow$  (iv).

*Proof* (i)  $\Leftrightarrow$  (ii): See [5, Theorem 18.15]. (ii)  $\Rightarrow$  (iii): Clearly  $g^* \circ L^*$  is Fréchet differentiable. Now let  $(x, y) \in X \times X$  and suppose that  $\beta > 0$  is a Lipschitz constant of  $\nabla g^*$ . It follows from [5, Corollary 16.53] that  $\|\nabla(g^* \circ L^*)x - \nabla(g^* \circ L^*)y\| = \|L \circ (\nabla g^*) \circ L^*x - L \circ (\nabla g^*) \circ L^*y\| = \|L((\nabla g^* \circ L^*)x - (\nabla g^* \circ L^*)y)\| \leq \|L\| \|(\nabla g^* \circ L^*)x - (\nabla g^* \circ L^*)y\| \leq \beta \|L\| \|L^*x - L^*y\| \leq \beta \|L\| \|L^*\| \|x - y\|$ . (iii)  $\Leftrightarrow$  (iv): Use the equivalence of (i) and (ii) applied with  $g$  replaced by  $(g^* \circ L^*)^*$ .  $\square$

## Appendix 3

We start by recalling the following well-known fact.

**Fact 13.8** *Let  $f: X \rightarrow ]-\infty, +\infty]$  be convex, lower semicontinuous and proper and let  $\gamma > 0$ . Then the following hold:*

- (i)  $(\partial f)^{-1} = \partial f^*$ .
- (ii)  $\text{Prox}_{\gamma f} + \text{Prox}_{(\gamma f)^*} = \text{Id}$ .

*Proof* (i): See, e.g., [27, Remark on page 216] or [23, Théorème 3.1].

(ii): See, e.g., [5, Theorem 14.3(iii)].  $\square$

**Lemma 13.5** *Let  $\gamma > 0$ . The Douglas–Rachford method given in (13.9) applied to the ordered pair  $(\gamma f, \gamma g)$  with a starting point  $x_0 \in X$  to solve (13.8) can be rewritten as:*



$$y_n = \text{Prox}_{\gamma f} x_n \quad (13.49a)$$

$$x_{n+1} = y_n - \text{Prox}_{(\gamma g)^*}(2y_n - x_n). \quad (13.49b)$$

*Proof* Using (13.9a), (13.10), and Fact 13.8(ii) applied with  $f$  replaced by  $g$  we have

$$\begin{aligned} x_{n+1} &= x_n - \text{Prox}_{\gamma f} x_n + \text{Prox}_{\gamma g}(2 \text{Prox}_{\gamma f} x_n - x_n) = x_n - y_n + \text{Prox}_{\gamma g}(2y_n - x_n) \\ &= x_n - y_n + 2y_n - x_n - \text{Prox}_{(\gamma g)^*}(2y_n - x_n) = y_n - \text{Prox}_{(\gamma g)^*}(2y_n - x_n), \end{aligned} \quad (13.50)$$

and the conclusion follows.  $\square$

## Appendix 4

**Proposition 13.4** *Let  $(x, y, z) \in X \times Y \times Z$  and let  $B$  and  $\tilde{f}$  be defined as in (13.39) and (13.41). Then the following hold:*

- (i)  $B^*y = (A^*y, C^*y)$ .
- (ii)  $\text{dom } \tilde{f} = \text{dom } f \times \{0\}$ .
- (iii)  $(\forall (x, z) \in \text{dom } \tilde{f})$  we have  $z = 0$  and  $B(x, z) = Ax$ .
- (iv)  $B(\text{dom } \tilde{f}) = A(\text{dom } f)$ .
- (v)  $0 \in \text{sri}(\text{dom } g - B(\text{dom } \tilde{f}))$ .
- (vi)  $\text{argmin}(\tilde{f} + g \circ B) = \text{argmin}(f + g \circ A) \times \{0\} \neq \emptyset$ .
- (vii)  $\text{Prox}_{\tilde{f}}(x, z) = (\text{Prox}_f x, 0)$ .
- (viii)  $\text{Prox}_{(\tau g \circ B)^*} = \tau B^* \text{Prox}_{\sigma g^*}(\sigma B)$ .

*Proof* (i): This clearly follows from (13.39). (ii): It follows from (13.41) that  $\text{dom } \tilde{f} = \text{dom } f \times \text{dom } \iota_{\{0\}} = \text{dom } f \times \{0\}$ . (iii): The claim that  $z = 0$  follows from (ii). Now combine with (13.39). (iv): Combine (ii) and (iii). (v): Combine (iv) and (13.34). (vi): We have

$$\text{argmin}(\tilde{f} + g \circ B) = \text{zer}(\partial \tilde{f} + B^* \circ \partial g \circ B) \quad (13.51a)$$

$$= \text{zer}(\partial f \times N_{\{0\}} + ((A^* \circ \partial g \circ A) \times (C^* \circ \partial g \circ C))) \quad (13.51b)$$

$$= (\text{zer}(\partial f + A^* \circ \partial g \circ A)) \times (\text{zer}(N_{\{0\}} + C^* \circ \partial g \circ C)), \quad (13.51c)$$

where (13.51a) follows from (v) and (13.4) applied with  $(f, g, L)$  replaced by  $(g, \tilde{f}, B)$ , and (13.51b) follows from (13.39) and (13.41). Therefore,  $(x, z) \in \text{argmin}(\tilde{f} + g \circ B) \Leftrightarrow [z = 0 \text{ and } x \in \text{zer}(\partial f + A^* \circ \partial g \circ A)] \Leftrightarrow (x, z) \in$

$\operatorname{argmin}(f + g \circ A) \times \{0\}$ . Now combine with (13.4). (vii): Combine (13.41) and [5, Proposition 23.18]. (viii): Indeed, Proposition 13.3(viii) implies

$$\operatorname{Prox}_{(\tau g \circ B)^*} = B^*(BB^* + (\tau \partial g)^*)^{-1}B = B^*(\sigma^{-1}\tau^{-1}\operatorname{Id} + \partial g^* \circ \tau^{-1}\operatorname{Id})^{-1}B \quad (13.52a)$$

$$= B^*(\sigma^{-1}(\operatorname{Id} + \sigma \partial g^*)\tau^{-1} \circ \operatorname{Id})^{-1}B = \tau B^* \operatorname{Prox}_{\sigma g^*}(\sigma B). \quad (13.52b)$$

□

## References

1. Attouch, H., Brézis, H.: Duality for the sum of convex functions in general Banach spaces. In: *Aspects of Mathematics and Its Applications* **34**, pp. 125–133. North-Holland, Amsterdam (1986)
2. Attouch, H., Théra, M.: A general duality principle for the sum of two operators. *J. Convex Anal.* **3**, 1–24 (1996)
3. Bauschke, H.H., Combettes, P.L.: A Dykstra-like algorithm for two monotone operators. *Pacific J. Optim.* **4**, 383–391 (2008)
4. Bauschke, H.H., Boţ, R.I., Hare, W.L., Moursi, W.M.: Attouch–Théra duality revisited: paramonotonicity and operator splitting. *J. Approx. Th.* **164**, 1065–1084 (2012)
5. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Second Edition. Springer (2017)
6. Bauschke, H.H., Koch, V.R.: Projection methods: Swiss Army knives for solving feasibility and best approximation problems with halfspaces. In: *Infinite Products of Operators and Their Applications*, *Contemp. Math.* vol. 636, pp. 1–40 (2012)
7. Beck, A.: *First-Order Methods in Optimization*, SIAM (2017)
8. Boţ, R.I., Csetnek, E.R.: ADMM for monotone operators: convergence analysis and rates. *Advances Comp. Math.* **45**, 327–359 (2019)
9. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* **3**, 1–122 (2011)
10. Boyle, J.P., Dykstra, R.L.: A method for finding projections onto the intersection of convex sets in Hilbert spaces. *Lecture Notes in Statistics* **37**, 28–47 (1986)
11. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**, 120–145 (2011)
12. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization* **53**, 475–504 (2004)
13. Briceño-Arias, L.M., Combettes, P.L.: A monotone + skew splitting model for composite monotone inclusions in duality. *SIAM J. Optim.* **21**, 1230–1250 (2011)
14. Combettes, P.L., Pesquet, J.-C.: Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal.* **20**, 307–330 (2012)
15. Combettes, P.L., Pesquet, J.-C.: A proximal decomposition method for solving convex variational inverse problems. *Inverse Problems* **24**, article 065014 (2008)
16. Combettes, P.L., Pesquet, J.-C.: Proximal splitting methods in signal processing. In: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, vol. 49, pp. 185–212. Springer, New York (2011)

17. Combettes, P.L., Vũ, B.-C.: Variable metric forward–backward splitting with applications to monotone inclusions in duality. *Optimization* **63**, 1289–1318 (2014)
18. Condat, L.: A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Th. Appl.* **158**, 460–479 (2013)
19. Deutsch, F.: *Best Approximation in Inner Product Spaces*, Springer (2001)
20. Eckstein, J.: *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, Ph.D. thesis, MIT (1989)
21. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Prog.* **55**, 293–318 (1992)
22. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, vol. 15, pp. 299–331. North-Holland, Amsterdam (1983)
23. Gossez, J.-P.: Opérateurs monotones non linéaires dans les espaces de Banach non réflexifs. *J. Math. Anal. Appl.* **34**, 371–395 (1971)
24. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Numer. Anal.* **16**, 964–979 (1979)
25. O’Connor, D., Vandenberghe, L.: On the equivalence of the primal–dual hybrid gradient method and Douglas–Rachford splitting. *Math. Prog. (Ser. A)* (2018), <https://doi.org/10.1007/s10107-018-1321-1>.
26. Riesz, F., Sz.-Nagy, B.: *Functional Analysis*. Dover paperback (1990)
27. Rockafellar, R.T.: On the maximal monotonicity of subdifferential mappings. *Pacific J. Math.* **33**, 209–216 (1970)
28. Stoer, J., Bulirsch, R.: *Introduction to Numerical Analysis*. Third Edition. Springer-Verlag (2002)
29. Tseng, P.: Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM J. Con. Optim.* **29**, 119–138 (1991)
30. Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.* **38**, 667–681 (2013)