Monotone Operator Theory in Convex Optimization*

Patrick L. Combettes

North Carolina State University, Department of Mathematics, Raleigh, NC 27695-8205, USA

Abstract. Several aspects of the interplay between monotone operator theory and convex optimization are presented. The crucial role played by monotone operators in the analysis and the numerical solution of convex minimization problems is emphasized. We review the properties of subdifferentials as maximally monotone operators and, in tandem, investigate those of proximity operators as resolvents. In particular, we study new transformations which map proximity operators to proximity operators, and establish connections with self-dual classes of firmly nonexpansive operators. In addition, new insights and developments are proposed on the algorithmic front.

Keywords. Firmly nonexpansive operator, monotone operator, operator splitting, proximal algorithm, proximity operator, proximity-preserving transformation, self-dual class, subdifferential.

1 Introduction and historical overview

In this paper, we examine various facets of the role of monotone operator theory in convex optimization and of the interplay between the two fields. Throughout, $\mathcal{H}$ is a real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle$, associated norm $\|\cdot\|$, and identity operator Id. To put our discussion in proper perspective, we first provide an historical account and highlight some key results (see Section 2 for notation).

Monotone operator theory is a fertile area of nonlinear analysis which emerged in 1960 in independent papers by Kačurovskiĭ, Minty, and Zarantonello. Let D be a nonempty subset of $\mathcal{H}$, let $A\colon D \to \mathcal{H}$, and let $B\colon D \to \mathcal{H}$. Extending the ordering of functions on the real line which results from the comparison of their increments, Zarantonello [122] declared B is slower than A if

$$(\forall x \in D)(\forall y \in D) \quad \langle x - y \mid Ax - Ay \rangle \geqslant \langle x - y \mid Bx - By \rangle, \tag{1.1}$$

which is denoted by $A \succcurlyeq B$. He then called A (isotonically) monotone if $A \succcurlyeq 0$, that is,

$$(\forall x \in D)(\forall y \in D) \quad \langle x - y \mid Ax - Ay \rangle \geqslant 0, \tag{1.2}$$

and supra-unitary if $A \succcurlyeq \mathrm{Id}$. An instance of the latter notion can be found in [64]. In modern language, it corresponds to that of 1-strong monotonicity. An important result of [122] is the following.

Theorem 1.1 (Zarantonello) *Let $A\colon \mathcal{H} \to \mathcal{H}$ be monotone and Lipschitzian. Then* $\operatorname{ran}(\mathrm{Id} + A) = \mathcal{H}$.

Monotonicity captures two well-known concepts. First, if $\mathcal{H} = \mathbb{R}$, a function $A\colon D \to \mathbb{R}$ is monotone if and only if it is increasing, that is,

$$(\forall x \in D)(\forall y \in D) \quad x < y \quad \Rightarrow \quad Ax \leqslant Ay. \tag{1.3}$$

*Contact author: P. L. Combettes, plc@math.ncsu.edu, phone:+1 (919) 515 2671. This work was supported by the National Science Foundation under grant CCF-1715671.

The second connection is with linear functional analysis: if $A\colon \mathcal{H} \to \mathcal{H}$ is linear and bounded, then it is monotone if and only if it is positive, that is,

$$(\forall x \in \mathcal{H}) \quad \langle x \mid Ax\rangle \geqslant 0. \tag{1.4}$$

In particular, if a bounded linear operator $A\colon \mathcal{H} \to \mathcal{H}$ is skew, that is $A^* = -A$, then it is monotone since

$$(\forall x \in \mathcal{H}) \quad \langle x \mid Ax\rangle = 0. \tag{1.5}$$

Regarding (1.3), a standard fact about a differentiable convex function on an open interval of $\mathbb{R}$ is that its derivative is increasing. This property, which is already mentioned in Jensen's 1906 foundational paper [70], was extended in 1960 by Kačurovskiĭ [71], who came up with the notion of monotonicity (1.2), discussed strong monotonicity, and observed that the gradient of a differentiable convex function $f\colon \mathcal{H} \to \mathbb{R}$ is monotone (see also [118]). In a paper submitted in 1960, Minty [85] also called $A\colon D \to \mathcal{H}$ monotone if it satisfies (1.2), and maximally monotone if it cannot be extended to a strictly larger domain while preserving (1.2). Although, strictly speaking, his definitions dealt with single-valued operators, he established results on monotone relations that naturally cover extensions to what we now call set-valued operators. According to Browder [31], who initiated the study of set-valued monotone operators in Banach spaces, the Hilbertian setting was developed by Minty in unpublished notes. A set-valued operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone if

$$(\forall x \in \mathcal{H})(\forall u \in \mathcal{H}) \quad \big[\, (x,u) \in \operatorname{gra} A \quad \Leftrightarrow \quad (\forall (y,v) \in \operatorname{gra} A)\ \ \langle x-y \mid u-v\rangle \geqslant 0 \,\big]. \tag{1.6}$$

In other words, A is monotone and there exists no monotone operator $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ distinct from A such that $\operatorname{gra} A \subset \operatorname{gra} B$. A key result of [85] is the following theorem, which can be viewed as an extension of Theorem 1.1 since a continuous monotone operator $A\colon \mathcal{H} \to \mathcal{H}$ is maximally monotone.

Theorem 1.2 (Minty) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a monotone operator. Then A is maximally monotone if and only if* $\operatorname{ran}(\mathrm{Id} + A) = \mathcal{H}$.

The paper [85] also establishes an important connection between monotonicity and nonexpansiveness, which we state in the following form.

Theorem 1.3 [13, Prop. 4.4 and Cor. 23.9] *Let $T\colon \mathcal{H} \to \mathcal{H}$. Then the following are equivalent:*

(i) *T is firmly nonexpansive, i.e.,* [32]

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\|^2 + \|(\mathrm{Id} - T)x - (\mathrm{Id} - T)y\|^2 \leqslant \|x-y\|^2. \tag{1.7}$$

(ii) $R = 2T - \mathrm{Id}$ *is nonexpansive, i.e.,* $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Rx - Ry\| \leqslant \|x-y\|$.

(iii) *There exists a maximally monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ such that T is the resolvent of A, i.e.,*

$$T = J_A, \quad \text{where} \quad J_A = (\mathrm{Id} + A)^{-1}. \tag{1.8}$$

From the onset, monotone operator theory impacted areas such as partial differential equations, evolution equations and inclusions, and nonlinear equations; see for instance [25, 29, 30, 60, 72, 73, 77, 86, 109, 117, 123, 127]. In particular, in such problems, it turned out to provide efficient tools to derive existence results. Standard references on the modern theory of monotone operators are [13, 25, 99, 110]. From a modeling standpoint, monotone operator theory constitutes a powerful framework that reduces many problems in nonlinear analysis to the simple formulation

$$\text{find } x \in \operatorname{zer} A = \big\{x \in \mathcal{H} \,\big|\, 0 \in Ax\big\} = \operatorname{Fix} J_A, \quad \text{where} \quad A\colon \mathcal{H} \to 2^{\mathcal{H}} \text{ is maximally monotone.} \tag{1.9}$$

The most direct connection between monotone operator theory and optimization is obtained through the subdifferential of a proper function $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$, i.e., the operator [91, 92, 101]

$$\partial f\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \big\{u \in \mathcal{H} \,\big|\, (\forall y \in \mathcal{H})\;\; \langle y - x \mid u\rangle + f(x) \leqslant f(y)\big\}. \tag{1.10}$$

This operator is easily seen to be monotone. In addition, from the standpoint of minimization, a straightforward yet fundamental consequence of (1.10) is Fermat's rule. It states that, for every proper function $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$,

$$\operatorname{Argmin} f = \operatorname{zer} \partial f. \tag{1.11}$$

The maximality of the subdifferential was first investigated by Minty [87] for certain classes of convex functions, and then by Moreau [94] in full generality.

Theorem 1.4 (Moreau) *Let $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be a proper lower semicontinuous convex function. Then ∂f is maximally monotone.*

One way to prove Moreau's theorem is to use Theorem 1.2; see [13, Theorem 21.2] or [25, Example 2.3.4]. Interestingly, Moreau's proof in [94] did not rely on Theorem 1.2 but on proximal calculus. The proximity operator of a function $f \in \Gamma_0(\mathcal{H})$ is [89]

$$\operatorname{prox}_f\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \underset{y \in \mathcal{H}}{\operatorname{argmin}} \left(f(y) + \frac{1}{2}\|x - y\|^2\right). \tag{1.12}$$

This operator is intimately linked to the subdifferential operator. Indeed, let $f \in \Gamma_0(\mathcal{H})$. Then

$$(\forall x \in \mathcal{H})(\forall u \in \mathcal{H}) \quad u \in \partial f(x) \quad \Leftrightarrow \quad x = \operatorname{prox}_f(x + u). \tag{1.13}$$

Alternatively,

$$(\forall x \in \mathcal{H})(\forall p \in \mathcal{H}) \quad p = \operatorname{prox}_f x \quad \Leftrightarrow \quad x - p \in \partial f(p), \tag{1.14}$$

which entails, using (1.10), that prox_f is firmly nonexpansive. Furthermore, (1.11) and (1.14) imply that $\operatorname{Argmin} f = \operatorname{Fix} \operatorname{prox}_f$. Since fixed points of firmly nonexpansive operators can be constructed by successive approximations [32, 97], a conceptual algorithm for finding a minimizer of f is

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_f x_n. \tag{1.15}$$

This scheme was first studied by Martinet in the early 1970s [82, 83], and a special case in the context of quadratic programming appeared in [19, Sect. 5.8]. Though of limited practical use, this so-called proximal point algorithm occupies nonetheless a central place in convex minimization schemes because it embraces many fundamental ideas and connections that have inspired much more efficient and broadly applicable minimization algorithms in the form of proximal splitting methods [13, 48, 63]. The methodology underlying these algorithms is to solve structured convex minimization problems using only the proximity operators of the individual functions present in the model.

Moreau's motivations for introducing the proximity operator (1.12) came from nonsmooth mechanics [90, 93, 95]. In recent years proximity operators have become prominent in convex optimization theory. For instance, they play a central theoretical role in [13]. On the application side, their increasing presence is particularly manifest in the broad area of data processing, where they were introduced in [52] and have since proven very effective in the modeling and the numerical solution of a vast array of problems in disciplines such as signal processing, image recovery, machine learning, and computational statistics; see for instance [18, 23, 33, 36, 41, 44, 46, 48, 50, 56, 66, 69, 98, 119, 114, 128].

At first glance, it may appear that the theory of subdifferentials and proximity operators forms a self-contained corpus of theoretical and algorithmic tools which is sufficient to deal with convex optimization problems, and that the broader concepts of monotone operators and resolvents play only a peripheral role in such problems. A goal of this paper is to show that monotone operator theory occupies a central position in convex optimization, and that many advances in the latter would not have been possible without it. Conversely, we shall see that some algorithmic developments in monotonicity methods have directly benefited from convex minimization methodologies. We shall also examine certain aspects of the gap that separates the two theories. Section 2 covers notation and background. Section 3 studies subdifferentials as maximally monotone operators and proximity operators as resolvents, discussing characterizations, new proximity-preserving transformations, and self-dual classes. Section 4 focuses on the use of monotone operator theory in analyzing and solving convex optimization problems, and it proposes new insights and developments.

2 Notation and background

We follow the notation of [13], where one will find a detailed account of the following notions. The direct Hilbert sum of $\mathcal{H}$ and a real Hilbert space $\mathcal{G}$ is denoted by $\mathcal{H} \oplus \mathcal{G}$. Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. We denote by $\operatorname{gra} A = \big\{(x,u) \in \mathcal{H} \times \mathcal{H} \;\big|\; u \in Ax\big\}$ the graph of A, by $\operatorname{dom} A = \big\{x \in \mathcal{H} \;\big|\; Ax \neq \varnothing\big\}$ the domain of A, by $\operatorname{ran} A = \big\{u \in \mathcal{H} \;\big|\; (\exists x \in \mathcal{H})\, u \in Ax\big\}$ the range of A, by $\operatorname{zer} A = \big\{x \in \mathcal{H} \;\big|\; 0 \in Ax\big\}$ the set of zeros of A, and by A^{-1} the inverse of A, i.e., the set-valued operator with graph $\big\{(u,x) \in \mathcal{H} \times \mathcal{H} \;\big|\; u \in Ax\big\}$. The parallel sum of A and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$, and the parallel composition of A by $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$ are, respectively,

$$A \,\Box\, B = (A^{-1} + B^{-1})^{-1} \quad \text{and} \quad L \triangleright A = \big(L \circ A^{-1} \circ L^*\big)^{-1}. \tag{2.1}$$

The resolvent of A is $J_A = (\mathrm{Id} + A)^{-1} = A^{-1} \,\Box\, \mathrm{Id}$. The set of fixed points of an operator $T\colon \mathcal{H} \to \mathcal{H}$ is $\operatorname{Fix} T = \big\{x \in \mathcal{H} \;\big|\; Tx = x\big\}$. The set of global minimizers of a function $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ is denoted by $\operatorname{Argmin} f$ and, if it is a singleton, its unique element is denoted by $\operatorname{argmin} f$. We denote by $\Gamma_0(\mathcal{H})$ the class of lower semicontinuous convex functions $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ such that $\operatorname{dom} f = \big\{x \in \mathcal{H} \;\big|\; f(x) < +\infty\big\} \neq \varnothing$. Now let $f \in \Gamma_0(\mathcal{H})$. The conjugate of f is the function $f^* \in \Gamma_0(\mathcal{H})$ defined by $f^*\colon u \mapsto \sup_{x\in\mathcal{H}}(\langle x \mid u\rangle - f(x))$. The subdifferential of f is defined in (1.10), and its inverse is $(\partial f)^{-1} = \partial f^*$. The proximity operator prox_f of f is defined in (1.12). We say that f is ν-strongly convex for some $\nu \in \left]0,+\infty\right[$ if $f - \nu\|\cdot\|^2/2$ is convex. The infimal convolution of f and $g \in \Gamma_0(\mathcal{H})$ is

$$f \,\Box\, g\colon \mathcal{H} \to [-\infty,+\infty] : x \mapsto \inf_{y\in\mathcal{H}} \big(f(y) + g(x-y)\big). \tag{2.2}$$

Let C be a convex subset of $\mathcal{H}$. The interior of C is denoted by $\operatorname{int} C$, the boundary of C by $\operatorname{bdry} C$, the indicator function of C by ι_C, the distance function to C by d_C, the support function of C by σ_C and, if C is nonempty and closed, the projection operator onto C by proj_C, i.e., $\operatorname{proj}_C = \operatorname{prox}_{\iota_C}$. A point $x \in C$ is in the strong relative interior of C, denoted by $\operatorname{sri} C$, if the cone generated by $C - x$ is a closed vector subspace of $\mathcal{H}$. We define

- $\mathcal{B}(\mathcal{H},\mathcal{G}) = \big\{T\colon \mathcal{H} \to \mathcal{G} \;\big|\; T \text{ is linear and bounded}\big\}$ and $\mathcal{B}(\mathcal{H}) = \mathcal{B}(\mathcal{H},\mathcal{H})$.
- $\mathcal{N}(\mathcal{H}) = \big\{T\colon \mathcal{H} \to \mathcal{H} \;\big|\; T \text{ is nonexpansive}\big\}$.

- $\mathcal{F}(\mathcal{H}) = \{T\colon \mathcal{H} \to \mathcal{H} \mid T \text{ is firmly nonexpansive}\}$.
- $\mathcal{M}(\mathcal{H}) = \{A\colon \mathcal{H} \to 2^{\mathcal{H}} \mid A \text{ is maximally monotone}\}$.
- $\mathcal{S}(\mathcal{H}) = \{A\colon \mathcal{H} \to 2^{\mathcal{H}} \mid (\exists\, f \in \Gamma_0(\mathcal{H}))\ A = \partial f\}$.
- $\mathcal{J}(\mathcal{H}) = \{T\colon \mathcal{H} \to \mathcal{H} \mid (\exists\, A \in \mathcal{M}(\mathcal{H}))\ T = J_A\}$.
- $\mathcal{P}(\mathcal{H}) = \{T\colon \mathcal{H} \to \mathcal{H} \mid (\exists\, f \in \Gamma_0(\mathcal{H}))\ T = \operatorname{prox}_f\}$.
- $\mathcal{K}(\mathcal{H}) = \{T\colon \mathcal{H} \to \mathcal{H} \mid T = \operatorname{proj}_K \text{ for some nonempty closed convex cone } K \subset \mathcal{H}\}$.
- $\mathcal{V}(\mathcal{H}) = \{T\colon \mathcal{H} \to \mathcal{H} \mid T = \operatorname{proj}_V \text{ for some closed vector space } V \subset \mathcal{H}\}$.

Facts mentioned in Section 1 are summarized by the inclusions

$$\mathcal{S}(\mathcal{H}) \subset \mathcal{M}(\mathcal{H}),\ \mathcal{V}(\mathcal{H}) \subset \mathcal{K}(\mathcal{H}) \subset \mathcal{P}(\mathcal{H}) \subset \mathcal{J}(\mathcal{H}) = \mathcal{F}(\mathcal{H}) \subset \mathcal{N}(\mathcal{H}),\ \text{and } \mathcal{F}(\mathcal{H}) \subset \mathcal{M}(\mathcal{H}). \tag{2.3}$$

We have

$$\big(\forall A \in \mathcal{M}(\mathcal{H})\big) \quad A^{-1} \in \mathcal{M}(\mathcal{H}) \quad \text{and} \quad J_{A^{-1}} + J_A = A \,\Box\, \mathrm{Id} + A^{-1} \,\Box\, \mathrm{Id} = \mathrm{Id}\,. \tag{2.4}$$

Theorem 2.1 (Moreau's decomposition [89, 91, 94]) *Let $f \in \Gamma_0(\mathcal{H})$ and set $q = \|\cdot\|^2/2$. Then $f \,\Box\, q + f^* \,\Box\, q = q$ and $\operatorname{prox}_f = J_{\partial f} = \nabla(f+q)^* = \nabla(f^* \,\Box\, q) = (\partial f^*) \,\Box\, \mathrm{Id} = \mathrm{Id} - \operatorname{prox}_{f^*}$.*

The next result brings together ideas from [6] and [94].

Theorem 2.2 ([12]) *Let $h\colon \mathcal{H} \to \mathbb{R}$ be continuous and convex, and set $f = h^* - q$, where $q = \|\cdot\|^2/2$. Then the following are equivalent:*

(i) *h is Fréchet differentiable on $\mathcal{H}$ and $\nabla h \in \mathcal{N}(\mathcal{H})$.*

(ii) *h is Fréchet differentiable on $\mathcal{H}$ and $\nabla h \in \mathcal{F}(\mathcal{H})$.*

(iii) *$q - h$ is convex.*

(iv) *$h^* - q$ is convex.*

(v) *$f \in \Gamma_0(\mathcal{H})$ and $h = f^* \,\Box\, q = q - f \,\Box\, q$.*

(vi) *$f \in \Gamma_0(\mathcal{H})$ and $\operatorname{prox}_f = \nabla h = \mathrm{Id} - \operatorname{prox}_{f^*}$.*

Lemma 2.3 [13, Prop. 2.58] *Let $f\colon \mathcal{H} \to \mathbb{R}$ be Gâteaux differentiable, let $L \in \mathcal{B}(\mathcal{H})$, and suppose that $\nabla f = L$. Then $L = L^*$, $f\colon x \mapsto f(0) + (1/2)\langle Lx \mid x\rangle$, and f is twice Fréchet differentiable.*

3 Subdifferentials as monotone operators

As seen in Section 1, from a convex optimization perspective, the subdifferential and the proximity operators of a function in $\Gamma_0(\mathcal{H})$ constitute, respectively, prime examples of maximally monotone and firmly nonexpansive operators. In this section with discuss some structural differences between $\mathcal{S}(\mathcal{H})$ and $\mathcal{M}(\mathcal{H})$, and between $\mathcal{P}(\mathcal{H})$ and $\mathcal{J}(\mathcal{H})$.

3.1 Characterization of subdifferentials

If $\mathcal{H} = \mathbb{R}$, then $\mathcal{M}(\mathcal{H}) = \mathcal{S}(\mathcal{H})$; see [103, Sect. 24] or [13, Cor. 22.23]. In general, however, this singular situation no longer manifests itself. For instance if $0 \neq A \in \mathcal{B}(\mathcal{H})$ is skew (see (1.5)), then $A \in \mathcal{M}(\mathcal{H})$ since A is monotone and continuous, but Lemma 2.3 asserts that it is not a gradient since it is not self-adjoint; it can therefore not be in $\mathcal{S}(\mathcal{H})$. A complete characterization of subdifferentials as maximally monotone operators was given by Rockafellar in [102]. An operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is cyclically monotone if, for every integer $n \geqslant 2$, every $(x_1, \ldots, x_{n+1}) \in \mathcal{H}^{n+1}$, and every $(u_1, \ldots, u_n) \in \mathcal{H}^n$,

$$\big[\,(x_1,u_1) \in \operatorname{gra} A, \ldots, (x_n,u_n) \in \operatorname{gra} A,\; x_{n+1} = x_1\,\big] \quad \Rightarrow \quad \sum_{i=1}^{n} \langle x_{i+1} - x_i \mid u_i\rangle \leqslant 0. \tag{3.1}$$

In this case, A is called maximally cyclically monotone if there exists no cyclically monotone operator $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ such that $\operatorname{gra} B$ properly contains $\operatorname{gra} A$.

Theorem 3.1 (Rockafellar) $\mathcal{S}(\mathcal{H}) = \big\{A \in \mathcal{M}(\mathcal{H}) \;\big|\; A \text{ is maximally cyclically monotone}\big\}$.

The question of representing a maximally monotone operator as the sum of a subdifferential and a remainder component is a challenging one. In the case of a monotone matrix A such a decomposition is obtained by writing A as the sum of its symmetric part (hence a gradient) and its antisymmetric part. This observation motivated Asplund [2] to investigate the decomposition of $A \in \mathcal{M}(\mathcal{H})$ as

$$A = G + B, \quad \text{where } G \in \mathcal{S}(\mathcal{H}) \text{ and } B \in \mathcal{M}(\mathcal{H}) \text{ is acyclic.} \tag{3.2}$$

Here, acyclic means that if $B = \partial g + C$ for some $g \in \Gamma_0(\mathcal{H})$ and some $C \in \mathcal{M}(\mathcal{H})$, then g is affine on $\operatorname{dom} A$. A sufficient condition for Alspund's cyclic+acyclic decomposition (3.2) to exist for $A \in \mathcal{M}(\mathcal{H})$ is that $\operatorname{int}\operatorname{dom} A \neq \varnothing$ [20]. Acyclic operators are not easy to apprehend, which shows that the notion of a maximally monotone operator remains only partially understood. A simpler decomposition was investigated in [21, 22] by imposing that B in (3.2) be the restriction of a skew operator in $\mathcal{B}(\mathcal{H})$. Thus, the so-called Borwein-Wiersma decomposition of $A \in \mathcal{M}(\mathcal{H})$ is

$$A = G + B, \quad \text{where } G \in \mathcal{S}(\mathcal{H}) \text{ and } B = S|_{\operatorname{dom} B}, \text{ with } S \in \mathcal{B}(\mathcal{H}) \text{ and } S^* = -S. \tag{3.3}$$

If $A \in \mathcal{M}(\mathcal{H})$ and $\operatorname{gra} A$ is a vector subspace, then A admits a Borwein-Wiersma decomposition if and only if $\operatorname{dom} A \subset \operatorname{dom} A^*$ [17, Thm. 5.1]. Another viewpoint on the distinction between a general maximally monotone operator and a subdifferential is presented in [126].

3.2 Characterizations of proximity operators

Exploring a different facet of the discussion of Section 3.1, we focus in this section on some properties of the class of proximity operators as a subset of that of firmly nonexpansive operators. We first review characterization results and then study the closedness of $\mathcal{P}(\mathcal{H})$ under various transformations.

A first natural question that arises is how to characterize those firmly nonexpansive operators which are proximity operators. As mentioned in Section 3.1, on the real line things are straightforward: since $\mathcal{M}(\mathbb{R}) = \mathcal{S}(\mathbb{R})$, Theorem 1.3 tells us $\mathcal{F}(\mathbb{R}) = \mathcal{P}(\mathbb{R})$. Alternatively, $T\colon \mathbb{R} \to \mathbb{R}$ belongs to $\mathcal{P}(\mathbb{R})$ if and only if it is nonexpansive and increasing [46]. In general, the characterization of subdifferential operators given in Theorem 3.1, together with Theorem 1.3, suggests introducing a cyclic version of (1.7) to achieve this goal. This leads to the following characterization.

Proposition 3.2 [7] *Let $T \in \mathcal{F}(\mathcal{H})$. Then $T \in \mathcal{P}(\mathcal{H})$ if and only if, for every integer $n \geqslant 2$ and every $(x_1, \ldots, x_{n+1}) \in \mathcal{H}^{n+1}$ such that $x_{n+1} = x_1$, we have $\sum_{i=1}^{n} \langle x_i - Tx_i \mid Tx_i - Tx_{i+1} \rangle \geqslant 0$.*

For our purposes, a more readily exploitable characterization is the following result due to Moreau (see also Theorem 2.2).

Theorem 3.3 [94] *Let $T \in \mathcal{N}(\mathcal{H})$ and let $q = \|\cdot\|^2/2$. Then $T \in \mathcal{P}(\mathcal{H})$ if and only if there exists a differentiable convex function $h \colon \mathcal{H} \to \mathbb{R}$ such that $T = \nabla h$. In this case, $T = \operatorname{prox}_f$, where $f = h^* - q$.*

Corollary 3.4 [94] *Let $T \in \mathcal{B}(\mathcal{H})$ be such that $\|T\| \leqslant 1$. Then $T \in \mathcal{P}(\mathcal{H})$ if and only if T is positive and self-adjoint.*

3.3 Proximity-preserving transformations

A transformation which preserves firm nonexpansiveness may not be proximity-preserving in the sense that it may not produce a proximity operator when applied to proximity operators. Here are two examples.

Example 3.5 (composition-based transformations) Transformations involving compositions are unlikely to be proximity-preserving for a simple reason: in the linear case, Corollary 3.4 imposes that such a transformation preserve self-adjointness. However, a product of symmetric matrices may not be symmetric. A standard example is the Douglas-Rachford splitting operator $T_{A,B}$ associated with two operators A and B in $\mathcal{M}(\mathcal{H})$ [13], which will arise in (4.6). In general,

$$T_{A,B} = J_A \circ (2J_B - \mathrm{Id}\,) + \mathrm{Id}\, - J_B = \frac{(2J_A - \mathrm{Id}\,) \circ (2J_B - \mathrm{Id}\,) + \mathrm{Id}}{2} \in \mathcal{J}(\mathcal{H}) \smallsetminus \mathcal{P}(\mathcal{H}). \tag{3.4}$$

The fact that $T_{A,B} \in \mathcal{J}(\mathcal{H})$ follows from the equivalence (i)⇔(ii) in Theorem 1.3. On the other hand, examples when $T_{A,B} \notin \mathcal{P}(\mathcal{H})$ for $A \in \mathcal{S}(\mathcal{H})$ and $B \in \mathcal{S}(\mathcal{H})$ can be easily constructed when J_A and J_B are 2×2 matrices as explained above. In fact, when A and B are linear relations in $\mathcal{S}(\mathbb{R}^N)$ ($N \geqslant 2$), the genericity of (3.4) is established in [16].

Example 3.6 (Spingarn's partial inverse) Let $A \in \mathcal{M}(\mathcal{H})$ and let V be a closed vector subspace of $\mathcal{H}$. The partial inverse of A with respect to V is the operator $A_V \colon \mathcal{H} \to 2^{\mathcal{H}}$ with graph

$$\operatorname{gra} A_V = \big\{ (\operatorname{proj}_V x + \operatorname{proj}_{V^\perp} u, \operatorname{proj}_V u + \operatorname{proj}_{V^\perp} x) \;\big|\; (x, u) \in \operatorname{gra} A \big\}. \tag{3.5}$$

This operator, which was introduced by Spingarn in [111], can be regarded as an intermediate object between A and A^{-1}. As shown in [111], $A \in \mathcal{M}(\mathcal{H}) \Leftrightarrow A_V \in \mathcal{M}(\mathcal{H})$. Therefore, by Theorem 1.3, $A \in \mathcal{M}(\mathcal{H}) \Leftrightarrow J_{A_V} \in \mathcal{J}(\mathcal{H})$. However,

$$A \in \mathcal{S}(\mathcal{H}) \quad \not\Rightarrow \quad J_{A_V} \in \mathcal{P}(\mathcal{H}). \tag{3.6}$$

To see this suppose that $\mathcal{H} = \mathbb{R}^2$, let $V = \big\{ (\xi_1, \xi_2) \in \mathcal{H} \;\big|\; \xi_1 = \xi_2 \big\}$, and let $A = \partial f$, where $f \colon (\xi_1, \xi_2) \mapsto \xi_1^2/2 + \xi_1\xi_2 + \xi_2^2$. Then, for every $(\xi_1, \xi_2) \in \mathcal{H}$, $\partial f(\xi_1, \xi_2) = (\xi_1 + \xi_2, \xi_1 + 2\xi_2)$, and we obtain

$$A_V(\xi_1, \xi_2) = (2\xi_1 + \xi_2, -\xi_1 + 2\xi_2) \quad \text{and} \quad J_{A_V}(\xi_1, \xi_2) = \frac{1}{10}(3\xi_1 - \xi_2, \xi_1 + 3\xi_2). \tag{3.7}$$

Thus $J_{A_V}^* \neq J_{A_V}$ and Corollary 3.4 implies that $J_{A_V} \notin \mathcal{P}(\mathcal{H})$.

Let us start with some simple proximity-preserving transformations.

Proposition 3.7 *Let $T \in \mathcal{P}(\mathcal{H})$. Then the following hold:*

(i) $\mathrm{Id} - T \in \mathcal{P}(\mathcal{H})$.

(ii) *Let $z \in \mathcal{H}$. Then $z + T(\cdot - z) \in \mathcal{P}(\mathcal{H})$.*

(iii) $-T \circ (-\mathrm{Id}) \in \mathcal{P}(\mathcal{H})$.

(iv) $J_T \in \mathcal{P}(\mathcal{H})$.

Proof. Let $f \in \Gamma_0(\mathcal{H})$ be such that $T = \mathrm{prox}_f$, and set $q = \|\cdot\|^2/2$.

(i): Set $g = f^*$. Then $g \in \Gamma_0(\mathcal{H})$ and Theorem 2.1 states that $\mathrm{prox}_g = \mathrm{Id} - T$.

(ii): Set $g = f(\cdot - z)$. Then $g \in \Gamma_0(\mathcal{H})$ and $\mathrm{prox}_g = z + T(\cdot - z)$ [52].

(iii): Set $g = f(-\cdot)$. Then $g \in \Gamma_0(\mathcal{H})$ and $\mathrm{prox}_g = -T \circ (-\mathrm{Id})$ [52].

(iv): Since $\mathcal{P}(\mathcal{H}) \subset \mathcal{J}(\mathcal{H}) \subset \mathcal{M}(\mathcal{H})$, J_T is well defined. Now set $g = f^* \,\square\, q$. Then $g \in \Gamma_0(\mathcal{H})$ and $g = (f + q)^*$. Thus, by Theorem 2.1, $J_T = T^{-1} \,\square\, \mathrm{Id} = (\mathrm{Id} + \partial f) \,\square\, \mathrm{Id} = \partial(f + q) \,\square\, \mathrm{Id} = (\partial g^*) \,\square\, \mathrm{Id} = \mathrm{prox}_g$. ☐

Proposition 3.8 *Let $f \in \Gamma_0(\mathcal{H})$, let $\mathcal{G}$ be a real Hilbert space, and let $M \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ be such that $0 \in \mathrm{sri}\,(\mathrm{ran}\, M^* - \mathrm{dom}\, f)$ and $MM^* - \mathrm{Id}_{\mathcal{G}}$ is positive. Set $q_{\mathcal{H}} = \|\cdot\|^2_{\mathcal{H}}/2$ and $q_{\mathcal{G}} = \|\cdot\|^2_{\mathcal{G}}/2$. Then $M \triangleright \mathrm{prox}_f \in \mathcal{P}(\mathcal{G})$. More specifically, $M \triangleright \mathrm{prox}_f = \mathrm{prox}_\varphi$, where $\varphi = (f + q_{\mathcal{H}}) \circ M^* - q_{\mathcal{G}}$.*

Proof. Set $\varphi = (f+q) \circ M^* - q$. The assumptions imply that $f \circ M^* \in \Gamma_0(\mathcal{G})$ and that $q_{\mathcal{H}} \circ M^* - q_{\mathcal{G}} \colon \mathcal{G} \to \mathbb{R}$ is convex and continuous. Consequently, $\varphi \in \Gamma_0(\mathcal{G})$ and, using (2.1) and [13, Cor. 16.53(i)],

$$\begin{aligned} M \triangleright \mathrm{prox}_f &= \big(M \circ (\mathrm{Id}_{\mathcal{H}} + \partial f) \circ M^*\big)^{-1} \\ &= \big(M \circ \partial(f + q_{\mathcal{H}}) \circ M^*\big)^{-1} \\ &= \Big(\partial\big((f + q_{\mathcal{H}}) \circ M^*\big)\Big)^{-1} \\ &= \Big(\mathrm{Id}_{\mathcal{G}} + \partial\big((f + q_{\mathcal{H}}) \circ M^* - q_{\mathcal{G}}\big)\Big)^{-1} \\ &= \mathrm{prox}_\varphi, \end{aligned} \tag{3.8}$$

as claimed. ☐

We now describe a composite proximity-preserving transformation.

Proposition 3.9 *Let I be a nonempty finite set and put $q = \|\cdot\|^2_{\mathcal{H}}/2$. For every $i \in I$, let $\omega_i \in \left]0, +\infty\right[$, let $\mathcal{G}_i$ be a real Hilbert space with identity operator Id_i, put $q_i = \|\cdot\|^2_{\mathcal{G}_i}/2$, let $\mathcal{K}_i$ be a real Hilbert space, let $L_i \in \mathcal{B}(\mathcal{H}, \mathcal{G}_i) \smallsetminus \{0\}$, let $M_i \in \mathcal{B}(\mathcal{K}_i, \mathcal{G}_i) \smallsetminus \{0\}$, let $f_i \in \Gamma_0(\mathcal{G}_i)$, let $g_i \in \Gamma_0(\mathcal{G}_i)$, and let $h_i \in \Gamma_0(\mathcal{K}_i)$. Suppose that $\sum_{i \in I} \omega_i \|L_i\|^2 \leqslant 1$ and that, for every $i \in I$,*

$$0 \in \mathrm{sri}\,\big(\mathrm{dom}\, h_i^* - M_i^*(\mathrm{dom}\, f_i \cap \mathrm{dom}\, g_i^*)\big) \quad \text{and} \quad 0 \in \mathrm{sri}\,(\mathrm{dom}\, f_i - \mathrm{dom}\, g_i^*). \tag{3.9}$$

Set

$$T = \sum_{i \in I} \omega_i L_i^* \circ \Big(\mathrm{prox}_{f_i} \,\square\, \big(\partial g_i \,\square\, (M_i \triangleright \partial h_i)\big)\Big) \circ L_i. \tag{3.10}$$

Then $T \in \mathcal{P}(\mathcal{H})$. More specifically,

$$T = \mathrm{prox}_f, \quad \text{where} \quad f = \left(\sum_{i \in I} \omega_i \Big(\big(f_i + g_i^* + h_i^* \circ M_i^*\big)^* \,\square\, q_i\Big) \circ L_i\right)^* - q. \tag{3.11}$$

Proof. The fact that $f \in \Gamma_0(\mathcal{H})$ follows from standard convex analysis [13]. Now let $i \in I$. We derive from (2.1), (3.9), [13, Cor. 16.30 and Thm. 16.47(i)], and Theorem 3.3 that

$$\begin{aligned} \operatorname{prox}_{f_i} \square \big(\partial g_i \square (M_i \rhd \partial h_i)\big) &= \Big(\mathrm{Id}_i + \partial f_i + \big(\partial g_i \square (M_i \rhd \partial h_i)\big)^{-1}\Big)^{-1} \\ &= \Big(\mathrm{Id}_i + \partial f_i + (\partial g_i)^{-1} + M_i \circ (\partial h_i)^{-1} \circ M_i^*\Big)^{-1} \\ &= \Big(\mathrm{Id}_i + \partial f_i + \partial g_i^* + M_i \circ \partial h_i^* \circ M_i^*\Big)^{-1} \\ &= \Big(\mathrm{Id}_i + \partial\big(f_i + g_i^* + h_i^* \circ M_i^*\big)\Big)^{-1} \\ &= \operatorname{prox}_{f_i + g_i^* + h_i^* \circ M_i^*} \\ &= \nabla\Big(\big(f_i + g_i^* + h_i^* \circ M_i^*\big)^* \square q_i\Big). \end{aligned} \tag{3.12}$$

Since $\operatorname{prox}_{f_i + g_i^* + h_i^* \circ M_i^*} \in \mathcal{N}(\mathcal{H})$,

$$\nabla\Big(\Big(\big(f_i + g_i^* + h_i^* \circ M_i^*\big)^* \square q_i\Big) \circ L_i\Big) = L_i^* \circ \operatorname{prox}_{f_i + g_i^* + h_i^* \circ M_i^*} \circ L_i \tag{3.13}$$

has Lipschitz constant $\|L_i\|^2$. Altogether,

$$T = \sum_{i \in I} \omega_i L_i^* \circ \Big(\operatorname{prox}_{f_i} \square \big(\partial g_i \square (M_i \rhd \partial h_i)\big)\Big) \circ L_i = \nabla\left(\sum_{i \in I} \omega_i \Big(\big(f_i + g_i^* + h_i^* \circ M_i^*\big)^* \square q_i\Big) \circ L_i\right) \tag{3.14}$$

has Lipschitz constant $\sum_{i \in I} \omega_i \|L_i\|^2 \leqslant 1$. In view of Theorem 3.3, the proof is complete. ☐

Remark 3.10 Let us highlight some special cases of Proposition 3.9. $\mathcal{G}$ is a real Hilbert space.

(i) Let $(T_i)_{i \in I}$ be a finite family in $\mathcal{P}(\mathcal{H})$ and let $(\omega_i)_{i \in I}$ be a finite family in $]0,1]$ such that $\sum_{i \in I} \omega_i = 1$. Then $\sum_{i \in I} \omega_i T_i \in \mathcal{P}(\mathcal{H})$. This result is due to Moreau [91]. Connections with the proximal average are discussed in [13].

(ii) In (i), taking $I = \{1,2\}$, $T_1 = T$, $T_2 = \mathrm{Id}$, and $\omega_1 = \lambda \in]0,1[$ yields $\mathrm{Id} + \lambda(T - \mathrm{Id}) \in \mathcal{P}(\mathcal{H})$. The fact that the under-relaxation of a proximity operator is a proximity operator appears in [52]. More precisely, it is shown there that if $T = \operatorname{prox}_h$ for some $h \in \Gamma_0(\mathcal{H})$, then $\mathrm{Id} + \lambda(T - \mathrm{Id}) = \operatorname{prox}_f$, where $f = h \square (\lambda q)/(1-\lambda)$, which can now be seen as a consequence of (3.11).

(iii) Let T_1 and T_2 be in $\mathcal{P}(\mathcal{H})$. Then $(T_1 - T_2 + \mathrm{Id})/2 \in \mathcal{P}(\mathcal{H})$. Indeed, Proposition 3.7(i) asserts that $\mathrm{Id} - T_2 \in \mathcal{P}(\mathcal{H})$ and then (i) that the average of T_1 and $\mathrm{Id} - T_2$ is also in $\mathcal{P}(\mathcal{H})$.

(iv) Let $T \in \mathcal{P}(\mathcal{G})$ and let $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$ be such that $\|L\| \leqslant 1$. Then $L^* \circ T \circ L \in \mathcal{P}(\mathcal{H})$.

(v) Let $T \in \mathcal{P}(\mathcal{H})$ and let V be a closed vector subspace of $\mathcal{H}$. Then it follows from (iv) that $\operatorname{proj}_V \circ T \circ \operatorname{proj}_V \in \mathcal{P}(\mathcal{H})$.

(vi) Suppose that $u \in \mathcal{H}$ satisfies $0 < \|u\| \leqslant 1$ and let $R \in \mathcal{P}(\mathbb{R})$. Set $L\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \langle x \mid u\rangle$ and $T\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto (R\langle x \mid u\rangle)u$. Then (iv) yields $T \in \mathcal{P}(\mathcal{H})$.

(vii) Let $M \in \mathcal{B}(\mathcal{G},\mathcal{H}) \smallsetminus \{0\}$, let $f \in \Gamma_0(\mathcal{H})$, let $g \in \Gamma_0(\mathcal{H})$, and let $h \in \Gamma_0(\mathcal{G})$ be such that $0 \in \operatorname{sri}(\operatorname{dom} h^* - M^*(\operatorname{dom} f \cap \operatorname{dom} g^*))$ and $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g^*)$. Then $\operatorname{prox}_f \square (\partial g \square (M \rhd \partial h)) \in \mathcal{P}(\mathcal{H})$. More specifically, $\operatorname{prox}_f \square (\partial g \square (M \rhd \partial h)) = \operatorname{prox}_{f + g^* + h^* \circ M^*}$.

(viii) In (vii), suppose that, in addition, $g = \varphi^* \,\square\, q_{\mathcal{H}}$ and $h = \psi^* \,\square\, q_{\mathcal{G}}$, where $\varphi \in \Gamma_0(\mathcal{H})$ and $\psi \in \Gamma_0(\mathcal{G})$. Then $\partial g = \{\text{prox}_\varphi\}$, $\partial h = \{\text{prox}_\psi\}$, and we conclude that $\text{prox}_f \,\square\, (\text{prox}_\varphi \,\square\, (M \triangleright \text{prox}_\psi)) \in \mathcal{P}(\mathcal{H})$. More specifically,

$$\text{prox}_f \,\square\, \big(\text{prox}_\varphi \,\square\, (M \triangleright \text{prox}_\psi)\big) = \text{prox}_{f+\varphi+(\psi+q_{\mathcal{G}})\circ M^*+q_{\mathcal{H}}} = \text{prox}_{(f+\varphi+(\psi+q_{\mathcal{G}})\circ M^*)/2}(\cdot/2) \tag{3.15}$$

has Lipschitz constant $1/2$.

(ix) In (vii), suppose that, in addition, $\mathcal{G} = \mathcal{H}$, $M = \text{Id}$, $g = \varphi^* \,\square\, q$, and $h = \iota_{\{0\}}$, where $\varphi \in \Gamma_0(\mathcal{H})$. Then $\partial g = \{\text{prox}_\varphi\}$, $\partial h = \{0\}^{-1}$, and we conclude that $\text{prox}_f \,\square\, \text{prox}_\varphi \in \mathcal{P}(\mathcal{H})$. More specifically,

$$\text{prox}_f \,\square\, \text{prox}_\varphi = \text{prox}_{f+\varphi+q} = \text{prox}_{(f+\varphi)/2}(\cdot/2) \tag{3.16}$$

has Lipschitz constant $1/2$. This result appears in [13, Cor. 25.35].

Proposition 3.9 allows us to interpret some algorithms as simple instances of the standard proximal point algorithm (1.15) for convex minimization.

Example 3.11 Let K be a closed convex cone in $\mathcal{H}$ with polar cone $K^\ominus$, let V be a closed vector subspace of $\mathcal{H}$, and set

$$f = \left(\frac{1}{2} d_{K^\ominus}^2 \circ \text{proj}_V\right)^* - \frac{\|\cdot\|^2}{2} \tag{3.17}$$

and $T = \text{proj}_V \circ \text{proj}_K \circ \text{proj}_V$. Then it follows from Proposition 3.9 (see also Remark 3.10(v)) that $T = \text{prox}_f$. Now let $x_0 \in V$ and consider the proximal point iterations $(\forall n \in \mathbb{N})$ $x_{n+1} = \text{prox}_f x_n$. Then the sequence $(x_n)_{n\in\mathbb{N}}$ is identical to that produced by the alternating projection algorithm $(\forall n \in \mathbb{N})$ $x_{n+1} = (\text{proj}_V \circ \text{proj}_K)x_n$. In [67], a specific choice of x_0, K, and V (the latter being a closed hyperplane) lead to a sequence $(x_n)_{n\in\mathbb{N}}$ that was shown to converge weakly but not strongly to the unique point in $K \cap V$, namely 0. In turn, (3.17) is a new example of a function for which the proximal point algorithm converges weakly but not strongly. Alternative constructions can be found in [14, 65].

Example 3.12 Let $(C_i)_{i\in I}$ be a finite family of nonempty closed convex subsets of $\mathcal{H}$. The convex feasibility problem is to find a point in $\bigcap_{i\in I} C_i$. When this problem has no solution, a situation that arises frequently in signal recovery due to inaccurate prior knowledge or measurement errors [38], one must find a surrogate minimization problem. Let us note that the standard method of periodic projections used in consistent problems is of little value here as the limit cycles it generates do not minimize any function [5]. Let $x_0 \in \mathcal{H}$ and $\varepsilon \in \left]0,1\right[$. In [38], it was proposed to minimize $(1/2)\sum_{i\in I}\omega_i d_{C_i}^2$, where $(\omega_i)_{i\in I}$ are in $\left]0,1\right]$ and satisfy $\sum_{i\in I}\omega_i = 1$, via the parallel projection method

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\left(\sum_{i\in I}\omega_i \text{proj}_{C_i} x_n - x_n\right), \quad \text{where} \quad \varepsilon \leqslant \lambda_n \leqslant (2-\varepsilon). \tag{3.18}$$

Now set $f = (\sum_{i\in I}\omega_i(\sigma_{C_i} \,\square\, q))^* - q$ and apply Proposition 3.9 with $(\forall i \in I)$ $\mathcal{G}_i = \mathcal{K}_i = \mathcal{H}$, $L_i = M_i = \text{Id}$, $f_i = \iota_{C_i}$, and $g_i = h_i = \iota_{\{0\}}$. Then $\text{prox}_f = \sum_{i\in I}\omega_i\text{proj}_{C_i}$, and (3.18) therefore turns out to be just a relaxed instance of Martinet's proximal point algorithm (1.15).

As noted in Example 3.5, a composition of proximity operators is usually not a proximity operator. Likewise, the sum of two proximity operators may not be in $\mathcal{N}(\mathcal{H})$ and therefore not in $\mathcal{P}(\mathcal{H})$. The following propositions provide some exceptions. We start with the identity $\text{prox}_{f_1}\circ\text{prox}_{f_2} = \text{prox}_{f_1+f_2}$, which is also discussed in special cases in [46, 47, 52, 121].

Proposition 3.13 *Let T_1 and T_2 be in $\mathcal{P}(\mathcal{H})$, say $T_1 = \operatorname{prox}_{f_1}$ and $T_2 = \operatorname{prox}_{f_2}$ for some f_1 and f_2 in $\Gamma_0(\mathcal{H})$. Suppose that $\operatorname{dom} f_1 \cap \operatorname{dom} f_2 \neq \varnothing$ and that one of the following holds:*

(i) $\mathcal{H} = \mathbb{R}$.

(ii) $(\forall x \in \operatorname{dom} \partial f_2)\ \partial f_2(x) \subset \partial f_2(T_1 x)$.

(iii) $(\forall (x,u) \in \operatorname{gra} \partial f_1)\ \partial f_2(x+u) \subset \partial f_2(x)$.

(iv) $0 \in \operatorname{sri}(\operatorname{dom} f_1 - \operatorname{dom} f_2)$ *and* $(\forall (x,u) \in \operatorname{gra} \partial f_1)\ \partial f_2(x) \subset \partial f_2(x+u)$.

Then $T_1 \circ T_2 \in \mathcal{P}(\mathcal{H})$. More specifically, in cases (ii)–(iv), $T_1 \circ T_2 = \operatorname{prox}_{f_1+f_2}$.

Proof. (i): A function $T \colon \mathbb{R} \to \mathbb{R}$ belongs to $\mathcal{P}(\mathbb{R})$ if and only if it is nonexpansive and increasing [46]. Since the composition of nonexpansive and increasing functions is likewise, we obtain the claim.

(ii)–(iv): [13, Prop. 24.18]. □

Proposition 3.14 *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\phi \in \Gamma_0(\mathbb{R})$ be even, set $\varphi = \phi \circ \|\cdot\| + \sigma_C$, $T_1 = \operatorname{prox}_{\phi\circ\|\cdot\|}$, and $T_2 = \operatorname{prox}_{\sigma_C}$. Then $T_1 \circ T_2 \in \mathcal{P}(\mathcal{H})$. More specifically, $T_1 \circ T_2 = \operatorname{prox}_{\varphi}$.*

Proof. Let $x \in \mathcal{H}$. If ϕ is constant, then $T_1 = \mathrm{Id}$ and the result is trivially true. We therefore assume otherwise, which allows us to derive from [27, Prop. 2.2] that

$$\operatorname{prox}_{\varphi} x = \begin{cases} \dfrac{\operatorname{prox}_{\phi} d_C(x)}{d_C(x)}(x - \operatorname{proj}_C x), & \text{if } d_C(x) > \max \operatorname{Argmin} \phi; \\ x - \operatorname{proj}_C x, & \text{if } d_C(x) \leqslant \max \operatorname{Argmin} \phi. \end{cases} \tag{3.19}$$

For $C = \{0\}$, this yields

$$T_1 x = \begin{cases} \dfrac{\operatorname{prox}_{\phi}\|x\|}{\|x\|} x, & \text{if } \|x\| > \max \operatorname{Argmin} \phi; \\ x, & \text{if } \|x\| \leqslant \max \operatorname{Argmin} \phi. \end{cases} \tag{3.20}$$

Since Theorem 2.1 yields $\mathrm{Id} - \operatorname{proj}_C = \operatorname{prox}_{\sigma_C}$, using (3.19) and (3.20), we get

$$\begin{aligned} \operatorname{prox}_{\varphi} x &= \begin{cases} \dfrac{\operatorname{prox}_{\phi}\|\operatorname{prox}_{\sigma_C} x\|}{\|\operatorname{prox}_{\sigma_C} x\|} \operatorname{prox}_{\sigma_C} x, & \text{if } \|\operatorname{prox}_{\sigma_C} x\| > \max \operatorname{Argmin} \phi; \\ \operatorname{prox}_{\sigma_C} x, & \text{if } \|\operatorname{prox}_{\sigma_C} x\| \leqslant \max \operatorname{Argmin} \phi \end{cases} \\ &= T_1(T_2 x). \end{aligned} \tag{3.21}$$

□

Remark 3.15 Proposition 3.14 has important applications.

(i) It follows from [46, Prop. 3.2(v)] that, if ϕ is differentiable at 0 with $\phi'(0) = 0$, then $\operatorname{prox}_{\varphi}$, which can be implemented explicitly via (3.19), is a proximal thresholder on C: $(\forall x \in \mathcal{H})\ \operatorname{prox}_{\varphi} x = 0 \Leftrightarrow x \in C$.

(ii) Let $\gamma \in]0,+\infty[$, let K be a nonempty closed convex cone in $\mathcal{H}$, let C be the polar cone of K, and let $x \in \mathcal{H}$. Upon setting $\phi = \gamma|\cdot|$ in Proposition 3.14 and using [13, Examp. 24.20], we obtain (see [44, Lemma 2.2] for a different derivation)

$$\operatorname{prox}_{\gamma\|\cdot\|+\iota_K} x = \big(\operatorname{prox}_{\gamma\|\cdot\|} \circ \operatorname{proj}_K\big)x = \begin{cases} \dfrac{\|\operatorname{proj}_K x\| - \gamma}{\|\operatorname{proj}_K x\|}\operatorname{proj}_K x, & \text{if } \|\operatorname{proj}_K x\| > \gamma; \\ 0, & \text{if } \|\operatorname{proj}_K x\| \leqslant \gamma. \end{cases} \tag{3.22}$$

On the other hand, setting $\phi = \iota_{[-\gamma,\gamma]}$ in Proposition 3.14 and using [13, Examp. 3.18], we obtain (see [9, Sec. 7] for different derivations)

$$\operatorname{proj}_{B(0;\gamma)\cap K} x = \big(\operatorname{proj}_{B(0;\gamma)} \circ \operatorname{proj}_K\big)x = \begin{cases} \dfrac{\gamma}{\|\operatorname{proj}_K x\|}\operatorname{proj}_K x, & \text{if } \|\operatorname{proj}_K x\| > \gamma; \\ \operatorname{proj}_K x, & \text{if } \|\operatorname{proj}_K x\| \leqslant \gamma. \end{cases} \tag{3.23}$$

Proposition 3.16 *Set $q = \|\cdot\|^2/2$, and let T_1 and T_2 be in $\mathcal{P}(\mathcal{H})$, say $T_1 = \operatorname{prox}_{f_1}$ and $T_2 = \operatorname{prox}_{f_2}$ for some f_1 and f_2 in $\Gamma_0(\mathcal{H})$. Suppose that $0 \in \operatorname{sri}(\operatorname{dom} f_1^* - \operatorname{dom} f_2^*)$ and that*

$$(f_1^* + f_2^*) \,\square\, q = f_1^* \,\square\, q + f_2^* \,\square\, q. \tag{3.24}$$

Then $T_1 + T_2 \in \mathcal{P}(\mathcal{H})$. More specifically, $T_1 + T_2 = \operatorname{prox}_{f_1 \,\square\, f_2}$.

Proof. It follows from [13, Prop. 15.7(i)] that $f_1 \,\square\, f_2 \in \Gamma_0(\mathcal{H})$. In addition, we derive from Theorem 2.1, (3.24), and [13, Prop. 13.24(i)] that $T_1+T_2 = \nabla\big(f_1^* \,\square\, q\big) + \nabla\big(f_2^* \,\square\, q\big) = \nabla\big(f_1^* \,\square\, q + f_2^* \,\square\, q\big) = \nabla\big((f_1^* + f_2^*) \,\square\, q\big) = \nabla\big((f_1 \,\square\, f_2)^* \,\square\, q\big) = \operatorname{prox}_{f_1 \,\square\, f_2}$, as claimed. ☐

Remark 3.17 Let C_1 and C_2 be nonempty closed convex subsets of $\mathcal{H}$, and set $f_1 = \iota_{C_1}$ and $f_2 = \iota_{C_2}$. Then the conclusion of Proposition 3.16 is that $\operatorname{proj}_{C_1} + \operatorname{proj}_{C_2} = \operatorname{proj}_{C_1+C_2}$. This property is discussed in [125] (for cones), in [13, Prop. 29.6], and in the recently posted paper [10].

3.4 Self-dual classes of firmly nonexpansive operators

Let us call a subclass $\mathcal{T}(\mathcal{H})$ of $\mathcal{J}(\mathcal{H})$ *self-dual* if $(\forall T \in \mathcal{T}(\mathcal{H}))\ \mathrm{Id} - T \in \mathcal{T}(\mathcal{H})$. This property plays an important role in our paper.

It is clear from (1.7) that $\mathcal{J}(\mathcal{H})$ is self-dual. This can also be recovered from Theorem 1.3 and (2.4). As seen in Proposition 3.7(i), $\mathcal{P}(\mathcal{H})$ is also self-dual. Now let $T \in \mathcal{K}(\mathcal{H})$. Then there exists a nonempty closed convex cone $K \subset \mathcal{H}$ such that $T = \operatorname{proj}_K$ and Moreau's conical decomposition expresses the projector onto the polar cone $K^{\ominus}$ as $\operatorname{proj}_{K^{\ominus}} = \mathrm{Id} - \operatorname{proj}_K$ [88]. This shows that $\mathcal{K}(\mathcal{H})$ is self-dual. Likewise, it follows from the standard Beppo Levi orthogonal decomposition of $\mathcal{H}$ [78] that the class $\mathcal{V}(\mathcal{H})$ of projectors onto closed vector subspaces of $\mathcal{H}$ is self-dual. We thus obtain the nested self-dual classes

$$\mathcal{V}(\mathcal{H}) \subset \mathcal{K}(\mathcal{H}) \subset \mathcal{P}(\mathcal{H}) \subset \mathcal{J}(\mathcal{H}). \tag{3.25}$$

Self-duality properties were investigated in [15], where other classes were identified and studied in depth. In particular, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Then A is paramonotone if and only if A^{-1} is [13, Prop. 22.2(i)]. As a result, it follows from (2.4) that the class $\mathcal{J}_{\mathrm{para}}(\mathcal{H})$ of resolvents of paramonotone maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ is self-dual [15] and since subdifferentials are paramonotone, we have $\mathcal{P}(\mathcal{H}) \subset \mathcal{J}_{\mathrm{para}}(\mathcal{H}) \subset \mathcal{J}(\mathcal{H})$. Likewise, since A is 3^* monotone if and only if A^{-1} is [13, Prop. 25.19(i)], and since subdifferentials are 3^* monotone, the class $\mathcal{J}_{3^*}(\mathcal{H})$ of resolvents of 3^* monotone maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ is self-dual and satisfies $\mathcal{P}(\mathcal{H}) \subset \mathcal{J}_{3^*}(\mathcal{H}) \subset \mathcal{J}(\mathcal{H})$.

Although our primary objective in Section 3.3 was to investigate transformations on the class $\mathcal{P}(\mathcal{H})$, similar questions could be asked about other self-dual classes. In this spirit, Zarantonello [125] has studied some transformations in $\mathcal{K}(\mathcal{H})$. Let T_1 and T_2 be in $\mathcal{K}(\mathcal{H})$, say $T_1 = \operatorname{proj}_{K_1}$ and $T_2 = \operatorname{proj}_{K_2}$. In connection with Proposition 3.13, he has shown that $\{T_1 \circ T_2, T_2 \circ T_1\} \subset \mathcal{K}(\mathcal{H})$ if and only if $T_2 \circ T_1 = T_1 \circ T_2$, in which case $T_1 \circ T_2 = \operatorname{proj}_{K_1 \cap K_2}$ [124]. On the other hand, in this context, the conclusion of Proposition 3.16, which states that $T_1 + T_2 = \operatorname{proj}_{K_1 + K_2} \in \mathcal{K}(\mathcal{H})$, is discussed in [125].

The proximity-preserving transformations studied in Section 3.3 have natural resolvent-preserving counterparts. For instance, mimicking the pattern of Remark 3.10(vii) and using [13, Thm. 25.3], one shows that, if $T = J_A \in \mathcal{J}(\mathcal{H})$, $B \in \mathcal{M}(\mathcal{H})$, $M \in \mathcal{B}(\mathcal{G},\mathcal{H}) \smallsetminus \{0\}$, and $C \in \mathcal{M}(\mathcal{G})$, then $T \,\square\, (B \,\square\, (M \triangleright C)) = J_{A+B^{-1}+M\circ C^{-1}\circ M^*} \in \mathcal{J}(\mathcal{H})$ provided that the cones generated by $\operatorname{dom} C^{-1} - M^*(\operatorname{dom} A \cap \operatorname{dom} B^{-1})$ and by $\operatorname{dom} A - \operatorname{dom} B^{-1}$ are closed vector subspaces.

4 Monotone operators in convex optimization

In this section we present several examples of maximally monotone operators which are not subdifferentials and which play fundamental and indispensable roles in the analysis and the numerical solution of convex optimization problems. We preface these examples with a brief overview of classical splitting methods [13] which depend less critically on monotone operator theory.

4.1 The interplay between splitting methods for convex optimization and monotone inclusion problems

The proximal point algorithm (1.15) was first developed for convex optimization. It was extended in [106] to solve the inclusion problem (1.9) for an operator $A \in \mathcal{M}(\mathcal{H})$ such that $\operatorname{zer} A \neq \varnothing$ via the iteration

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = J_{\gamma_n A} x_n, \quad \text{where} \quad \gamma_n \in \left]0, +\infty\right[. \tag{4.1}$$

However, the algorithmic theory for the case of monotone inclusions does not subsume that for the case of convex optimization. Thus, as shown in [26], the weak convergence of $(x_n)_{n\in\mathbb{N}}$ to a point in A holds when $\sum_{n\in\mathbb{N}} \gamma_n^2 = +\infty$, and this condition can be weakened to $\sum_{n\in\mathbb{N}} \gamma_n = +\infty$ if $A \in \mathcal{S}(\mathcal{H})$ (see also [65] for finer properties in the subdifferential case). This is explained by the fact that, given $z \in \operatorname{zer} A$, (4.1) and Theorem 1.3 yield

$$(\forall n \in \mathbb{N}) \quad \|x_{n+1} - z\|^2 \leqslant \|x_n - z\|^2 - \|J_{\gamma_n A} x_n - x_n\|^2 \tag{4.2}$$

for a general $A \in \mathcal{M}(\mathcal{H})$ while, when $A = \partial f$ for some $f \in \Gamma_0(\mathcal{H})$, it can be sharpened to

$$(\forall n \in \mathbb{N}) \quad \|x_{n+1} - z\|^2 \leqslant \|x_n - z\|^2 - \|J_{\gamma_n A} x_n - x_n\|^2 - 2\gamma_n\big(f(x_{n+1}) - \inf f(\mathcal{H})\big). \tag{4.3}$$

Going back to the discussion of Section 3, this sheds a different light on the differences between $\mathcal{S}(\mathcal{H})$ and $\mathcal{M}(\mathcal{H})$. Naturally, the applicability of (4.1) depends on the ease of implementation of the resolvents $(J_{A_n})_{n\in\mathbb{N}}$. A more structured inclusion problem is the following.

Problem 4.1 Let $A \in \mathcal{M}(\mathcal{H})$ and $B \in \mathcal{M}(\mathcal{H})$ be such that $0 \in \operatorname{ran}(A + B)$. Find a zero of $A + B$.

Generally speaking, when replacing monotone operators by subdifferentials in certain inclusion problems, one recovers a convex minimization problem provided some constraint qualification holds [13]. In this regard, we shall also consider the following convex optimization problem.

Problem 4.2 Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that $0 \in \operatorname{ran}(\partial f + \partial g)$. Find a minimizer of $f+g$ over $\mathcal{H}$.

There are three classical methods for solving Problem 4.1, which we present here in simple forms (see [28, 34, 53] and the references therein for refinements). All three methods produce a sequence $(x_n)_{n\in\mathbb{N}}$ which converges weakly to a zero of $A+B$ [13, 115, 116], but they involve different assumptions on B. Let us stress that the importance of these three splitting methods is not only historical: many seemingly different splitting methods are just, explicitly or implicitly, reformulations of these basic schemes in alternate settings (e.g., product spaces, dual spaces, primal-dual spaces, renormed spaces, or a combination thereof); see [3, 4, 28, 40, 41, 42, 49, 51, 54, 58, 59, 76, 107, 116, 120] and the references therein for specific examples.

- **Forward-backward splitting.** In Problem 4.1, suppose that $B\colon \mathcal{H} \to \mathcal{H}$ and that $\beta^{-1}B \in \mathfrak{F}(\mathcal{H})$ for some $\beta \in \left]0,+\infty\right[$. Let $x_0 \in \mathcal{H}$ and $\varepsilon \in \left]0, 2/(\beta+1)\right[$, and iterate

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} \varepsilon \leqslant \gamma_n \leqslant (2-\varepsilon)/\beta \\ y_n = x_n - \gamma_n B x_n \\ x_{n+1} = J_{\gamma_n A} y_n. \end{array}\right.\end{array} \tag{4.4}$$

  In view of Theorem 2.2, the assumptions on B translate into the fact that, in Problem 4.2, $f_2\colon \mathcal{H} \to \mathbb{R}$ is differentiable and that $B = \nabla f_2$ is β-Lipschitzian, while $A = \partial f_1$.

- **Tseng's forward-backward-forward splitting.** In Problem 4.1, suppose that $B\colon \mathcal{H} \to \mathcal{H}$ is β-Lipschitzian for some $\beta \in \left]0,+\infty\right[$. Let $x_0 \in \mathcal{H}$ and $\varepsilon \in \left]0, 1/(\beta+1)\right[$, and iterate

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} \varepsilon \leqslant \gamma_n \leqslant (1-\varepsilon)/\beta \\ y_n = x_n - \gamma_n B x_n \\ p_n = J_{\gamma_n A} y_n \\ q_n = p_n - \gamma_n B p_n \\ x_{n+1} = x_n - y_n + q_n. \end{array}\right.\end{array} \tag{4.5}$$

- **Douglas-Rachford splitting.** Let $x_0 \in \mathcal{H}$ and $\gamma \in \left]0,+\infty\right[$, and iterate

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} x_n = J_{\gamma B} y_n \\ z_n = J_{\gamma A}(2x_n - y_n) \\ y_{n+1} = y_n + z_n - x_n. \end{array}\right.\end{array} \tag{4.6}$$

Historically, the forward-backward method grew out of the projected gradient method in convex optimization [79], and the first version for Problem 4.1 was proposed in [84]. Another example of a monotone operator splitting method that evolved from convex optimization is Dykstra's method [11], which was first devised for indicator functions in [24]. By contrast, the forward-backward-forward [116] and Douglas-Rachford [80] methods were developed directly for Problem 4.1, and then specialized to Problem 4.2. In principle, however, even though monotone inclusions provide a more synthetic and natural framework, it is possible (at least *a posteriori*) to derive their convergence in the scenario of Problem 4.2 from optimization concepts only, without invoking monotone operator theory. Nonetheless, non-subdifferential maximally monotone operators may still be at play. For instance, note that in the Douglas-Rachford algorithm (4.6), we have

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = Ty_n, \quad \text{where} \quad T = \frac{(2J_{\gamma A} - \mathrm{Id}\,) \circ (2J_{\gamma B} - \mathrm{Id}\,) + \mathrm{Id}}{2}. \tag{4.7}$$

Upon invoking (3.4) and Theorem 1.3, we see that $T \in \mathcal{J}(\mathcal{H})$ and that there exists $C \in \mathcal{M}(\mathcal{H})$ such that $T = J_C$, namely $C = T^{-1} - \mathrm{Id}$. Hence

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = J_C y_n, \quad \text{where} \quad C = T^{-1} - \mathrm{Id}\ . \tag{4.8}$$

In other words, $(y_n)_{n\in\mathbb{N}}$ is produced by an instance of the proximal point algorithm (4.1) and, in this sense, the dynamics of the Douglas-Rachford algorithm are implicitly governed by a maximally monotone operator (as seen in Example 3.5, this operator is typically not in $\mathcal{S}(\mathcal{H})$, even if A and B are). This observation, which was made in [58], has actually a much more general scope. Indeed, as shown in [39], several operator splitting algorithms are driven by successive approximations of an averaged operator $T\colon \mathcal{H} \to \mathcal{H}$, i.e., an operator of the form $T = (1-\alpha)\,\mathrm{Id} + \alpha R$, where $R\colon \mathcal{H} \to \mathcal{H}$ is nonexpansive and $\alpha \in \left]0,1\right[$ (further examples are found in more recent papers such as [55] and [100]). We derive from Theorem 1.3 that there exists $C \in \mathcal{M}(\mathcal{H})$ (and $C \notin \mathcal{S}(\mathcal{H})$ in general) such that $R = 2J_C - \mathrm{Id}$, namely $C = ((R + \mathrm{Id}\,)/2)^{-1} - \mathrm{Id}$. Therefore, $T = \mathrm{Id} + 2\alpha(J_C - \mathrm{Id}\,)$. In turn, a sequence $(x_n)_{n\in\mathbb{N}}$ produced by the successive approximations of T is generated implicitly by the relaxed resolvent iteration

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda(J_C x_n - x_n), \quad \text{where} \quad \lambda = 2\alpha \quad \text{and} \quad C = \left(\mathrm{Id} + \frac{1}{2\alpha}(T - \mathrm{Id}\,)\right)^{-1} - \mathrm{Id}\ . \tag{4.9}$$

For example, let us consider the forward-backward algorithm (4.4) with a fixed proximal parameter $\gamma \in \left]0, 2/\beta\right[$. Then

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n \quad \text{where} \quad T = J_{\gamma A} \circ (\mathrm{Id} - \gamma B). \tag{4.10}$$

Furthermore, T is averaged with constant $\alpha = 2/(4-\beta\gamma)$ [53]. Altogether, the forward-backward iteration (4.10) is an instance of the relaxed proximal point algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda(J_C x_n - x_n), \quad \text{where} \quad \lambda = \frac{4}{4-\beta\gamma}$$
$$\text{and} \quad C = \left(\frac{\beta\gamma}{4}\,\mathrm{Id} + \frac{4-\beta\gamma}{4}\big(J_{\gamma A} \circ (\mathrm{Id} - \gamma B)\big)\right)^{-1} - \mathrm{Id}\ . \tag{4.11}$$

4.2 Rockafellar's saddle function operator

The following result is due to Rockafellar [103, 104] (he actually used a somewhat more general notion of closedness, made precise in these papers, for the function $\mathcal{L}$).

Theorem 4.3 (Rockafellar) *Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be real Hilbert spaces, let $\mathcal{L}\colon \mathcal{H}_1 \oplus \mathcal{H}_2 \to [-\infty, +\infty]$ be such that, for every $x_1 \in \mathcal{H}_1$ and every $x_2 \in \mathcal{H}_2$, $-\mathcal{L}(x_1, \cdot) \in \Gamma_0(\mathcal{H}_2)$ and $\mathcal{L}(\cdot, x_2) \in \Gamma_0(\mathcal{H}_1)$. Set*

$$(\forall x_1 \in \mathcal{H}_1)(\forall x_2 \in \mathcal{H}_2) \quad \boldsymbol{A}(x_1, x_2) = \partial\mathcal{L}(\cdot, x_2)(x_1) \times \partial(-\mathcal{L}(x_1, \cdot))(x_2). \tag{4.12}$$

Then $\boldsymbol{A} \in \mathcal{M}(\mathcal{H}_1 \oplus \mathcal{H}_2)$ and

$$\operatorname{zer} \boldsymbol{A} = \big\{(x_1, x_2) \in \mathcal{H}_1 \oplus \mathcal{H}_2 \;\big|\; \mathcal{L}(x_1, x_2) = \inf \mathcal{L}(\mathcal{H}_1, x_2) = \sup \mathcal{L}(x_1, \mathcal{H}_2)\big\} \tag{4.13}$$

is the set of saddle points of $\mathcal{L}$.

A geometrical interpretation of (4.12) is that $(u_1,u_2)\in\boldsymbol{A}(x_1,x_2)$ if and only if (x_1,x_2) is a saddle point of the convex-concave function $(x_1',x_2')\mapsto\mathcal{L}(x_1',x_2')-\langle x_1'\mid u_1\rangle+\langle x_2'\mid u_2\rangle$. The maximally monotone operator $\boldsymbol{A}$ of (4.12) is deeply rooted in convex optimization due to the foundational role it plays in Lagrangian theory and duality schemes [13, 103, 105, 107]. Yet, as the following example shows, it is not a subdifferential.

Example 4.4 In Theorem 4.3, set $\mathcal{H}_1=\mathcal{H}_2=\mathbb{R}$ and $\mathcal{L}\colon(x_1,x_2)\mapsto x_1^2-x_1x_2$. Then (4.12) yields $\boldsymbol{A}\colon(x_1,x_2)\mapsto(2x_1-x_2,x_1)$. Thus, $\boldsymbol{A}\in\mathcal{B}(\mathbb{R}^2)$ is positive and not self-adjoint. It follows from Lemma 2.3 that $\boldsymbol{A}\in\mathcal{M}(\mathcal{H})\smallsetminus\mathcal{S}(\mathcal{H})$.

The idea of using the proximal point algorithm (4.1) with the operator $\boldsymbol{A}$ of (4.12) to find a saddle point of $\mathcal{L}$ was proposed by Rockafellar in [106]. In [107], he applied it to the concrete problem of minimizing a convex function subject to convex inequality constraints, using the ordinary Lagrangian as a saddle function. The resulting algorithm is known as the proximal method of multipliers.

4.3 Spingarn's partial inverse operator

Let $A\in\mathcal{M}(\mathcal{H})$, let V be a closed vector subspace of $\mathcal{H}$, and let the partial inverse of A with respect to V be the operator $A_V\in\mathcal{M}(\mathcal{H})$ defined in (3.5). As discussed in [111], problems of the form

$$\text{find}\;\; x\in V \;\;\text{and}\;\; u\in V^{\perp} \;\;\text{such that}\;\; u\in Ax \tag{4.14}$$

can be solved by applying the proximal point algorithm (4.1) to A_V; this method is known as the method of partial inverses, and it has strong connections with the Douglas-Rachford algorithm [58, 74, 81]. For instance, if $A=\partial f$ for some $f\in\Gamma_0(\mathcal{H})$ such that f admits a minimizer over V and $0\in\operatorname{sri}(V-\operatorname{dom}f)$, (4.14) reduces to finding a solution of the Fenchel dual pair

$$\underset{x\in V}{\text{minimize}}\;\; f(x) \quad\text{and}\quad \underset{u\in V^{\perp}}{\text{minimize}}\;\; f^*(u). \tag{4.15}$$

In this case, given $x_0\in V$ and $u_0\in V^{\perp}$, the method of partial inverses iterates

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l} y_n=\operatorname{prox}_f(x_n+u_n)\\ v_n=x_n+u_n-y_n\\ (x_{n+1},u_{n+1})=(\operatorname{proj}_V y_n,\operatorname{proj}_{V^{\perp}}v_n),\end{array}\right.\end{array} \tag{4.16}$$

and the sequences $(x_n)_{n\in\mathbb{N}}$ and $(u_n)_{n\in\mathbb{N}}$ converge weakly to a solution to the primal and dual problems in (4.15) [13, Prop. 28.2]. This algorithm has many applications in convex optimization, e.g., [68, 75, 76, 111, 112, 113]. It also constitutes the basic building block of the progressive hedging algorithm in stochastic programming [108]. Thus, despite its apparent simplicity, this partial inverse approach is quite powerful and it can tackle the following primal-dual problem.

Problem 4.5 Let I be a nonempty finite set, and let $(\mathcal{H}_i)_{i\in I}$ and $\mathcal{G}$ be real Hilbert spaces. Let $r\in\mathcal{G}$, let $g\in\Gamma_0(\mathcal{G})$, and, for every $i\in I$, let $z_i\in\mathcal{H}_i$, let $f_i\in\Gamma_0(\mathcal{H}_i)$, and let $L_i\in\mathcal{B}(\mathcal{H}_i,\mathcal{G})$. Solve the primal problem

$$\underset{(\forall i\in I)\; x_i\in\mathcal{H}_i}{\text{minimize}}\;\; \sum_{i\in I}\big(f_i(x_i)-\langle x_i\mid z_i\rangle\big)+g\Bigg(\sum_{i\in I}L_ix_i-r\Bigg), \tag{4.17}$$

together with the dual problem

$$\underset{v\in\mathcal{G}}{\text{minimize}}\;\; \sum_{i\in I}f_i^*(z_i-L_i^*v)+g^*(v)+\langle v\mid r\rangle. \tag{4.18}$$

It is shown in [1] that, when applied to a version of (4.14) suitably reformulated in a product space, (4.16) yields a proximal splitting algorithm that solves Problem 4.5 and employs the operators $(\mathrm{prox}_{f_i})_{i\in I}$, prox_g, $(L_i)_{i\in I}$, and $(L_i^*)_{i\in I}$ separately.

The backbone of all the above-mentioned applications of the method of partial inverses to convex optimization is the partial inverse of an operator in $\mathcal{S}(\mathcal{H})$. As seen in Example 3.6, this maximally monotone operator is not in $\mathcal{S}(\mathcal{H})$ in general.

4.4 Primal-dual algorithm for mixed composite minimization

We re-examine through the lens of the maximally monotone saddle function operator (4.12) a mixed composite minimization problem proposed and studied in [49] with different tools.

Problem 4.6 Let $f \in \Gamma_0(\mathcal{H})$, let $h\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable with a μ-Lipschitzian gradient for some $\mu \in \left]0,+\infty\right[$, let $\mathcal{G}$ be a real Hilbert space, let $g \in \Gamma_0(\mathcal{G})$, and let $\ell \in \Gamma_0(\mathcal{G})$ be $1/\nu$-strongly convex for some $\nu \in \left]0,+\infty\right[$. Suppose that $0 \neq L \in \mathcal{B}(\mathcal{H},\mathcal{G})$ and that

$$0 \in \mathrm{ran}\,\big(\partial f + L^* \circ (\partial g \,\Box\, \partial \ell) \circ L + \nabla h\big). \tag{4.19}$$

Consider the problem

$$\underset{x\in\mathcal{H}}{\mathrm{minimize}}\;\; f(x) + (g \,\Box\, \ell)(Lx) + h(x), \tag{4.20}$$

and the dual problem

$$\underset{v\in\mathcal{G}}{\mathrm{minimize}}\;\; \big(f^* \,\Box\, h^*\big)(-L^*v) + g^*(v) + \ell^*(v). \tag{4.21}$$

From a numerical standpoint, solving (4.20) is challenging as it involves five objects (four functions, three of which are nonsmooth, and a linear operator), while traditional proximal splitting techniques are limited to two objects; see (4.4)–(4.6). In [49], Problem 4.6 was analyzed and solved as an instance of a more general primal-dual inclusion problem involving monotone operators, which was reformulated as that of finding a zero of the sum of two operators in $\mathcal{M}(\mathcal{H}\oplus\mathcal{G})$. Let us stress that, even in the special case of Problem 4.6, this inclusion problem still involves operators which are not subdifferentials. To see this, we now propose an alternative derivation of the results of [49, Sect. 4] using the saddle function formalism of Theorem 4.3. Following the same pattern as in [105, Examp. 11] (with the conventions of [13, Prop. 19.20]), we define the Lagrangian of Problem 4.6 as

$$\begin{aligned}
&\mathcal{L}\colon \mathcal{H}\oplus\mathcal{G} \to [-\infty,+\infty]\\
&\qquad (x,v) \mapsto \begin{cases} -\infty, & \text{if } x \in \mathrm{dom}\, f \text{ and } v \notin \mathrm{dom}\, g^* \cap \mathrm{dom}\, \ell^*;\\ f(x) + h(x) + \langle Lx \mid v\rangle - g^*(v) - \ell^*(v), & \text{if } x \in \mathrm{dom}\, f \text{ and } v \in \mathrm{dom}\, g^* \cap \mathrm{dom}\, \ell^*;\\ +\infty, & \text{if } x \notin \mathrm{dom}\, f,\end{cases}
\end{aligned} \tag{4.22}$$

and observe that it satisfies the assumptions of Theorem 4.3. In turn, using standard subdifferential calculus [13], we deduce that the associated maximally monotone operator $\boldsymbol{A}$ of (4.12) is

$$\begin{aligned}
\boldsymbol{A}\colon \mathcal{H}\oplus\mathcal{G} \to 2^{\mathcal{H}\oplus\mathcal{G}}\colon (x,v) &\mapsto \partial\mathcal{L}(\cdot,v)(x) \times \partial(-\mathcal{L}(x,\cdot))(v)\\
&= \big(\partial f(x) + \nabla h(x) + L^*v, \partial g^*(v) + \nabla \ell^*(v) - Lx\big).
\end{aligned} \tag{4.23}$$

It is noteworthy that this operator admits a Borwein-Wiersma decomposition (3.3), namely

$$\boldsymbol{A} = \partial\boldsymbol{\varphi} + \boldsymbol{S}, \quad \text{where} \quad \begin{cases} \boldsymbol{\varphi} \colon \mathcal{H} \oplus \mathcal{G} \to \left]-\infty, +\infty\right] : (x, v) \mapsto f(x) + h(x) + g^*(v) + \ell^*(v) \\ \boldsymbol{S} \colon \mathcal{H} \oplus \mathcal{G} \to \mathcal{H} \oplus \mathcal{G} \colon (x, v) \mapsto (L^* v, -Lx). \end{cases} \tag{4.24}$$

Here $\boldsymbol{\varphi} \in \Gamma_0(\mathcal{H} \oplus \mathcal{G})$ and $\boldsymbol{S} \in \mathcal{B}(\mathcal{H} \oplus \mathcal{G})$ is nonzero and skew, which shows that $\boldsymbol{A} \notin \mathcal{S}(\mathcal{H} \oplus \mathcal{G})$ by virtue of Lemma 2.3. By Theorem 4.3, a zero (x, v) of $\boldsymbol{A}$ is a saddle point of $\mathcal{L}$, which implies that (x, v) solves Problem 4.6, i.e., x solves (4.20) and v solves (4.21). However, the decomposition (4.24) does not lend itself easily to splitting methods as they would require computing $J_{\partial\boldsymbol{\varphi}} = \operatorname{prox}_{\boldsymbol{\varphi}} = \operatorname{prox}_{f+h} \times \operatorname{prox}_{g^*+\ell^*}$, which does not admit a closed form expression in general. A more judicious decomposition of $\boldsymbol{A}$ is

$$\boldsymbol{A} = \partial\boldsymbol{f} + \boldsymbol{B}, \quad \text{where} \quad \begin{cases} \boldsymbol{f} \colon \mathcal{H} \oplus \mathcal{G} \to \left]-\infty, +\infty\right] : (x, v) \mapsto f(x) + g^*(v) \\ \boldsymbol{B} \colon \mathcal{H} \oplus \mathcal{G} \to \mathcal{H} \oplus \mathcal{G} \colon (x, v) \mapsto (\nabla h(x) + L^* v, \nabla \ell^*(v) - Lx). \end{cases} \tag{4.25}$$

Note that $\boldsymbol{f} \in \Gamma_0(\mathcal{H} \oplus \mathcal{G})$ and that computing $J_{\partial\boldsymbol{f}} = \operatorname{prox}_{\boldsymbol{f}} = \operatorname{prox}_f \times \operatorname{prox}_{g^*}$ requires only the ability to compute prox_f and $\operatorname{prox}_{g^*} = \operatorname{Id} - \operatorname{prox}_g$. Furthermore [49],

$$\boldsymbol{B} \in \mathcal{M}(\mathcal{H} \oplus \mathcal{G}) \text{ is monotone and } \beta\text{-Lipschitzian with } \beta = \max\{\mu, \nu\} + \|L\|. \tag{4.26}$$

This structure makes the task of finding a zero of $\boldsymbol{A}$ amenable to the forward-backward-forward algorithm (4.5), which requires one evaluation of $\operatorname{prox}_{\gamma\boldsymbol{f}}$ and two evaluations of $\boldsymbol{B}$ at each iteration. As seen in Section 4.1, given $\varepsilon \in \left]0, 1/(\beta+1)\right[$, the forward-backward-forward algorithm constructs a sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ which converges weakly to a point in $\operatorname{zer} \boldsymbol{A}$ via the recursion

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \varepsilon \leqslant \gamma_n \leqslant (1-\varepsilon)/\beta \\ \boldsymbol{y}_n = \boldsymbol{x}_n - \gamma_n \boldsymbol{B}\boldsymbol{x}_n \\ \boldsymbol{p}_n = \operatorname{prox}_{\gamma_n \boldsymbol{f}} \boldsymbol{y}_n \\ \boldsymbol{q}_n = \boldsymbol{p}_n - \gamma_n \boldsymbol{B}\boldsymbol{p}_n \\ \boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \boldsymbol{y}_n + \boldsymbol{q}_n. \end{array} \right. \end{array} \tag{4.27}$$

Now set $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = (x_n, v_n)$, $\boldsymbol{y}_n = (y_{1,n}, y_{2,n})$, $\boldsymbol{p}_n = (p_{1,n}, p_{2,n})$, and $\boldsymbol{q}_n = (q_{1,n}, q_{2,n})$. Then, in view of (4.25), (4.27) assumes the form of the primal-dual method of [49, Sect. 4], namely

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \varepsilon \leqslant \gamma_n \leqslant (1-\varepsilon)/\beta \\ y_{1,n} = x_n - \gamma_n\big(\nabla h(x_n) + L^* v_n\big) \\ y_{2,n} = v_n + \gamma_n\big(Lx_n - \nabla \ell^*(v_n)\big) \\ p_{1,n} = \operatorname{prox}_{\gamma_n f} y_{1,n} \\ p_{2,n} = \operatorname{prox}_{\gamma_n g^*} y_{2,n} \\ q_{1,n} = p_{1,n} - \gamma_n\big(\nabla h(p_{1,n}) + L^* p_{2,n}\big) \\ q_{2,n} = p_{2,n} + \gamma_n\big(Lp_{1,n} - \nabla \ell^*(p_{2,n})\big) \\ x_{n+1} = x_n - y_{1,n} + q_{1,n} \\ v_{n+1} = v_n - y_{2,n} + q_{2,n}. \end{array} \right. \end{array} \tag{4.28}$$

We conclude that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a solution x to (4.20) and that $(v_n)_{n\in\mathbb{N}}$ converges weakly to a solution v to (4.21).

Remark 4.7 Let us make a few observations regarding Problem 4.6 and the iterative method (4.28).

(i) Algorithm (4.28) achieves full splitting of the functions and of the linear operators. In addition, all the smooth functions are activated via explicit gradient steps, while the nonsmooth ones are activated via their proximity operator.

(ii) The special case when $\ell = \iota_{\{0\}}$ and $h = 0$ leads to the monotone+skew decomposition approach of [28]. As discussed in [28, Rem. 2.9], in this case the use Douglas-Rachford algorithm (4.6) can also be contemplated since the resolvents $J_{\gamma\partial \boldsymbol{f}}$ and $J_{\gamma \boldsymbol{S}}$ of the operators in (4.24) have explicit forms.

(iii) In [120], Problem 4.6 is also written as that of finding a zero of $\boldsymbol{A}$ in (4.25). However, it is then reformulated in a new Hilbert space obtained by suitably renorming $\mathcal{H} \times \mathcal{G}$. This formulation yields an equivalent inclusion problem for an operator which can be decomposed as the sum of two maximally monotone operators amenable to forward-backward splitting (see Problem 4.1 and (4.4)) and, *in fine*, an algorithm which also achieves full splitting (see [41, 54, 66] for related work). A special case of this framework is the algorithm proposed in [35].

(iv) The construction of algorithm (4.28) revolves around the problem of finding a zero of the operator $\boldsymbol{A} \in \mathcal{M}(\mathcal{H}\oplus\mathcal{G}) \smallsetminus \mathcal{S}(\mathcal{H}\oplus\mathcal{G})$ of (4.25). It is not clear how this, or any of the splitting algorithms mentioned in (i)–(iii), could have been devised using only subdifferential tools.

4.5 Lagrangian formulations of composite problems

We consider a special case of Problem 4.6 which corresponds to the standard Fenchel-Rockafellar duality framework.

Problem 4.8 Let $f \in \Gamma_0(\mathcal{H})$, let $\mathcal{G}$ be a real Hilbert space, and let $g \in \Gamma_0(\mathcal{G})$. Suppose that $0 \neq L \in \mathcal{B}(\mathcal{H},\mathcal{G})$ and that $0 \in \operatorname{ran}(\partial f + L^* \circ \partial g \circ L)$. The objective is to solve the primal problem

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + g(Lx) \tag{4.29}$$

as well as the dual problem

$$\underset{v\in\mathcal{G}}{\text{minimize}} \;\; f^*(-L^*v) + g^*(v). \tag{4.30}$$

We have already discussed in Remark 4.7(ii) monotone operator-based algorithms to solve (4.29)–(4.30). Alternatively, set $\boldsymbol{\mathcal{H}} = \mathcal{H} \oplus \mathcal{G}$, $\boldsymbol{f} \colon \boldsymbol{\mathcal{H}} \to \left]-\infty,+\infty\right] : (x,y) \mapsto f(x) + g(y)$, and $\boldsymbol{L} \colon \boldsymbol{\mathcal{H}} \to \mathcal{G} \colon (x,y) \mapsto Lx - y$. Then (4.29) is equivalent to minimizing $\boldsymbol{f}$ over $\ker \boldsymbol{L}$. The Lagrangian for this type of problem is $\mathcal{L} \colon \boldsymbol{\mathcal{H}} \oplus \mathcal{G} \to \left]-\infty,+\infty\right] : (\boldsymbol{x}, v) \mapsto \boldsymbol{f}(\boldsymbol{x}) + \langle \boldsymbol{L}\boldsymbol{x} \mid v\rangle$ [105, Examp. 4'] (see also [13, Prop. 19.21]) and the associated maximally monotone operator $\boldsymbol{A}$ of (4.12) is defined at $(\boldsymbol{x}, v) \in \boldsymbol{\mathcal{H}}\oplus\mathcal{G}$ to be $\boldsymbol{A}(\boldsymbol{x}, v) = (\partial \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{L}^*v, -\boldsymbol{L}\boldsymbol{x})$. Thus, solving (4.29)–(4.30) is equivalent to finding a zero of the operator $\boldsymbol{A} \in \mathcal{M}(\mathcal{H}\oplus\mathcal{G}\oplus\mathcal{G}) \smallsetminus \mathcal{S}(\mathcal{H}\oplus\mathcal{G}\oplus\mathcal{G})$ defined by

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{G})(\forall v \in \mathcal{G}) \quad \boldsymbol{A}(x,y,v) = \big(\partial f(x) + L^*v, \partial g(y) - v, -Lx + y\big). \tag{4.31}$$

In [57], this problem is approached by splitting $\boldsymbol{A}$ as

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{G})(\forall v \in \mathcal{G}) \quad \boldsymbol{A}(x,y,v) = \big(\partial f(x) + L^*v, 0, -Lx\big) + \big(0, \partial g(y) - v, y\big). \tag{4.32}$$

Given $\gamma \in]0,+\infty[$, $\mu_1 \in \mathbb{R}$, $\mu_2 \in \mathbb{R}$, $x_0 \in \mathcal{H}$, $y_0 \in \mathcal{G}$, and $v_0 \in \mathcal{G}$, applying the Douglas-Rachford algorithm (4.6) to this decomposition leads to the algorithm [57]

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
x_{n+1} \in \underset{x \in \mathcal{H}}{\text{Argmin}} \left(f(x) + \langle Lx \mid v_n \rangle + \dfrac{1}{2\gamma}\|Lx - y_n\|^2 + \dfrac{\gamma\mu_1^2}{2}\|x - x_n\|^2 \right) \\
y_{n+1} = \underset{y \in \mathcal{G}}{\text{argmin}} \left(g(y) - \langle y \mid v_n \rangle + \dfrac{1}{2\gamma}\|Lx_{n+1} - y\|^2 + \dfrac{\gamma\mu_2^2}{2}\|y - y_n\|^2 \right) \\
v_{n+1} = v_n + \gamma^{-1}\big(Lx_{n+1} - y_{n+1}\big).
\end{array}
\right.
\end{array}
\tag{4.33}
$$

When $\mu_1 = \mu_2 = 0$, this scheme corresponds to the alternating direction method of multipliers (ADMM) [48, 61, 62, 63] and, just like it, requires a potentially complex minimization involving f and L jointly to construct x_{n+1} (see [58, 59] for connections between ADMM and the Douglas-Rachford algorithm). To circumvent this issue and obtain a method that does split f, g, and L, let us decompose $\boldsymbol{A}$ as $\boldsymbol{A} = \boldsymbol{M} + \boldsymbol{S}$, where

$$
(\forall x \in \mathcal{H})(\forall y \in \mathcal{G})(\forall v \in \mathcal{G}) \quad
\begin{cases}
\boldsymbol{M}(x, y, v) = \big(\partial f(x), \partial g(y), 0\big) \\
\boldsymbol{S}(x, y, v) = \big(L^* v, -v, -Lx + y\big).
\end{cases}
\tag{4.34}
$$

Applying (4.5) to this subdifferential+skew decomposition in $\mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G}$, we obtain the following algorithm, which employs prox_f, prox_g, L, and L^*.

Proposition 4.9 *Consider the setting of Problem 4.8 and let $(x_0, y_0, v_0) \in \mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G}$. Iterate*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\varepsilon \leqslant \gamma_n \leqslant (1 - \varepsilon)/\sqrt{1 + \|L\|^2} \\
r_n = \gamma_n(Lx_n - y_n) \\
p_n = \text{prox}_{\gamma_n f}\big(x_n - \gamma_n L^* v_n\big) \\
q_n = \text{prox}_{\gamma_n g}\big(y_n + \gamma_n v_n\big) \\
x_{n+1} = p_n - \gamma_n L^* r_n \\
y_{n+1} = q_n + \gamma_n r_n \\
v_{n+1} = v_n + \gamma_n\big(Lp_n - q_n\big).
\end{array}
\right.
\end{array}
\tag{4.35}
$$

Then $(x_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ converge weakly to solutions to (4.29) and (4.30), respectively.

Proof. This is an application of [28, Thm. 2.5(ii)] to the maximally monotone operator $\boldsymbol{M}$ and the monotone and Lipschitzian operator $\boldsymbol{S}$ of (4.34). Note that the Lipschitz constant of $\boldsymbol{S}$ is $\|\boldsymbol{S}\| = \sqrt{1 + \|L\|^2}$ and that $(\forall n \in \mathbb{N})$ $J_{\gamma_n \boldsymbol{M}} = \text{prox}_{\gamma_n f} \times \text{prox}_{\gamma_n g} \times \mathrm{Id}$. Thus, using elementary algebraic manipulations, (4.5) reduces to (4.35). □

Let us note that (4.35) bears a certain resemblance with the algorithm

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\varepsilon \leqslant \gamma_n \leqslant (1 - \varepsilon)\min\{1, 1/\|L\|\}/2 \\
p_n = v_n + \gamma_n(Lx_n - y_n) \\
x_{n+1} = \text{prox}_{\gamma_n f}\big(x_n - \gamma_n L^* p_n\big) \\
y_{n+1} = \text{prox}_{\gamma_n g}\big(y_n + \gamma_n p_n\big) \\
v_{n+1} = v_n + \gamma_n\big(Lx_{n+1} - y_{n+1}\big),
\end{array}
\right.
\end{array}
\tag{4.36}
$$

proposed in a finite-dimensional setting in [37].

5 Closing remarks

The constant interactions between convex optimization and monotone operator theory have greatly benefited both fields. On the numerical side, spectacular advances have been made in the last years in the area of splitting algorithms to solve complex structured problems. While many methods have been obtained by recasting classical algorithms in product spaces, often with the help of duality arguments, recent proposals such as that of [43] rely on different paradigms and make asynchronous and block-iterative implementations possible. Despite the relative maturity of the field, there remain plenty of exciting open problems, and we can mention only a few here. For instance, on the theoretical side, duality for monotone inclusions is based on rather rudimentary principles, whereby dual solutions exist if and only if primal solution exist, and it does not match the more subtle results from Fenchel-Rockafellar duality in classical convex optimization. On the algorithmic front, splitting based on Bregman distances is still in its infancy. This framework is motivated by the need to solve problems in Banach spaces, where standard notions of resolvent and proximity operators are no longer appropriate, but also by numerical considerations in basic Euclidean spaces since some proximity operators may be easier to implement in Bregman form or some functions may have more exploitable properties when examined through Bregman distances [8, 45, 96]. As a final word, let us emphasize that a monumental achievement of Browder, Kačurovskiĭ, Minty, Moreau, Rockafellar, and Zarantonello was to build, within the unchartered field of nonlinear analysis, structured and fertile areas that extended ideas from classical linear functional analysis. It remains a huge challenge to delimit and construct such areas in the vast world of nonconvex/nonmonotone problems, that would preserve enough structure to support a solid and meaningful theory and, at the same time, lend itself to the development of powerful algorithms that would produce more than just local solutions.

References

[1] M. A. Alghamdi, A. Alotaibi, P. L. Combettes, and N. Shahzad, A primal-dual method of partial inverses for composite inclusions, *Optim. Lett.,* vol. 8, pp. 2271–2284, 2014.

[2] E. Asplund, A monotone convergence theorem for sequences of nonlinear mappings, in *Nonlinear Functional Analysis, Part 1,* (F. E. Browder, ed.), pp. 1–9. AMS, Providence, RI, 1970.

[3] H. Attouch, L. M. Briceño-Arias, and P. L. Combettes, A parallel splitting method for coupled monotone inclusions, *SIAM J. Control Optim.*, vol. 48, pp. 3246–3270, 2010.

[4] H. Attouch, J. Peypouquet, and P. Redont, Backward-forward algorithms for structured monotone inclusions in Hilbert spaces, *J. Math. Anal. Appl.,* vol. 457, pp. 1095–1117, 2018.

[5] J.-B. Baillon, P. L. Combettes, and R. Cominetti, There is no variational characterization of the cycles in the method of periodic projections, *J. Funct. Anal.,* vol. 262, pp. 400–408, 2012.

[6] J.-B. Baillon and G. Haddad, Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones, *Israel J. Math.,* vol. 26, pp. 137–150, 1977.

[7] S. Bartz, H. H. Bauschke, J. M. Borwein, S. Reich, and X. Wang, Fitzpatrick functions, cyclic monotonicity and Rockafellar's antiderivative, *Nonlinear Anal.,* vol. 66, pp. 1198–1223, 2007.

[8] H. H. Bauschke, J. Bolte, and M. Teboulle, A descent lemma beyond Lipschitz gradient continuity: First-order methods revisited and applications, *Math. Oper. Res.,* vol. 42, pp. 330–348, 2017.

[9] H. H. Bauschke, M. N. Bui, and X. Wang, Projecting onto the intersection of a cone and a sphere, 2017. https://arxiv.org/pdf/1708.00585

[10] H. H. Bauschke, M. N. Bui, and X. Wang, On the sum of projectors onto convex sets, 2018. https://arxiv.org/pdf/1802.02287

[11] H. H. Bauschke and P. L. Combettes, A Dykstra-like algorithm for two monotone operators, *Pac. J. Optim.,* vol. 4, pp. 383–391, 2008.

[12] H. H. Bauschke and P. L. Combettes, The Baillon-Haddad theorem revisited, *J. Convex Anal.,* vol. 17, pp. 781–787, 2010.

[13] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces,* 2nd ed. Springer, New York, 2017.

[14] H. H. Bauschke, E. Matoušková, and S. Reich, Projection and proximal point methods: Convergence results and counterexamples, *Nonlinear Anal.,* vol. 56, pp. 715–738, 2004.

[15] H. H. Bauschke, S. M. Moffat, and X. Wang, Firmly nonexpansive mappings and maximally monotone operators: Correspondence and duality, *Set-Valued Var. Anal.,* vol. 20, pp. 131–153, 2012.

[16] H. H. Bauschke, J. Schaad, and X. Wang, On Douglas-Rachford operators that fail to be proximal mappings, *Math. Program.,* vol. B168, pp. 55–61, 2018.

[17] H. H. Bauschke, X. Wang, and L. Yao, On Borwein-Wiersma decompositions of monotone linear relations, *SIAM J. Optim.,* vol. 20, pp. 2636–2652, 2010.

[18] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.,* vol. 2, pp. 183–202, 2009.

[19] R. Bellman, R. E. Kalaba, and J. A. Lockett, *Numerical Inversion of the Laplace Transform: Applications to Biology, Economics Engineering, and Physics.* Elsevier, New York, 1966.

[20] J. M. Borwein, Maximal monotonicity via convex analysis, *J. Convex Anal.,* vol. 13, pp. 561–586, 2006.

[21] J. M. Borwein, Asplund decompositions of monotone operators, *ESAIM: Proc.,* vol. 17, pp. 19–25, 2007.

[22] J. M. Borwein and H. Wiersma, Asplund decomposition of monotone operators, *SIAM J. Optim.,* vol. 18, pp. 946–960, 2007.

[23] R. I. Boţ and C. Hendrich, Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization, *J. Math. Imaging Vision,* vol. 49, pp. 551–568, 2014.

[24] J. P. Boyle and R. L. Dykstra, A method for finding projections onto the intersection of convex sets in Hilbert spaces, *Lect. Notes Stat.* vol. 37, pp. 28–47, 1986.

[25] H. Brézis, *Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert.* North-Holland/Elsevier, New York, 1973.

[26] H. Brézis and P. L. Lions, Produits infinis de résolvantes, *Israel J. Math.,* vol. 29, pp. 329–345, 1978.

[27] L. M. Briceño-Arias and P. L. Combettes, Convex variational formulation with smooth coupling for multicomponent signal decomposition and recovery, *Numer. Math. Theory Methods Appl.,* vol. 2, pp. 485–508, 2009.

[28] L. M. Briceño-Arias and P. L. Combettes, A monotone+skew splitting model for composite monotone inclusions in duality, *SIAM J. Optim.,* vol. 21, pp. 1230–1250, 2011.

[29] F. E. Browder, The solvability of non-linear functional equations, *Duke Math. J.,* vol. 30, pp. 557–566, 1963.

[30] F. E. Browder, Variational boundary value problems for quasi-linear elliptic equations of arbitrary order, *Proc. Natl. Acad. Sci. USA,* vol. 50, pp. 31–37, 1963.

[31] F. E. Browder, Multi-valued monotone nonlinear mappings and duality mappings in Banach spaces, *Trans. Amer. Math. Soc.,* vol. 118, pp. 338–351, 1965.

[32] F. E. Browder, Convergence theorems for sequences of nonlinear operators in Banach spaces, *Math Z.,* vol. 100, pp. 201–225, 1967.

[33] C. L. Byrne, *Iterative Optimization in Inverse Problems.* CRC Press, Boca Raton, FL, 2014.

[34] A. Chambolle and C. Dossal, On the convergence of the iterates of the "Fast iterative shrinkage/thresholding algorithm", *J. Optim. Theory Appl.,* vol. 166, pp. 968–982, 2015.

[35] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *J. Math. Imaging Vision,* vol. 40, pp. 120–145, 2011.

[36] C. Chaux, M. El-Gheche, J. Farah, J.-C. Pesquet, and B. Pesquet-Popescu, A parallel proximal splitting method for disparity estimation from multicomponent images under illumination variation, *J. Math. Imaging Vision,* vol. 47, pp. 167–178, 2013.

[37] G. Chen and M. Teboulle, A proximal-based decomposition method for convex minimization problems, *Math. Program.,* vol. 64, pp. 81–101, 1994.

[38] P. L. Combettes, Inconsistent signal feasibility problems: Least-squares solutions in a product space, *IEEE Trans. Signal Process.,* vol. 42, pp. 2955–2966, 1994.

[39] P. L. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, *Optimization,* vol. 53, pp. 475–504, 2004.

[40] P. L. Combettes, Iterative construction of the resolvent of a sum of maximal monotone operators, *J. Convex Anal.,* vol. 16, pp. 727–748, 2009.

[41] P. L Combettes, L. Condat, J.-C. Pesquet, and B. C. Vũ, A forward-backward view of some primal-dual optimization methods in image recovery, *Proc. IEEE Int. Conf. Image Process.*, Paris, France, Oct. 27–30, 2014, pp. 4141–4145.

[42] P. L. Combettes, Dinh Dũng, and B. C. Vũ, Dualization of signal recovery problems, *Set-Valued Var. Anal.,* vol. 18, pp. 373–404, 2010.

[43] P. L. Combettes and J. Eckstein, Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions, *Math. Program.,* vol. B168, pp. 645–672, 2018.

[44] P. L. Combettes and C. L. Müller, Perspective functions: Proximal calculus and applications in high-dimensional statistics, *J. Math. Anal. Appl.,* vol. 457, pp. 1283–1306, 2018.

[45] P. L. Combettes and Q. V. Nguyen, Solving composite monotone inclusions in reflexive Banach spaces by constructing best Bregman approximations from their Kuhn-Tucker set, *J. Convex Anal.,* vol. 23, pp. 481–510, 2016.

[46] P. L. Combettes and J.-C. Pesquet, Proximal thresholding algorithm for minimization over orthonormal bases, *SIAM J. Optim.,* vol. 18, pp. 1351–1376, 2007.

[47] P. L. Combettes and J.-C. Pesquet, A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery, *IEEE J. Selected Topics Signal Process.,* vol. 1, pp. 564–574, 2007.

[48] P. L. Combettes and J.-C. Pesquet, Proximal splitting methods in signal processing, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering,* pp. 185–212. Springer, New York, 2011.

[49] P. L. Combettes and J.-C. Pesquet, Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators, *Set-Valued Var. Anal.*, vol. 20, pp. 307–330, 2012.

[50] P. L. Combettes, S. Salzo, and S. Villa, Consistent learning by composite proximal thresholding, *Math. Program.,* vol. B167, pp. 99–127, 2018.

[51] P. L. Combettes and B. C. Vũ, Variable metric forward-backward splitting with applications to monotone inclusions in duality, *Optimization,* vol. 63, pp. 1289–1318, 2014.

[52] P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.,* vol. 4, pp. 1168–1200, 2005.

[53] P. L. Combettes and I. Yamada, Compositions and convex combinations of averaged nonexpansive operators, *J. Math. Anal. Appl.,* vol. 425, pp. 55–70, 2015.

[54] L. Condat, A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms, *J. Optim. Theory Appl.,* vol. 158, pp. 460–479, 2013.

[55] D. Davis and W. Yin, A three-operator splitting scheme and its optimization applications, *Set-Valued Var. Anal.,* vol. 25, pp, 829–858, 2017.

[56] J. Duchi and Y. Singer, Efficient online and batch learning using forward backward splitting, *J. Mach. Learn. Res.,* vol. 10, pp. 2899–2934, 2009.

[57] J. Eckstein, Some saddle-function splitting methods for convex programming, *Optim. Methods Softw.,* vol. 4, pp. 75–83, 1994.

[58] J. Eckstein and D. P. Bertsekas, On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Program.,* vol. 55, pp. 293–318, 1992.

[59] D. Gabay, Applications of the method of multipliers to variational inequalities, in: M. Fortin and R. Glowinski (eds.), *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary Value Problems,* pp. 299–331. North-Holland, Amsterdam, 1983.

[60] N. Ghoussoub, *Self-Dual Partial Differential Systems and Their Variational Principles.* Springer, New York, 2009.

[61] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics.* Philadelphia: SIAM, 1989.

[62] R. Glowinski and A. Marrocco, Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité, d'une classe de problèmes de Dirichlet non linéaires, *RAIRO Anal. Numer.,* vol. 2, pp. 41–76, 1975.

[63] R. Glowinski, S. J. Osher, and W. Yin (eds.), *Splitting Methods in Communication, Imaging, Science, and Engineering.* Springer, New York, 2016.

[64] M. Golomb, Zur Theorie der nichtlinearen Integralgleichungen, Integralgleichungssysteme und allgemeinen Funktionalgleichungen, *Math. Z.,* vol. 39, pp. 45–75, 1935.

[65] O. Güler, On the convergence of the proximal point algorithm for convex minimization, *SIAM J. Control Optim.*, vol. 29, pp. 403–419, 1991.

[66] B. He and X. Yuan, Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective, *SIAM J. Imaging Sci.,* vol. 5, pp. 119–149, 2012.

[67] H. S. Hundal, An alternating projection that does not converge in norm, *Nonlinear Anal.*, vol. 57, pp. 35–61, 2004.

[68] H. Idrissi, O. Lefebvre, and C. Michelot, Applications and numerical convergence of the partial inverse method, *Lecture Notes in Math.,* vol. 1405, pp. 39–54, 1989.

[69] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, Proximal methods for hierarchical sparse coding, *J. Machine Learning Res.,* vol. 12, pp. 2297–2334, 2011.

[70] J. L. W. V. Jensen, Sur les fonctions convexes et les inégalités entre les valeurs moyennes, *Acta Math.,* vol. 30, pp. 175–193, 1906.

[71] R. I. Kačurovskiĭ, Monotone operators and convex functionals, *Uspekhi Mat. Nauk,* vol. 15, pp. 213–215, 1960.

[72] R. I. Kačurovskiĭ, Nonlinear monotone operators in Banach spaces, *Uspekhi Mat. Nauk,* vol. 23, pp. 121–168, 1968; English translation: *Russian Math. Surveys,* vol. 23, pp. 117–165, 1968.

[73] Y. Kōmura, Nonlinear semi-groups in Hilbert space, *J. Math. Soc. Japan,* vol. 19, pp. 493–507, 1967.

[74] J. Lawrence and J. E. Spingarn, On fixed points of non-expansive piecewise isometric mappings, *Proc. London Math. Soc.,* vol. 55, pp. 605–624, 1987.

[75] B. Lemaire, The proximal algorithm, in: *New methods in Optimization and Their Industrial Uses,* (J. P. Penot, Ed.), International Series of Numerical Mathematics, vol. 87, pp. 73–87. Birkhäuser, Boston, MA, 1989.

[76] A. Lenoir and Ph. Mahey, A survey on operator splitting and decomposition of convex programs, *RAIRO-Oper. Res.,* vol. 51, pp. 17–41, 2017.

[77] J. Leray and J.-L. Lions, Quelques résultats de Višik sur les problèmes elliptiques nonlinéaires par les méthodes de Minty-Browder, *Bull. Soc. Math. France,* vol. 93, pp. 97–107, 1965.

[78] B. Levi, Sul principio di Dirichlet, *Rend. Circ. Mat. Palermo,* vol. 22, pp. 293–359, 1906.

[79] E. S. Levitin and B. T. Polyak, Constrained minimization methods, *U.S.S.R. Comput. Math. Math. Phys.*, vol. 6, pp. 1–50, 1966.

[80] P.-L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1979.

[81] Ph. Mahey, S. Oualibouch, and Pham Dinh Tao, Proximal decomposition on the graph of a maximal monotone operator, *SIAM J. Optim.,* vol. 5, pp. 454–466, 1995.

[82] B. Martinet, Régularisation d'inéquations variationnelles par approximations successives, *Rev. Fr. Inform. Rech. Oper.,* vol. 4, pp. 154–158, 1970.

[83] B. Martinet, Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox, *C. R. Acad. Sci. Paris,* vol. A274, pp. 163–165, 1972.

[84] B. Mercier, *Topics in Finite Element Solution of Elliptic Problems* (Lectures on Mathematics, no. 63). Tata Institute of Fundamental Research, Bombay, 1979.

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

[86] G. J. Minty, On a "monotonicity" method for the solution of nonlinear equations in Banach spaces, *Proc. Natl. Acad. Sci. USA,* vol. 50, pp. 1038–1041, 1963.

[87] G. J. Minty, On the monotonicity of the gradient of a convex function, *Pac. J. Math.,* vol. 14, pp. 243–247, 1964.

[88] J. J. Moreau, Décomposition orthogonale d'un espace hilbertien selon deux cônes mutuellement polaires, *C. R. Acad. Sci. Paris Sér. A Math.,* vol. 255, pp. 238–240, 1962.

[89] J. J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris,* vol. A255, pp. 2897–2899, 1962.

[90] J. J. Moreau, Les liaisons unilatérales et le principe de Gauss, *C. R. Acad. Sci. Paris,* vol. A256, pp. 871–874, 1963.

[91] J. J. Moreau, Propriétés des applications "prox", *C. R. Acad. Sci. Paris,* vol. A256, pp. 1069–1071, 1963.

[92] J. J. Moreau, Fonctionnelles sous-différentiables, *C. R. Acad. Sci. Paris,* vol. A257, pp. 4117–4119, 1963.

[93] J. J. Moreau, Sur la naissance de la cavitation dans une conduite, *C. R. Acad. Sci. Paris,* vol. A259, pp. 3948–3950, 1964.

[94] J. J. Moreau, Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. France,* vol. 93, pp. 273–299, 1965.

[95] J. J. Moreau, Quadratic programming in mechanics: Dynamics of one-sided constraints, *SIAM J. Control,* vol. 4, pp. 153–158, 1966.

[96] Q. V. Nguyen, Forward-backward splitting with Bregman distances, *Vietnam J. Math.,* vol. 45, pp. 519–539, 2017.

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

[98] N. Papadakis, G. Peyré, and E. Oudet, Optimal transport with proximal splitting, *SIAM J. Imaging Sci.,* vol. 7, pp. 212–238, 2014.

[99] R. R. Phelps, *Convex Functions, Monotone Operators and Differentiability,* 2nd ed. Springer-Verlag, Berlin, 1993.

[100] H. Raguet, J. Fadili, and G. Peyré, A generalized forward-backward splitting, *SIAM J. Imaging Sci.,* vol. 6, pp. 1199–1226, 2013.

[101] R. T. Rockafellar, *Convex Functions and Dual Extremum Problems.* Ph.D. thesis, Harvard University, 1963.

[102] R. T. Rockafellar, Characterization of the subdifferentials of convex functions, *Pac. J. Math.,* vol. 17, pp. 497–510. 1966.

[103] R. T. Rockafellar, Monotone operators associated with saddle-functions and minimax problems, in: *Nonlinear Functional Analysis, Part 1,* (F. E. Browder, ed.), pp. 241–250. AMS, Providence, RI, 1970.

[104] R. T. Rockafellar, Saddle points and convex analysis, in: *Differential Games and Related Topics,* H. W. Kuhn and G. P. Szegö (eds.), pp. 109–127. North-Holland, New York, 1971.

[105] R. T. Rockafellar, *Conjugate Duality and Optimization.* SIAM, Philadelphia, PA 1974.

[106] R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.,* vol. 14, pp. 877–898, 1976.

[107] R. T. Rockafellar, Augmented Lagrangians and applications of the proximal point algorithm in convex programming, *Math. Oper. Res.,* vol. 1, pp. 97–116, 1976.

[108] R. T. Rockafellar and R. J. B. Wets, Scenarios and policy aggregation in optimization under uncertainty, *Math. Oper. Res.,* vol. 16, pp. 1–29, 1991.

[109] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*. Amer. Math. Soc., Providence, RI, 1997.

[110] S. Simons, *From Hahn-Banach to Monotonicity,* 2nd ed. Springer-Verlag, Berlin, 2008.

[111] J. E. Spingarn, Partial inverse of a monotone operator, *Appl. Math. Optim.*, vol. 10, pp. 247–265, 1983.

[112] J. E. Spingarn, Applications of the method of partial inverses to convex programming: Decomposition, *Math. Program.* 32, pp. 199–223, 1985.

[113] J. E. Spingarn, A projection method for least-squares solutions to overdetermined systems of linear inequalities, *Linear Algebra Appl.,* vol. 86, pp. 211–236, 1987.

[114] S. Sra, S. Nowozin, and S. J. Wright (eds.), *Optimization for Machine Learning.* MIT Press, Cambridge, MA, 2012.

[115] P. Tseng, Applications of a splitting algorithm to decomposition in convex programming and variational inequalities, *SIAM J. Control Optim.,* vol. 29, pp. 119–138, 1991.

[116] P. Tseng, A modified forward-backward splitting method for maximal monotone mappings, *SIAM J. Control Optim.*, vol. 38, pp. 431–446, 2000.

[117] M. M. Vaĭnberg, *Variational Method and Method of Monotone Operators in the Theory of Nonlinear Equations.* Nauka, Moskow, 1972. English translation: Wiley, NY, 1973.

[118] M. M. Vaĭnberg and R. I. Kačurovskiĭ, On the variational theory of nonlinear operators and equations, *Dokl. Akad. Nauk SSSR,* vol. 129, pp. 1199–1202, 1959.

[119] S. Vaiter, G. Peyré, and J. Fadili, Model consistency of partly smooth regularizers, *IEEE Trans. Inform. Theory,* vol. 64, pp. 1725–1737, 2018.

[120] B. C. Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators, *Adv. Comput. Math.,* vol. 38, pp. 667–681, 2013.

[121] Y.-L. Yu, On decomposing the proximal map, *Adv. Neural Inform. Process. Syst.,* vol. 16, pp. 91–99, 2013.

[122] E. H. Zarantonello, Solving functional equations by contractive averaging, Mathematical Research Center technical summary report no. 160, University of Wisconsin, Madison, 1960.

[123] E. H. Zarantonello (ed.), *Contributions to Nonlinear Functional Analysis.* Academic Press, New York, 1971.

[124] E. H. Zarantonello, The product of commuting conical projections is a conical projection, *Proc. Amer. Math. Soc.,* vol. 38, pp. 591–594, 1973.

[125] E. H. Zarantonello, L'algèbre des projecteurs coniques, *Lecture Notes in Econom. and Math. Systems,* vol. 102, pp. 232–243, 1974.

[126] E. H. Zarantonello, La structure des opérateurs monotones, *Rev. Un. Mat. Argentina,* vol. 34, pp. 90–96, 1988.

[127] E. Zeidler, *Nonlinear Functional Analysis and Its Applications II/B – Nonlinear Monotone Operators*. Springer-Verlag, New York, 1990.

[128] X. Zeng and M. Figueiredo, Solving OSCAR regularization problems by fast approximate proximal splitting algorithms, *Digital Signal Process.,* vol. 31, pp. 124–135, 2014.