The geometry of monotone operator splitting methods

Patrick L. Combettes*

North Carolina State University, Department of Mathematics
Raleigh, NC 27695-8205, USA
plc@math.ncsu.edu

Abstract. We propose a geometric framework to describe and analyze a wide array of operator splitting methods for solving monotone inclusion problems. The initial inclusion problem, which typically involves several operators combined through monotonicity-preserving operations, is seldom solvable in its original form. We embed it in an auxiliary space, where it is associated with a surrogate monotone inclusion problem with a more tractable structure and which allows for easy recovery of solutions to the initial problem. The surrogate problem is solved by successive projections onto half-spaces containing its solution set. The outer approximation half-spaces are constructed by using the individual operators present in the model separately. This geometric framework is shown to encompass traditional methods as well as state-of-the-art asynchronous block-iterative algorithms, and its flexible structure provides a pattern to design new ones.

Contents

*This work was supported by the National Science Foundation under grant CCF-2211123.

1 Introduction

Throughout, $\mathcal{H}$ is a real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle$ and $2^{\mathcal{H}}$ stands for the power set of $\mathcal{H}$. Our main focus is on the following monotone inclusion problem.

Problem 1.1 Let $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a monotone operator, that is,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})(\forall x^* \in Mx)(\forall y^* \in My) \quad \langle x - y \mid x^* - y^* \rangle \geqslant 0. \tag{1.1}$$

The task is to find $x \in \mathcal{H}$ such that $0 \in Mx$.

Monotone inclusion problems are intimately linked to the birth of nonlinear analysis. They first appeared as a powerful models to establish existence, uniqueness, and stability results for various nonlinear problems [87, 205, 239, 403, 405]. Over the past six decades, monotone inclusion models have penetrated almost all areas of mathematics and its applications. Nowadays, Problem 1.1 models a broad range of equilibria in areas such as dynamical systems [2], ill-posed problems [4], domain decomposition methods [6, 18, 21], circuit theory [11, 107, 108, 109, 211], machine learning [15, 149, 232, 382], evolution equations [17, 70, 353], partial differential equations [29, 71, 123, 206, 304, 353, 406], signal processing [45, 144, 152, 318], image processing [47, 114, 153, 210, 311], game theory [48, 57, 77, 98, 124, 189, 190, 203], network flow problems [54, 92, 341, 342], equilibrium theory [73, 140, 296], mean-field games [80, 81], control theory [83, 84, 101, 165, 356], data science [116, 148, 391], optimization [131, 179, 215, 373, 374], statistics [141, 394], neural networks [147, 389, 395], traffic equilibrium [158, 196], systems theory [162, 166], mechanics [194, 278], optimal transportation [302], and minimax theory [335].

Early numerical solution methods to solve Problem 1.1 can be found in [12, 88, 89, 246, 262, 312, 354, 378, 379, 403, 404]. These methods are of the explicit Euler type, meaning that, at iteration n, the update x_{n+1} is determined by finding a point in Mx_n. An alternative method, which first appeared in [259] and then in more detail in [339], is the proximal point algorithm, where the update is obtained through the implicit relation $x_n - x_{n+1} \in Mx_{n+1}$. Such approaches have limited potential since they can be directly implemented only in specific situations. For instance, the Euler step methods of [88, 89, 90] impose certain properties on M and asymptotically vanishing step sizes, which is detrimental to numerical stability and speed of convergence. On the other hand, the proximal point algorithm requires explicit expressions for the resolvent of M, which is seldom possible. In most problems, however, M has a complex structure and it is typically expressed in terms of monotonicity-preserving operations involving simpler operators. The principle governing *splitting methods* is to devise algorithms in which each of the elementary operators arising in the decomposition of M are used individually, hence breaking up Problem 1.1 into tasks that are more manageable.

The first monotone operator splitting methods arose in the late 1970s and were motivated by applications in mechanics and partial differential equations [194, 209, 278]. The three main algorithms that dominated the field were designed for problems in which

$$M = A + B, \tag{1.2}$$

where $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ are maximally monotone: the forward-backward method [277], the Douglas–Rachford method [265], and Tseng's forward-backward-forward method [375]. In recent years, the field of monotone operator splitting algorithms has benefited from a new impetus, fueled by the emerging application areas mentioned above and their demand for solving efficiently increasingly complex large-dimensional problems. Thus, duality techniques have arisen to address composite models of the form

$$M = A + L^* \circ B \circ L, \tag{1.3}$$

where L is a linear operator from $\mathcal{H}$ to a Hilbert space $\mathcal{G}$ and $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ are maximally monotone [76]. These techniques have been further developed to devise splitting algorithms for the more structured model [62, 145, 387]

$$M = A + \sum_{k=1}^{p} L_k^* \circ \left(B_k^{-1} + D_k^{-1}\right)^{-1} \circ L_k + C, \tag{1.4}$$

where each linear operator L_k maps $\mathcal{H}$ to a Hilbert space $\mathcal{G}_k$, and the operators $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, $B_k\colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$, $D_k\colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$, and $C\colon \mathcal{H} \to \mathcal{H}$ are maximally monotone. Splitting algorithms for models which are more finely structured than (1.4) have also been proposed as well as multivariate versions that capture coupled systems of monotone inclusions; see [97] and the references therein. On a different front, block-iterative algorithms, which allow for the activation of only a subgroup of operators present in the model at a given iteration, have also been developed [93, 97, 136, 237]. At the same time, a multitude of splitting algorithms tailored to specific models have been elaborated. For instance, if $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ are maximally monotone and $C\colon \mathcal{H} \to \mathcal{H}$ is cocoercive, splitting algorithms have been proposed in [161, 321] for the decomposition $M = A + B + C$ and in particular in [79] if $B\colon \mathcal{H} \to \mathcal{H}$ is Lipschitzian and in [249] if $B\colon \mathcal{H} \to \mathcal{H}$ is linear and bounded.

Given the abundance of activity in monotone operator splitting techniques, it is important to identify general structures and principles, as well as possible bonds between algorithm design methodologies in order not only to simplify and clarify the state of the art, but also to facilitate the developments of new methods in the future. From the outset, fixed point theory has been a tool of choice to achieve this goal. For instance, it has played an important role in the analysis of the proximal point algorithm [248, 275, 339]. In [127], fixed point iterations of averaged operators were shown to provide a convenient framework to investigate the asymptotic behavior of classical splitting algorithms such as the forward-backward, backward-backward, Douglas–Rachford, and Peaceman–Rachford algorithms. Further applications of averaged operator iterations to design and analyze splitting methods can be found in [82, 114, 137, 148, 154, 156, 161, 321, 322, 323, 348, 392]. Fixed point modeling is also a central algorithmic development tool in recent works such as [14, 79, 272]. In spite of these achievements, fixed point methods seem less well suited to capture in simple terms the most flexible splitting methods such as the block-iterative asynchronous methods of [93, 97, 136, 237], which were built using geometric arguments. The purpose of the present paper is to provide a standardized pattern for building and analyzing splitting methods around the following geometric framework. It comprises an embedding step, where the initial Problem 1.1 is replaced by a more tractable surrogate inclusion problem in an auxiliary space $\mathbf{X}$ from which the solutions to the original problem can be easily recovered. The second step is an iterative process in which the current iterate is projected onto a closed half-space that serves as an outer approximation to the surrogate solution set.

Framework 1.2 Geometric algorithmic template for solving Problem 1.1.

(i) **Embedding:** Find a real Hilbert space $\mathbf{X}$, a maximally monotone operator $\mathcal{M}\colon \mathbf{X} \to 2^{\mathbf{X}}$, and an operator $\mathcal{T}\colon \mathbf{X} \to \mathcal{H}$ such that $\mathcal{T}(\operatorname{zer}\mathcal{M}) \subset \operatorname{zer} M$. We call $(\mathbf{X}, \mathcal{M}, \mathcal{T})$ an *embedding* of Problem 1.1.

(ii) **Iterations:**

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots\\
\left\lfloor
\begin{array}{l}
\mathbf{H}_n \text{ is a closed half-space of } \mathbf{X} \text{ such that } \operatorname{zer}\mathcal{M} \subset \mathbf{H}_n\\
\mathbf{x}_{n+1} \text{ is a relaxed projection of } \mathbf{x}_n \text{ onto } \mathbf{H}_n.
\end{array}
\right.
\end{array}
\tag{1.5}
$$

In optimization, the use of half-spaces as outer approximations to the solution set goes back to the cutting plane methods of [121, 243, 258]; see also [250, 383, 402]. In monotone inclusion problems, modeling iterations as successive projections onto separating half-spaces occurs in several papers [35, 125, 358, 359]. We aim at showing that Framework 1.2 is sufficiently broad and flexible to encompass a wide array of existing methods while providing a template to create new ones. It will allow us to derive in a unified fashion simple proofs of existing convergence results. It will also make it possible to establish seamlessly strongly convergent variants of these algorithms. The proofs we provide are new, and so are some of the results.

The remainder of the paper is organized as follows. To make our presentation self-contained, Section 2 covers the necessary mathematical background on monotone operator theory. It also contains various examples of maximally monotone operators and a detailed history of the field. In Section 3, we present several models for decomposing M in Problem 1.1. These decompositions will generate the embeddings required in Framework 1.2 and form the backbone of the splitting methods discussed in the paper. The geometric principles underlying our approach are presented in Section 4, where the main convergence theorems are laid out. In Section 5, we study the proximal point algorithm and explore several of its facets. In Sections 6, 7, and 8, we study, respectively, the Douglas–Rachford, forward-backward-forward, and forward-backward methods through the lens of Framework 1.2 and capture a broad range of algorithms and applications by embedding them in bigger spaces. Block-iterative Kuhn–Tucker and saddle projective splitting methods are addressed in Sections 9 and 10, respectively. Finally, several extensions and variants of the results are discussed in Section 11.

2 Monotone operators

2.1 Notation and basic definitions

The material of this section can be found in [37].

2.1.1 General notation

$\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces, $\mathcal{B}(\mathcal{H},\mathcal{G})$ is the space of bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$, $\mathcal{B}(\mathcal{H})=\mathcal{B}(\mathcal{H},\mathcal{H})$, and $\mathcal{H}\oplus\mathcal{G}$ denotes the Hilbert direct sum of $\mathcal{H}$ and $\mathcal{G}$. The identity operator of $\mathcal{H}$ is denoted by $\mathrm{Id}_{\mathcal{H}}$, its scalar product by $\langle\cdot\mid\cdot\rangle_{\mathcal{H}}$, and the associated norm by $\|\cdot\|_{\mathcal{H}}$ (the subscripts will be omitted when the context is clear). The weak convergence of a sequence $(x_n)_{n\in\mathbb{N}}$ to x is denoted by $x_n\rightharpoonup x$, whereas $x_n\to x$ denotes its strong convergence; the set of weak sequential cluster points of $(x_n)_{n\in\mathbb{N}}$ is denoted by $\mathfrak{W}(x_n)_{n\in\mathbb{N}}$.

2.1.2 Sets

Let C be a subset of $\mathcal{H}$. The interior of C is $\operatorname{int}C$, the *indicator function* of C is

$$\iota_C\colon\mathcal{H}\to\left]-\infty,+\infty\right]\colon x\mapsto\begin{cases}0, & \text{if } x\in C;\\ +\infty, & \text{otherwise,}\end{cases}\tag{2.1}$$

the *support function* of C is

$$\sigma_C\colon \mathcal{H} \to [-\infty,+\infty] : x^* \mapsto \sup_{x\in C} \langle x \mid x^*\rangle, \tag{2.2}$$

and the *distance function* to C is

$$d_C\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \inf_{y\in C} \|x-y\|. \tag{2.3}$$

Suppose that C is convex. We denote by $\operatorname{cone} C$ the smallest cone that contains C and by $\operatorname{sri} C$ the *strong relative interior* of C, i.e.,

$$\operatorname{sri} C = \big\{x \in C \mid \operatorname{cone}(-x+C) \text{ is a closed vector subspace of } \mathcal{H}\big\}. \tag{2.4}$$

If $\mathcal{H}$ is finite-dimensional, $\operatorname{sri} C$ coincides with the *relative interior* $\operatorname{ri} C$ of C, i.e., the interior of C relative to the smallest affine subspace of $\mathcal{H}$ containing C. Suppose that C is nonempty, closed, and convex. For every $x \in \mathcal{H}$,

$$\operatorname{proj}_C x \text{ is the unique point in } C \text{ such that } d_C(x) = \|x - \operatorname{proj}_C x\|. \tag{2.5}$$

This process defines the *projection operator* $\operatorname{proj}_C\colon \mathcal{H} \to \mathcal{H}$ of C. The simple case of a closed half-space is central to our approach.

Example 2.1 ([37, Example 29.20]) Let $u^* \in \mathcal{H}$, let $\eta \in \mathbb{R}$, and suppose that $H = \big\{z \in \mathcal{H} \mid \langle z \mid u^*\rangle \leqslant \eta\big\} \neq \varnothing$. Let $x \in \mathcal{H}$ and set

$$d = \begin{cases} \dfrac{\langle x \mid u^*\rangle - \eta}{\|u^*\|^2}u^*, & \text{if } \langle x \mid u^*\rangle > \eta;\\ 0, & \text{otherwise.}\end{cases} \tag{2.6}$$

Then $\operatorname{proj}_H x = x - d$.

2.1.3 Functions

The set of 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 $\operatorname{argmin}_{x\in\mathcal{H}} f(x)$. The *infimal convolution* of $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ and $h\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ is

$$f \,\square\, h\colon \mathcal{H} \to [-\infty,+\infty] : x \mapsto \inf_{y\in\mathcal{H}} \big(f(y) + h(x-y)\big). \tag{2.7}$$

We denote by $\Gamma_0(\mathcal{H})$ the class of functions $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ which are lower semicontinuous, convex, and such that $\operatorname{dom} f = \big\{x \in \mathcal{H} \mid f(x) < +\infty\big\} \neq \varnothing$. Let $f \in \Gamma_0(\mathcal{H})$. The *conjugate* of f is

$$\Gamma_0(\mathcal{H}) \ni f^*\colon x^* \mapsto \sup_{x\in\mathcal{H}} \big(\langle x \mid x^*\rangle - f(x)\big). \tag{2.8}$$

For every $x \in \mathcal{H}$,

$$\operatorname{prox}_f x \text{ is the unique minimizer over } \mathcal{H} \text{ of } y \mapsto f(y) + \frac{1}{2}\|x-y\|^2. \tag{2.9}$$

This process defines the *proximity operator* $\operatorname{prox}_f \colon \mathcal{H} \to \mathcal{H}$ of f. We have

$$(\forall \gamma \in \left]0,+\infty\right[)(\forall x \in \mathcal{H}) \quad x = \operatorname{prox}_{\gamma f} x + \gamma \operatorname{prox}_{f^*/\gamma} \big(x/\gamma\big). \tag{2.10}$$

The *Moreau envelope* of f of parameter $\gamma \in \left]0,+\infty\right[$ is

$$^{\gamma}f = f \,\square \left(\frac{1}{2\gamma}\|\cdot\|^2\right). \tag{2.11}$$

2.1.4 Set-valued operators

Let $M \colon \mathcal{H} \to 2^{\mathcal{H}}$. The *graph* of M is

$$\operatorname{gra} M = \big\{(x,x^*) \in \mathcal{H} \times \mathcal{H} \mid x^* \in Mx\big\}. \tag{2.12}$$

The *inverse* of M is the operator $M^{-1} \colon \mathcal{H} \to 2^{\mathcal{H}}$ defined through the relation

$$\big(\forall (x,x^*) \in \mathcal{H} \times \mathcal{H}\big) \quad x^* \in Mx \quad \Leftrightarrow \quad x \in M^{-1}x^*. \tag{2.13}$$

Thus,

$$\operatorname{gra} M^{-1} = \big\{(x^*,x) \in \mathcal{H} \times \mathcal{H} \mid (x,x^*) \in \operatorname{gra} M\big\}. \tag{2.14}$$

The set of *fixed points* of M is

$$\operatorname{Fix} M = \big\{x \in \mathcal{H} \mid x \in Mx\big\}, \tag{2.15}$$

the set of *zeros* of M is

$$\operatorname{zer} M = M^{-1}0 = \big\{x \in \mathcal{H} \mid 0 \in Mx\big\}, \tag{2.16}$$

and the *resolvent* of M is the operator

$$J_M = (\operatorname{Id} + M)^{-1}. \tag{2.17}$$

In other words,

$$(\forall x \in \mathcal{H})(\forall p \in \mathcal{H}) \quad p \in J_M x \;\Leftrightarrow\; (p, x-p) \in \operatorname{gra} M \tag{2.18}$$

and therefore

$$\operatorname{zer} M = \operatorname{Fix} J_M. \tag{2.19}$$

We have

$$(\forall\gamma \in \,]0,+\infty[)(\forall x \in \mathcal{H}) \quad x - J_{\gamma M}x = \gamma\, J_{M^{-1}/\gamma}\big(x/\gamma\big). \tag{2.20}$$

The *Yosida approximation* of index $\gamma \in \,]0,+\infty[$ of M is

$$^{\gamma}M = \frac{\mathrm{Id} - J_{\gamma M}}{\gamma} = \big(\gamma\mathrm{Id} + M^{-1}\big)^{-1} = \big(J_{\gamma^{-1}M^{-1}}\big) \circ \gamma^{-1}\mathrm{Id} \tag{2.21}$$

and it satisfies

$$\mathrm{zer}\, M = \mathrm{zer}\, {}^{\gamma}M. \tag{2.22}$$

The *domain* of M is

$$\mathrm{dom}\, M = \big\{x \in \mathcal{H} \mid Mx \neq \varnothing\big\} \tag{2.23}$$

and the *range* of M is

$$\mathrm{ran}\, M = \bigcup_{x\in\mathrm{dom}\, M} Mx = \big\{x^* \in \mathcal{H} \mid (\exists\, x \in \mathrm{dom}\, M)\ x^* \in Mx\big\}. \tag{2.24}$$

We have

$$\mathrm{dom}\, M^{-1} = \mathrm{ran}\, M \quad \text{and} \quad \mathrm{ran}\, M^{-1} = \mathrm{dom}\, M. \tag{2.25}$$

If, for some $x \in \mathcal{H}$, Mx is a singleton, we let Mx denote its single element. We say that M is *injective* if $(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ Mx \cap My \neq \varnothing \Rightarrow x = y$. Finally, given $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, $B\colon \mathcal{G} \to 2^{\mathcal{G}}$, $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$, and $\alpha \in \mathbb{R}$, we set

$$A + \alpha L^* \circ B \circ L\colon \begin{array}{rcl} \mathcal{H} & \to & 2^{\mathcal{H}} \\ x & \mapsto & \big\{x^* + \alpha L^* y^* \mid x^* \in Ax \text{ and } y^* \in B(Lx)\big\}. \end{array} \tag{2.26}$$

2.1.5 Monotone operators

Let $M\colon \mathcal{H} \to 2^{\mathcal{H}}$. Then M is *monotone* if

$$\big(\forall(x,x^*) \in \mathrm{gra}\, M\big)\big(\forall(y,y^*) \in \mathrm{gra}\, M\big) \quad \langle x - y \mid x^* - y^*\rangle \geqslant 0 \tag{2.27}$$

and *maximally monotone* if, further, there exists no monotone operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ such that $\mathrm{gra}\, M \subset \mathrm{gra}\, A \neq \mathrm{gra}\, M$, that is (see Figure 2.1),

$$\begin{aligned} &\big(\forall(x,x^*) \in \mathcal{H} \times \mathcal{H}\big) \\ &\quad \big[\ (x,x^*) \in \mathrm{gra}\, M \ \Leftrightarrow\ \big(\forall(y,y^*) \in \mathrm{gra}\, M\big)\ \langle x - y \mid x^* - y^*\rangle \geqslant 0\ \big]. \end{aligned} \tag{2.28}$$

We have

$$M \text{ maximally monotone } \Rightarrow \text{ zer } M \text{ is closed and convex.} \tag{2.29}$$

Let $\beta \in \left]0,+\infty\right[$. Then M is *β-strongly monotone* if $M - \beta\mathrm{Id}$ is monotone, that is,

$$\big(\forall (x,x^*) \in \operatorname{gra} M\big)\big(\forall (y,y^*) \in \operatorname{gra} M\big) \quad \langle x-y \mid x^*-y^*\rangle \geqslant \beta\|x-y\|^2. \tag{2.30}$$

Now let D be a nonempty subset of $\mathcal{H}$, let $\alpha \in \left]0,+\infty\right[$, and let $M\colon D \to \mathcal{H}$. Then M is *nonexpansive* if

$$(\forall x \in D)(\forall y \in D) \quad \|Mx - My\| \leqslant \|x-y\|, \tag{2.31}$$

α-averaged if $\alpha \leqslant 1$ and $\mathrm{Id} + \alpha^{-1}(M - \mathrm{Id})$ is nonexpansive, *α-cocoercive* if M^{-1} is α-strongly monotone, that is,

$$(\forall x \in D)(\forall y \in D) \quad \langle x-y \mid Mx - My\rangle \geqslant \alpha\|Mx-My\|^2, \tag{2.32}$$

and *firmly nonexpansive* if it is 1-cocoercive. Alternatively,

$$M \text{ is firmly nonexpansive } \Leftrightarrow \ 2M - \mathrm{Id} \text{ is nonexpansive.} \tag{2.33}$$

The following result is known as the Baillon–Haddad theorem.

Lemma 2.2 ([26, Corollaire 10]) *Let $\alpha \in \left]0,+\infty\right[$ and let $f\colon \mathcal{H} \to \mathbb{R}$ be convex, Fréchet differentiable, and such that ∇f is $1/\alpha$-Lipschitzian. Then ∇f is α-cocoercive.*

2.2 History

Monotonicity goes back to classical calculus and the notion of an increasing real-valued function defined on an interval $D \subset \mathbb{R}$, i.e., a function $f\colon D \to \mathbb{R}$ that satisfies

$$(\forall x \in D)(\forall y \in D) \quad \big(x-y\big)\big(f(x)-f(y)\big) \geqslant 0. \tag{2.34}$$

The special properties enjoyed by such functions have long been recognized; see for instance [159, 195, 222]. The monotonicity condition (2.34) is also tied to the infancy of the theory of convex functions. Thus, it was shown in [233] that, if D is open and $g\colon D \to \mathbb{R}$ is a twice differentiable function with derivative f, then (2.34) implies that g is convex. On the numerical side, (2.34) is an important property in connection with solving iteratively the root finding problem [303]

$$\text{find } \ x \in D \ \text{ such that } \ f(x) = 0. \tag{2.35}$$

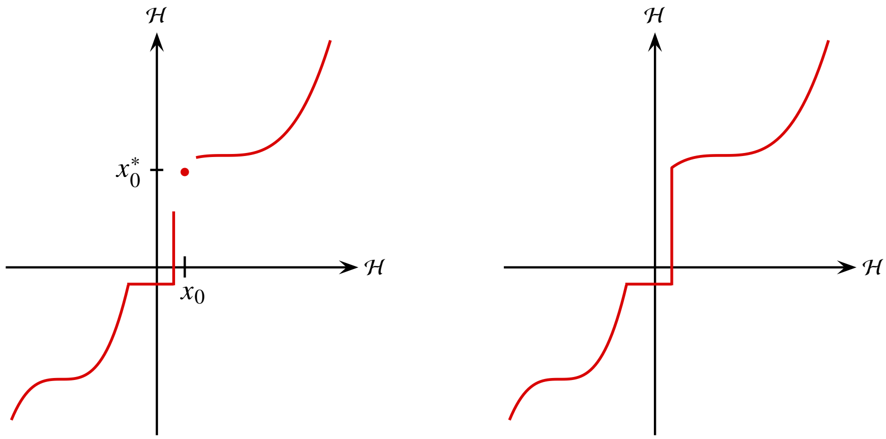

Figure 2.1: Left: Graph of a monotone, but not maximally monotone, operator: the point (x_0, x_0^*) can be added to the graph and the resulting graph remains monotone. Right: Graph of a maximally monotone operator: adding any point to the graph does not preserve its monotonicity.

Monotone operators on $\mathbb{R}$ also appeared in nonlinear circuit theory in the 1940s in the form of quasi-linear resistors [171, 172, 173]. A quasi-linear resistor is a two-pole circuit element characterized by the property that the current going through it increases smoothly with the voltage across it. In other words, the transformation underlying its current-voltage characteristic is differentiable and increasing. Dipoles with monotonic characteristics were further investigated in [280]. To study networks involving a broader range of devices, this concept was extended by Minty in [281, 282] to maximally monotone set-valued transformations on $\mathbb{R}$ (see Figure 2.2 and [103] for examples). Interestingly, as will be discussed shortly, Minty turned out to be one of the founders of monotone operator theory. For further relevant early work on the connections between monotone operators and network theory, see [52, 163] and, for more abstract ramifications, see [166, 341].

Another precursor of monotonicity is found in linear functional analysis, where a linear operator $M\colon \mathcal{H} \supset D \to \mathcal{H}$ is declared accretive if [241]

$$(\forall x \in D) \quad \langle x \mid Mx \rangle \geqslant 0. \tag{2.36}$$

In this context, the notion of a maximally accretive operator was introduced in [314]. Accretive operators are also central to passive linear network theory [50, 401]. One of the first instances of (2.36) in electrical networks is the current-voltage

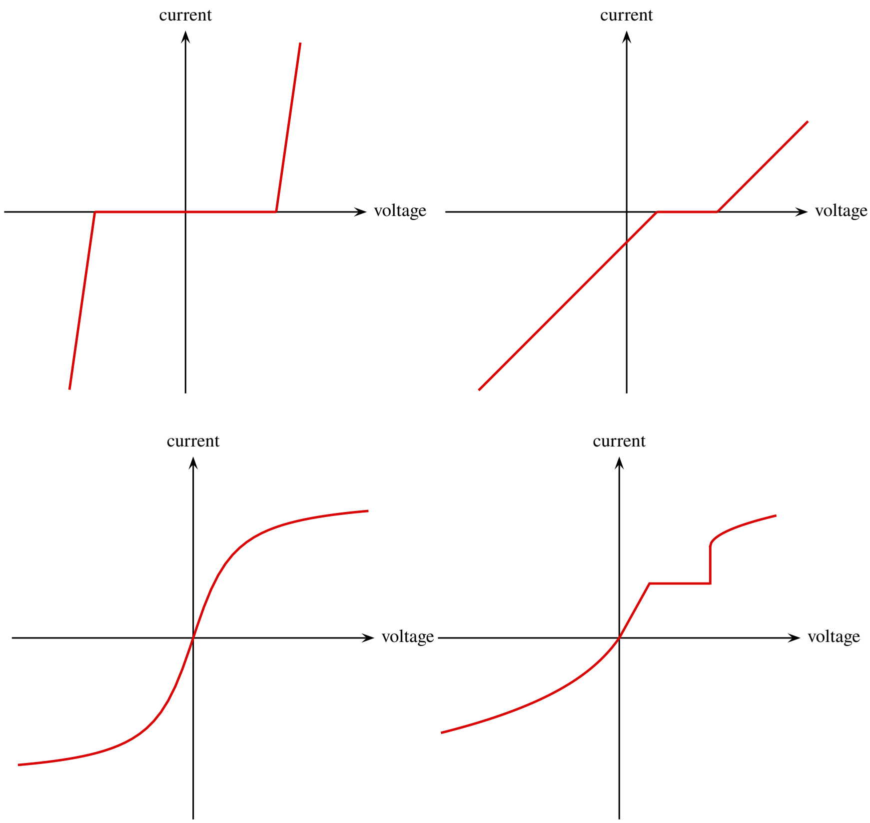

Figure 2.2: Current-voltage characteristics of quasi-linear resistors as monotone operators from $\mathbb{R}$ to $2^{\mathbb{R}}$. Top left: breakdown diodes in series [327]. Top right: breakdown diode and resistance in series [327]. Bottom left: anode-dynode beam-deflection tube [327]. Bottom right: the maximally monotone current-voltage characteristic of [282].

transformation of the four-pole circuit element known as an ideal gyrator [370].

The above notions of increasing functions and positive operators can be brought together by considering an operator $M\colon \mathcal{H} \supset D \to \mathcal{H}$ such that

$$(\forall x \in D)(\forall x \in D) \quad \langle x - y \mid Mx - My \rangle \geqslant 0. \tag{2.37}$$

Instances of (2.37) appear implicitly in [216] and, more explicitly, in [376, 377] in connection with the existence of solutions to Hammerstein integral equations; see also [217] for more general types of equations. Another instance, which corresponds to what is now called strict monotonicity, appears in [91], where $\mathcal{H}$ is the standard Euclidean space. The systematic study of operators satisfying (2.37) started in 1960 an opened an important new chapter of nonlinear functional analysis. Three independent papers submitted that year are associated with the birth of monotone operator theory.

- In an article submitted in February 1960, Kačurovskiĭ [239] called *monotone* an operator that satisfies (2.37). This paper concerned the monotonicity of the gradient of a differentiable convex function (see also [381]) and the existence of solutions to certain nonlinear equations. It also introduced strongly monotone operators.

- In a technical report completed in June 1960, Zarantonello called (2.37) an (isotonically) monotonicity property and discussed supra-unitary (in modern language, strongly monotone) operators. In connection with the solution of nonlinear equations, an important result of [403] is that, if $M\colon \mathcal{H} \to \mathcal{H}$ is monotone and Lipschitzian, then $\mathrm{Id} + M$ is surjective.

- In an article submitted in December 1960, Minty [283] also called $M\colon D \to \mathcal{H}$ monotone if it satisfies (2.37). In addition, he introduced the fundamental concept of maximal monotonicity and established key connections with non-expansive operators. Although, strictly speaking, his definitions dealt with single-valued operators, he established results on monotone relations that naturally suggest extensions to the set-valued case (1.1). According to Browder [86], who initiated the study of set-valued monotone operators in Banach spaces, the Hilbertian setting was worked out by Minty in unpublished notes.

Accounts of the history of the development of monotone operator theory in the 1960s can be found in [58], [87], [240], [263, Section 2.12], [286], and [380, Chapter VI]. In that period, the main mathematical areas of applications were nonlinear equations, partial differential equations, boundary-value problems, non-expansive semigroups, convex analysis, evolution equations, and variational inequalities; see [68, 85, 87, 205, 245, 257, 292, 386, 405] and their bibliographies. At the same time, monotonicity continued to be used in the analysis of networks and systems, for instance in [399, 400], where it is known as incremental positiveness;

see also [162] where monotonicity is called incremental passivity. The main use of monotone operators was to establish existence, uniqueness, or stability results in a variety of nonlinear problems in analysis.

2.3 Examples of maximally monotone operators

The following example concerns single-valued operators; Examples 2.4–2.10 follow from it [37, Chapter 20].

Example 2.3 ([284, Lemma 1]) Let $A\colon \mathcal{H} \to \mathcal{H}$ be monotone and *hemicontinuous* (in particular, continuous) in the sense that

$$\big(\forall(x,y,z)\in\mathcal{H}^3\big) \qquad \lim_{0<\alpha\downarrow 0} \langle z \mid A(x+\alpha y)\rangle = \langle z \mid Ax\rangle. \tag{2.38}$$

Then A is maximally monotone.

Example 2.4 Let $T\colon \mathcal{H} \to \mathcal{H}$ be nonexpansive and let $\alpha \in [-1,1]$. Then $\mathrm{Id}+\alpha T$ is maximally monotone. In particular, set $A = \mathrm{Id} - T$. Then A is maximally monotone and $\operatorname{zer} A = \operatorname{Fix} T$.

Example 2.5 Let $A\colon \mathcal{H} \to \mathcal{H}$ be cocoercive. Then A is maximally monotone.

Example 2.6 Let $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and set $A = J_M$. Then A is maximally monotone and $\operatorname{zer} A = \operatorname{zer} M^{-1}$.

Example 2.7 Let $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\gamma \in \left]0,+\infty\right[$, and set $A = {}^{\gamma}M$ (see (2.21)). Then A is γ-cocoercive, hence maximally monotone, and $\operatorname{zer} A = \operatorname{zer} M$.

Example 2.8 Let $f \in \Gamma_0(\mathcal{H})$ and set $A = \operatorname{prox}_f$. Then A is maximally monotone.

Example 2.9 Let C be a nonempty closed convex subset of $\mathcal{H}$ and set $A = \operatorname{proj}_C$. Then A is maximally monotone.

Example 2.10 Let $A \in \mathcal{B}(\mathcal{H})$ be a *skew* operator, i.e., $A^* = -A$. Then A is maximally monotone.

Here is an elementary example of a maximally monotone set-valued operator on the real line.

Example 2.11 Let $a \in \mathbb{R}$ and $b \in \mathbb{R}$ be such that $a < b$, let $f : [a, b] \to \mathbb{R}$ be increasing (see (2.34)), and define

$$(\forall x \in \mathbb{R}) \quad Ax = \begin{cases} \varnothing, & \text{if } \ x \notin [a,b]; \\ \left]-\infty, f(a)\right], & \text{if } \ x = a; \\ \left[f(b), +\infty\right[, & \text{if } \ x = b; \\ \left[\sup f([a,x[), \inf f(]x,b])\right], & \text{if } \ x \in \left]a,b\right[. \end{cases} \tag{2.39}$$

Then A is maximally monotone.

The following example is a central result in variational methods (see [285, Corollary p. 244] for a special case).

Example 2.12 ([291]) Let $f : \mathcal{H} \to \left]-\infty, +\infty\right]$ be proper. Then the *subdifferential*

$$\partial f : \mathcal{H} \to 2^{\mathcal{H}} : x \mapsto \left\{ x^* \in \mathcal{H} \mid (\forall y \in \mathcal{H}) \ \ \langle y - x \mid x^* \rangle + f(x) \leqslant f(y) \right\} \tag{2.40}$$

of f is monotone and (*Fermat's rule*) $\operatorname{zer} \partial f = \operatorname{Argmin} f$. If $f \in \Gamma_0(\mathcal{H})$, then ∂f is maximally monotone and $(\partial f)^{-1} = \partial f^*$.

Example 2.13 ([334, Theorem 24.3]) Let $A : \mathbb{R} \to 2^{\mathbb{R}}$ be maximally monotone. Then there exists $f \in \Gamma_0(\mathbb{R})$ such that $A = \partial f$.

Example 2.14 Let C be a nonempty convex subset of $\mathcal{H}$. Then, setting $f = \iota_C$ in Example 2.12, we conclude that the *normal cone* operator

$$\begin{aligned} N_C = \partial \iota_C : \mathcal{H} &\to 2^{\mathcal{H}} \\ x &\mapsto \begin{cases} \left\{ x^* \in \mathcal{H} \mid (\forall y \in C) \ \langle y - x \mid x^* \rangle \leqslant 0 \right\}, & \text{if } \ x \in C; \\ \varnothing, & \text{otherwise} \end{cases} \end{aligned} \tag{2.41}$$

of C is monotone and that it is maximally monotone if C is closed, in which case $(N_C)^{-1} = \partial \sigma_C$.

Example 2.15 Let V be a closed vector subspace of $\mathcal{H}$. Then it follows from Example 2.14 that

$$N_V : \mathcal{H} \to 2^{\mathcal{H}} : x \mapsto \begin{cases} V^{\perp}, & \text{if } \ x \in V; \\ \varnothing, & \text{otherwise} \end{cases} \tag{2.42}$$

is maximally monotone and $(N_V)^{-1} = N_{V^{\perp}}$.

The next two examples involve the Laplacian operator and are central to partial differential equations [19, 29, 69, 206, 406].

Example 2.16 ([19, Theorem 17.2.10]) Let Ω be a nonempty bounded open subset of $\mathbb{R}^N$, suppose that $\mathcal{H} = L^2(\Omega)$, and set

$$A\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \begin{cases} -\Delta x, & \text{if } x \in H_0^1(\Omega) \text{ and } \Delta x \in \mathcal{H};\\ \varnothing, & \text{otherwise.} \end{cases} \tag{2.43}$$

Then it follows from Example 2.12 that A is maximally monotone as the subdifferential of the function

$$f\colon \mathcal{H} \to \left]-\infty,+\infty\right] : x \mapsto \begin{cases} \dfrac{1}{2}\displaystyle\int_\Omega \|\nabla x(\omega)\|^2 d\omega, & \text{if } \ x \in H_0^1(\Omega);\\ +\infty, & \text{otherwise,} \end{cases} \tag{2.44}$$

which is in $\Gamma_0(\mathcal{H})$. In addition, if $\operatorname{bdry}\Omega$ is of class $\mathscr{C}^2$, then $\operatorname{dom}\partial f = H^2(\Omega) \cap H_0^1(\Omega)$.

Example 2.17 ([19, Section 17.2.9]) Let Ω be a nonempty bounded open subset of $\mathbb{R}^N$ such that $\operatorname{bdry}\Omega$ is of class $\mathscr{C}^2$, let $\partial/\partial\nu$ denote the outward normal derivative to $\operatorname{bdry}\Omega$, suppose that $\mathcal{H} = L^2(\Omega)$, let $h \in \mathcal{H}$, and set

$$\begin{aligned} A\colon \ & \mathcal{H} \to 2^{\mathcal{H}}\\ & x \ \mapsto \begin{cases} -\Delta x - h, & \text{if } x \in H^2(\Omega) \text{ and } \partial x/\partial\nu = 0 \text{ a.e. on } \operatorname{bdry}\Omega;\\ \varnothing, & \text{otherwise.} \end{cases} \end{aligned} \tag{2.45}$$

Then it follows from Example 2.12 that A is maximally monotone as the subdifferential of the function

$$\begin{aligned} f\colon \ & \mathcal{H} \to \left]-\infty,+\infty\right]\\ & x \ \mapsto \begin{cases} \dfrac{1}{2}\displaystyle\int_\Omega \|\nabla x(\omega)\|^2 d\omega - \int_\Omega x(\omega)h(\omega)d\omega, & \text{if } \ x \in H^1(\Omega);\\ +\infty, & \text{otherwise,} \end{cases} \end{aligned} \tag{2.46}$$

which is in $\Gamma_0(\mathcal{H})$.

The next scenario arises in the study of evolution equations by monotonicity methods [69, 70, 353, 406].

Example 2.18 ([69, Example 4], [353, Chapter IV], [406, Chapter 32]) Let H be a separable real Hilbert space, let $T \in \left]0,+\infty\right[$, and suppose that $\mathcal{H} = L^2([0,T];\mathsf{H})$. For every $y \in \mathcal{H}$, the function $x\colon [0,T] \to \mathsf{H}\colon t \mapsto \int_0^t y(s)ds$ is differentiable a.e. on $\left]0,T\right[$ with $x' = y$ a.e. Define

$$H^1\big([0,T];\mathsf{H}\big) = \big\{x \in \mathcal{H} \mid x' \in L^2\big([0,T];\mathsf{H}\big)\big\}, \tag{2.47}$$

let $\mathsf{x}_0 \in \mathsf{H}$, and set

$$A\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \begin{cases} \{x'\}, & \text{if } x \in H^1\big([0,T];\mathsf{H}\big) \text{ and } x(0) = \mathsf{x}_0; \\ \varnothing, & \text{otherwise} \end{cases} \tag{2.48}$$

and

$$B\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \begin{cases} \{x'\}, & \text{if } x \in H^1\big([0,T];\mathsf{H}\big) \text{ and } x(0) = x(T); \\ \varnothing, & \text{otherwise.} \end{cases} \tag{2.49}$$

Then A and B are maximally monotone.

Example 2.19 ([70, Exemple 2.3.3]) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, let H be a separable real Hilbert space, let $\mathsf{A}\colon \mathsf{H} \to 2^{\mathsf{H}}$ be maximally monotone, and set $\mathcal{H} = L^2((\Omega, \mathcal{F}, \mu); \mathsf{H})$. Define $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ via

$$\begin{aligned}(\forall x \in \mathcal{H})(\forall x^* \in \mathcal{H}) \quad (x, x^*) \in \operatorname{gra} A \quad &\Leftrightarrow \\ \text{for } \mu\text{-almost every } \omega \in \Omega, \quad &\big(x(\omega), x^*(\omega)\big) \in \operatorname{gra} \mathsf{A}\end{aligned} \tag{2.50}$$

and suppose that one of the following holds:

(i) $\mu(\Omega) < +\infty$.

(ii) $\mathsf{0} \in \mathsf{A}\mathsf{0}$.

Then A is maximally monotone.

We now turn to an equilibrium problem in the sense of [56].

Example 2.20 ([13, Theorem 3.5]) Let C be a nonempty closed convex subset of $\mathcal{H}$ and suppose that $F\colon C \times C \to \mathbb{R}$ satisfies the following:

(i) $(\forall x \in C)\ F(x,x) = 0$.

(ii) $(\forall x \in C)(\forall y \in C)\ F(x,y) + F(y,x) \leqslant 0$.

(iii) For every $x \in C$, $F(x,\cdot)\colon C \to \mathbb{R}$ is lower semicontinuous and convex.

(iv) $(\forall x \in C)(\forall y \in C)(\forall z \in C)\ \varlimsup_{0<\varepsilon\to 0} F\big((1-\varepsilon)x + \varepsilon z, y\big) \leqslant F(x,y)$.

Set

$$\begin{aligned}&A\colon \mathcal{H} \to 2^{\mathcal{H}} \\ &\quad x \mapsto \begin{cases} \big\{x^* \in \mathcal{H} \mid (\forall y \in C)\ F(x,y) + \langle x - y \mid x^*\rangle \geqslant 0\big\}, & \text{if } x \in C; \\ \varnothing, & \text{otherwise.} \end{cases}\end{aligned} \tag{2.51}$$

Then A is maximally monotone and $\operatorname{zer} A = \big\{x \in C \mid (\forall y \in C)\ F(x,y) \geqslant 0\big\}$ is the set of *equilibria* of F.

We conclude with an example in the theory of saddle functions.

Example 2.21 ([335, Theorem 3]) Let $F\colon \mathcal{H}\oplus\mathcal{G} \to [-\infty,+\infty]$ be a *saddle function*, i.e., a convex-concave function which is proper and closed in the sense of [335, 336] (for instance, for every $x \in \mathcal{H}$ and every $y \in \mathcal{G}$, $-F(x,\cdot) \in \Gamma_0(\mathcal{G})$ and $F(\cdot,y) \in \Gamma_0(\mathcal{H})$). Set

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{G}) \quad A(x,y) = \partial F(\cdot,y)(x) \times \partial\big(-F(x,\cdot)\big)(y). \tag{2.52}$$

Then A is maximally monotone and

$$\operatorname{zer} A = \big\{(x,y) \in \mathcal{H} \oplus \mathcal{G} \mid F(x,y) = \inf F(\mathcal{H},y) = \sup F(x,\mathcal{G})\big\} \tag{2.53}$$

is the set of *saddle points* of F.

The following illustration is set in the powerful perturbation framework of Rockafellar [333, 335, 338] (see also [238]), which provides a systematic tool to construct duality frameworks in minimization problems.

Example 2.22 Let $\mathcal{V}$ be a real Hilbert space, let $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ be a proper function, and consider the primal problem

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x). \tag{2.54}$$

Let $F\colon \mathcal{H}\oplus\mathcal{V} \to \left]-\infty,+\infty\right]$ be a *perturbation* of f, i.e., $(\forall x \in \mathcal{H})\; f(x) = F(x,0)$. The associated *Lagrangian* is

$$\mathscr{L}_F\colon \mathcal{H} \oplus \mathcal{V} \mapsto [-\infty,+\infty] : (x,v^*) \mapsto \inf_{v\in\mathcal{V}} \big(F(x,v) - \langle v \mid v^*\rangle\big), \tag{2.55}$$

the associated *dual problem* is

$$\underset{v^*\in\mathcal{V}}{\text{minimize}} \;\; \sup_{x\in\mathcal{H}} \big(-\mathscr{L}_F(x,v^*)\big), \tag{2.56}$$

and the associated *saddle operator* is

$$\mathcal{S}_F\colon \mathcal{H}\oplus\mathcal{V} \to 2^{\mathcal{H}\oplus\mathcal{V}}\colon (x,v^*) \mapsto \partial\big(\mathscr{L}_F(\cdot,v^*)\big)(x)\times\partial\big(-\mathscr{L}_F(x,\cdot)\big)(v^*). \tag{2.57}$$

It follows from Example 2.21 that $\mathcal{S}_F$ is maximally monotone. In addition, if $(x,v^*) \in \operatorname{zer}\mathcal{S}_F$, then x solves (2.54) and v^* solves (2.56).

2.4 Basic theory

2.4.1 Operations preserving maximal monotonicity

The examples of Section 2.3 can be combined in various fashions to create maximally monotone operators.

Lemma 2.23 ([37, Proposition 20.22]) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $z \in \mathcal{H}$, let $u \in \mathcal{H}$, and let $\gamma \in \,]0,+\infty[$. Then A^{-1} and $x \mapsto u + \gamma A(x+z)$ are maximally monotone.*

Lemma 2.24 ([37, Proposition 23.18]) *Let $(\mathcal{H}_i)_{i\in I}$ be a finite family of real Hilbert spaces, set*

$$\boldsymbol{\mathcal{H}} = \bigoplus_{i\in I} \mathcal{H}_i, \tag{2.58}$$

and, for every $i \in I$, let $A_i\colon \mathcal{H}_i \to 2^{\mathcal{H}_i}$ be maximally monotone. Set

$$\boldsymbol{A}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}\colon (x_i)_{i\in I} \mapsto \times_{i\in I} A_i x_i. \tag{2.59}$$

Then $\boldsymbol{A}$ is maximally monotone.

Lemma 2.25 *Let $\beta \in \,]0,+\infty[$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, let $U \in \mathcal{B}(\mathcal{H})$ be self-adjoint and β-strongly monotone, and let $\mathcal{X}$ be the real Hilbert space obtained by endowing $\mathcal{H}$ with the scalar product $(x,y) \mapsto \langle Ux \mid y\rangle$. Then the following hold:*

(i) $\operatorname{zer}(U^{-1} \circ A) = \operatorname{zer} A$.

(ii) *Suppose that $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone. Then $U^{-1} \circ A\colon \mathcal{X} \to 2^{\mathcal{X}}$ is maximally monotone.*

(iii) *Let $\alpha \in \,]0,+\infty[$ and suppose that $A\colon \mathcal{H} \to \mathcal{H}$ is α-cocoercive. Then $U^{-1} \circ A\colon \mathcal{X} \to 2^{\mathcal{X}}$ is $\alpha\beta$-cocoercive.*

Proof. (i) is clear and (ii) is proved in [151, Lemma 3.7(i)].

(iii): Take $(x,y) \in \mathcal{H} \times \mathcal{H}$. Then

$$\begin{aligned}\langle x - y \mid (U^{-1} \circ A)x - (U^{-1} \circ A)y\rangle_{\mathcal{X}} &= \langle x - y \mid Ax - Ay\rangle_{\mathcal{H}}\\ &\geqslant \alpha\|Ax - Ay\|^2_{\mathcal{H}}.\end{aligned} \tag{2.60}$$

However, $\|U^{-1}x\|^2_{\mathcal{X}} = \langle x \mid U^{-1}x\rangle_{\mathcal{H}} \leqslant \|U\|^{-1}\,\|x\|^2_{\mathcal{H}}$ and $\|U\|^{-1} \leqslant \beta^{-1}$ [241, Section VI.2.6]. ☐

Lemma 2.26 ([37, Theorem 25.3], [59, Section 24], [308, Corollary 4.2(a)]) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone, let $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$, and suppose that*

$$\operatorname{cone}\big(L(\operatorname{dom} A) - \operatorname{dom} B\big) \text{ is a closed vector subspace of } \mathcal{G}. \tag{2.61}$$

Then $A + L^* \circ B \circ L$ is maximally monotone.

Lemma 2.27 ([37, Corollary 25.5]) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and such that one of the following holds:*

(i) $\operatorname{cone}(\operatorname{dom} A - \operatorname{dom} B)$ *is a closed vector subspace of $\mathcal{H}$.*

(ii) $\operatorname{dom} B = \mathcal{H}$.

(iii) $\operatorname{dom} A \cap \operatorname{int}\operatorname{dom} B \neq \varnothing$.

Then $A + B$ is maximally monotone.

Lemma 2.28 ([8, Theorem 2.1]) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone and such that* $\operatorname{dom} B = \mathcal{H}$ *and $A - B$ is monotone. Then $A - B$ is maximally monotone.*

Lemma 2.29 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Define the* parallel sum *of A and B as*

$$A \,\square\, B = \big(A^{-1} + B^{-1}\big)^{-1} \tag{2.62}$$

and suppose that $\operatorname{cone}(\operatorname{ran} A - \operatorname{ran} B)$ *is a closed vector subspace of $\mathcal{H}$. Then $A \,\square\, B$ is maximally monotone.*

Proof. This follows from (2.25), Lemma 2.23, and Lemma 2.27(i). ☐

Lemma 2.30 ([46, Lemma 2.2]) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$. Define the* parallel composition *of A with L as*

$$L \triangleright A = \big(L \circ A^{-1} \circ L^*\big)^{-1}. \tag{2.63}$$

Suppose that

$$\operatorname{cone}\big(\operatorname{ran} A - L^*(\operatorname{ran} B)\big) \text{ is a closed vector subspace of } \mathcal{H}. \tag{2.64}$$

Then $(L \triangleright A) \,\square\, B$ is a maximally monotone operator from $\mathcal{G}$ to $2^{\mathcal{G}}$.

Example 2.31 ([132, Proposition 4.5(i)–(ii)]) Let $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$ be such that $\|L\| \leqslant 1$ and let $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone. Define the *resolvent composition* of B with L as

$$L \diamond B = L^* \triangleright (B + \mathrm{Id}_{\mathcal{G}}) - \mathrm{Id}_{\mathcal{H}} \tag{2.65}$$

and the *resolvent cocomposition* of B with L as $L \blacklozenge B = (L \diamond B^{-1})^{-1}$. Then $L \diamond B$ and $L \blacklozenge B$ are maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$.

Example 2.32 Let $0 < p \in \mathbb{N}$, let $(\omega_k)_{1 \leqslant k \leqslant p}$ be a family in $]0,1]$ such that $\sum_{k=1}^{p} \omega_k = 1$, and let $(A_k)_{1 \leqslant k \leqslant p}$ be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$. Then the *resolvent average*

$$\left(\sum_{k=1}^{p} \omega_k J_{A_k}\right)^{-1} - \mathrm{Id}_{\mathcal{H}} \tag{2.66}$$

is maximally monotone. This result was originally established in [30, Proposition 2.7] and derived from Example 2.31 in [132, Remark 4.10(ii)].

Example 2.33 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator 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

$$\operatorname{gra} A_V = \big\{(\operatorname{proj}_V x + \operatorname{proj}_{V^\perp} x^*, \operatorname{proj}_V x^* + \operatorname{proj}_{V^\perp} x) \mid (x,x^*) \in \operatorname{gra} A\big\}. \tag{2.67}$$

This construction was introduced in [362], which contains the following (see [362, Section 2]):

(i) A_V is maximally monotone.

(ii) Let $x \in \mathcal{H}$. Then $x \in \operatorname{zer} A_V \Leftrightarrow (\operatorname{proj}_V x, \operatorname{proj}_{V^\perp} x) \in \operatorname{gra} A$.

2.4.2 Resolvent

In terms of solving inclusion problems, the resolvent of (2.17) is the most important operator attached to a monotone operator A. First, as seen in (2.18), it can be employed as a device to generate points in the graph of A. Second, as seen in (2.19), its fixed point set coincides with the set of zeros of A. Third, resolvents provide an effective bridge between the theory of nonexpansive operators and that of monotone operators. This connection goes back to the theory of semigroups of linear nonexpansive operators. The following result, essentially due to Minty [283], establishes such a connection in the nonlinear case. It states in particular that the resolvent of a maximally monotone operator is a firmly nonexpansive operator which is defined everywhere.

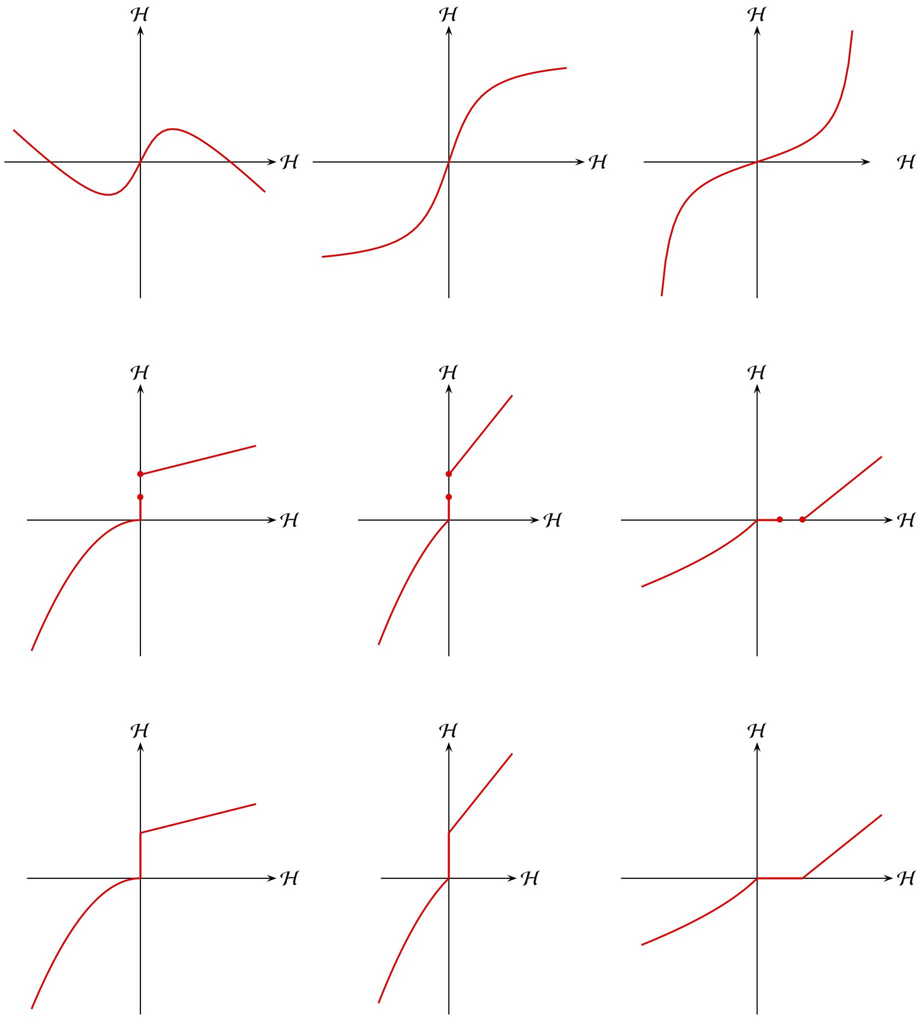

Figure 2.3: Illustration of Minty's theorem (Lemma 2.34). From left to right on each row: graph of A, graph of $\mathrm{Id} + A$, and graph of J_A. Top: A is not monotone: $\operatorname{ran}(\mathrm{Id} + A) = \operatorname{dom} J_A \neq \mathcal{H}$ and J_A is not firmly nonexpansive. Middle: A is monotone but not maximally monotone: J_A is firmly nonexpansive but $\operatorname{ran}(\mathrm{Id}+A) = \operatorname{dom} J_A \neq \mathcal{H}$. Bottom: A is maximally monotone: J_A is firmly nonexpansive with $\operatorname{ran}(\mathrm{Id} + A) = \operatorname{dom} J_A = \mathcal{H}$.

Lemma 2.34 ([37, Proposition 23.8]) *Let D be a nonempty subset of $\mathcal{H}$, let $T\colon D\to\mathcal{H}$, and set $A=T^{-1}-\mathrm{Id}$. Then the following hold (see Figure 2.3):*

(i) $D=\operatorname{ran}(\mathrm{Id}+A)$ *and* $T=J_A$.

(ii) *T is firmly nonexpansive if and only if A is monotone.*

(iii) *T is firmly nonexpansive and $D=\mathcal{H}$ if and only if A is maximally monotone.*

Here are a few examples of resolvents that will be explicitly needed; see [37, 122, 144] for additional examples with closed form expressions and, in particular, instances of proximity operators.

Example 2.35 ([291, Proposition 6.a]) Let $f\in\Gamma_0(\mathcal{H})$. Then $J_{\partial f}=\operatorname{prox}_f$.

Example 2.36 ([290, Exemple p. 2897]) Let C be a nonempty closed convex subset of $\mathcal{H}$. Then $J_{N_C}=\operatorname{prox}_{\iota_C}=\operatorname{proj}_C$.

Example 2.37 ([37, Proposition 23.18]) Let $0<m\in\mathbb{N}$, let $(\mathcal{H}_i)_{1\leqslant i\leqslant m}$ be real Hilbert spaces, set

$$\boldsymbol{\mathcal{H}}=\bigoplus_{i=1}^{m}\mathcal{H}_i, \tag{2.68}$$

and, for every $i\in\{1,\ldots,m\}$, let $A_i\colon\mathcal{H}_i\to 2^{\mathcal{H}_i}$ be maximally monotone. Set

$$\boldsymbol{A}\colon\boldsymbol{\mathcal{H}}\to 2^{\boldsymbol{\mathcal{H}}}\colon (x_i)_{1\leqslant i\leqslant m}\mapsto \bigtimes_{1\leqslant i\leqslant m} A_i x_i. \tag{2.69}$$

Then $\boldsymbol{A}$ is maximally monotone (Lemma 2.24) and

$$J_{\boldsymbol{A}}\colon\boldsymbol{\mathcal{H}}\to\boldsymbol{\mathcal{H}}\colon (x_i)_{1\leqslant i\leqslant m}\mapsto \big(J_{A_i}x_i\big)_{1\leqslant i\leqslant m}. \tag{2.70}$$

Example 2.38 Let $A\colon\mathcal{H}\to 2^{\mathcal{H}}$ be maximally monotone, let V be a closed vector subspace of $\mathcal{H}$, and let A_V be the partial inverse of Example 2.33. In addition, let $x\in\mathcal{H}$ and $p\in\mathcal{H}$. Then

$$p=J_{A_V}x \quad\Leftrightarrow\quad \operatorname{proj}_V p+\operatorname{proj}_{V^\perp}(x-p)=J_A x. \tag{2.71}$$

Proof. This is implicitly in [362, Section 4]; see [7, Lemma 2.2] for a proof. ◻

Example 2.39 ([151, Lemmas 3.7(iii) and 3.1]) As in Lemma 2.25(ii), $A\colon\mathcal{H}\to 2^{\mathcal{H}}$ is maximally monotone, $U\in\mathcal{B}(\mathcal{H})$ is self-adjoint and strongly monotone, and $\mathcal{X}$ is the real Hilbert space obtained by endowing $\mathcal{H}$ with the scalar product $(x,y)\mapsto\langle Ux\mid y\rangle$. Then $J_{U^{-1}\circ A}=(U+A)^{-1}\circ U$.

Example 2.40 ([132, Propositions 1.2 and 4.1(v)]) Let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ be such that $\|L\| \leqslant 1$, let $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone, and consider the resolvent compositions of Example 2.31. Then

$$J_{L \diamond B} = L^* \circ J_B \circ L \quad \text{and} \quad J_{L \blacklozenge B} = \mathrm{Id}_{\mathcal{H}} - L^* \circ L + L^* \circ J_B \circ L. \tag{2.72}$$

2.4.3 Warped resolvents

A generalization of the notion of a resolvent is the following.

Definition 2.41 ([95, Definition 1.1]) Let D be a nonempty subset of $\mathcal{H}$, let $U\colon D \to \mathcal{H}$, and let $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ be such that $\operatorname{ran} U \subset \operatorname{ran}(U+M)$ and $U+M$ is injective. The *warped resolvent* of M with kernel U is $J_M^U = (U+M)^{-1} \circ U\colon D \to D$.

The properties of warped resolvent generalize those of classical ones. In this respect, here is an extension of (2.18)–(2.19).

Lemma 2.42 *Let D and E be nonempty subsets of $\mathcal{H}$, let $U\colon D \to \mathcal{H}$, let $C\colon E \to \mathcal{H}$, and let $W\colon \mathcal{H} \to 2^{\mathcal{H}}$ be such that* $\operatorname{ran} U \subset \operatorname{ran}(U + W + C)$ *and $U + W + C$ is injective. Then the following hold:*

(i) *Let $x \in D$ and $p \in D$. Then $p = J_{W+C}^U x \Leftrightarrow (p, Ux - Up - Cp) \in \operatorname{gra} W$.*

(ii) $\operatorname{Fix} J_{W+C}^U = D \cap \operatorname{zer}(W + C)$.

Proof. Note that $J_{W+C}^U\colon D \to D$ is well defined.

(i): $p = J_{W+C}^U x \Leftrightarrow p = (U + W + C)^{-1}(Ux) \Leftrightarrow Ux \in Up + Wp + Cp \Leftrightarrow Ux - Up - Cp \in Wp$.

(ii): Let $x \in \mathcal{H}$. Then (i) yields $x = J_{W+C}^U x \Leftrightarrow [x \in D \text{ and } (x, -Cx) \in \operatorname{gra} W] \Leftrightarrow [x \in D \text{ and } x \in \operatorname{zer}(W + C)]$. ☐

An instance of a warped resolvent with a linear kernel appears in Example 2.39, where $D = \mathcal{H}$ and $U \in \mathcal{B}(\mathcal{H})$ is a self-adjoint strongly monotone operator. Self-adjoint monotone operators which are not strongly monotone have also been used as kernels; see [65, 392]. The next example features a monotone kernel in $\mathcal{B}(\mathcal{H})$ which is not self-adjoint.

Example 2.43 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone, and suppose that $0 \neq L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Set $\mathbf{X} = \mathcal{H} \oplus \mathcal{G}$ and

$$\begin{cases} \mathcal{K}\colon \mathbf{X} \to 2^{\mathbf{X}}\colon (x, y^*) \mapsto (Ax + L^* y^*) \times (B^{-1} y^* - Lx) \\ U\colon \mathbf{X} \to \mathbf{X}\colon (x, y^*) \mapsto (x - L^* y^*, Lx + y^*). \end{cases} \tag{2.73}$$

As will be seen in Lemma 3.8, $\mathcal{K}$ is the Kuhn–Tucker operator associated with the problem of finding a zero of $A + L^* \circ B \circ L$. It follows from (2.73) that

$$J_{\mathcal{K}}^{\boldsymbol{U}} \colon \mathbf{X} \to \mathbf{X} \colon (x, y^*) \mapsto \big(J_A(x - L^* y^*), J_{B^{-1}}(Lx + y^*)\big), \tag{2.74}$$

whereas $J_{\mathcal{K}}$ is typically intractable.

The next examples employ nonlinear kernels.

Example 2.44 Let $M \colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and such that $\operatorname{zer} M \neq \varnothing$, let $f \colon \mathcal{H} \to \left]-\infty, +\infty\right]$ be a Legendre function such that $\operatorname{dom} M \subset \operatorname{int}\operatorname{dom} f$, and set $D = \operatorname{int}\operatorname{dom} f$ and $U = \nabla f$. Then it follows from [33, Corollary 3.14(ii)] that $J_M^U \colon D \to D$ is a well-defined warped resolvent, called the D-resolvent of M. It is an essential tool in the study of algorithms based on Bregman distances which goes back to [67, 106, 176, 368].

Example 2.45 Let $A \colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B \colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let $f \in \Gamma_0(\mathcal{H})$ be essentially smooth [33]. Suppose that $D = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A$ is a nonempty subset of $\operatorname{int}\operatorname{dom} B$, that B is single-valued on $\operatorname{int}\operatorname{dom} B$, that ∇f is strictly monotone on D, and that $(\nabla f - B)(D) \subset \operatorname{ran}(\nabla f + A)$. Set $M = A + B$ and $U \colon D \to \mathcal{H} \colon x \mapsto \nabla f(x) - Bx$. Then the warped resolvent coincides with the Bregman forward-backward operator $J_M^U = (\nabla f + A)^{-1} \circ (\nabla f - B)$ investigated in [96], where it is shown to capture a construction found in [326] and known as the *auxiliary principle*. In the case when A and B are subdifferentials, J_M^U is the operator studied in [299] and, in Euclidean spaces, in [31]. Scenarios in which J_M^U is more manageable than J_M are discussed in [31, 96, 267, 299, 326, 369].

Example 2.46 Let $A \colon \mathcal{H} \to 2^{\mathcal{H}}$, let $C \colon \mathcal{H} \to \mathcal{H}$ be cocoercive, let $Q \colon \mathcal{H} \to \mathcal{H}$ be monotone and Lipschitzian, and let $\gamma \in \left]0, +\infty\right[$. The underlying problem is to find a point in $\operatorname{zer}(A + C + Q)$ and we recover the *nonlinear forward-backward operator* of [207] as a warped resolvent as follows. Set $M = \gamma(A + C + Q)$, let $K \colon \mathcal{H} \to \mathcal{H}$ be strongly monotone and Lipschitzian, and set $U = K - \gamma(C + Q)$. Then $J_M^U = (K + \gamma A)^{-1} \circ (K - \gamma(C + Q))$, which is the operator driving the algorithms of [207].

Remark 2.47 If B is cocoercive and $f = \|\cdot\|^2/2$ in Example 2.45, or if $K = \mathrm{Id}$ and $Q = 0$ and $C = B$ in Example 2.46, then $J_M^U = J_{\gamma A} \circ (\mathrm{Id} - \gamma B)$. This operator will arise in the forward-backward algorithm of Section 8.

Lemma 2.48 *Let $Q \colon \mathcal{H} \to \mathcal{H}$ be Lipschitzian with constant $\beta \in \left]0, +\infty\right[$, let $K \colon \mathcal{H} \to \mathcal{H}$ be strongly monotone with constant $\alpha \in \left]0, +\infty\right[$, let $\varepsilon \in \left]0, \alpha\right[$, and set $U = K - \gamma Q$. Then the following hold:*

(i) *Let* $\gamma \in \left]0, (\alpha - \varepsilon)/\beta\right]$. *Then* U *is* ε*-strongly monotone.* ([95, Lemma 5.1(i)])

(ii) *Suppose that* $\alpha = 1$ *and* $K = \mathrm{Id}$*, and let* $\gamma \in \left]0, (1 - \varepsilon)/\beta\right]$*, Then* U *is cocoercive with constant* $1/(2 - \varepsilon)$. ([95, Lemma 5.1(ii)])

(iii) *Suppose that* $\alpha = 1$*,* $K = \mathrm{Id}$*, and* Q *is* $1/\beta$*-cocoercive, and let* $\gamma \in \left]0, 2/\beta\right[$*. Then* U *is* $\gamma\beta/2$*-averaged, hence nonexpansive.* ([127, Lemma 2.3])

2.4.4 Topological properties

We record key properties of the graphs of monotone operators.

Lemma 2.49 ([37, Proposition 20.38(ii)]) *Let* $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ *be maximally monotone. Then* $\operatorname{gra} M$ *is sequentially closed in* $\mathcal{H}^{\mathrm{weak}} \times \mathcal{H}^{\mathrm{strong}}$*, i.e., for every sequence* $(x_n, x_n^*)_{n\in\mathbb{N}}$ *in* $\operatorname{gra} M$ *and every* $(x, x^*) \in \mathcal{H} \times \mathcal{H}$*, if* $x_n \rightharpoonup x$ *and* $x_n^* \to x^*$*, then* $(x, x^*) \in \operatorname{gra} M$.

Lemma 2.50 ([37, Corollary 26.6]) *Let* $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ *and* $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ *be maximally monotone, let* $(x_n, x_n^*)_{n\in\mathbb{N}}$ *be a sequence in* $\operatorname{gra} A$*, let* $(y_n, y_n^*)_{n\in\mathbb{N}}$ *be a sequence in* $\operatorname{gra} B$*, let* $x \in \mathcal{H}$*, and let* $x^* \in \mathcal{H}$*. Suppose that*

$$x_n \rightharpoonup x, \quad x_n^* \rightharpoonup x^*, \quad x_n - y_n \to 0, \quad \text{and} \quad x_n^* + y_n^* \to 0. \tag{2.75}$$

Then $x \in \operatorname{zer}(A + B)$*,* $-x^* \in \operatorname{zer}(-A^{-1} \circ (-\mathrm{Id}) + B^{-1})$*,* $(x, x^*) \in \operatorname{gra} A$*, and* $(x, -x^*) \in \operatorname{gra} B$.

2.4.5 Subdifferentials

The subdifferential operator of Example 2.12 is an essential tool in variational analysis.

Lemma 2.51 ([37, Proposition 16.6 and Theorem 16.47(i)]) *Let* $f \in \Gamma_0(\mathcal{H})$*,* $g \in \Gamma_0(\mathcal{G})$*, and* $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ *be such that* $(L(\operatorname{dom} f)) \cap \operatorname{dom} g \neq \varnothing$*. Then the following hold:*

(i) $\operatorname{zer}(\partial f + L^* \circ (\partial g) \circ L) \subset \operatorname{zer} \partial(f + g \circ L) = \operatorname{Argmin}(f + g \circ L)$.

(ii) *Suppose that one of the following is satisfied:*

  (a) $0 \in \operatorname{sri}(L(\operatorname{dom} f) - \operatorname{dom} g)$.

  (b) $L(\operatorname{dom} f) - \operatorname{dom} g$ *is a closed vector subspace of* $\mathcal{G}$.

  (c) $\operatorname{dom} g = \mathcal{G}$.

  (d) $\mathcal{G}$ *is finite-dimensional and* $(\operatorname{ri} L(\operatorname{dom} f)) \cap (\operatorname{ri} \operatorname{dom} g) \neq \varnothing$.

  Then $\partial(f + g \circ L) = \partial f + L^* \circ (\partial g) \circ L$.

3 Structured monotone inclusions

Our master problem is the following two-operator inclusion.

Problem 3.1 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. The objective is to

$$\text{find} \;\; x \in \mathcal{H} \;\; \text{such that} \;\; 0 \in Ax + Bx. \tag{3.1}$$

3.1 Two-operator formulations

We provide problem formulations which correspond to specific choices of the operators A and B in Problem 3.1 from the examples of Section 2.3.

Problem 3.2 In Problem 3.1, let $f \in \Gamma_0(\mathcal{H})$, set $A = \partial f$, and suppose that B is at most single-valued. Then (3.1) reduces to the variational inequality problem [263]

$$\text{find} \;\; x \in \mathcal{H} \;\; \text{such that} \;\; (\forall y \in \mathcal{H}) \;\; \langle x - y \mid Bx \rangle + f(x) \leqslant f(y). \tag{3.2}$$

Problem 3.3 In Problem 3.2, let C be a nonempty closed convex subset of $\mathcal{H}$ and set $f = \iota_C$. Then (3.2) reduces to the standard variational inequality problem [192, 244]

$$\text{find} \;\; x \in C \;\; \text{such that} \;\; (\forall y \in C) \;\; \langle x - y \mid Bx \rangle \leqslant 0. \tag{3.3}$$

Problem 3.4 In Problem 3.3, suppose that C is a cone with dual cone $C^{\oplus}$. Then (3.3) reduces to the *complementarity problem* [190]

$$\text{find} \;\; x \in C \;\; \text{such that} \;\; x \perp Bx \;\; \text{and} \;\; Bx \in C^{\oplus}. \tag{3.4}$$

Problem 3.5 In Problem 3.1, let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$, and set $A = \partial f$ and $B = \partial g$. Suppose that one of the following holds:

(i) $0 \in \operatorname{sri}(\operatorname{dom} f - \operatorname{dom} g)$.

(ii) $g\colon \mathcal{H} \to \mathbb{R}$ is differentiable.

Then the objective is to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + g(x). \tag{3.5}$$

Problem 3.6 In Problem 3.5, let C be a nonempty closed convex subset of $\mathcal{H}$ and set $f = \iota_C$. Suppose that one of the following holds:

(i) $0 \in \operatorname{sri}(C - \operatorname{dom} g)$.

(ii) $g\colon \mathcal{H} \to \mathbb{R}$ is differentiable.

Then the objective is to

$$\underset{x \in C}{\text{minimize}} \;\; g(x). \tag{3.6}$$

3.2 Composite problems

We start by presenting a duality framework for monotone inclusions introduced in [308, 330, 331] (see [5, 23, 180, 196, 197, 278, 293, 329] for special cases).

Problem 3.7 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone, and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. The objective is to solve the primal inclusion

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + L^*\big(B(Lx)\big) \tag{3.7}$$

together with the dual inclusion

$$\text{find } y^* \in \mathcal{G} \text{ such that } 0 \in -L\big(A^{-1}(-L^*y^*)\big) + B^{-1}y^*. \tag{3.8}$$

Lemma 3.8 ([76, Propositions 2.7 and 2.8]) *In the setting of Problem 3.7, let* $\mathbf{X} = \mathcal{H} \oplus \mathcal{G}$, *let* Z *and* Z^* *be the sets of solutions to* (3.7) *and* (3.8), *respectively, and set*

$$\begin{cases} \boldsymbol{M}\colon \mathbf{X} \to 2^{\mathbf{X}}\colon (x, y^*) \mapsto Ax \times B^{-1}y^* \\ \boldsymbol{S}\colon \mathbf{X} \to \mathbf{X}\colon (x, y^*) \mapsto (L^*y^*, -Lx). \end{cases} \tag{3.9}$$

Define the Kuhn–Tucker operator *of Problem 3.7 as*

$$\mathcal{K} = \boldsymbol{M} + \boldsymbol{S} \tag{3.10}$$

and the set of Kuhn–Tucker points *as* $\operatorname{zer}\mathcal{K}$. *Then the following hold:*

(i) $\boldsymbol{M}$ *is maximally monotone.*

(ii) $\boldsymbol{S} \in \mathcal{B}(\mathbf{X})$ *is skew and maximally monotone, with* $\|\boldsymbol{S}\| = \|L\|$.

(iii) $\mathcal{K}$ *is maximally monotone.*

(iv) $\operatorname{zer}\mathcal{K}$ *is a closed convex subset of* $Z \times Z^*$ *in* $\mathbf{X}$.

(v) *(see also [180, 308, 330])* $Z \neq \varnothing \Leftrightarrow \operatorname{zer}\mathcal{K} \neq \varnothing \Leftrightarrow Z^* \neq \varnothing$.

The best known instance for Problem 3.7 is the classical Fenchel–Rockafellar duality framework [332].

Problem 3.9 Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ be such that

$$0 \in \operatorname{sri}\big(L(\operatorname{dom} f) - \operatorname{dom} g\big). \tag{3.11}$$

Set $A = \partial f$ and $B = \partial g$ in Problem 3.7. Then it follows from Lemma 2.51 that (3.7) is the primal problem

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + g(Lx), \tag{3.12}$$

(3.8) is the *Fenchel–Rockafellar dual* problem

$$\underset{y^*\in\mathcal{G}}{\text{minimize}} \;\; f^*(-L^*y^*) + g^*(y^*), \tag{3.13}$$

and (3.10) yields the Kuhn–Tucker operator

$$\mathcal{K}\colon (x,y^*) \mapsto \big(\partial f(x) + L^*y^*\big) \times \big(-Lx + \partial g^*(y^*)\big). \tag{3.14}$$

Problem 3.10 Let V be a closed vector subspace of $\mathcal{H}$ and let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Then, in the case when $\mathcal{G} = \mathcal{H}$ and $L = \mathrm{Id}$, the Kuhn–Tucker operator (3.10) associated with the operators N_V and A is

$$\mathcal{K}\colon \mathcal{H}\oplus\mathcal{H} \to 2^{\mathcal{H}\oplus\mathcal{H}}\colon (x,x^*) \mapsto \big(N_Vx + x^*\big) \times \big(A^{-1}x^* - x\big). \tag{3.15}$$

In view of Example 2.15, the problem of finding a zero of the maximally monotone operator $\mathcal{K}$ reduces to

$$\text{find}\;\; x\in V \;\;\text{and}\;\; x^*\in V^{\perp} \;\;\text{such that}\;\; x^*\in Ax. \tag{3.16}$$

This formulation was first considered by Spingarn in [362].

An extension of Problem 3.7 involving several linearly composed terms is the following.

Problem 3.11 Let $0 < p \in \mathbb{N}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and, for every $k \in \{1\ldots,p\}$, let $\mathcal{G}_k$ be a real Hilbert space, let $B_k\colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ be maximally monotone, and let $L_k \in \mathcal{B}(\mathcal{H},\mathcal{G}_k)$. The objective is to solve the primal inclusion

$$\text{find}\;\; x\in\mathcal{H} \;\;\text{such that}\;\; 0 \in Ax + \sum_{k=1}^{p} L_k^*\big(B_k(L_kx)\big) \tag{3.17}$$

together with the dual inclusion

$$\text{find}\;\; y_1^*\in\mathcal{G}_1,\ldots,y_p^*\in\mathcal{G}_p \;\;\text{such that}$$
$$\left(\exists\, x\in A^{-1}\left(-\sum_{k=1}^{p} L_k^*y_k^*\right)\right)(\forall k\in\{1,\ldots,p\})\;\; L_kx \in B_k^{-1}y_k^*. \tag{3.18}$$

Lemma 3.12 *In the setting of Problem 3.11, set* $\mathbf{X} = \mathcal{H}\oplus\mathcal{G}_1\oplus\cdots\oplus\mathcal{G}_p$ *and let* Z *and* Z^* *be the sets of solutions to (3.17) and (3.18), respectively. Define the* Kuhn–Tucker operator *of Problem 3.11 as*

$$\mathcal{K}\colon \mathbf{X} \to 2^{\mathbf{X}}\colon (x,y_1^*,\ldots,y_p^*) \mapsto$$
$$\left(Ax + \sum_{k=1}^{p} L_k^*y_k^*\right) \times \big(-L_1x + B_1^{-1}y_1^*\big) \times\cdots\times \big(-L_px + B_p^{-1}y_p^*\big) \tag{3.19}$$

and the set of Kuhn–Tucker points *as* $\operatorname{zer}\mathcal{K}$. *Then the following hold:*

(i) $\mathcal{K}$ *is maximally monotone.*

(ii) $\operatorname{zer}\mathcal{K}$ *is a closed convex subset of* $Z \times Z^*$ *in* $\mathbf{X}$.

(iii) $Z \neq \varnothing \Leftrightarrow \operatorname{zer}\mathcal{K} \neq \varnothing \Leftrightarrow Z^* \neq \varnothing$.

Proof. Similar to that of Lemma 3.8. ☐

An alternative angle on Problem 3.9 is provided by the Lagrangian approach of Example 2.22. Set $\boldsymbol{f} : \mathcal{H} \oplus \mathcal{G} \to \left]-\infty,+\infty\right] : \boldsymbol{x} = (x,y) \mapsto f(x) + g(y)$, $\boldsymbol{L} : \mathcal{H} \oplus \mathcal{G} \to \mathcal{G} : (x,y) \mapsto Lx - y$, and $\mathbf{X} = \mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G}$. Then the primal problem (3.12) is equivalent to

$$\underset{\boldsymbol{x} \in \ker \boldsymbol{L}}{\text{minimize}} \;\; \boldsymbol{f}(\boldsymbol{x}) \tag{3.20}$$

and a standard perturbation function for it is [338, Example 4'] (see also [37, Proposition 19.21])

$$\boldsymbol{F} : \mathbf{X} \to \left]-\infty,+\infty\right] : (\boldsymbol{x}, v) \mapsto \boldsymbol{f}(\boldsymbol{x}) + \iota_{\{0\}}(\boldsymbol{L}\boldsymbol{x} + v). \tag{3.21}$$

We derive from (2.55) that the associated Lagrangian is

$$\mathscr{L}_{\boldsymbol{F}} : \mathbf{X} \to \left]-\infty,+\infty\right] : (\boldsymbol{x}, v^*) \mapsto \boldsymbol{f}(\boldsymbol{x}) + \langle \boldsymbol{L}\boldsymbol{x} \mid v^* \rangle, \tag{3.22}$$

from (2.56) that the associated dual problem is (3.13), and from (2.57) that the associated saddle operator is

$$\mathcal{S}_{\boldsymbol{F}} : \mathbf{X} \to 2^{\mathbf{X}} : (\boldsymbol{x}, v^*) \mapsto (\partial \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{L}^* v^*) \times \{-\boldsymbol{L}\boldsymbol{x}\}, \tag{3.23}$$

i.e.,

$$\begin{array}{rccl} \mathcal{S}_{\boldsymbol{F}} : & \mathbf{X} & \to & 2^{\mathbf{X}} \\ & (x,y,v^*) & \mapsto & \big(\partial f(x) + L^* v^*\big) \times \big(\partial g(y) - v^*\big) \times \{-Lx + y\}. \end{array} \tag{3.24}$$

We saw in Example 2.22 that, if $(x,y,v^*) \in \operatorname{zer}\mathcal{S}_{\boldsymbol{F}}$, then x solves the primal problem (3.12) and v^* solves the dual problem (3.13). A version of this result for Problem 3.7 is the following where, although there is no notion of a Lagrangian, we can introduce a saddle operator.

Lemma 3.13 *In the setting of Problem 3.7, set* $\mathbf{X} = \mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G}$ *and let* Z *and* Z^* *be the sets of solutions to (3.7) and (3.8), respectively. Define the Kuhn–Tucker operator* $\mathcal{K}$ *as in (3.10) and define the* saddle operator *of Problem 3.7 as*

$$\begin{array}{rccl} \mathcal{S} : & \mathbf{X} & \to & 2^{\mathbf{X}} \\ & (x,y,v^*) & \mapsto & (Ax + L^* v^*) \times (By - v^*) \times \{-Lx + y\}. \end{array} \tag{3.25}$$

Then the following hold:

(i) *$\mathcal{S}$ is maximally monotone.*

(ii) $\operatorname{zer}\mathcal{S}$ *is closed and convex.*

(iii) *Suppose that* $(x, y, v^*) \in \operatorname{zer}\mathcal{S}$. *Then* $(x, v^*) \in \operatorname{zer}\mathcal{K} \subset Z \times Z^*$.

(iv) $Z^* \neq \varnothing \Leftrightarrow \operatorname{zer}\mathcal{S} \neq \varnothing \Leftrightarrow \operatorname{zer}\mathcal{K} \neq \varnothing \Leftrightarrow Z \neq \varnothing$.

Proof. A special case of [97, Proposition 1(i)–(v)(a)]. □

3.3 Examples of embeddings in Framework 1.2

Example 3.14 Suppose that it is computationally feasible solve Problem 1.1 directly in the original space $\mathcal{H}$. Then an embedding of Problem 1.1 is just $(\mathcal{H}, M, \mathrm{Id})$.

Example 3.15 Let $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $U \in \mathcal{B}(\mathcal{H})$ be a self-adjoint strongly monotone operator, let $\mathbf{X}$ be the real Hilbert space obtained by endowing $\mathcal{H}$ with the scalar product $(x, y) \mapsto \langle Ux \mid y\rangle$, let $\mathcal{M} = U^{-1} \circ M$, and set $\mathcal{T} = \mathrm{Id}$. Then it follows from Lemma 2.25(i)–(ii) that $(\mathbf{X}, \mathcal{M}, \mathcal{T})$ is an embedding of Problem 1.1.

Example 3.16 Let $\alpha \in \left]0, 1\right]$ and let $T\colon \mathcal{H} \to \mathcal{H}$ be α-averaged. In Problem 1.1, suppose that $M = \mathrm{Id} - T$ (see Example 2.4) and set

$$\mathbf{X} = \mathcal{H}, \quad \mathcal{M} = \left(\mathrm{Id} + \frac{1}{2\alpha}(T - \mathrm{Id})\right)^{-1} - \mathrm{Id}, \quad \text{and} \quad \mathcal{T} = \mathrm{Id}. \tag{3.26}$$

Then $(\mathbf{X}, \mathcal{M}, \mathcal{T})$ is an embedding of Problem 1.1. Indeed, since $\mathrm{Id} + \alpha^{-1}(T - \mathrm{Id})$ is nonexpansive, we derive from [37, Proposition 4.4] that $\mathrm{Id} + (2\alpha)^{-1}(T - \mathrm{Id})$ is firmly nonexpansive and hence from Lemma 2.34(iii) that $\mathcal{M}$ is maximally monotone, with $\operatorname{zer}\mathcal{M} = \operatorname{zer} M = \operatorname{Fix} T$.

Example 3.17 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let $\gamma \in \left]0, +\infty\right[$. Let

$$\mathbf{X} = \mathcal{H}, \quad \mathcal{M} = \big(J_{\gamma A} \circ (2J_{\gamma B} - \mathrm{Id}) + \mathrm{Id} - J_{\gamma B}\big)^{-1} - \mathrm{Id}, \quad \text{and} \quad \mathcal{T} = J_{\gamma B}. \tag{3.27}$$

Then it follows from [179, Section 4] that $(\mathbf{X}, \mathcal{M}, \mathcal{T})$ is an embedding of Problem 3.1. In this setting, we actually have $\mathcal{T}(\operatorname{zer}\mathcal{M}) = \operatorname{zer} M$ [127, Lemma 2.6(iii)].

Example 3.18 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Let $\mathbf{X} = \mathcal{H} \oplus \mathcal{H}$, $\mathcal{M}\colon \mathbf{X} \to 2^{\mathbf{X}}\colon (x, x^*) \mapsto (Ax + x^*) \times (-x + B^{-1}x^*)$, and $\mathcal{T}\colon \mathbf{X} \to \mathcal{H}\colon (x, x^*) \mapsto x$. Then applying Lemma 3.8 with $\mathcal{G} = \mathcal{H}$ and $L = \mathrm{Id}$ shows that $(\mathbf{X}, \mathcal{M}, \mathcal{T})$ is an embedding of Problem 3.1. This embedding is implicitly present in the projective splitting algorithm of [181], which is therefore an instance of Framework 1.2.

We now discuss structured inclusion problems that offer greater modeling flexibility by involving three or more operators. The principle of a splitting algorithm, which is to involve each operator individually, faces a serious challenge in the presence of such formulations. Indeed, since inclusion is a binary relation, for reasons discussed in [76, 129] and analyzed in more depth in [346], it is not possible to split problems that involve more than two set-valued operators. A purpose of Framework 1.2 is to circumvent this fundamental limitation by seeking more tractable reformulations in bigger spaces.

Example 3.19 Let $0 < p \in \mathbb{N}$ and, for every $k \in \{1, \ldots, p\}$, let $A_k : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. The problem is to

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in \sum_{k=1}^{p} A_k x. \tag{3.28}$$

Let $\mathbf{X}$ be the p-fold Hilbert direct sum $\mathcal{H}^p$ and set

$$\begin{cases} \boldsymbol{V} = \big\{(x_1, \ldots, x_p) \in \mathbf{X} \mid x_1 = \cdots = x_p\big\} \\ \boldsymbol{A} : \mathbf{X} \to 2^{\mathbf{X}} : (x_1, \ldots, x_p) \mapsto A_1 x_1 \times \cdots \times A_p x_p \\ \mathfrak{M} = \boldsymbol{A} + N_{\boldsymbol{V}} \\ \mathfrak{T} : \mathbf{X} \to \mathcal{H} : (x_1, \ldots, x_p) \mapsto x_1. \end{cases} \tag{3.29}$$

Then

$$\boldsymbol{V}^{\perp} = \left\{ (x_1^*, \ldots, x_p^*) \in \mathbf{X} \;\middle|\; \sum_{k=1}^{p} x_k^* = 0 \right\} \tag{3.30}$$

and it follows from Example 2.15 that $(\mathbf{X}, \mathfrak{M}, \mathfrak{T})$ is an embedding of (3.28). This setting to split the sum of $p > 2$ monotone operators was introduced by Spingarn in [362, Section 5] (see also [218]). It reduces the p-operator problem (3.28) to the two-operator inclusion $\boldsymbol{0} \in \boldsymbol{A}\boldsymbol{x} + N_{\boldsymbol{V}}\boldsymbol{x}$. The idea of rephrasing multi-operator problems in product spaces finds its roots in convex feasibility problems [315, 316], where the problem of finding a point in the intersection $\bigcap_{k=1}^{p} C_k$ of closed convex subsets $(C_k)_{1 \leqslant k \leqslant p}$ of $\mathcal{H}$ is associated with that of finding a point in $\boldsymbol{C} \cap \boldsymbol{V}$ in $\mathbf{X}$, where $\boldsymbol{C} = C_1 \times \cdots \times C_p$.

Example 3.20 In the setting of Problem 3.7, set $\mathbf{X} = \mathcal{H} \oplus \mathcal{G}$, define $\boldsymbol{M}$ and $\boldsymbol{S}$ as in (3.9), let $\mathfrak{K} = \boldsymbol{M} + \boldsymbol{S}$ be the Kuhn–Tucker operator of (3.10), and let $\mathfrak{T} : \mathbf{X} \to \mathcal{H} : (x, y^*) \mapsto x$. Then, in view of Lemma 3.8(iv), $(\mathbf{X}, \mathfrak{K}, \mathfrak{T})$ is an embedding of (3.7). This embedding, which underlies the *monotone+skew* framework of [76], reduces Problem 3.7, which involves three operators in the primal space $\mathcal{H}$ (namely, A, B, and L), to a problem in $\mathbf{X}$ that involves the two operators $\boldsymbol{M}$ and $\boldsymbol{S}$.

Example 3.21 In the setting of Problem 3.11, set $\mathbf{X} = \mathcal{H} \oplus \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p$, let $\mathcal{K}$ be the Kuhn–Tucker operator of (3.19), and let

$$\mathcal{T} \colon \mathbf{X} \to \mathcal{H} \colon (x, y_1^*, \ldots, y_p^*) \mapsto x. \tag{3.31}$$

Then it follows from Lemma 3.12(ii) that $(\mathbf{X}, \mathcal{K}, \mathcal{T})$ is an embedding of (3.17).

Next, we consider an embedding for strongly monotone problems.

Example 3.22 Let $\rho \in \left]0, +\infty\right[$, let $0 < p \in \mathbb{N}$, let $z \in \mathcal{H}$, and let $A \colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. For every $k \in \{1, \ldots, p\}$, let $B_k \colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ and $D_k \colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ be maximally monotone, and suppose that $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$. The problem is to

$$\text{find } x \in \mathcal{H} \text{ such that } z \in Ax + \sum_{k=1}^{p} L_k^* \big((B_k \,\Box\, D_k)(L_k x)\big) + \rho x. \tag{3.32}$$

Let $\mathbf{X} = \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p$, let

$$\begin{aligned} \mathcal{M} \colon \quad \mathbf{X} \quad &\to 2^{\mathbf{X}} \\ (y_1^*, \ldots, y_p^*) &\mapsto \left(-L_1\left(J_{A/\rho}\left(\frac{1}{\rho}\left(z - \sum_{k=1}^{p} L_k^* y_k^*\right)\right)\right) + B_1^{-1} y_1^* + D_1^{-1} y_1^*\right) \\ &\quad \times \cdots \times \left(-L_p\left(J_{A/\rho}\left(\frac{1}{\rho}\left(z - \sum_{k=1}^{p} L_k^* y_k^*\right)\right)\right) + B_p^{-1} y_p^* + D_p^{-1} y_p^*\right), \end{aligned} \tag{3.33}$$

and let

$$\mathcal{T} \colon \mathbf{X} \to \mathcal{H} \colon (y_1^*, \ldots, y_p^*) \mapsto J_{A/\rho}\left(\frac{1}{\rho}\left(z - \sum_{k=1}^{p} L_k^* y_k^*\right)\right). \tag{3.34}$$

Then it follows from [151, Proposition 5.2(iii)] that $(\mathbf{X}, \mathcal{M}, \mathcal{T})$ is an embedding of (3.32).

Our last example concerns an embedding based on a saddle operator.

Example 3.23 In the setting of Problem 3.7, set $\mathbf{X} = \mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G}$, let $\mathcal{S}$ be the saddle operator of (3.25), and let $\mathcal{T} \colon \mathbf{X} \to \mathcal{H} \colon (x, y, v^*) \mapsto x$. Then it follows from Lemma 3.13(iii) that $(\mathbf{X}, \mathcal{S}, \mathcal{T})$ is an embedding of (3.7).

Additional examples of embeddings will be provided by Examples 7.9, 9.8, and 10.4.

4 Two geometric convergence principles

4.1 Overview

The methodology of Framework 1.2 is to identify a target set Z in a suitable Hilbert space in such a way that every point in Z yields a solution to the original problem of interest. The algorithms we shall consider are Fejérian in the sense that every iteration brings the current iterate closer to every point in Z.

4.2 Fejér monotone scheme

Let us first recall some basic facts about weak and strong convergence in Hilbert spaces.

Lemma 4.1 [37, Section 2.5] *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and let $x \in \mathcal{H}$. Then the following hold:*

(i) *Let Z be a nonempty subset of $\mathcal{H}$. Suppose that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset Z$ and that, for every $z \in Z$, $(\|x_n - z\|)_{n\in\mathbb{N}}$ converges. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.*

(ii) $x_n \rightharpoonup x \;\Leftrightarrow\; \big[\, (x_n)_{n\in\mathbb{N}}$ *is bounded and* $\mathfrak{W}(x_n)_{n\in\mathbb{N}} = \{x\} \,\big]$.

(iii) $x_n \to x \;\Leftrightarrow\; \big[\, x_n \rightharpoonup x$ *and* $\varlimsup \|x_n\| \leqslant \|x\| \,\big]$.

Theorem 4.2 *Let Z be a nonempty closed convex subset of $\mathcal{H}$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence of relaxation parameters in $]0,2[$, and let $x_0 \in \mathcal{H}$. Iterate (see Figure 4.1)*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
H_n \textit{ is a closed half-space such that } Z \subset H_n \\
p_n = \operatorname{proj}_{H_n} x_n \\
x_{n+1} = x_n + \lambda_n (p_n - x_n).
\end{array}
\right.
\end{array}
\tag{4.1}
$$

Then the following hold:

(i) *Fejér monotonicity:* $(\forall z \in Z)(\forall n \in \mathbb{N})\ \|x_{n+1} - z\| \leqslant \|x_n - z\|$.

(ii) $\sum_{n\in\mathbb{N}} \lambda_n (2 - \lambda_n) \|p_n - x_n\|^2 < +\infty$.

(iii) *Suppose that* $\sup_{n\in\mathbb{N}} \lambda_n < 2$. *Then* $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$.

(iv) *Suppose that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset Z$. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.*

Proof. Let $z \in Z$. Then, for every $n \in \mathbb{N}$, $H_n = \{u \in \mathcal{H} \mid \langle u - p_n \mid x_n - p_n \rangle \leqslant 0\}$ and, since $z \in H_n$, (4.1) yields

$$\begin{aligned}
\|x_{n+1} - z\|^2 &= \|x_n - z\|^2 + 2\lambda_n \langle x_n - z \mid p_n - x_n \rangle + \lambda_n^2 \|p_n - x_n\|^2 \\
&= \|x_n - z\|^2 - \lambda_n(2-\lambda_n)\|p_n - x_n\|^2 + 2\lambda_n \langle z - p_n \mid x_n - p_n \rangle \\
&\leqslant \|x_n - z\|^2 - \lambda_n(2-\lambda_n)\|p_n - x_n\|^2 \qquad (4.2)\\
&= \|x_n - z\|^2 - \frac{2-\lambda_n}{\lambda_n}\|x_{n+1} - x_n\|^2 \qquad (4.3)\\
&\leqslant \|x_n - z\|^2. \qquad (4.4)
\end{aligned}$$

(i): See (4.4).

(ii): Fix $N \in \mathbb{N}$. Then (4.2) yields

$$\sum_{n=0}^{N} \lambda_n(2-\lambda_n)\|p_n - x_n\|^2 \leqslant \|x_0 - z\|^2 \tag{4.5}$$

and we conclude by letting $N \to +\infty$.

(ii)⇒(iii): This follows from (4.3).

(iv): In view of (i), $(\|x_n - z\|)_{n\in\mathbb{N}}$ converges. The claim therefore follows from Lemma 4.1(i). ☐

Remark 4.3 In 1922, Fejér [191] studied the following problem: given a nonempty closed set $Z \subset \mathbb{R}^N$ and a point $y \notin Z$, can one find a point $x \in \mathbb{R}^N$ such that

$$(\forall z \in Z) \quad \|x - z\| < \|y - z\|. \tag{4.6}$$

This led Motzkin and Schoenberg to adopt in [294] the terminology *Fejér monotone* to describe sequences satisfying property (i) in Theorem 4.2. In their paper (see also [3]), an algorithm was developed to solve systems of linear inequalities in $\mathbb{R}^N$ by successive projections onto the half-spaces defining the polyhedral solution set Z, and Fejér monotonicity was shown to be an adequate tool to study the convergence of this algorithm. Further analysis of Fejér monotonicity was proposed in [66, 186, 187, 324, 325] and nowadays it constitutes a central tool to analyze the asymptotic behavior of various algorithms [37].

Remark 4.4 In general, the convergence of $(x_n)_{n\in\mathbb{N}}$ to $x \in Z$ in Theorem 4.2(iv) is only weak and, even if it were strong, there exists no rate of convergence on $(\|x_n - x\|)_{n\in\mathbb{N}}$, even in Euclidean spaces [39, 220, 397]. In particular, achieving a linear rate of convergence, that is, securing the existence of $\kappa \in\,]0, +\infty[$ and $\rho \in\,]0, 1[$ such that

$$(\forall n \in \mathbb{N}) \quad \|x_n - x\| \leqslant \kappa\rho^n, \tag{4.7}$$

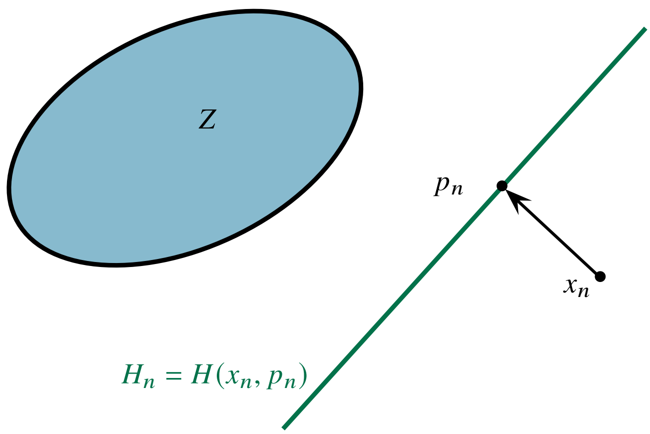

Figure 4.1: Iteration n of the Fejérian algorithm (4.1).

requires stringent additional assumptions on the problem. In our inclusion context, a typical assumption is strong monotonicity; see [37, Proposition 26.16] for an example. In the broader context of Theorem 4.2(i), it is clear that $(d_C(x_n))_{n\in\mathbb{N}}$ decreases and that, for every $n \in \mathbb{N}$ and $m \in \mathbb{N}$, $\|x_n - x_{n+m}\| \leqslant \|x_n - \text{proj}_C\, x_n\| + \|x_{n+m} - \text{proj}_C\, x_n\| \leqslant 2d_C(x_n)$. Hence, (4.7) will hold with $\kappa = 2d_C(x_0)$ if the decreasing property can be strengthened to $(\forall n \in \mathbb{N})\ d_C(x_{n+1}) \leqslant \rho d_C(x_n)$.

Remark 4.5 The implementation of (4.1) is said to be *unrelaxed* if $(\forall n \in \mathbb{N})$ $\lambda_n = 1$.

4.3 Haugazeau-like scheme

Theorem 4.2 guarantees only weak convergence to an unspecified point in Z and, as will be seen on several occasions later, strong convergence fails in general (many of these examples will be based on a scenario of [230] concerning the method of alternating projections). However, in some infinite-dimensional applications in areas such as inverse problems, control, mechanics, PDEs, optics, and analog computing, weak convergence does not offer sufficient guarantees and strong convergence is required. The geometric approach described in this section emanates from ideas found in the work of Haugazeau on the convex feasibility problem [224, 225]. It will provide strong convergence to a specific point in Z, namely the projection of the initial point onto Z. This means that the resulting algorithm is also of interest, even in Euclidean spaces, as a best approximation method.

The following technical fact will be employed repeatedly.

Lemma 4.6 ([225, Théorème 3-1]; see also [37, Corollary 29.25]) *Let* $(x, y, z) \in \mathcal{H}^3$. *Define*

$$H(x, y) = \big\{z \in \mathcal{H} \mid \langle z - y \mid x - y\rangle \leqslant 0\big\}, \tag{4.8}$$

$C = H(x, y) \cap H(y, z)$*, and, if* $C \neq \varnothing$,

$$\mathsf{Q}(x, y, z) = \operatorname{proj}_C x. \tag{4.9}$$

Set $\chi = \langle x - y \mid y - z\rangle$, $\mu = \|x - y\|^2$, $\nu = \|y - z\|^2$, *and* $\rho = \mu\nu - \chi^2$. *Then exactly one of the following holds:*

(i) $\rho = 0$ *and* $\chi < 0$, *in which case* $C = \varnothing$.

(ii) $[\,\rho = 0$ and $\chi \geqslant 0\,]$ or $\rho > 0$, in which case $C \neq \varnothing$ and

$$\mathsf{Q}(x, y, z) = \begin{cases} z, & \text{if } \rho = 0 \text{ and } \chi \geqslant 0;\\ x + (1 + \chi/\nu)(z - y), & \text{if } \rho > 0 \text{ and } \chi\nu \geqslant \rho;\\ y + (\nu/\rho)\big(\chi(x - y) + \mu(z - y)\big), & \text{if } \rho > 0 \text{ and } \chi\nu < \rho. \end{cases} \tag{4.10}$$

The essential components of the following theorem are found in the unpublished thesis of Haugazeau [225] (see [224] for a preliminary variant), where he considered the specific problem of projecting a point onto the intersection of finitely many sets using their individual projection operators cyclically.

Theorem 4.7 *Let* Z *be a nonempty closed convex subset of* $\mathcal{H}$, *let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence of relaxation parameters in* $]0, 1]$, *and let* $x_0 \in \mathcal{H}$. *Iterate (see Figure 4.2)*

$$\begin{array}{l} \textit{for } n = 0, 1, \ldots\\ \left\lfloor \begin{array}{l} H_n \textit{ is a closed half-space such that } Z \subset H_n\\ p_n = \operatorname{proj}_{H_n} x_n\\ r_n = x_n + \lambda_n(p_n - x_n)\\ x_{n+1} = \mathsf{Q}(x_0, x_n, r_n). \end{array} \right. \end{array} \tag{4.11}$$

Then the sequence $(x_n)_{n\in\mathbb{N}}$ *is well defined and the following hold:*

(i) $(\forall n \in \mathbb{N})\ Z \subset H(x_0, x_n) \cap H(x_n, r_n)$.

(ii) $(\exists\, \ell \in [0, +\infty[)\ \|x_n - x_0\| \uparrow \ell \leqslant d_Z(x_0)$.

(iii) $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$.

(iv) $\sum_{n\in\mathbb{N}} \lambda_n^2 \|p_n - x_n\|^2 < +\infty$.

(v) *Suppose that* $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset Z$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly to* $\operatorname{proj}_Z x_0$.

Proof. First, recall that the projector onto a nonempty closed convex subset D of $\mathcal{H}$ is characterized by [37, Theorem 3.16]

$$(\forall x \in \mathcal{H}) \quad \operatorname{proj}_D x \in D \quad \text{and} \quad D \subset H(x, \operatorname{proj}_D x). \tag{4.12}$$

We also observe that (4.11) implies that

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad H(x_n, p_n)&\\ &= \big\{z \in \mathcal{H} \mid \langle z - p_n \mid x_n - r_n\rangle \leqslant 0\big\}\\ &= \big\{z \in \mathcal{H} \mid \langle z - r_n \mid x_n - r_n\rangle \leqslant \langle p_n - r_n \mid x_n - r_n\rangle\big\}\\ &= \big\{z \in \mathcal{H} \mid \langle z - r_n \mid x_n - r_n\rangle \leqslant -\lambda_n(1-\lambda_n)\|x_n - p_n\|^2\big\}\\ &\subset H(x_n, r_n).\end{aligned} \tag{4.13}$$

(i): Let $n \in \mathbb{N}$ be such that x_n exists. It follows from (4.11) and (4.13) that $Z \subset H_n = H(x_n, p_n) \subset H(x_n, r_n)$. It is therefore enough to show that $Z \subset H(x_0, x_n)$. This inclusion certainly holds for $n = 0$ since $H(x_0, x_0) = \mathcal{H}$. Furthermore, it follows from (4.12) and (4.11) that

$$\begin{aligned} Z \subset H(x_0, x_n) \quad &\Rightarrow \quad Z \subset H(x_0, x_n) \cap H(x_n, r_n)\\ &\Rightarrow \quad Z \subset H\big(x_0, \mathsf{Q}(x_0, x_n, r_n)\big)\\ &\Leftrightarrow \quad Z \subset H(x_0, x_{n+1}),\end{aligned} \tag{4.14}$$

which establishes the assertion by induction. This also shows that $H(x_0, x_n) \cap H(x_n, r_n) \neq \varnothing$ and hence that x_{n+1} is well defined.

(ii)–(iii): Let $n \in \mathbb{N}$. By construction, $x_{n+1} = \mathsf{Q}(x_0, x_n, r_n) \in H(x_0, x_n) \cap H(x_n, r_n)$. Consequently, since x_n is the projection of x_0 onto $H(x_0, x_n)$ and $x_{n+1} \in H(x_0, x_n)$, we have $\|x_0 - x_n\| \leqslant \|x_0 - x_{n+1}\|$. On the other hand, since $\operatorname{proj}_Z x_0 \in Z \subset H(x_0, x_n)$, we have $\|x_0 - x_n\| \leqslant \|x_0 - \operatorname{proj}_Z x_0\|$. It follows that $(\|x_0 - x_k\|)_{k\in\mathbb{N}}$ converges to some $\ell \in [0, \|x_0 - \operatorname{proj}_Z x_0\|]$, which establishes (ii), and that

$$\lim \|x_0 - x_k\| \leqslant \|x_0 - \operatorname{proj}_Z x_0\|. \tag{4.15}$$

However, since $x_{n+1} \in H(x_0, x_n)$, we have

$$\begin{aligned}\|x_{n+1} - x_n\|^2 &\leqslant \|x_{n+1} - x_n\|^2 + 2\langle x_{n+1} - x_n \mid x_n - x_0\rangle\\ &= \|x_0 - x_{n+1}\|^2 - \|x_0 - x_n\|^2.\end{aligned} \tag{4.16}$$

Hence,

$$\sum_{k=0}^{n} \|x_{k+1} - x_k\|^2 \leqslant \|x_0 - x_{n+1}\|^2 \leqslant \|x_0 - \operatorname{proj}_Z x_0\|^2 \tag{4.17}$$

and therefore

$$\sum_{k\in\mathbb{N}} \|x_{k+1} - x_k\|^2 < +\infty. \tag{4.18}$$

(iv): For every $n \in \mathbb{N}$, we derive from the inclusion $x_{n+1} \in H(x_n, r_n)$ that

$$\begin{aligned}\|r_n - x_n\|^2 &\leqslant \|x_{n+1} - r_n\|^2 + \|x_n - r_n\|^2 \\ &\leqslant \|x_{n+1} - r_n\|^2 + 2\langle x_{n+1} - r_n \mid r_n - x_n\rangle + \|x_n - r_n\|^2 \\ &= \|x_{n+1} - x_n\|^2.\end{aligned} \tag{4.19}$$

Hence, by (iii) and (4.11),

$$\sum_{n\in\mathbb{N}} \lambda_n^2 \|p_n - x_n\|^2 = \sum_{n\in\mathbb{N}} \|r_n - x_n\|^2 < +\infty. \tag{4.20}$$

(v): Let us note that (ii) implies that $(x_n)_{n\in\mathbb{N}}$ is bounded. Now let $x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$. Then, by weak lower semicontinuity of $\|\cdot\|$ [37, Lemma 2.42] and (ii),

$$\|x_0 - x\| \leqslant \underline{\lim} \|x_0 - x_{k_n}\| \leqslant \|x_0 - \operatorname{proj}_Z x_0\| = \inf_{z\in Z} \|x_0 - z\|. \tag{4.21}$$

Hence, since $x \in Z$, $x = \operatorname{proj}_Z x_0$ is the only weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ and it follows from Lemma 4.1(ii) that $x_n \rightharpoonup \operatorname{proj}_Z x_0$. In turn, (ii) yields

$$\|x_0 - \operatorname{proj}_Z x_0\| \leqslant \underline{\lim} \|x_0 - x_n\| = \lim \|x_0 - x_n\| \leqslant \|x_0 - \operatorname{proj}_Z x_0\|. \tag{4.22}$$

Thus, $x_0 - x_n \rightharpoonup x_0 - \operatorname{proj}_Z x_0$ and $\|x_0 - x_n\| \to \|x_0 - \operatorname{proj}_Z x_0\|$. We therefore derive from Lemma 4.1(iii) that $x_0 - x_n \to x_0 - \operatorname{proj}_Z x_0$, i.e., $x_n \to \operatorname{proj}_Z x_0$. □

4.4 Graph-based cuts

We consider the problem of finding a zero of a maximally monotone operator $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ decomposed as $M = W + C$, where $W\colon \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone and $C\colon \mathcal{H} \to \mathcal{H}$ is cocoercive, using the geometric principles of Theorems 4.2 and 4.7. To this end, we shall construct half-spaces by selecting points in the graph of W. Let us start with a weak convergence result.

Theorem 4.8 *Let $\alpha \in \,]0, +\infty[$, let $W\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $C\colon \mathcal{H} \to \mathcal{H}$ be α-cocoercive and such that $Z = \operatorname{zer}(W + C) \neq \varnothing$, let $x_0 \in \mathcal{H}$, and*

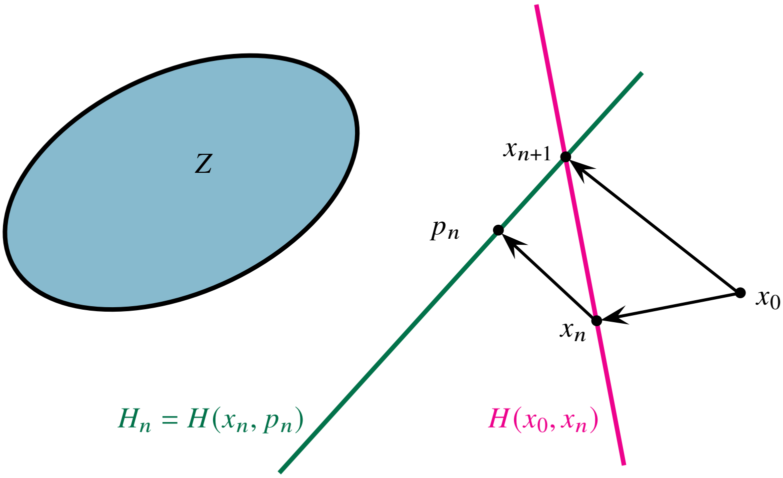

Figure 4.2: Iteration n of the Haugazeau-like algorithm (4.11) with $\lambda_n = 1$.

let $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,2[$. *Iterate*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
(w_n, w_n^*) \in \operatorname{gra} W, \;\; q_n \in \mathcal{H} \\
t_n^* = w_n^* + Cq_n \\
\delta_n = \langle x_n - w_n \mid t_n^* \rangle - \|w_n - q_n\|^2/(4\alpha) \\
d_n = \begin{cases} \dfrac{\delta_n}{\|t_n^*\|^2} t_n^*, & \textit{if } \delta_n > 0; \\ 0, & \textit{otherwise} \end{cases} \\
x_{n+1} = x_n - \lambda_n d_n.
\end{array}
\right.
\end{array}
\tag{4.23}
$$

Then the following hold:

(i) $(x_n)_{n\in\mathbb{N}}$ *is bounded.*

(ii) $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n)\|d_n\|^2 < +\infty$.

(iii) *Suppose that* $w_n - x_n \rightharpoonup 0$, $w_n - q_n \to 0$, *and* $t_n^* \to 0$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* Z.

Proof. We first observe that (4.23) is well defined since $(\forall n \in \mathbb{N})\ \delta_n > 0 \Rightarrow t_n^* \neq 0$. It follows from Example 2.5 and Lemma 2.27(ii) that

$$
W + C \text{ is maximally monotone,} \tag{4.24}
$$

and hence from (2.29) that Z is a nonempty closed convex subset of $\mathcal{H}$. Set

$$(\forall n \in \mathbb{N}) \quad H_n = \left\{ z \in \mathcal{H} \;\middle|\; \langle z - w_n \mid t_n^* \rangle \leqslant \frac{\|w_n - q_n\|^2}{4\alpha} \right\} \tag{4.25}$$

and let $z \in Z$. For every $n \in \mathbb{N}$, since $(z, -Cz) \in \operatorname{gra} W$ and $(w_n, w_n^*) \in \operatorname{gra} W$, it results from the monotonicity of W that $\langle w_n - z \mid w_n^* + Cz \rangle \geqslant 0$. Hence, since C is α-cocoercive,

$$\begin{aligned}
(\forall n \in \mathbb{N}) \quad & \langle z - w_n \mid t_n^* \rangle \\
&= \langle z - w_n \mid w_n^* + Cq_n \rangle \\
&\leqslant \langle z - w_n \mid Cq_n - Cz \rangle \qquad (4.26)\\
&= \langle q_n - w_n \mid Cq_n - Cz \rangle + \langle z - q_n \mid Cq_n - Cz \rangle \\
&\leqslant \langle q_n - w_n \mid Cq_n - Cz \rangle - \alpha \|Cq_n - Cz\|^2 \qquad (4.27)\\
&= 2\left\langle \frac{q_n - w_n}{\sqrt{4\alpha}} \;\middle|\; \sqrt{\alpha}(Cq_n - Cz) \right\rangle - \left\|\sqrt{\alpha}(Cq_n - Cz)\right\|^2 \\
&= \frac{\|w_n - q_n\|^2}{4\alpha} - \left\| \sqrt{\alpha}(Cq_n - Cz) + \frac{w_n - q_n}{\sqrt{4\alpha}} \right\|^2 \\
&\leqslant \frac{\|w_n - q_n\|^2}{4\alpha}. \qquad (4.28)
\end{aligned}$$

This shows that $(\forall n \in \mathbb{N})\ Z \subset H_n$. In addition, it results from (4.23) and Example 2.1 that

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \big(\operatorname{proj}_{H_n} x_n - x_n\big), \tag{4.29}$$

which corresponds to the setting of Theorem 4.2.

(i): This follows from Theorem 4.2(i).

(ii): This follows from Theorem 4.2(ii).

(iii): Let $x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$. Then $w_{k_n} = x_{k_n} + (w_{k_n} - x_{k_n}) \rightharpoonup x$. On the other hand, since C is $1/\alpha$-Lipschitzian,

$$\|w_n^* + Cw_n\| = \|t_n^* + Cw_n - Cq_n\| \leqslant \|t_n^*\| + \frac{\|w_n - q_n\|}{\alpha} \to 0. \tag{4.30}$$

In addition, since $(w_n, w_n^*)_{n\in\mathbb{N}}$ is in $\operatorname{gra} W$, $(w_n, w_n^* + Cw_n)_{n\in\mathbb{N}}$ is in $\operatorname{gra}(W + C)$. It then follows from (4.24) and Lemma 2.49 that $x \in Z$. We conclude by invoking Theorem 4.2(iv). ☐

We now turn to strong convergence.

Theorem 4.9 *Let $\alpha \in]0,+\infty[$, let $W\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $C\colon \mathcal{H} \to \mathcal{H}$ be α-cocoercive and such that $Z = \operatorname{zer}(W+C) \neq \varnothing$, let $x_0 \in \mathcal{H}$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1]$. Iterate*

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
(w_n, w_n^*) \in \operatorname{gra} W,\;\; q_n \in \mathcal{H}\\
t_n^* = w_n^* + Cq_n\\
\delta_n = \langle x_n - w_n \mid t_n^* \rangle - \|w_n - q_n\|^2/(4\alpha)\\
d_n = \begin{cases} \dfrac{\delta_n}{\|t_n^*\|^2} t_n^*, & \text{if } \delta_n > 0;\\ 0, & \text{otherwise} \end{cases}\\
r_n = x_n - \lambda_n d_n\\
x_{n+1} = \mathsf{Q}(x_0, x_n, r_n),
\end{array}
\right.
\end{array}
\tag{4.31}
$$

where Q is defined in Lemma 4.6. Then the following hold:

(i) *$(x_n)_{n\in\mathbb{N}}$ is bounded.*

(ii) $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$.

(iii) $\sum_{n\in\mathbb{N}} \lambda_n^2 \|d_n\|^2 < +\infty$.

(iv) *Suppose that $w_n - x_n \rightharpoonup 0$, $w_n - q_n \to 0$, and $t_n^* \to 0$. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to $\operatorname{proj}_Z x_0$.*

Proof. Define $(H_n)_{n\in\mathbb{N}}$ as in (4.25) and note that (4.28) yields $Z \subset \bigcap_{n\in\mathbb{N}} H_n$. Furthermore, we derive from (4.31) and Example 2.1 that $(\forall n \in \mathbb{N})$ $r_n = x_n + \lambda_n(\operatorname{proj}_{H_n} x_n - x_n)$. This places us in the setting of Theorem 4.7.

(i): This follows from Theorem 4.7(ii).

(ii): See Theorem 4.7(iii).

(iii): This follows from Theorem 4.7(iv).

(iv): As in the proof of Theorem 4.8(iii), $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset Z$. The claim follows from Theorem 4.7(v). ☐

In the absence of the cocoercive operator C, we can choose $(q_n)_{n\in\mathbb{N}} = (w_n)_{n\in\mathbb{N}}$ in (4.23) and (4.31), and Theorems 4.8 and 4.9 simplify as follows.

Proposition 4.10 *Let $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that $Z = \operatorname{zer} M \neq \varnothing$, let $x_0 \in \mathcal{H}$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$. Iterate*

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
(m_n, m_n^*) \in \operatorname{gra} M\\
d_n = \begin{cases} \dfrac{\langle x_n - m_n \mid m_n^* \rangle}{\|m_n^*\|^2} m_n^*, & \text{if } \langle x_n - m_n \mid m_n^* \rangle > 0;\\ 0, & \text{otherwise} \end{cases}\\
x_{n+1} = x_n - \lambda_n d_n.
\end{array}
\right.
\end{array}
\tag{4.32}
$$

Then the following hold:

(i) $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n)\|d_n\|^2 < +\infty$.

(ii) *Suppose that* $m_n - x_n \rightharpoonup 0$ *and* $m_n^* \to 0$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* Z.

Proposition 4.11 *Let* $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ *be a maximally monotone operator such that* $Z = \operatorname{zer} M \neq \varnothing$, *let* $x_0 \in \mathcal{H}$, *and let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,1]$. *Iterate*

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
(m_n, m_n^*) \in \operatorname{gra} M\\
d_n = \begin{cases} \dfrac{\langle x_n - m_n \mid m_n^*\rangle}{\|m_n^*\|^2} m_n^*, & \text{if } \langle x_n - m_n \mid m_n^*\rangle > 0;\\ 0, & \text{otherwise}\end{cases}\\
r_n = x_n - \lambda_n d_n\\
x_{n+1} = \mathsf{Q}(x_0, x_n, r_n),
\end{array}
\right.
\end{array}
\tag{4.33}
$$

where Q *is defined in Lemma 4.6. Then the following hold:*

(i) $\sum_{n\in\mathbb{N}} \lambda_n^2\|d_n\|^2 < +\infty$.

(ii) *Suppose that* $m_n - x_n \rightharpoonup 0$ *and* $m_n^* \to 0$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges to strongly to* $\operatorname{proj}_Z x_0$.

4.5 Warped resolvent cuts

Algorithms (4.23) and (4.31) are conceptual in the sense that they do not provide an explicit mechanism to find points in the graph of W. In this section, we propose implementable versions that pick points in $\operatorname{gra} W$ using the warped resolvents of Lemma 2.42.

Theorem 4.12 *Let* $\alpha \in\,]0,+\infty[$, *let* $W\colon \mathcal{H} \to 2^{\mathcal{H}}$ *be maximally monotone, let* $C\colon \mathcal{H} \to \mathcal{H}$ *be* α*-cocoercive and such that* $Z = \operatorname{zer}(W + C) \neq \varnothing$, *let* $x_0 \in \mathcal{H}$, *and let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,2[$. *Further, for every* $n \in \mathbb{N}$, *let* $U_n\colon \mathcal{H} \to \mathcal{H}$ *be an operator such that* $\operatorname{ran} U_n \subset \operatorname{ran}(U_n + W + C)$ *and* $U_n + W + C$ *is injective.*

Iterate

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
w_n = J_{W+C}^{U_n} x_n \\
w_n^* = U_n x_n - U_n w_n - C w_n \\
q_n \in \mathcal{H} \\
t_n^* = w_n^* + C q_n \\
\delta_n = \langle x_n - w_n \mid t_n^* \rangle - \|w_n - q_n\|^2/(4\alpha) \\
d_n = \begin{cases} \dfrac{\delta_n}{\|t_n^*\|^2} t_n^*, & \textit{if } \delta_n > 0; \\ 0, & \textit{otherwise} \end{cases} \\
x_{n+1} = x_n - \lambda_n d_n.
\end{array}
\right.
\end{array}
\tag{4.34}
$$

Then the following hold:

(i) $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n)\|d_n\|^2 < +\infty$.

(ii) *Suppose that one of the following is satisfied:*

(a) $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$ *and* $(\|d_n\|)_{n\in\mathbb{N}}$ *converges;*

(b) $\inf_{n\in\mathbb{N}} \lambda_n > 0$ *and* $\sup \lambda_n < 2$*;*

together with one of the following:

(c) $w_n - x_n \rightharpoonup 0$*,* $U_n w_n - U_n x_n \to 0$*, and* $w_n - q_n \to 0$*;*

(d) $q_n - x_n \to 0$ *and there exist* $\beta_1 \in \,]1/(4\alpha), +\infty[$ *and* $\beta_2 \in \,]0, +\infty[$ *such that the kernels* $(U_n)_{n\in\mathbb{N}}$ *are* β_1*-strongly monotone and* β_2*-Lipschitzian.*

Then $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* Z*.*

Proof. Lemma 2.42(i) indicates that (4.34) is governed by the scenario of Theorem 4.8.

(i): See Theorem 4.8(ii).

(ii): A consequence of (i) under (ii)(a) or (ii)(b) is that

$$\|d_n\| \to 0. \tag{4.35}$$

Indeed, the claim is clear under (ii)(b) whereas, under (ii)(a), we have $\underline{\lim}\, \|d_n\| = 0$ and therefore $\lim \|d_n\| = 0$. Next, let us assume that (ii)(c) holds. Then it follows from (4.34) and (2.32) that

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad \|t_n^*\| &= \|U_n w_n - U_n x_n + C w_n - C q_n\| \\
&\leqslant \|U_n w_n - U_n x_n\| + \|C w_n - C q_n\| \qquad (4.36)\\
&\leqslant \|U_n w_n - U_n x_n\| + \frac{\|w_n - q_n\|}{\alpha} \\
&\to 0. \qquad (4.37)
\end{aligned}
$$

In view of Theorem 4.8(iii), the claim is established. It remains to show that (ii)(d)$\Rightarrow$(ii)(c). Because the operators $(U_n + W + C)_{n\in\mathbb{N}}$ are β_1-strongly monotone, the operators $(U_n + W + C)^{-1}_{n\in\mathbb{N}}$ are β_1-cocoercive, hence $1/\beta_1$-Lipschitzian. Consequently, since the operators $(U_n)_{n\in\mathbb{N}}$ are β_2-Lipschitzian, the operators $(J^{U_n}_{W+C})_{n\in\mathbb{N}}$ are β_2/β_1-Lipschitzian. Now let $z \in Z$. Then we derive from (4.34) and Lemma 2.42(ii) that

$$(\forall n \in \mathbb{N})\quad \|w_n - z\| = \left\|J^{U_n}_{W+C}x_n - J^{U_n}_{W+C}z\right\| \leqslant \frac{\beta_2}{\beta_1}\|x_n - z\|. \tag{4.38}$$

Appealing to Theorem 4.8(i), we infer that $(w_n)_{n\in\mathbb{N}}$ is bounded. Thus, since $q_n - x_n \to 0$ and C is $1/\alpha$-Lipschitzian, the sequences

$$(\|w_n - x_n\|)_{n\in\mathbb{N}},\ (\|w_n - q_n\|)_{n\in\mathbb{N}},\ \text{and } (\|Cw_n - Cq_n\|)_{n\in\mathbb{N}} \text{ are bounded.} \tag{4.39}$$

However, (4.36) entails that

$$(\forall n \in \mathbb{N})\quad \|t^*_n\| \leqslant \beta_2\|w_n - x_n\| + \frac{\|w_n - q_n\|}{\alpha}, \tag{4.40}$$

which verifies that $(\|t^*_n\|)_{n\in\mathbb{N}}$ is bounded. In turn, (4.34) and (4.35) imply that

$$\overline{\lim}\,\delta_n \leqslant \lim \|t^*_n\|\,\|d_n\| = 0. \tag{4.41}$$

Moreover, for every $n \in \mathbb{N}$, (4.34) yields

$$\begin{aligned}
\delta_n &= \langle w_n - x_n \mid U_n w_n - U_n x_n\rangle + \langle w_n - x_n \mid Cw_n - Cq_n\rangle - \frac{\|w_n - q_n\|^2}{4\alpha}\\
&\geqslant \beta_1\|w_n - x_n\|^2 + \langle w_n - q_n \mid Cw_n - Cq_n\rangle + \langle q_n - x_n \mid Cw_n - Cq_n\rangle\\
&\quad - \frac{\|w_n - q_n\|^2}{4\alpha}\\
&\geqslant \beta_1\Big(\|w_n - q_n\|^2 + 2\langle w_n - q_n \mid q_n - x_n\rangle + \|q_n - x_n\|^2\Big)\\
&\quad + \alpha\|Cw_n - Cq_n\|^2 + \langle q_n - x_n \mid Cw_n - Cq_n\rangle - \frac{\|w_n - q_n\|^2}{4\alpha}\\
&\geqslant \left(\beta_1 - \frac{1}{4\alpha}\right)\|w_n - q_n\|^2 + \beta_1\Big(2\langle w_n - q_n \mid q_n - x_n\rangle + \|q_n - x_n\|^2\Big)\\
&\quad + \langle q_n - x_n \mid Cw_n - Cq_n\rangle\\
&\geqslant \left(\beta_1 - \frac{1}{4\alpha}\right)\|w_n - q_n\|^2\\
&\quad + \|q_n - x_n\|\big(\beta_1\|q_n - x_n\| - 2\beta_1\|w_n - q_n\| + \|Cw_n - Cq_n\|\big).
\end{aligned} \tag{4.42}$$

Therefore, since $\|q_n - x_n\| \to 0$, it follows from (4.39) and (4.41) that $w_n - q_n \to 0$ and hence that $w_n - x_n \to 0$. Since

$$\|U_n w_n - U_n x_n\| \leqslant \beta_2\|w_n - x_n\| \leqslant \beta_2(\|w_n - q_n\| + \|q_n - x_n\|) \to 0, \tag{4.43}$$

the proof is complete. ◻

Remark 4.13 In the special case when $C = 0$, $(q_n)_{n\in\mathbb{N}} = (w_n)_{n\in\mathbb{N}}$, and conditions (ii)(b) and (ii)(c) are satisfied, Theorem 4.12(ii) is closely related to [95, Theorem 4.2(ii)].

We conclude this section with the strongly convergent best approximation companion algorithm resulting from Theorem 4.9.

Theorem 4.14 *Let $\alpha \in \left]0,+\infty\right[$, let $W\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $C\colon \mathcal{H} \to \mathcal{H}$ be α-cocoercive and such that $Z = \operatorname{zer}(W + C) \neq \varnothing$, let $x_0 \in \mathcal{H}$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0, 1\right]$. Further, for every $n \in \mathbb{N}$, let $U_n\colon \mathcal{H} \to \mathcal{H}$ be an operator such that $\operatorname{ran} U_n \subset \operatorname{ran}(U_n + W + C)$ and $U_n + W + C$ is injective. Iterate*

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots\\
\left\lfloor
\begin{array}{l}
w_n = J^{U_n}_{W+C} x_n\\
w_n^* = U_n x_n - U_n w_n - C w_n\\
q_n \in \mathcal{H}\\
t_n^* = w_n^* + C q_n\\
\delta_n = \langle x_n - w_n \mid t_n^* \rangle - \|w_n - q_n\|^2/(4\alpha)\\
d_n = \begin{cases} \dfrac{\delta_n}{\|t_n^*\|^2} t_n^*, & \text{if } \delta_n > 0;\\ 0, & \text{otherwise} \end{cases}\\
r_n = x_n - \lambda_n d_n\\
x_{n+1} = \mathsf{Q}(x_0, x_n, r_n),
\end{array}
\right.
\end{array}
\tag{4.44}
$$

where Q is defined in Lemma 4.6. Then the following hold:

(i) $\sum_{n\in\mathbb{N}} \lambda_n^2 \|d_n\|^2 < +\infty$.

(ii) *Suppose that one of the following is satisfied:*

  (a) $\sum_{n\in\mathbb{N}} \lambda_n^2 = +\infty$ *and* $(\|d_n\|)_{n\in\mathbb{N}}$ *converges;*

  (b) $\inf_{n\in\mathbb{N}} \lambda_n > 0$*;*

  together with one of the following:

  (c) $w_n - x_n \rightharpoonup 0$, $U_n w_n - U_n x_n \to 0$, *and* $w_n - q_n \to 0$*;*

  (d) $q_n - x_n \to 0$ *and there exist* $\beta_1 \in \left]1/(4\alpha), +\infty\right[$ *and* $\beta_2 \in \left]0, +\infty\right[$ *such that the kernels* $(U_n)_{n\in\mathbb{N}}$ *are* β_1*-strongly monotone and* β_2*-Lipschitzian.*

  Then $(x_n)_{n\in\mathbb{N}}$ *converges strongly to* $\operatorname{proj}_Z x_0$.

Proof. In view of Lemma 2.42(i), (4.44) is an instance of (4.31) and we shall therefore employ Theorem 4.9.

(i): See Theorem 4.9(iii).

(ii): It follows from (i) and (4.44) that $d_n \to 0$. Indeed, this is evident under (ii)(b) whereas, under (ii)(a), we have $\underline{\lim}\, \|d_n\| = 0$ and therefore $\lim \|d_n\| = 0$. Let us now assume that (ii)(c) holds. Then (4.37) is satisfied and we obtain the assertion by invoking Theorem 4.9(iv). Finally, to show that (ii)(d)⇒(ii)(c), we remark that Theorem 4.9(i) asserts that $(x_n)_{n\in\mathbb{N}}$ is bounded. Hence, we follow the same pattern as in the proof of Theorem 4.12(ii)(d) to conclude. ☐

5 The proximal point algorithm

5.1 Preview

The proximal point algorithm is an implicit method to construct a zero of a maximally monotone operator which goes back to a quadratic programming method proposed in [49, Section 5.8]. In the nonlinear case, it first appeared in Lieutaud's work [259] (this fact seems to have been overlooked in the literature, see Remark 6.1), then in [274, 275] for subdifferentials and in [339] for the general case. Iteration n of the unrelaxed form of the algorithm can be interpreted as a backward Euler discretization of the Cauchy problem [24, Section 3.2] (see Example 2.18)

$$\begin{cases} x(0) = x_0 \\ -x'(t) \in Mx(t), \ \text{for a.e. } t \in \left]0,+\infty\right[\end{cases} \tag{5.1}$$

with time step $\gamma_n \in \left]0,+\infty\right[$, that is,

$$\frac{x_n - x_{n+1}}{\gamma_n} \in Mx_{n+1} \tag{5.2}$$

or, equivalently, $x_{n+1} = J_{\gamma_n M}x_n$.

5.2 Fejérian algorithm

The following theorem, which brings together results from [72, 179, 197, 214, 253, 274, 275, 339], will be derived from Theorem 4.12.

Theorem 5.1 *Let $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that $Z = \operatorname{zer} M \neq \varnothing$, let $x_0 \in \mathcal{H}$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,2\right[$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,+\infty\right[$. Iterate*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(J_{\gamma_n M}x_n - x_n\big) \tag{5.3}$$

and suppose that one of the following holds:

(i) $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$ *and* $(\forall n \in \mathbb{N})\ \gamma_n = 1$.

(ii) $\sum_{n\in\mathbb{N}} \gamma_n^2 = +\infty$ *and* $(\forall n \in \mathbb{N})\ \lambda_n = 1$.

(iii) $\inf_{n\in\mathbb{N}} \lambda_n > 0$, $\sup_{n\in\mathbb{N}} \lambda_n < 2$, *and* $\inf_{n\in\mathbb{N}} \gamma_n > 0$.

Then $\|J_{\gamma_n M}x_n - x_n\|/\gamma_n \to 0$ *and* $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* Z.

Proof. Let us apply Theorem 4.12 with

$$C = 0 \quad\text{and}\quad (\forall n \in \mathbb{N})\quad U_n = \gamma_n^{-1}\mathrm{Id} \quad\text{and}\quad q_n = w_n. \tag{5.4}$$

We derive from (2.19) that the variables of the iterations (4.34) satisfy

$$(\forall n \in \mathbb{N})\quad t_n^* = \frac{x_n - w_n}{\gamma_n},\ \delta_n = \gamma_n\|t_n^*\|^2,\ \text{and}\ d_n = x_n - w_n. \tag{5.5}$$

Thus, the sequence $(x_n)_{n\in\mathbb{N}}$ produced by (5.3) coincides with that of (4.34). In turn, Theorem 4.12(i) yields

$$\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n)\|d_n\|^2 < +\infty. \tag{5.6}$$

We now show that one of conditions (ii)(a)–(ii)(b) and one of conditions (ii)(c)–(ii)(d) of Theorem 4.12(ii) are fulfilled in each scenario. We also recall from (4.35) that (ii)(a) and (ii)(b) in Theorem 4.12 each imply that

$$d_n \to 0. \tag{5.7}$$

(i): Let us check that conditions (ii)(a) and (ii)(d) are fulfilled. For (ii)(a), it is enough to show that $(\|d_n\|)_{n\in\mathbb{N}}$ decreases. To this end, set $T = 2J_M - \mathrm{Id}$. Then Lemma 2.34(iii) and (2.33) assert that T is nonexpansive. Therefore, (5.5) yields

$$\begin{aligned}(\forall n \in \mathbb{N})\quad 2\|d_{n+1}\| &= \|Tx_{n+1} - x_{n+1}\| \\ &= \|Tx_{n+1} - Tx_n + (1-\lambda_n/2)(Tx_n - x_n)\| \\ &\leqslant \|x_{n+1} - x_n\| + (1-\lambda_n/2)\|Tx_n - x_n\| \\ &= (\lambda_n/2)\|Tx_n - x_n\| + (1-\lambda_n/2)\|Tx_n - x_n\| \\ &= 2\|d_n\|,\end{aligned} \tag{5.8}$$

as desired. For (ii)(d), note that (5.7) and (5.5) imply that $q_n - x_n = w_n - x_n = -d_n \to 0$. In addition, it is clear from (5.4) that $(U_n)_{n\in\mathbb{N}}$ satisfies the required conditions with $\beta_1 = \beta_2 = 1$.

(ii): Condition (ii)(b) holds. To show that (ii)(c) holds as well, we first infer from (5.5) and (5.6) that $\sum_{n\in\mathbb{N}} \gamma_n^2\|t_n^*\|^2 < +\infty$ and hence that $w_n - x_n = -\gamma_n t_n^* \to 0$. Furthermore, since $\sum_{n\in\mathbb{N}} \gamma_n^2 = +\infty$, $\underline{\lim}\,\|t_n^*\| = 0$. On the other hand, $(\forall n \in \mathbb{N})$

$t_n^* = \gamma_n^{-1}(x_n - w_n) = \gamma_n^{-1}(x_n - x_{n+1})$. Hence, using (2.18), the monotonicity of M, and the Cauchy–Schwarz inequality, we obtain

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad 0 &\leqslant \langle w_n - w_{n+1} \mid t_n^* - t_{n+1}^* \rangle / \gamma_{n+1} \\
&= \langle x_{n+1} - x_{n+2} \mid t_n^* - t_{n+1}^* \rangle / \gamma_{n+1} \\
&= \langle t_{n+1}^* \mid t_n^* - t_{n+1}^* \rangle \\
&= \langle t_{n+1}^* \mid t_n^* \rangle - \|t_{n+1}^*\|^2 \\
&\leqslant \|t_{n+1}^*\| \big(\|t_n^*\| - \|t_{n+1}^*\| \big),
\end{aligned} \tag{5.9}
$$

which shows that $(\|t_n^*\|)_{n\in\mathbb{N}}$ decreases. Altogether, $U_n x_n - U_n w_n = t_n^* \to 0$.

(iii): Condition (ii)(b) is assumed. Let us check (ii)(c). Since (5.5) and (5.6) yield $\sum_{n\in\mathbb{N}} \gamma_n^2 \|t_n^*\|^2 < +\infty$, we have $x_n - w_n = \gamma_n t_n^* \to 0$. Finally, since $\inf_{n\in\mathbb{N}} \gamma_n > 0$, $U_n x_n - U_n w_n = t_n^* \to 0$.

We conclude the proof by noting that in all three cases above we have $\|J_{\gamma_n M} x_n - x_n\| / \gamma_n = \|t_n^*\| \to 0$. □

Remark 5.2 Let $f \in \Gamma_0(\mathcal{H})$ and suppose that $M = \partial f$ in Theorem 5.1. Then, as seen in Example 2.12, M is maximally monotone and $Z = \operatorname{Argmin} f$. In this case, the condition on $(\gamma_n)_{n\in\mathbb{N}}$ in Theorem 5.1(ii) can be improved to $\sum_{n\in\mathbb{N}} \gamma_n = +\infty$ [72, Théorème 9].

5.3 Haugazeau-like algorithm

We employ Theorem 4.14 to obtain a strongly convergent variant of the proximal point algorithm; see [35, 360] for related results. Examples of proximal point iterations that fail to converge strongly are constructed in [41, 131, 221].

Theorem 5.3 *Let $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that $Z = \operatorname{zer} M \neq \varnothing$, let $x_0 \in \mathcal{H}$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1]$ such that $\inf_{n\in\mathbb{N}} \lambda_n > 0$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$ such that $\inf_{n\in\mathbb{N}} \gamma_n > 0$. Iterate*

$$
(\forall n \in \mathbb{N}) \quad x_{n+1} = \mathsf{Q}\big(x_0, x_n, x_n + \lambda_n (J_{\gamma_n M} x_n - x_n)\big), \tag{5.10}
$$

where Q is defined in Lemma 4.6. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to $\operatorname{proj}_Z x_0$.

Proof. In Theorem 4.14, set $C = 0$ and $(\forall n \in \mathbb{N})$ $U_n = \gamma_n^{-1}\mathrm{Id}$ and $q_n = w_n$. Then (5.5) holds and the sequence $(x_n)_{n\in\mathbb{N}}$ produced by (5.10) coincides with that of (4.44). In turn, Theorem 4.14(i) yields $\sum_{n\in\mathbb{N}} \lambda_n^2 \|d_n\|^2 < +\infty$. Therefore, $x_n - w_n = d_n \to 0$ and $U_n x_n - U_n w_n = \gamma_n^{-1} d_n \to 0$. This confirms that condition (ii)(c) in Theorem 4.14(ii) is fulfilled. Since condition (ii)(b) holds by assumption, the proof is complete. □

5.4 Special cases and variants

As mentioned in Section 1, direct implementations of the proximal point algorithm are limited due to the potential difficulty of evaluating the resolvents in (5.3) and (5.10). As we shall see in this section, the proximal point framework can nonetheless be an effective device to establish indirectly the convergence of algorithms that can be identified, possibly in a different space, as an instance of (5.3). Early examples in the context of inequality-constrained minimization problems are found in [340], where a dual application of an approximate proximal point algorithm was shown to yield a method of multipliers (also called the augmented Lagrangian method) that extends some classical ones from [228] and [319] (see also [337]). A primal-dual quadratically perturbed variant of this algorithm, known as the proximal method of multipliers, was also introduced in [340] as an application of an approximate proximal point algorithm to find saddle points of the Lagrangian (see also [343, 352] and their bibliographies for recent work along these lines). The applications described below reduce to implementations of the proximal point algorithm that feature full operator splitting when several linear and nonlinear operators are present in the original problem.

5.4.1 The Euler method

We derive from the proximal point algorithm a (forward) Euler method to find a zero of a cocoercive operator.

Proposition 5.4 *Let* $\alpha \in \left]0,+\infty\right[$ *and let* $B\colon \mathcal{H} \to \mathcal{H}$ *be* α*-cocoercive, with* $\operatorname{zer} B \neq \varnothing$*. Let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $\left]0,2\alpha\right[$ *such that* $\sum_{n\in\mathbb{N}} \gamma_n(2\alpha-\gamma_n) = +\infty$ *and let* $x_0 \in \mathcal{H}$*. Iterate*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n B x_n. \tag{5.11}$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\operatorname{zer} B$*.*

Proof. Set $M = (\mathrm{Id} - \alpha B)^{-1} - \mathrm{Id}$. Since αB is firmly nonexpansive with domain $\mathcal{H}$, $\mathrm{Id} - \alpha B$ is likewise and Lemma 2.34(iii) asserts that M is maximally monotone. On the other hand, $\operatorname{zer} M = \operatorname{zer} B$, $J_M = \mathrm{Id} - \alpha B$, and hence (5.11) becomes

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (J_M x_n - x_n), \quad \text{where} \quad \lambda_n = \gamma_n/\alpha \in \left]0,2\right[. \tag{5.12}$$

Thus, since $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, the claim follows from Theorem 5.1(i). ☐

Remark 5.5 As just shown, the Euler method (5.11) is an instance of the proximal point algorithm (5.3). Conversely, we can interpret the proximal point iterations in the format

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (J_M x_n - x_n), \quad \text{where} \quad \lambda_n \in \left]0,2\right[\tag{5.13}$$

as an instance of (5.11). Indeed, let $M\colon\mathcal{H}\to 2^{\mathcal{H}}$ be maximally monotone and set $B={}^1M$ and $(\forall n\in\mathbb{N})$ $\gamma_n=\lambda_n$. Then, as seen in Example 2.7, $\operatorname{zer} M=\operatorname{zer} B$ and B is 1-cocoercive, while (2.21) implies that (5.13) reduces to (5.11).

The following example is about the gradient method (see [102, 157] for the premises of this algorithm).

Example 5.6 Let $\alpha\in\,]0,+\infty[$ and let $g\colon\mathcal{H}\to\mathbb{R}$ be convex, differentiable, and such that ∇g is $1/\alpha$-Lipschitzian, with $\operatorname{Argmin} g\neq\varnothing$. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2\alpha[$ such that $\sum_{n\in\mathbb{N}}\gamma_n(2\alpha-\gamma_n)=+\infty$ and let $x_0\in\mathcal{H}$. Iterate

$$(\forall n\in\mathbb{N})\quad x_{n+1}=x_n-\gamma_n\nabla g(x_n). \tag{5.14}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Argmin} g$.

Proof. Combine Lemma 2.2 and Proposition 5.4. □

As noted in [38, Remark 4.8(ii)] in the context of Example 5.6, the convergence in Proposition 5.4 can fail to be strong. The next result, which guarantees strong convergence, is obtained by defining M and $(\lambda_n)_{n\in\mathbb{N}}$ as in the proof of Proposition 5.4 and using Theorem 5.3.

Proposition 5.7 *Let $\alpha\in\,]0,+\infty[$ and let $B\colon\mathcal{H}\to\mathcal{H}$ be α-cocoercive, with* $\operatorname{zer} B\neq\varnothing$. *Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,\alpha]$ such that* $\inf_{n\in\mathbb{N}}\gamma_n>0$ *and let $x_0\in\mathcal{H}$. Iterate*

$$(\forall n\in\mathbb{N})\quad x_{n+1}=\mathsf{Q}\big(x_0,x_n,x_n-\gamma_nBx_n\big), \tag{5.15}$$

where Q is defined in Lemma 4.6. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to $\operatorname{proj}_{\operatorname{zer} B} x_0$.

5.4.2 Fixed point problem

We address the basic problem of constructing a fixed point of a nonexpansive operator $T\colon\mathcal{H}\to\mathcal{H}$. The following result is derived as an instance of the proximal point algorithm of Theorem 5.1 via the embedding of Example 3.16.

Proposition 5.8 *Let $\alpha\in\,]0,1]$ and let $T\colon\mathcal{H}\to\mathcal{H}$ be α*-averaged. *Suppose that* $\operatorname{Fix} T\neq\varnothing$, *let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1/\alpha[$ such that $\sum_{n\in\mathbb{N}}\lambda_n(1-\alpha\lambda_n)=+\infty$, and let $x_0\in\mathcal{H}$. Iterate*

$$(\forall n\in\mathbb{N})\quad x_{n+1}=x_n+\lambda_n(Tx_n-x_n). \tag{5.16}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{Fix} T$.

Proof. We use the embedding of Example 3.16. Define $\mathcal{M}$ as in (3.26) and note that $J_{\mathcal{M}} = \mathrm{Id} + (2\alpha)^{-1}(T - \mathrm{Id})$. We therefore rewrite (5.16) as

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \mu_n(J_{\mathcal{M}}x_n - x_n), \quad \text{where} \quad \mu_n = 2\alpha\lambda_n \in \left]0, 2\right[. \tag{5.17}$$

Then $\sum_{n\in\mathbb{N}} \mu_n(2 - \mu_n) = +\infty$ and, appealing to Theorem 5.1(i), we conclude that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer} \mathcal{M} = \operatorname{Fix} T$. ☐

In the case when $\alpha = 1$, Proposition 5.8 is due to Groetsch [219] and (5.16) is known as the *Krasnosel'skiĭ–Mann iteration*, owing to its connection with iterative schemes proposed in [247] and [273], and it is a pillar of nonlinear numerical functional analysis [37, 104, 168]. Here is a strongly convergent variant derived from Theorem 5.3 (see [204] for an example of the failure of strong convergence in Proposition 5.8).

Proposition 5.9 *Let* $\alpha \in \left]0, 1\right]$ *and let* $T\colon \mathcal{H} \to \mathcal{H}$ *be* α*-averaged. Suppose that* $\operatorname{Fix} T \neq \varnothing$*, let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $\left]0, 1/(2\alpha)\right]$ *such that* $\inf_{n\in\mathbb{N}} \lambda_n > 0$*, and let* $x_0 \in \mathcal{H}$*. Iterate*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \mathsf{Q}\big(x_0, x_n, x_n + \lambda_n(Tx_n - x_n)\big), \tag{5.18}$$

where Q *is defined in Lemma 4.6. Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly to* $\operatorname{proj}_{\operatorname{Fix} T} x_0$.

Proof. Define $\mathcal{M}$ as in (3.26), argue as in the proof of Proposition 5.8 to observe that (5.18) is an instance of (5.10), and conclude by invoking Theorem 5.3. ☐

5.4.3 Resolvent compositions

We focus on the inclusion problem of [132, Section 6], which is modeled by resolvent compositions (see Example 2.40) and solvable via the proximal point algorithm.

Proposition 5.10 *Suppose that* $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ *satisfies* $0 < \|L\| \leqslant 1$*, let* $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ *be maximally monotone, let* $V \neq \{0\}$ *be a closed vector subspace of* $\mathcal{H}$*, and let* $\gamma \in \left]0, +\infty\right[$*. Let* S *be the set of solutions to the problem*

$$\text{find } x \in V \text{ such that } 0 \in B(Lx) \tag{5.19}$$

and let Z *be the set of solutions to the problem*

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in \big(\operatorname{proj}_V \diamond \big(L \blacklozenge (\gamma B)\big)\big)x. \tag{5.20}$$

Then (5.20) *is an exact relaxation of* (5.19) *in the sense that* $S \neq \varnothing \Rightarrow Z = S$. *Now assume that* $Z \neq \varnothing$, *let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,2[$ *such that* $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, *and let* $x_0 \in V$. *Iterate*

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
y_n = Lx_n \\
q_n = J_{\gamma B} y_n - y_n \\
z_n = L^* q_n \\
x_{n+1} = x_n + \lambda_n \operatorname{proj}_V z_n.
\end{array}
\right.
\end{array}
\tag{5.21}
$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* Z.

Proof. The exact relaxation claim is established in [132, Theorem 6.3(v)]. Now set $M = \operatorname{proj}_V \diamond (L \blacklozenge (\gamma B))$ and note that $\|\operatorname{proj}_V\| = 1$ and $\operatorname{proj}_V^* = \operatorname{proj}_V$. Hence, it follows from Example 2.31 that M is maximally monotone and from Example 2.40 that $J_M = \operatorname{proj}_V \circ (\mathrm{Id}_{\mathcal{H}} - L^* \circ L + L^* \circ J_{\gamma B} \circ L) \circ \operatorname{proj}_V$. Altogether, the convergence result follows from Theorem 5.1(i) ☐

Here is a strongly convergent algorithm based on the Haugazeau variant.

Proposition 5.11 *Suppose that* $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ *satisfies* $0 < \|L\| \leqslant 1$, *let* $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ *be maximally monotone, let* $V \neq \{0\}$ *be a closed vector subspace of* $\mathcal{H}$, *and let* $\gamma \in \,]0,+\infty[$. *Suppose that the set* Z *of solutions to the problem*

$$
\text{find } x \in \mathcal{H} \text{ such that } 0 \in \big(\operatorname{proj}_V \diamond \big(L \blacklozenge (\gamma B)\big)\big)x
\tag{5.22}
$$

is not empty. Let $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,1]$ *such that* $\inf_{n\in\mathbb{N}} \lambda_n > 0$, *and let* $x_0 \in V$. *Iterate*

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
y_n = Lx_n \\
q_n = J_{\gamma B} y_n - y_n \\
z_n = L^* q_n \\
x_{n+1} = \mathsf{Q}\big(x_0, x_n, x_n + \lambda_n \operatorname{proj}_V z_n\big),
\end{array}
\right.
\end{array}
\tag{5.23}
$$

where Q *is defined in Lemma* 4.6. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly to* $\operatorname{proj}_Z x_0$.

Proof. Arguing as in the proof of Proposition 5.10, this is an application of Theorem 5.3 with $M = \operatorname{proj}_V \diamond (L \blacklozenge (\gamma B))$ and $(\forall n \in \mathbb{N})\ \gamma_n = 1$. ☐

Below we recover the relaxation framework of [153] for signal reconstruction in the presence of possibly inconsistent nonlinear observations.

Example 5.12 Let $0 < p \in \mathbb{N}$, let $\gamma \in \,]0,+\infty[$, and let $V \neq \{0\}$ be a closed vector subspace of $\mathcal{H}$. For every $k \in \{1, \ldots, p\}$, let $\mathcal{G}_k$ be a real Hilbert space, let

$L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$, let $\omega_k \in \left]0,+\infty\right[$, let $F_k \colon \mathcal{G}_k \to \mathcal{G}_k$ be firmly nonexpansive, and let $r_k \in \mathcal{G}_k$. Consider the nonlinear reconstruction problem [153, Problem 1.1]

$$\text{find } x \in V \text{ such that } \big(\forall k \in \{1,\ldots,p\}\big) \quad F_k(L_k x) = r_k \tag{5.24}$$

and the relaxed variational inequality problem [153, Problem 1.3]

$$\text{find } x \in V \text{ such that } \sum_{k=1}^{p} \omega_k L_k^*\big(F_k(L_k x) - r_k\big) \in V^{\perp}. \tag{5.25}$$

Suppose that $0 < \sum_{k=1}^{p} \omega_k \|L_k\|^2 \leqslant 1$ and that (5.25) admits solutions. Let $x_0 \in V$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,2\right[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, and iterate

$$\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor\begin{array}{l}
\text{for } k = 1,\ldots,p\\
\left\lfloor\begin{array}{l}
y_{k,n} = L_k x_n\\
q_{k,n} = r_k - F_k y_{k,n}
\end{array}\right.\\
z_n = \sum_{k=1}^{p} \omega_k L_k^* q_{k,n}\\
x_{n+1} = x_n + \lambda_n \operatorname{proj}_V z_n.
\end{array}\right.
\end{array} \tag{5.26}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a solution to (5.25).

Proof. Let $\mathcal{G}$ be the standard product vector space $\mathcal{G}_1 \times \cdots \times \mathcal{G}_p$, with generic element $\boldsymbol{y} = (y_k)_{1\leqslant k\leqslant p}$, and equipped with the scalar product $(\boldsymbol{y}, \boldsymbol{y}') \mapsto \sum_{k=1}^{p} \omega_k \langle y_k \mid y'_k\rangle$. Further, set $\boldsymbol{L} \colon \mathcal{H} \to \mathcal{G} \colon x \mapsto (L_1 x,\ldots,L_p x)$ and

$$\boldsymbol{B} \colon \mathcal{G} \to 2^{\mathcal{G}} \colon \boldsymbol{y} \mapsto \big((\mathrm{Id}-F_1+r_1)^{-1}y_1 - y_1\big)\times\cdots\times\big((\mathrm{Id}-F_p+r_p)^{-1}y_p - y_p\big). \tag{5.27}$$

In this setting, (5.24) is a realization of (5.19), (5.25) of (5.20), and (5.26) of (5.21) (see [132, Example 6.10] for details). The claim therefore results from Proposition 5.10. ☐

5.4.4 The method of partial inverses

We go back to a formulation already touched upon in Problem 3.10. Given a maximally monotone operator $A \colon \mathcal{H} \to 2^{\mathcal{H}}$ and a closed vector subspace V of $\mathcal{H}$, Spingarn considered in [362] the problem

$$\text{find } x \in V \text{ and } x^* \in V^{\perp} \text{ such that } x^* \in Ax \tag{5.28}$$

and solved it by applying the proximal point algorithm to the partial inverse A_V (see Example 2.33). The resulting algorithm is called the *method of partial inverses*. The following is a relaxed version of the convergence result of [362, Theorem 4.1(i)] (see [7, Theorem 2.4]).

Theorem 5.13 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let V be a closed vector subspace of $\mathcal{H}$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$. Suppose that (5.28) has solutions, let $x_0 \in V$, let $x_0^* \in V^\perp$, and iterate*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
p_n = J_A(x_n + x_n^*) \\
p_n^* = x_n + x_n^* - p_n \\
x_{n+1} = x_n - \lambda_n \operatorname{proj}_V p_n^* \\
x_{n+1}^* = x_n^* - \lambda_n \operatorname{proj}_{V^\perp} p_n.
\end{array}
\right.
\end{array}
\tag{5.29}
$$

Then the following hold:

(i) $\operatorname{proj}_V p_n - x_n \to 0$ *and* $\operatorname{proj}_{V^\perp} p_n^* - x_n^* \to 0$.

(ii) *There exists a solution (x, x^*) to (5.28) such that $x_n \rightharpoonup x$ and $x_n^* \rightharpoonup x^*$.*

Proof. Set

$$
(\forall n \in \mathbb{N}) \quad z_n = x_n + x_n^* \tag{5.30}
$$

and note that, since $(x_n)_{n\in\mathbb{N}}$ lies in V and $(x_n^*)_{n\in\mathbb{N}}$ lies in $V^\perp$, (5.29) can be rewritten as

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
p_n = J_A(x_n + x_n^*) \\
p_n^* = x_n + x_n^* - p_n \\
x_{n+1} = x_n + \lambda_n(\operatorname{proj}_V p_n - x_n) \\
x_{n+1}^* = x_n^* + \lambda_n(\operatorname{proj}_{V^\perp} p_n^* - x_n^*).
\end{array}
\right.
\end{array}
\tag{5.31}
$$

Thus,

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad & \operatorname{proj}_V \left(\frac{z_{n+1} - z_n}{\lambda_n} + z_n \right) + \operatorname{proj}_{V^\perp} \left(z_n - \left(\frac{z_{n+1} - z_n}{\lambda_n} + z_n \right) \right) \\
&= \operatorname{proj}_V \left(\frac{z_{n+1} - z_n}{\lambda_n} + z_n \right) + \operatorname{proj}_{V^\perp} \left(\frac{z_n - z_{n+1}}{\lambda_n} \right) \\
&= \operatorname{proj}_V \left(\frac{x_{n+1} - x_n}{\lambda_n} + x_n \right) + \operatorname{proj}_{V^\perp} \left(\frac{x_n^* - x_{n+1}^*}{\lambda_n} \right) \\
&= \operatorname{proj}_V p_n + \operatorname{proj}_{V^\perp} (x_n^* - p_n^*) \\
&= \operatorname{proj}_V p_n + \operatorname{proj}_{V^\perp} (p_n - x_n) \\
&= p_n \\
&= J_A z_n.
\end{aligned}
\tag{5.32}
$$

Hence, it follows from (5.30), (5.31), and Example 2.38 that

$$(\forall n \in \mathbb{N}) \quad z_{n+1} = z_n + \lambda_n \big(J_{A_V} z_n - z_n\big). \tag{5.33}$$

Altogether, we derive from Theorem 5.1(i) that

$$J_{A_V} z_n - z_n \to 0 \tag{5.34}$$

and that there exists $z \in \operatorname{zer} A_V$ such that

$$z_n \rightharpoonup z. \tag{5.35}$$

(i): In view of (5.31), (5.30), Example 2.38, and (5.34), we have

$$\operatorname{proj}_V p_n - x_n = \operatorname{proj}_V (J_{A_V} z_n) - x_n = \operatorname{proj}_V \big(J_{A_V} z_n - z_n\big) \to 0 \tag{5.36}$$

and

$$x_n^* - \operatorname{proj}_{V^\perp} p_n^* = \operatorname{proj}_{V^\perp} (p_n - x_n) = \operatorname{proj}_{V^\perp} J_A z_n = \operatorname{proj}_{V^\perp} \big(z_n - J_{A_V} z_n\big) \to 0. \tag{5.37}$$

(ii): As seen above $z \in \operatorname{zer} A_V$. Now set $(x, x^*) = (\operatorname{proj}_V z, \operatorname{proj}_{V^\perp} z)$. Then Example 2.33(ii) guarantees that (x, x^*) solves (5.28). In addition, since proj_V and $\operatorname{proj}_{V^\perp}$ are linear and continuous, they are weakly continuous. We conclude that $x_n = \operatorname{proj}_V z_n \rightharpoonup \operatorname{proj}_V z = x$ and $x_n^* = \operatorname{proj}_{V^\perp} z_n \rightharpoonup \operatorname{proj}_{V^\perp} z = x^*$. ☐

Example 5.14 In Theorem 5.13, let $f \in \Gamma_0(\mathcal{H})$ be such that $0 \in \operatorname{sri}(\operatorname{dom} f - V)$, set $A = \partial f$, and suppose that f admits minimizers over V. Then (5.28) amounts to finding a solution to the Fenchel dual pair

$$\underset{x \in V}{\text{minimize}} \;\; f(x) \quad \text{and} \quad \underset{x^* \in V^\perp}{\text{minimize}} \;\; f^*(x^*). \tag{5.38}$$

In this case, given $x_0 \in V$ and $x_0^* \in V^\perp$, the method of partial inverses (5.29) iterates

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} p_n = \operatorname{prox}_f (x_n + x_n^*) \\ p_n^* = x_n + x_n^* - p_n \\ x_{n+1} = x_n - \lambda_n \operatorname{proj}_V p_n^* \\ x_{n+1}^* = x_n^* - \lambda_n \operatorname{proj}_{V^\perp} p_n \end{array} \right. \end{array} \tag{5.39}$$

and Theorem 5.13(ii) guarantees that there exists a primal-dual solution (x, x^*) of (5.38) such that $x_n \rightharpoonup x$ and $x_n^* \rightharpoonup x^*$.

Algorithm (5.29) has many applications in convex optimization, e.g., [231, 253, 256, 309, 362, 363, 364]. As shown in [344], it also constitutes the basic building block of the *progressive hedging algorithm* in stochastic programming [345].

Although the method of partial inverses (5.29) is presented in the context of the simple problem (5.28), it has far reaching ramifications. We present below an application proposed in [7], where it is applied to Problem 3.11. In terms of Framework 1.2, this approach can be seen as a rephrasing of Problem 3.11 as an instance of (5.28) in $\mathbf{X} = \mathcal{H} \oplus \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p$.

Proposition 5.15 *Let $0 < p \in \mathbb{N}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and, for every $k \in \{1\dots,p\}$, let $\mathcal{G}_k$ be a real Hilbert space, let $B_k\colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ be maximally monotone, and let $L_k \in \mathcal{B}(\mathcal{H},\mathcal{G}_k)$. Suppose that the set Z of solutions to the inclusion*

$$\textit{find}\;\; x \in \mathcal{H} \;\;\textit{such that}\;\; 0 \in Ax + \sum_{k=1}^{p} L_k^*\big(B_k(L_k x)\big) \tag{5.40}$$

is not empty and let Z^* be the set of solutions to the dual inclusion

$$\begin{aligned}&\textit{find}\;\; y_1^* \in \mathcal{G}_1,\dots,y_p^* \in \mathcal{G}_p \;\;\textit{such that}\\ &\qquad\qquad \left(\exists\, x \in A^{-1}\left(-\sum_{k=1}^{p} L_k^* y_k^*\right)\right)(\forall k \in \{1,\dots,p\})\;\; L_k x \in B_k^{-1} y_k^*.\end{aligned} \tag{5.41}$$

Let $x_0 \in \mathcal{H}$ and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$. Set

$$U = \left(\mathrm{Id} + \sum_{k=1}^{p} L_k^* \circ L_k\right)^{-1} \tag{5.42}$$

and, for every $k \in \{1,\dots,p\}$, let $y_{k,0}^* \in \mathcal{G}_k$ and set $y_{k,0} = L_k x_0$. Additionally, set

$$x_0^* = -\sum_{k=1}^{p} L_k^* y_{k,0}^*, \tag{5.43}$$

and iterate

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
p_n = J_A(x_n + x_n^*) \\
p_n^* = x_n + x_n^* - p_n \\
\textit{for } k = 1, \ldots, p \\
\left\lfloor
\begin{array}{l}
q_{k,n} = J_{B_k}(y_{k,n} + y_{k,n}^*) \\
q_{k,n}^* = y_{k,n} + y_{k,n}^* - q_{k,n}
\end{array}
\right. \\
t_n = U\big(p_n^* + \sum_{k=1}^{p} L_k^* q_{k,n}^*\big) \\
w_n = U\big(p_n + \sum_{k=1}^{p} L_k^* q_{k,n}\big) \\
x_{n+1} = x_n - \lambda_n t_n \\
x_{n+1}^* = x_n^* + \lambda_n(w_n - p_n) \\
\textit{for } k = 1, \ldots, p \\
\left\lfloor
\begin{array}{l}
y_{k,n+1} = y_{k,n} - \lambda_n L_k t_n \\
y_{k,n+1}^* = y_{k,n}^* + \lambda_n(L_k w_n - q_{k,n}).
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{5.44}
$$

Then there exist $x \in Z$ and $(y_k^*)_{1 \leqslant k \leqslant p} \in Z^*$ such that $x_n \rightharpoonup x$ and, for every $k \in \{1, \ldots, p\}$, $y_{k,n}^* \rightharpoonup y_k^*$.

Proof. Define

$$
\begin{cases}
\mathcal{G} = \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p \\
B \colon \mathcal{G} \to 2^{\mathcal{G}} \colon (y_1, \ldots, y_p) \mapsto B_1 y_1 \times \cdots \times B_p y_p \\
L \colon \mathcal{H} \to \mathcal{G} \colon x \mapsto (L_1 x, \ldots, L_p x)
\end{cases}
\tag{5.45}
$$

and note that $L^* \colon \mathcal{G} \to \mathcal{H} \colon (y_1^*, \ldots, y_p^*) \mapsto L_1^* y_1^* + \cdots + L_p^* y_p^*$. Moreover set, for every $n \in \mathbb{N}$, $q_n = (q_{k,n})_{1 \leqslant k \leqslant p}$, $q_n^* = (q_{k,n}^*)_{1 \leqslant k \leqslant p}$, $y_n = (y_{k,n})_{1 \leqslant k \leqslant p}$, and $y_n^* = (y_{k,n}^*)_{1 \leqslant k \leqslant p}$. In this setting, B is maximally monotone and $J_B \colon (y_k)_{1 \leqslant k \leqslant p} \mapsto (J_{B_k} y_k)_{1 \leqslant k \leqslant p}$ (Example 2.37), so that (5.44) can be rewritten as

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
p_n = J_A(x_n + x_n^*) \\
q_n = J_B(y_n + y_n^*) \\
p_n^* = x_n + x_n^* - p_n \\
q_n^* = y_n + y_n^* - q_n \\
t_n = U\big(p_n^* + L^* q_n^*\big) \\
w_n = U\big(p_n + L^* q_n\big) \\
x_{n+1} = x_n - \lambda_n t_n \\
y_{n+1} = y_n - \lambda_n L t_n \\
x_{n+1}^* = x_n^* + \lambda_n(w_n - p_n) \\
y_{n+1}^* = y_n^* + \lambda_n(L w_n - q_n).
\end{array}
\right.
\end{array}
\tag{5.46}
$$

Let us introduce

$$
\begin{cases}
\mathbf{X} = \mathcal{H} \oplus \mathcal{G} \\
\boldsymbol{V} = \big\{(x, y) \in \mathbf{X} \mid Lx = y\big\} \\
\boldsymbol{Z} = \big\{(x, y^*) \in \mathbf{X} \mid -L^* y^* \in Ax \;\text{ and }\; y^* \in B(Lx)\big\} \\
\boldsymbol{A} \colon \mathbf{X} \to 2^{\mathbf{X}} \colon (x, y) \mapsto Ax \times By \\
\boldsymbol{S} = \big\{(\boldsymbol{x}, \boldsymbol{x}^*) \in \boldsymbol{V} \times \boldsymbol{V}^\perp \mid \boldsymbol{x}^* \in \boldsymbol{A}\boldsymbol{x}\big\}
\end{cases}
\tag{5.47}
$$

and observe that

$$
\begin{cases}
\boldsymbol{V}^\perp = \big\{(x^*, y^*) \in \mathbf{X} \mid x^* = -L^* y^*\big\} \\
\boldsymbol{S} = \big\{\big((x, Lx), (-L^* y^*, y^*)\big) \in \mathbf{X} \times \mathbf{X} \mid (x, y^*) \in \boldsymbol{Z}\big\}.
\end{cases}
\tag{5.48}
$$

Then Lemma 3.12(iii) implies that

$$
\text{(5.40) admits solutions} \;\Leftrightarrow\; \boldsymbol{Z} \neq \varnothing \;\Leftrightarrow\; \boldsymbol{S} \neq \varnothing. \tag{5.49}
$$

Now define $(\forall n \in \mathbb{N})$ $\boldsymbol{p}_n = (p_n, q_n)$, $\boldsymbol{p}_n^* = (p_n^*, q_n^*)$, $\boldsymbol{x}_n = (x_n, y_n)$, and $\boldsymbol{x}_n^* = (x_n^*, y_n^*)$. Then $\boldsymbol{x}_0 \in \boldsymbol{V}$ and $\boldsymbol{x}_0^* \in \boldsymbol{V}^\perp$. Moreover, by Lemma 2.24 and Example 2.37, $\boldsymbol{A}$ is maximally monotone and

$$
(\forall n \in \mathbb{N}) \quad J_{\boldsymbol{A}}(\boldsymbol{x}_n + \boldsymbol{x}_n^*) = \big(J_A(x_n + x_n^*), J_B(y_n + y_n^*)\big). \tag{5.50}
$$

Furthermore, since $U = (\mathrm{Id} + L^* \circ L)^{-1}$, it follows from (5.47) and [37, Example 29.19] that

$$
(\forall n \in \mathbb{N}) \quad \mathrm{proj}_{\boldsymbol{V}^\perp} \boldsymbol{p}_n = \Big(p_n - U(p_n + L^* q_n), q_n - L\big(U(p_n + L^* q_n)\big)\Big) \tag{5.51}
$$

and

$$
(\forall n \in \mathbb{N}) \quad \mathrm{proj}_{\boldsymbol{V}} \boldsymbol{p}_n^* = \Big(U(p_n^* + L^* q_n^*), L\big(U(p_n^* + L^* q_n^*)\big)\Big). \tag{5.52}
$$

Combining (5.50), (5.51), and (5.52), we rewrite (5.46) as

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{p}_n = J_{\boldsymbol{A}}(\boldsymbol{x}_n + \boldsymbol{x}_n^*) \\
\boldsymbol{p}_n^* = \boldsymbol{x}_n + \boldsymbol{x}_n^* - \boldsymbol{p}_n \\
\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \lambda_n \,\mathrm{proj}_{\boldsymbol{V}} \boldsymbol{p}_n^* \\
\boldsymbol{x}_{n+1}^* = \boldsymbol{x}_n^* - \lambda_n \,\mathrm{proj}_{\boldsymbol{V}^\perp} \boldsymbol{p}_n.
\end{array}
\right.
\end{array}
\tag{5.53}
$$

In turn, Theorem 5.13(ii) implies that there exists $(\boldsymbol{x}, \boldsymbol{x}^*) \in \boldsymbol{S}$ such that $\boldsymbol{x}_n \rightharpoonup \boldsymbol{x}$ and $\boldsymbol{x}_n^* \rightharpoonup \boldsymbol{x}^*$. We then derive from (5.48) that there exists $(x, y^*) \in \boldsymbol{Z}$ such that $(x_n, y_n^*) \rightharpoonup (x, y^*)$. We complete the proof by invoking Lemma 3.12(ii). $\square$

5.4.5 Renorming

The potency of the proximal point algorithm can be further extended by setting it up in a renormed space. In terms of Framework 1.2, the guiding principle lies in the embedding of Example 3.15. Here is a weak convergence result.

Proposition 5.16 *Let $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that $Z = \operatorname{zer} M \neq \varnothing$, let $U \in \mathcal{B}(\mathcal{H})$ be a self-adjoint strongly monotone operator, and let $\mathcal{X}$ be the real Hilbert space obtained by endowing $\mathcal{H}$ with the scalar product $(x, y) \mapsto \langle Ux \mid y\rangle$. Let $x_0 \in \mathcal{H}$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 2[$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0, +\infty[$. Iterate*

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
u_n = \gamma_n^{-1} U x_n \\
p_n = \big(\gamma_n^{-1} U + M\big)^{-1} u_n \\
x_{n+1} = x_n + \lambda_n (p_n - x_n)
\end{array}
\right.
\end{array}
\tag{5.54}
$$

and suppose that one of the following holds:

(i) $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$ *and* $(\forall n \in \mathbb{N})\ \gamma_n = 1$.

(ii) $\sum_{n\in\mathbb{N}} \gamma_n^2 = +\infty$ *and* $(\forall n \in \mathbb{N})\ \lambda_n = 1$.

(iii) $\inf_{n\in\mathbb{N}} \lambda_n > 0$, $\sup_{n\in\mathbb{N}} \lambda_n < 2$, *and* $\inf_{n\in\mathbb{N}} \gamma_n > 0$.

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.

Proof. In view of Lemma 2.25(ii) and Example 2.39, (5.54) is just the proximal point algorithm (5.3) applied to the maximally monotone operator $U^{-1} \circ M$ in $\mathcal{X}$. Since weak convergences in $\mathcal{H}$ and $\mathcal{X}$ coincide, the claims follow from Lemma 2.25(i) and Theorem 5.1. ☐

Remark 5.17 In terms of the warped resolvent of Section 2.4.3, the update in (5.54) can be written as $x_{n+1} = x_n + \lambda_n(J^U_{\gamma_n M} x_n - x_n)$.

Likewise, Theorem 5.3 leads to a strongly convergent algorithm.

Proposition 5.18 *Let $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator such that $Z = \operatorname{zer} M \neq \varnothing$, let $U \in \mathcal{B}(\mathcal{H})$ be a self-adjoint strongly monotone operator, and let $\mathcal{X}$ be the real Hilbert space obtained by endowing $\mathcal{H}$ with the scalar product $(x, y) \mapsto \langle Ux \mid y\rangle$. Let $x_0 \in \mathcal{H}$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 1]$ such that $\inf_{n\in\mathbb{N}} \lambda_n > 0$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0, +\infty[$ such that $\inf_{n\in\mathbb{N}} \gamma_n > 0$.*

Iterate

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
u_n = \gamma_n^{-1} U x_n \\
p_n = \big(\gamma_n^{-1} U + M\big)^{-1} u_n \\
x_{n+1} = \mathsf{Q}\big(x_0, x_n, x_n + \lambda_n (p_n - x_n)\big),
\end{array}
\right.
\end{array}
\tag{5.55}
$$

where Q *is defined in Lemma 4.6. Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly to* $\operatorname{proj}_Z x_0$.

Proof. It follows from Lemma 2.25(ii) and Example 2.39 that applying the algorithm (5.10) to the maximally monotone operator $U^{-1} \circ M$ in $\mathcal{X}$ yields (5.55). Since strong convergences in $\mathcal{H}$ and $\mathcal{X}$ coincide, the assertion follows from Lemma 2.25(i) and Theorem 5.3. □

Although the inversion of the operators $(\gamma_n^{-1} U + M)_{n\in\mathbb{N}}$ in (5.54) and (5.55) may be intimidating, we show below that the renormed proximal point algorithm leads to important instances of fully executable splitting algorithms. First, we revisit a classical minimization problem and recover an algorithm known as the *proximal Landweber method*.

Example 5.19 Let $\varphi \in \Gamma_0(\mathcal{H})$, let $\mu \in \left]0, +\infty\right[$, and let $y \in \mathcal{G}$. Suppose that $0 \neq L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ and that the set Z of solutions to the optimization problem

$$
\underset{x\in\mathcal{H}}{\text{minimize}} \;\; \varphi(x) + \frac{\mu}{2}\|Lx - y\|^2 \tag{5.56}
$$

is not empty. Without loss of generality (rescale), assume that $\mu\|L\|^2 < 1$. Let $x_0 \in \mathcal{H}$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0, 2\right[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$, and iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
u_n = x_n - \mu L^*(L x_n) \\
p_n = \operatorname{prox}_\varphi(u_n + \mu L^* y) \\
x_{n+1} = x_n + \lambda_n (p_n - x_n).
\end{array}
\right.
\end{array}
\tag{5.57}
$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.

Proof. Set $f = \varphi - \mu\langle \cdot \mid L^* y\rangle$, $M = \partial(\varphi + \mu\|L \cdot - y\|^2/2) = \partial f + \mu L^* \circ L$, and $U = \mathrm{Id} - \mu L^* \circ L$. Then $f \in \Gamma_0(\mathcal{H})$, M is maximally monotone with $\operatorname{zer} M = Z$ by virtue of Example 2.12, $U \in \mathcal{B}(\mathcal{H})$ is self-adjoint and strongly monotone, and $(U + M)^{-1} = \operatorname{prox}_f = \operatorname{prox}_\varphi(\cdot + \mu L^* y)$. Consequently, (5.57) is the implementation of (5.54) with, for every $n \in \mathbb{N}$, $\gamma_n = 1$, and Proposition 5.16(i) brings the conclusion. □

Next, we return to the primal-dual composite inclusion framework of Problem 3.7 and approach it via Framework 1.2 where, as discussed in Example 3.20, the embedding is based on $\mathbf{X} = \mathcal{H} \oplus \mathcal{G}$ and the Kuhn–Tucker operator $\mathcal{K}$ of Lemma 3.8.

Example 5.20 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone, and let $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Suppose that the set Z of solutions to the primal inclusion

$$\text{find } \ x \in \mathcal{H} \ \text{ such that } \ 0 \in Ax + L^*\big(B(Lx)\big) \tag{5.58}$$

is not empty and let Z^* be the set of solutions to the dual inclusion

$$\text{find } \ y^* \in \mathcal{G} \ \text{ such that } \ 0 \in -L\big(A^{-1}(-L^*y^*)\big) + B^{-1}y^*. \tag{5.59}$$

Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 2[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$, let $x_0 \in \mathcal{H}$, let $y_0^* \in \mathcal{G}$, and let $\sigma \in \,]0, +\infty[$ and $\tau \in \,]0, +\infty[$ be such that $\tau\sigma\|L\|^2 < 1$. Iterate

$$\begin{array}{l}\text{for } n = 0, 1, \ldots\\ \left\lfloor \begin{array}{l} x_n^* = \tau L^* y_n^*\\ p_n = J_{\tau A}(x_n - x_n^*)\\ y_n = \sigma L(2p_n - x_n)\\ q_n^* = J_{\sigma B^{-1}}(y_n^* + y_n)\\ x_{n+1} = x_n + \lambda_n(p_n - x_n)\\ y_{n+1}^* = y_n^* + \lambda_n(q_n^* - y_n^*). \end{array}\right.\end{array} \tag{5.60}$$

Then there exist $x \in Z$ and $y^* \in Z^*$ such that $x_n \rightharpoonup x$ and $y_n^* \rightharpoonup y^*$.

Proof. Set $\mathbf{X} = \mathcal{H} \oplus \mathcal{G}$ and

$$\begin{cases}\mathcal{K}\colon \mathbf{X} \to 2^{\mathbf{X}}\colon (x, y^*) \mapsto \big(Ax + L^*y^*\big) \times \big(-Lx + B^{-1}y^*\big)\\ \boldsymbol{U}\colon \mathbf{X} \to \mathbf{X}\colon (x, y^*) \mapsto \big(\tau^{-1}x - L^*y^*, -Lx + \sigma^{-1}y^*\big).\end{cases} \tag{5.61}$$

As seen in Lemma 3.8(iii)–(iv), $\mathcal{K}$ is the maximally monotone Kuhn–Tucker operator associated with (5.58)–(5.59) and to prove the claim it is enough to show that $(x_n, y_n^*)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}\mathcal{K}$, which we shall derive from Proposition 5.16(i). It is clear that $\boldsymbol{U} \in \mathcal{B}(\mathbf{X})$ is self-adjoint. Now set $\beta = 1 - \sqrt{\sigma\tau}\|L\|$. Then, since $\tau\sigma\|L\|^2 < 1$, $\beta \in \,]0, 1[$ and, for every $(x, y^*) \in \mathbf{X}$, the Cauchy–Schwarz inequality yields

$$
\begin{aligned}
\langle \boldsymbol{U}(x,y^*) \mid (x,y^*)\rangle_{\mathbf{X}} &= \tau^{-1}\|x\|^2 - 2\langle Lx \mid y^*\rangle + \sigma^{-1}\|y^*\|^2 \\
&\geqslant \tau^{-1}\|x\|^2 - 2\sqrt{\tau\sigma}\|L\|\left\|\frac{x}{\sqrt{\tau}}\right\|\left\|\frac{y^*}{\sqrt{\sigma}}\right\| + \sigma^{-1}\|y^*\|^2 \\
&= \tau^{-1}\|x\|^2 - 2(1-\beta)\left\|\frac{x}{\sqrt{\tau}}\right\|\left\|\frac{y^*}{\sqrt{\sigma}}\right\| + \sigma^{-1}\|y^*\|^2 \\
&= \left(\left\|\frac{x}{\sqrt{\tau}}\right\| - \left\|\frac{y^*}{\sqrt{\sigma}}\right\|\right)^2 + 2\beta\left\|\frac{x}{\sqrt{\tau}}\right\|\left\|\frac{y^*}{\sqrt{\sigma}}\right\| \\
&= (1-\beta)\left(\left\|\frac{x}{\sqrt{\tau}}\right\| - \left\|\frac{y^*}{\sqrt{\sigma}}\right\|\right)^2 + \beta\left(\left\|\frac{x}{\sqrt{\tau}}\right\|^2 + \left\|\frac{y^*}{\sqrt{\sigma}}\right\|^2\right) \\
&\geqslant \beta\big(\tau^{-1}\|x\|^2 + \sigma^{-1}\|y^*\|^2\big) \\
&\geqslant \beta\min\{\tau^{-1},\sigma^{-1}\}\|(x,y^*)\|_{\mathbf{X}}^2,
\end{aligned}
\tag{5.62}
$$

which confirms that $\boldsymbol{U}$ is strongly monotone. It remains to show that (5.60) is a realization of (5.54) with the above operators $\boldsymbol{\mathcal{K}}$ and $\boldsymbol{U}$. Define $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = (x_n, y_n^*)$, $\boldsymbol{p}_n = (p_n, q_n^*)$, and $\boldsymbol{u}_n = \boldsymbol{U}\boldsymbol{x}_n$. Then we derive from (5.60) and (2.18) that

$$
(\forall n \in \mathbb{N}) \quad \begin{cases} x_n - p_n - \tau L^* y_n^* \in \tau A p_n \\ y_n^* - q_n^* + \sigma L(2p_n - x_n) \in \sigma B^{-1} q_n^* \end{cases} \tag{5.63}
$$

This yields $(\forall n \in \mathbb{N})$ $\boldsymbol{u}_n - \boldsymbol{U}\boldsymbol{p}_n \in \boldsymbol{\mathcal{K}}\boldsymbol{p}_n$, i.e., $\boldsymbol{p}_n = (\boldsymbol{U}+\boldsymbol{\mathcal{K}})^{-1}\boldsymbol{u}_n$. Altogether, (5.60) corresponds to the iteration

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
\boldsymbol{u}_n = \boldsymbol{U}\boldsymbol{x}_n \\
\boldsymbol{p}_n = \big(\boldsymbol{U}+\boldsymbol{\mathcal{K}}\big)^{-1}\boldsymbol{u}_n \\
\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \lambda_n(\boldsymbol{p}_n - \boldsymbol{x}_n),
\end{array}\right.
\end{array}
\tag{5.64}
$$

which is precisely (5.54) with $(\forall n \in \mathbb{N})$ $\gamma_n = 1$. ☐

Remark 5.21 Here are a few observations regarding Example 5.20.

(i) We have derived weak convergence from Proposition 5.16(i). Using items (ii) or (iii) in Proposition 5.16 leads to alternative forms of (5.60) involving proximal parameters $(\gamma_n)_{n\in\mathbb{N}}$.

(ii) It is straightforward to derive a strongly convergent best approximation variant of (5.60) from Proposition 5.18 by following the same pattern as in the proof of Example 5.20, i.e., applying (5.55) to the operators $\boldsymbol{\mathcal{K}}$ and $\boldsymbol{U}$ of (5.61).

(iii) Algorithm (5.60) can be adapted to Problem 3.11 by applying it to the setting of (5.45) and using Example 2.37.

(iv) Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$, and set $A = \partial f$ and $B = \partial g$ in Example 5.20, which corresponds to the primal-dual minimization setting of Problem 3.9. The specialization of Example 5.20 to this minimization problem appears in [155, Theorem 3.2], where (5.60) is called the *Chambolle–Pock algorithm* because it collapses to the algorithm proposed in [113, Algorithm I] in Euclidean spaces when $(\forall n \in \mathbb{N})\ \lambda_n = 1$ (see [156] for variations on this algorithm). The fact that the Chambolle–Pock algorithm is a renormed proximal point algorithm was first observed in [227].

6 Douglas–Rachford splitting

6.1 Preview

The Douglas–Rachford splitting algorithm is an implicit alternating direction method designed in [170] to solve the matrix equation $Ax + Bx = f$, where A and B are positive-definite matrices arising from the discretization of partial differentiation operators. It is described by the iteration process

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
x_{n+1/2} - x_n + Ax_{n+1/2} + Bx_n = f \\
x_{n+1} - x_n + Ax_{n+1/2} + Bx_{n+1} = f.
\end{array}
\right.
\end{array}
\tag{6.1}
$$

In 1968, Lieutaud [259] (see also [260]) proposed an infinite-dimensional nonlinear generalization of the method by showing that (6.1) can be extended to single-valued hemicontinuous monotone operators with $\operatorname{dom} A = \operatorname{dom} B = \mathcal{H}$. In particular, he established in [259] that, with the additional assumption that A or B is strongly monotone, $(x_n)_{n\in\mathbb{N}}$ converges strongly to some $x \in \mathcal{H}$ which satisfies $Ax + Bx = f$. The investigation of the method for general set-valued maximally monotone operators was initiated in [265], with subsequent improvements in [37, 42, 128, 179, 366]. See also [393] for further analysis.

To chart the path from the original Douglas–Rachford algorithm to its modern version for monotone set-valued operators, let us go back to the matrix setting. Upon eliminating the intermediate variables $(x_{n+1/2})_{n\in\mathbb{N}}$ in (6.1) and noting that $AJ_A = \operatorname{Id} - J_A$, we obtain

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad x_{n+1} &= J_B\big(x_n - AJ_A(x_n - Bx_n + f) + f\big) \\
&= J_B\big(Bx_n + J_A(x_n - Bx_n + f)\big).
\end{aligned}
\tag{6.2}
$$

Now set $(\forall n \in \mathbb{N})\ x_n = J_B y_n$. Then we derive from (6.2) that

$$(\forall n \in \mathbb{N}) \quad \begin{aligned} y_{n+1} &= BJ_B y_n + J_A(J_B y_n - BJ_B y_n + f) \\ &= y_n - J_B y_n + J_A(2J_B y_n - y_n + f), \end{aligned} \tag{6.3}$$

which leads to the recursion

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = J_B y_n \\ z_n = J_A(2x_n - y_n + f) \\ y_{n+1} = y_n + z_n - x_n. \end{array} \right. \end{array} \tag{6.4}$$

As noted in [265], unlike (6.1), this algorithm is well defined for arbitrary maximally monotone set-valued operators and is now referred to as the Douglas–Rachford splitting algorithm in this context.

Remark 6.1 In particular, upon setting $B = 0$ and $f = 0$ in (6.4) and assuming that $A : \mathcal{H} \to \mathcal{H}$ is hemicontinuous and strongly monotone, it follows from Lieutaud's result [259] that the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the recursion

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = J_A x_n \tag{6.5}$$

converges strongly to a zero of A. This is actually the first instance of convergence of the proximal point algorithm, which has been attributed to later work in the literature. The case when A and B are gradients of convex functions was also considered in [259] in connection with the minimization of the sum of two differentiable convex functions.

6.2 Weak convergence

We present results for a form of the Douglas–Rachford algorithm (6.4) which includes relaxation parameters and a dual inclusion problem.

Theorem 6.2 *Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ and $B : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 2[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$, and let $\gamma \in]0, +\infty[$. Suppose that the set Z of solutions to the inclusion*

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx \tag{6.6}$$

is not empty and let Z^* be the set of solutions to the dual problem

$$\text{find } x^* \in \mathcal{H} \text{ such that } 0 \in -A^{-1}(-x^*) + B^{-1}x^*. \tag{6.7}$$

Let $y_0 \in \mathcal{H}$ and iterate

$$\begin{array}{l} \textit{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = J_{\gamma B} y_n \\ x_n^* = \gamma^{-1}(y_n - x_n) \\ z_n = J_{\gamma A}(2x_n - y_n) \\ y_{n+1} = y_n + \lambda_n (z_n - x_n). \end{array} \right. \end{array} \tag{6.8}$$

Then there exists $y \in \mathcal{H}$ such that $y_n \rightharpoonup y$. Now set $x = J_{\gamma B} y$ and $x^* = {}^{\gamma}By$. Then the following hold:

(i) $x_n \rightharpoonup x \in Z$.

(ii) $x_n^* \rightharpoonup x^* \in Z^*$.

Proof. We rely on the embedding of Example 3.17. Set

$$R_{\gamma A} = 2J_{\gamma A} - \mathrm{Id}, \quad R_{\gamma B} = 2J_{\gamma B} - \mathrm{Id}, \quad \text{and} \quad \mathcal{M} = \left(\frac{R_{\gamma A} \circ R_{\gamma B} + \mathrm{Id}}{2} \right)^{-1} - \mathrm{Id}. \tag{6.9}$$

Then it follows from (2.33) and Lemma 2.34(iii) that $(R_{\gamma A} \circ R_{\gamma B} + \mathrm{Id})/2$ is firmly nonexpansive and that $\mathcal{M}$ is maximally monotone. In addition, [37, Proposition 26.1(iii)(b)] asserts that

$$\varnothing \neq Z = J_{\gamma B}(\operatorname{zer} \mathcal{M}), \tag{6.10}$$

while [37, Proposition 26.1(iii)(c)] asserts that

$$\varnothing \neq Z^* = {}^{\gamma}B(\operatorname{zer} \mathcal{M}). \tag{6.11}$$

Furthermore, we derive from (6.8) and (6.9) that

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = y_n + \frac{\lambda_n}{2}\big(R_{\gamma A}(R_{\gamma B} y_n) - y_n\big) = y_n + \lambda_n \big(J_{\mathcal{M}} y_n - y_n\big), \tag{6.12}$$

i.e., $(y_n)_{n \in \mathbb{N}}$ is constructed by the proximal point algorithm (5.3) for $\mathcal{M}$. Since (6.10) implies that $\operatorname{zer} \mathcal{M} \neq \varnothing$, Theorem 5.1(i) asserts that

$$J_{\mathcal{M}} y_n - y_n \to 0 \quad \text{and} \quad (\exists\, y \in \operatorname{zer} \mathcal{M}) \quad y_n \rightharpoonup y. \tag{6.13}$$

In turn, (6.10) yields $x = J_{\gamma B} y \in Z$, while (6.8) yields

$$z_n - x_n = J_{\gamma A}(2x_n - y_n) - x_n = J_{\mathcal{M}} y_n - y_n \to 0. \tag{6.14}$$

(i): Let us set

$$(\forall n \in \mathbb{N}) \quad z_n^* = \gamma^{-1}(2x_n - y_n - z_n). \tag{6.15}$$

Then (6.8) and (2.18) yield

$$(\forall n \in \mathbb{N}) \quad \begin{cases} (z_n, z_n^*) \in \operatorname{gra} A \\ (x_n, x_n^*) \in \operatorname{gra} B \\ x_n - z_n = \gamma(x_n^* + z_n^*). \end{cases} \tag{6.16}$$

Since Lemma 2.34(iii) asserts that $J_{\gamma B}$ is nonexpansive,

$$(\forall n \in \mathbb{N}) \quad \|x_n - x_0\| = \|J_{\gamma B} y_n - J_{\gamma B} y_0\| \leqslant \|y_n - y_0\|. \tag{6.17}$$

Hence, since $(y_n)_{n\in\mathbb{N}}$ is bounded, so is $(x_n)_{n\in\mathbb{N}}$. Now take $z \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup z$. Then it follows from (6.14), (6.13), (6.15), and (6.16) that

$$z_{k_n} \rightharpoonup z,\ z_{k_n}^* \rightharpoonup \gamma^{-1}(z-y),\ z_n - x_n \to 0,\ \text{and}\ z_n^* + x_n^* = \gamma^{-1}(x_n - z_n) \to 0. \tag{6.18}$$

In turn, Lemma 2.50 yields $z \in \operatorname{zer}(A+B) = Z$,

$$\big(z, \gamma^{-1}(z-y)\big) \in \operatorname{gra} A, \quad \text{and} \quad \big(z, \gamma^{-1}(y-z)\big) \in \operatorname{gra} B. \tag{6.19}$$

Hence, (2.18) implies that

$$z = J_{\gamma B} y. \tag{6.20}$$

Thus, $x = J_{\gamma B} y$ is the unique weak sequential cluster point of the bounded sequence $(x_n)_{n\in\mathbb{N}}$ and therefore, by Lemma 4.1(ii), $x_n \rightharpoonup x$.

(ii): We have $y_n \rightharpoonup y \in \operatorname{zer} \mathcal{M}$ and, by (i), $x_n \rightharpoonup x$. Hence, $x_n^* = \gamma^{-1}(y_n - x_n) \rightharpoonup \gamma^{-1}(y-x) = {}^{\gamma}By = x^*$. In view of (6.11), the proof is complete. ☐

Remark 6.3 The convergence result of [265] is that, for the unrelaxed scheme (6.4), $(y_n)_{n\in\mathbb{N}}$ converges weakly to a point $y \in \mathcal{H}$ such that $J_{\gamma B} y \in Z$ (see [127, 179] for the relaxed case). In the special case when $J_{\gamma B}$ is weakly sequentially continuous, as is the case when $\mathcal{H}$ is finite-dimensional, $x_n = J_{\gamma B} y_n \rightharpoonup J_{\gamma B} y \in Z$. The key fact that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}(A+B)$ without any further assumption was first proved in [366] in the unrelaxed case. Theorem 6.2 was established in [37, Theorem 26.11]. The component of the proof given above up to (6.13) exploits an idea from [179], that identifies the core iteration of (6.8) as an instantiation of the proximal point algorithm.

Remark 6.4 Connections between the Douglas–Rachford algorithms and the method of partial inverses of Section 5.4.4 are discussed in [251, Section 1]; see also [179, Section 5] and [271]. Let us show that we can actually derive Theorem 5.13(ii) from Theorem 6.2. Let $(x_n)_{n\in\mathbb{N}}$, $(x_n^*)_{n\in\mathbb{N}}$, $(p_n)_{n\in\mathbb{N}}$ and $(p_n^*)_{n\in\mathbb{N}}$ be the sequence generated by (5.29) and set $(\forall n \in \mathbb{N})$ $y_n = x_n + x_n^*$ and $z_n = \mathrm{proj}_V(2p_n - y_n)$. Then (5.29) yields

$$(\forall n \in \mathbb{N}) \quad \mathrm{proj}_V\, p_n^* + \mathrm{proj}_{V^\perp}\, p_n = \mathrm{proj}_V(y_n - p_n) + p_n - \mathrm{proj}_V\, p_n = p_n - z_n. \tag{6.21}$$

Altogether,

$$(\forall n \in \mathbb{N}) \quad p_n = J_A y_n, \; z_n = \mathrm{proj}_V(2p_n - y_n), \text{ and } y_{n+1} = y_n + \lambda_n(z_n - p_n). \tag{6.22}$$

In view of Example 2.36, this recursion is precisely that of (6.8) for the operators (N_V, A) with $\gamma = 1$. We therefore derive the following from Theorem 6.2: $(y_n)_{n\in\mathbb{N}}$ converges weakly to a point $y \in \mathcal{H}$ and, if we set $x = J_A y$ and $x^* = y - J_A y$, then $p_n \rightharpoonup x \in \mathrm{zer}(N_V + A)$ and, by Example 2.15, $p_n^* \rightharpoonup x^* \in \mathrm{zer}(N_{V^\perp} + A^{-1})$. Furthermore, (6.19)–(6.20) implies that $(x, -x^*) = (x, x-y) \in \mathrm{gra}\, N_V$ and $(x, x^*) = (x, y - x) \in \mathrm{gra}\, A$. Thus, Example 2.15 yields $(x, x^*) \in \mathrm{gra}\, N_V \cap \mathrm{gra}\, A$ and (x, x^*) therefore solves (5.28). Finally, since [128, Equation (11)] asserts that $J_A y = \mathrm{proj}_V\, y$ and since proj_V is weakly continuous, we have $x_n = \mathrm{proj}_V(x_n + x_n^*) = \mathrm{proj}_V\, y_n \rightharpoonup \mathrm{proj}_V\, y = x$ and $x_n^* = \mathrm{proj}_{V^\perp}\, y_n \rightharpoonup \mathrm{proj}_{V^\perp}\, y = y - \mathrm{proj}_V\, y = x^*$. Let us add that, in this setting, the operator $\mathcal{M}$ of (6.9) is just the partial inverse A_V.

Remark 6.5 The many application areas of the Douglas–Rachford algorithm (in its original two-operator form or transposed in product spaces) include road design [40], equilibrium problems [73], biostatistics [142], signal recovery [143], traffic theory [196], noise removal [365], and compressive sensing [398] (see also [261] for additional references).

6.3 Strong convergence

As shown in [94, Counterexample 2], the convergence of $(x_n)_{n\in\mathbb{N}}$ in Theorem 6.2(i) is only weak. The following version based on Theorem 5.3 furnishes strong convergence.

Theorem 6.6 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, suppose that $\mathrm{zer}(A + B) \neq \varnothing$, let $y_0 \in \mathcal{H}$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 1]$ such that $\inf_{n\in\mathbb{N}} \lambda_n > 0$, and let $\gamma \in\,]0, +\infty[$. Iterate*

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = J_{\gamma B} y_n \\ x_n^* = \gamma^{-1}(y_n - x_n) \\ z_n = J_{\gamma A}(2x_n - y_n) \\ y_{n+1} = \mathsf{Q}\big(y_0, y_n, y_n + \lambda_n(z_n - x_n)\big), \end{array} \right. \end{array} \tag{6.23}$$

where Q is defined in Lemma 4.6. Let Z and Z^* be the sets of solutions to (6.6) and (6.7), respectively. Then the following hold:

(i) $(x_n)_{n\in\mathbb{N}}$ *converges strongly to a point in* Z.

(ii) $(x_n^*)_{n\in\mathbb{N}}$ *converges strongly to a point in* Z^*.

Proof. Define $\mathcal{M}$ as in (6.9) and set $y = \mathrm{proj}_{\mathrm{zer}\,\mathcal{M}}\, y_0$, $x = J_{\gamma B} y$, and $x^* = \gamma^{-1}(y-x)$. Then it follows from (6.10) that $x \in Z$ and from (6.11) that $x^* \in Z^*$. Additionally, we derive from (6.23) that

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = \mathsf{Q}\big(y_0, y_n, y_n + \lambda_n(J_{\mathcal{M}} y_n - y_n)\big). \tag{6.24}$$

Hence, Theorem 5.3 yields $y_n \to y$ and, by continuity of $J_{\gamma B}$, $x_n = J_{\gamma B} y_n \to J_{\gamma B} y = x$. Finally, $x_n^* = \gamma^{-1}(y_n - x_n) \to \gamma^{-1}(y-x) = x^*$. ☐

Remark 6.7 The method of partial inverses of Theorem 5.13 may converge only weakly [94, Counterexample 4]. A strongly convergent version can be designed using Remark 6.4 and Theorem 6.6.

6.4 Special cases and variants

6.4.1 Minimization setting

We illustrate an application of the Douglas–Rachford algorithm to primal-dual minimization.

Example 6.8 Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$ be such that $Z = \mathrm{Argmin}\,(f+g) \neq \varnothing$ and $0 \in \mathrm{sri}(\mathrm{dom}\, f - \mathrm{dom}\, g)$. Set $Z^* = \mathrm{Argmin}\,(f^* \circ (-\mathrm{Id}) + g^*)$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$, let $\gamma \in \,]0,+\infty[$, let $y_0 \in \mathcal{H}$, and iterate

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
x_n = \mathrm{prox}_{\gamma g}\, y_n \\
x_n^* = \gamma^{-1}(y_n - x_n) \\
z_n = \mathrm{prox}_{\gamma f}(2x_n - y_n) \\
y_{n+1} = y_n + \lambda_n(z_n - x_n).
\end{array}
\right.
\end{array} \tag{6.25}$$

Then it follows from Problem 3.9, Example 2.35, and Theorem 6.2 that there exists $(x, x^*) \in Z \times Z^*$ such that $x_n \rightharpoonup x$ and $x_n^* \rightharpoonup x^*$.

Remark 6.9 Relations between the Douglas–Rachford algorithm (6.25) and other methods have been noted in the literature.

(i) It is observed in [155, Section 3.1.1] that the Douglas–Rachford algorithm (6.25) can be viewed as a limiting case of the Chambolle–Pock algorithm (see Remark 5.21(iv)) by implementing it in the case when $\mathcal{G} = \mathcal{H}$, $L = \mathrm{Id}$, and $\sigma = 1/\tau = \gamma$. Note, however, that this setting violates the condition $\tau\sigma\|L\|^2 < 1$ used to prove weak convergence of (5.60) in Example 5.20.

(ii) Consider the setting of Problem 3.9 and note that the primal minimization problem (3.12) is equivalent to

$$\underset{(x,y)\in \operatorname{gra} L}{\text{minimize}} \;\; f(x)+g(y). \tag{6.26}$$

The (unscaled) *augmented Lagrangian* associated with (6.26) is the saddle function (see Example 2.21) on $(\mathcal{H}\oplus\mathcal{G})\oplus\mathcal{G}$ defined as

$$\begin{aligned} F\colon \mathcal{H}\oplus\mathcal{G}\oplus\mathcal{G} &\to \left]-\infty,+\infty\right] \\ (x,y,v^*) &\mapsto f(x)+g(y)+\langle Lx-y \mid v^*\rangle+\frac{1}{2}\|Lx-y\|^2. \end{aligned} \tag{6.27}$$

Iteration n of the alternating-direction method of multipliers (ADMM) consists in minimizing F over x for y_n and v_n^* fixed to get x_n, then over y for x_n and v_n^* fixed to get y_{n+1}, and then applying a proximal maximization step with respect to the Lagrange multiplier v^* for x_n and y_{n+1} fixed to get v_{n+1}^*. It was originally proposed in [208], refined in [198], and further developed in [63, 179, 197, 209]. Given $y_0\in\mathcal{G}$ and $v_0^*\in\mathcal{G}$, ADMM iterates

$$\begin{array}{l} \text{for } n=0,1,\ldots \\ \left\lfloor \begin{array}{l} x_n \in \underset{x\in\mathcal{H}}{\operatorname{Argmin}} \left(f(x)+\langle Lx \mid v_n^*\rangle+\frac{1}{2}\|Lx-y_n\|^2\right) \\ d_n = Lx_n \\ y_{n+1} = \underset{y\in\mathcal{G}}{\operatorname{argmin}} \left(g(y)-\langle y \mid v_n^*\rangle+\frac{1}{2}\|d_n-y\|^2\right) \\ v_{n+1}^* = v_n^*+d_n-y_{n+1}. \end{array} \right. \end{array} \tag{6.28}$$

It should be emphasized that ADMM is not a splitting algorithm in our sense since the computation of x_n involves a minimization step which does not separate f and L, and can therefore be hard to execute. This step is also set-valued in general. Nonetheless, (6.28) can be interpreted as an application of the Douglas–Rachford algorithm (6.25) to the functions $f^*\circ(-L^*)$ (here again, note that f and L are not separated and that the typically non-explicit operator $\operatorname{prox}_{f^*\circ(-L^*)}$ intervenes) and g^* present in the dual problem (3.13) [197] (see also [179]). This is merely an algorithmic identification and not a claim that ADMM converges. Convergence requires more restrictions on the problem, for instance finite-dimensionality of $\mathcal{H}$ and $\mathcal{G}$ and invertibility of $L^*\circ L$ in [179, Section 5]. For further analysis, see [28, 60, 347].

6.4.2 Peaceman–Rachford splitting

The first implicit alternating direction method [55] to solve the positive-definite matrix equation $Ax+Bx=f$ is the Peaceman–Rachford algorithm [307] (see also

[169]). It is described by the iterative process

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_{n+1/2} - x_n + Ax_{n+1/2} + Bx_n = f \\ x_{n+1} - x_{n+1/2} + Ax_{n+1/2} + Bx_{n+1} = f. \end{array} \right. \end{array} \tag{6.29}$$

Using the same arguments used to transition from (6.1) to (6.4), we rewrite (6.29) as

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = J_B y_n \\ z_n = J_A(2x_n - y_n + f) \\ y_{n+1} = y_n + 2(z_n - x_n). \end{array} \right. \end{array} \tag{6.30}$$

The strong convergence of $(x_n)_{n\in\mathbb{N}}$ to a solution to the equation $Ax + Bx = f$, where A and B are single-valued hemicontinuous monotone operators such that $\operatorname{dom} A = \operatorname{dom} B = \mathcal{H}$ and B is strongly monotone, was established in [259] and, with the additional assumption that $\mathcal{H}$ is finite-dimensional and the operators are continuous, in [242].

Algorithm (6.30) was first considered for general maximally monotone set-valued operators A and B in [265]. In the presence of a scaling parameter $\gamma \in \left]0, +\infty\right[$ and taking $f = 0$ without loss of generality, the Peaceman–Rachford algorithm becomes

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} x_n = J_{\gamma B} y_n \\ z_n = J_{\gamma A}(2x_n - y_n) \\ y_{n+1} = y_n + 2(z_n - x_n), \end{array} \right. \end{array} \tag{6.31}$$

Upon defining $\mathcal{M}$ as in (6.9), we derive from (6.31) that

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = (2J_{\mathcal{M}} - \mathrm{Id})y_n. \tag{6.32}$$

We can view (6.31) as a limiting case of the Douglas–Rachford algorithm (6.8) in which the relaxation parameters $(\lambda_n)_{n\in\mathbb{N}}$ are allowed to be 2. This, of course, means that (6.31) operates outside of the setting of Theorem 5.1 and hence of the geometric framework of Theorem 4.2. As a result, the weak convergence of $(y_n)_{n\in\mathbb{N}}$ cannot be guaranteed without additional assumptions since (6.32) amounts to iterating a merely nonexpansive operator (see [265, Remark 6] for a counterexample). Strong convergence of $(x_n)_{n\in\mathbb{N}}$ to a point in $\operatorname{zer}(A + B)$ takes place when B is strongly monotone [265, Remark 2]. More generally, strong convergence occurs when B is uniformly monotone on bounded sets or when $\operatorname{int}\operatorname{Fix}(2J_{\gamma A} - \mathrm{Id})(2J_{\gamma B} - \mathrm{Id}) \neq \varnothing$ [128, Remark 2.2(iv)].

6.4.3 A three-operator splitting algorithm

An extension of the Douglas–Rachford algorithm (6.8) was proposed in [161] by adding a cocoercive operator to the inclusion (6.6).

Proposition 6.10 *Let $\tau \in \left]0,+\infty\right[$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let $C\colon \mathcal{H} \to \mathcal{H}$ be τ-cocoercive. Suppose that the set Z of solutions to the inclusion*

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx + Cx \tag{6.33}$$

is not empty and let Z^* be the set of solutions to the dual problem

$$\text{find } x^* \in \mathcal{H} \text{ such that } 0 \in -(A+C)^{-1}(-x^*) + B^{-1}x^*. \tag{6.34}$$

Let $\gamma \in \left]0,2\tau\right[$, set $\delta = 2-\gamma/(2\tau)$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,\delta\right[$ such that $\sum_{n\in\mathbb{N}} \lambda_n(\delta-\lambda_n) = +\infty$, and let $y_0 \in \mathcal{H}$. Iterate

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
x_n = J_{\gamma B}\, y_n\\
x_n^* = \gamma^{-1}(y_n - x_n)\\
r_n = y_n + \gamma C x_n\\
z_n = J_{\gamma A}(2x_n - r_n)\\
y_{n+1} = y_n + \lambda_n(z_n - x_n).
\end{array}
\right.
\end{array}
\tag{6.35}
$$

Then there exists $y \in \mathcal{H}$ such that $y_n \rightharpoonup y$. Now set $x = J_{\gamma B} y$ and $x^* = {}^{\gamma}B y$. Then the following hold:

(i) $x_n \rightharpoonup x \in Z$.

(ii) $x_n^* \rightharpoonup x^* \in Z^*$.

Proof. Remarkably, we can closely follow the proof of Theorem 6.2. The key additional facts established in [161, Proposition 2.1 and Lemma 2.2] are that, for $\alpha = 1/\delta$,

$$T = J_{\gamma A}\circ\big(2J_{\gamma B} - \mathrm{Id} - \gamma C\circ J_{\gamma B}\big) + \mathrm{Id} - J_{\gamma B} \text{ is } \alpha\text{-averaged and } Z = J_{\gamma B}(\operatorname{Fix} T). \tag{6.36}$$

We write the maximally monotone operator $\mathcal{M}$ of (3.26) as

$$\mathcal{M} = \left(\mathrm{Id} + \frac{1}{2\alpha}\Big(J_{\gamma A}\circ\big(2J_{\gamma B} - \mathrm{Id} - \gamma C\circ J_{\gamma B}\big) - J_{\gamma B}\Big)\right)^{-1} - \mathrm{Id} \tag{6.37}$$

and, in view of Example 3.16 and (6.36), work with the embedding $(\mathcal{H}, \mathcal{M}, J_{\gamma B})$ of (6.33). Then $\varnothing \neq Z = J_{\gamma B}(\operatorname{zer}\mathcal{M})$ and $(y_n)_{n\in\mathbb{N}}$ is produced by the proximal point algorithm $(\forall n \in \mathbb{N})$ $y_{n+1} = y_n + \mu_n(J_{\mathcal{M}} y_n - y_n)$, where $\mu_n = 2\alpha\lambda_n \in \left]0,2\right[$.

Using Theorem 5.1(i), we infer that $(y_n)_{n\in\mathbb{N}}$ converges weakly to a point $y \in \operatorname{zer}\mathcal{M}$ and that $J_{\mathcal{M}}y_n - y_n \to 0$. Hence, we derive from (6.36), (6.35), and (6.37) that

$$x = J_{\gamma B}y \in Z \;\; \textit{and} \;\; z_n - x_n = 2\alpha(J_{\mathcal{M}}y_n - y_n) \to 0, \tag{6.38}$$

and hence that

$$\|Cz_n - Cx_n\| \leqslant \alpha^{-1}\|z_n - x_n\| \to 0. \tag{6.39}$$

(i): Set $(\forall n \in \mathbb{N})$ $z_n^* = \gamma^{-1}(2x_n - z_n - r_n) + Cz_n$. In view of (6.35) and (2.18),

$$(\forall n \in \mathbb{N}) \quad \begin{cases} (z_n, z_n^*) \in \operatorname{gra}(A + C) \\ (x_n, x_n^*) \in \operatorname{gra} B \\ z_n^* + x_n^* = \gamma^{-1}(x_n - z_n) + Cz_n - Cx_n. \end{cases} \tag{6.40}$$

Next, fix $z \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup z$. Since $y_{k_n} \rightharpoonup y$, it follows from (6.38), (6.39), (6.40), and (6.35) that

$$z_{k_n} \rightharpoonup z,\; z_{k_n}^* \rightharpoonup \gamma^{-1}(z - y),\; z_n - x_n \to 0, \;\; \text{and} \;\; z_n^* + x_n^* \to 0. \tag{6.41}$$

By applying Lemma 2.50 to the maximally monotone operators $A + C$ (see Example 2.5 and Lemma 2.27(ii)) and B, we deduce from (6.40) and (6.41) that $z \in \operatorname{zer}(A + C + B) = Z$,

$$\big(z, \gamma^{-1}(z - y)\big) \in \operatorname{gra}(A + C), \quad \text{and} \quad \big(z, \gamma^{-1}(y - z)\big) \in \operatorname{gra} B. \tag{6.42}$$

In turn, (2.18) asserts that $z = J_{\gamma B}y$, making $x = J_{\gamma B}y$ the unique weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ which is bounded since $(y_n)_{n\in\mathbb{N}}$ is. By Lemma 4.1(ii), $x_n \rightharpoonup x$.

(ii): Since $y_n \rightharpoonup y$ and $x_n \rightharpoonup x$, we have $x_n^* = \gamma^{-1}(y_n - x_n) \rightharpoonup \gamma^{-1}(y - x) = {}^{\gamma}\!By = x^* \in Z^*$ by (6.11). ☐

Remark 6.11 Here are a few comments on Proposition 6.10.

(i) The conclusion of Proposition 6.10(i) was first established in [161, Theorem 2.1.1(b)] with a different proof. See also [321] for a discussion and connections with [322].

(ii) The duality result of Proposition 6.10(ii) is new.

(iii) A strongly convergent version of Proposition 6.10 can be obtained by adapting the proof of Theorem 6.6 to the presence of C, as was done above.

(iv) When $C = 0$, Proposition 6.10 produces the Douglas–Rachford setting of Theorem 6.2. When $B = 0$, (6.35) yields a special case of the forward-backward method of [154, Proposition 4.4(iii)] in which the proximal parameters are all equal to γ.

7 Tseng's forward-backward-forward splitting

7.1 Preview

In Section 5.4.1, we have discussed a Euler method for finding a zero of a single-valued operator $B\colon \mathcal{H} \to \mathcal{H}$ under a cocoercivity condition. Under the more general assumption that B is monotone and β-Lipschitzian, the Euler method is no longer appropriate, and we can use a scheme proposed by Antipin [12] and Korpelevič [246] that involves a double activation of the operator B. Specifically, in this method, $\gamma \in]0, 1/\beta[$ and $x_0 \in \mathcal{H}$ are fixed and we iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
b_n^* = \gamma B x_n \\
m_n = x_n - b_n^* \\
m_n^* = B m_n \\
x_{n+1} = x_n - \gamma m_n^*.
\end{array}
\right.
\end{array}
\tag{7.1}
$$

Clearly, the sequence $(m_n, m_n^*)_{n\in\mathbb{N}}$ lies gra B and it is straightforward to see that, by choosing $(\lambda_n)_{n\in\mathbb{N}}$ suitably in (4.32), we obtain (7.1). The convergence properties of the Antipin–Korpelevič method can therefore be deduced from the results of Section 4.4 applied to B.

Tseng's algorithm can be viewed as a generalization of (7.1) for the problem of finding a zero of $A + B$, where $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone and B is as above. It is called the forward-backward-forward algorithm because it performs a forward step on B, then a backward step on A, and finally another forward step on B. We are going to derive the convergence of Tseng's forward-backward-forward splitting algorithm from the principles of Section 4.4 and, more precisely, from the warped resolvent algorithm of Section 4.5.

7.2 Fejérian algorithm

We cast the forward-backward-forward algorithm as an instance of (4.34) and then prove its weak convergence via Theorem 4.12. This result was originally established in [375, Theorem 3.4(b)], where different arguments were used.

Theorem 7.1 *Let $\beta \in]0, +\infty[$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $B\colon \mathcal{H} \to \mathcal{H}$ be monotone and β-Lipschitzian, and suppose that $Z = \operatorname{zer}(A + B) \neq \varnothing$. Let $x_0 \in \mathcal{H}$, let $\varepsilon \in]0, 1/(\beta + 1)[$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (1 - \varepsilon)/\beta]$. Iterate*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
b_n^* = \gamma_n B x_n \\
m_n = J_{\gamma_n A}(x_n - b_n^*) \\
x_{n+1} = m_n - \gamma_n B m_n + b_n^*.
\end{array}
\right.
\end{array}
\tag{7.2}
$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.

Proof. Our objective is to apply Theorem 4.12 with

$$W = A + B,\ C = 0,\ \text{and}\ (\forall n \in \mathbb{N})\ \ U_n = \gamma_n^{-1}\mathrm{Id} - B \text{ and } q_n = w_n. \tag{7.3}$$

Since $C = 0$, let us rename $(w_n)_{n\in\mathbb{N}}$ as $(m_n)_{n\in\mathbb{N}}$. Example 2.3 and Lemma 2.27(ii) entail that W is maximally monotone. Moreover, a consequence of Lemma 2.48(i)–(ii) is that

$$(\forall n \in \mathbb{N})\ \ \gamma_n U_n \text{ is } \varepsilon\text{-strongly monotone and } 1/(2-\varepsilon)\text{-cocoercive}. \tag{7.4}$$

Additionally, we derive from [95, Proposition 3.9] that

$$(\forall n \in \mathbb{N})\quad \operatorname{ran} U_n \subset \operatorname{ran}(U_n + W + C)\ \ \text{and}\ \ U_n + W + C \text{ is injective}. \tag{7.5}$$

We also observe that

$$(\forall n \in \mathbb{N})\quad J^{U_n}_{W+C} = J^{U_n}_{A+B} = \big(\gamma_n^{-1}\mathrm{Id}+A\big)\circ\big(\gamma_n^{-1}\mathrm{Id}-B\big) = J_{\gamma_n A}\circ(\mathrm{Id}-\gamma_n B). \tag{7.6}$$

Hence, the variables of (4.34) in this setting become

$$(\forall n \in \mathbb{N})\quad \begin{cases} m_n = J_{\gamma_n A}(x_n - \gamma_n B x_n)\\ t_n^* = U_n x_n - U_n m_n\\ \delta_n = \langle m_n - x_n \mid U_n m_n - U_n x_n\rangle. \end{cases} \tag{7.7}$$

Now set

$$(\forall n \in \mathbb{N})\quad \lambda_n = \begin{cases} \dfrac{\gamma_n\|t_n^*\|^2}{\delta_n}, & \text{if } \delta_n > 0;\\ \varepsilon, & \text{otherwise.} \end{cases} \tag{7.8}$$

We derive from (7.4) that

$$(\forall n \in \mathbb{N})\quad \delta_n = \langle m_n - x_n \mid U_n m_n - U_n x_n\rangle \geqslant \beta\varepsilon\|m_n - x_n\|^2, \tag{7.9}$$

which implies that

$$(\forall n \in \mathbb{N})\quad \delta_n \leqslant 0 \ \Leftrightarrow\ m_n = x_n \ \Leftrightarrow\ t_n^* = 0. \tag{7.10}$$

A consequence of (7.4) is that, if $\delta_n > 0$,

$$\frac{\varepsilon}{\gamma_n} \leqslant \frac{\|U_n m_n - U_n x_n\|}{\|m_n - x_n\|} \leqslant \frac{\|U_n m_n - U_n x_n\|^2}{\langle m_n - x_n \mid U_n m_n - U_n x_n\rangle} \leqslant \frac{2-\varepsilon}{\gamma_n} \tag{7.11}$$

and we therefore obtain from (7.8) that

$$\lambda_n = \frac{\gamma_n \|U_n m_n - U_n x_n\|^2}{\langle m_n - x_n \mid U_n m_n - U_n x_n\rangle} \in [\varepsilon, 2-\varepsilon]. \tag{7.12}$$

Hence, (4.34) and (7.10) yield

$$(\forall n \in \mathbb{N}) \quad d_n = \frac{\gamma_n}{\lambda_n} t_n^*. \tag{7.13}$$

Consequently, the sequence $(x_n)_{n\in\mathbb{N}}$ produced by (7.2) coincides with that of (4.34). We therefore appeal to Theorem 4.12(ii) to conclude since its condition (ii)(b) holds thanks to (7.12), whereas its condition (ii)(d) holds thanks to (7.4) and the fact that $(\gamma_n)_{n\in\mathbb{N}}$ lies in $[\varepsilon, (1-\varepsilon)/\beta]$. ☐

7.3 Haugazeau-like algorithm

We present a strongly convergent best approximation version of the forward-backward-forward method based on Theorem 4.14.

Theorem 7.2 *Let $\beta \in \left]0, +\infty\right[$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $B\colon \mathcal{H} \to \mathcal{H}$ be monotone and β-Lipschitzian, and suppose that $Z = \operatorname{zer}(A+B) \neq \varnothing$. Let $x_0 \in \mathcal{H}$, let $\varepsilon \in \left]0, 1/(\beta+1)\right[$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (1-\varepsilon)/\beta]$. Iterate*

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} b_n^* = \gamma_n B x_n \\ m_n = J_{\gamma_n A}(x_n - b_n^*) \\ r_n = \dfrac{1}{2}\big(x_n + m_n - \gamma_n B m_n + b_n^*\big) \\ x_{n+1} = \mathsf{Q}(x_0, x_n, r_n), \end{array} \right. \end{array} \tag{7.14}$$

where Q is defined in Lemma 4.6. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to $\operatorname{proj}_Z x_0$.

Proof. We prove the claim as an application of Theorem 4.14 in the setting of (7.3). Let us use the same variables as in (7.7) and

$$(\forall n \in \mathbb{N}) \quad \lambda_n = \begin{cases} \dfrac{\gamma_n \|t_n^*\|^2}{2\delta_n}, & \text{if } \delta_n > 0; \\ \varepsilon/2, & \text{otherwise.} \end{cases} \tag{7.15}$$

Then, using the same arguments as in the proof of Theorem 4.12, we see that $(\lambda_n)_{n\in\mathbb{N}}$ lies in $[\varepsilon/2, 1]$ and that the sequence $(x_n)_{n\in\mathbb{N}}$ produced by (7.14) coincides with that of (4.44). Since conditions (ii)(b) and (ii)(d) in Theorem 4.14(ii) are fulfilled, we obtain the claim. ☐

7.4 Special cases and variants

7.4.1 The monotone+skew algorithm

The approach presented here was proposed in [76] to solve the monotone inclusion (3.7) and it was the first algorithm to fully split the operators A, B, and L. Its methodology conforms to the program of Framework 1.2: we use the embedding of Example 3.20 to transfer the initial 3-operator problem (3.7) in the primal space $\mathcal{H}$ to one involving the Kuhn–Tucker operator $\mathcal{K} = \boldsymbol{M}+\boldsymbol{S}$ of (3.10) in the larger primal-dual space $\mathsf{X} = \mathcal{H} \oplus \mathcal{G}$. The algorithmic strategy *per se* is then straightforward: since $\boldsymbol{M}$ is maximally monotone and $\boldsymbol{S}$ is monotone and Lipschitzian, we can apply Tseng's forward-backward-forward algorithm (Theorem 7.1) in X to find a Kuhn–Tucker point and hence a primal-dual solution.

We derive from Theorem 7.1 the weak convergence of the monotone+skew algorithm of [76, Theorem 3.1(ii)] (we can derive a strongly convergent version from Theorem 7.2 using the same arguments).

Proposition 7.3 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone, and assume that $0 \neq L \in \mathcal{B}(\mathcal{H},\mathcal{G})$. Suppose that the set Z of solutions to the primal inclusion*

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + L^*\big(B(Lx)\big) \tag{7.16}$$

is not empty and let Z^* be the set of solutions to the dual inclusion

$$\text{find } y^* \in \mathcal{G} \text{ such that } 0 \in -L\big(A^{-1}(-L^*y^*)\big) + B^{-1}y^*. \tag{7.17}$$

Let $x_0 \in \mathcal{H}$, let $y_0^* \in \mathcal{G}$, let $\varepsilon \in \left]0, 1/(\|L\|+1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (1-\varepsilon)/\|L\|\,]$, and set

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots\\
\left\lfloor
\begin{array}{l}
y_{1,n} = x_n - \gamma_n L^* y_n^*\\
y_{2,n}^* = y_n^* + \gamma_n L x_n\\
m_{1,n} = J_{\gamma_n A} y_{1,n}\\
m_{2,n}^* = J_{\gamma_n B^{-1}} y_{2,n}^*\\
q_{1,n} = m_{1,n} - \gamma_n L^* m_{2,n}^*\\
q_{2,n}^* = m_{2,n}^* + \gamma_n L m_{1,n}\\
x_{n+1} = x_n - y_{1,n} + q_{1,n}\\
y_{n+1}^* = y_n^* - y_{2,n}^* + q_{2,n}^*.
\end{array}
\right.
\end{array} \tag{7.18}$$

Then there exist $x \in Z$ and $y^* \in Z^*$ such that $-L^*y^* \in Ax$, $y^* \in B(Lx)$, $x_n \rightharpoonup x$, and $y_n^* \rightharpoonup y^*$.

Proof. Set $\mathbf{X} = \mathcal{H} \oplus \mathcal{G}$, define $\boldsymbol{M}$ and $\boldsymbol{S}$ as in (3.9), and set $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = (x_n, y_n^*)$, $\boldsymbol{y}_n = (y_{1,n}, y_{2,n}^*)$, $\boldsymbol{m}_n = (m_{1,n}, m_{2,n}^*)$, and $\boldsymbol{q}_n = (q_{1,n}, q_{2,n}^*)$, Then, in view of Example 2.37, (7.18) becomes

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{y}_n = \boldsymbol{x}_n - \gamma_n \boldsymbol{S}\boldsymbol{x}_n \\
\boldsymbol{m}_n = J_{\gamma_n \boldsymbol{M}} \boldsymbol{y}_n \\
\boldsymbol{q}_n = \boldsymbol{m}_n - \gamma_n \boldsymbol{S}\boldsymbol{m}_n \\
\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \boldsymbol{y}_n + \boldsymbol{q}_n,
\end{array}
\right.
\end{array}
\tag{7.19}
$$

which we rewrite as an instance of (7.2), namely,

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{b}_n^* = \gamma_n \boldsymbol{S}\boldsymbol{x}_n \\
\boldsymbol{m}_n = J_{\gamma_n \boldsymbol{M}} (\boldsymbol{x}_n - \boldsymbol{b}_n^*) \\
\boldsymbol{x}_{n+1} = \boldsymbol{m}_n - \gamma_n \boldsymbol{S}\boldsymbol{m}_n + \boldsymbol{b}_n^*.
\end{array}
\right.
\end{array}
\tag{7.20}
$$

It then follows from Theorem 7.1 and Lemma 3.8 that $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}(\boldsymbol{M} + \boldsymbol{S}) \subset Z \times Z^*$, as claimed. ☐

Remark 7.4 The methodology of Theorem 7.1 is to find a Kuhn–Tucker point, i.e., a zero of $\boldsymbol{M} + \boldsymbol{S}$. As noted in [76, Remark 2.9], this can also be achieved by using the Douglas–Rachford algorithm (6.8) which, upon setting $U = (\mathrm{Id} + \gamma^2 L^* \circ L)^{-1}$ and $V = (\mathrm{Id} + \gamma^2 L \circ L^*)^{-1}$, and taking $\gamma \in \left]0, +\infty\right[$ and a sequence $(\lambda_n)_{n\in\mathbb{N}}$ in $\left]0, 2\right[$ such that $\sum_{n\in\mathbb{N}} \lambda_n (2 - \lambda_n) = +\infty$, assumes the form

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
x_n = U(y_{1,n} - \gamma L^* y_{2,n}^*) \\
y_n^* = V(y_{2,n}^* + \gamma L y_{1,n}) \\
y_{1,n+1} = y_{1,n} + \lambda_n \big(J_{\gamma A}(2x_n - y_{1,n}) - x_n \big) \\
y_{2,n+1}^* = y_{2,n}^* + \lambda_n \big(J_{\gamma B^{-1}}(2y_n^* - y_{2,n}^*) - y_n^* \big).
\end{array}
\right.
\end{array}
\tag{7.21}
$$

Weak convergence of $(x_n, y_n^*)_{n\in\mathbb{N}}$ to a point in $Z \times Z^*$ follows from Theorem 6.2(i). The numerical effectiveness of (7.21) depends on the ease of implementation of the operators U and V. This approach was rediscovered in [300] in an image restoration application.

7.4.2 A Lagrangian approach to composite minimization

We revisit the setting of Problem 3.9, which was identified as an instance of Problem 3.7 and can therefore be solved using (7.18) with $A = \partial f$ and $B =$

∂g. Following [131, Section 4.5], we explore a different route which amounts to employing the embedding $(\mathbf{X}, \mathcal{S}_{\boldsymbol{F}}, \mathfrak{T})$, where $\mathbf{X} = \mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G}$,

$$\begin{aligned} \mathcal{S}_{\boldsymbol{F}}\colon \quad & \mathbf{X} && \to 2^{\mathbf{X}} \\ & (x, y, v^*) && \mapsto \big(\partial f(x) + L^* v^*\big) \times \big(\partial g(y) - v^*\big) \times \{-Lx + y\} \end{aligned} \tag{7.22}$$

is the saddle operator of (3.24), and $\mathfrak{T}\colon \mathbf{X} \to \mathcal{H}\colon (x, y, v^*) \mapsto x$. Let us write $\mathcal{S}_{\boldsymbol{F}} = \boldsymbol{M} + \boldsymbol{S}$, where

$$\begin{cases} \boldsymbol{M}\colon (x, y, v^*) \mapsto \partial f(x) \times \partial g(y) \times \{0\} \\ \boldsymbol{S}\colon (x, y, v^*) \mapsto \big(L^* v^*, -v^*, -Lx + y\big). \end{cases} \tag{7.23}$$

Then $\|\boldsymbol{S}\| = \sqrt{1 + \|L\|^2}$ and $(\forall n \in \mathbb{N})$ $J_{\gamma_n \boldsymbol{M}} = \operatorname{prox}_{\gamma_n f} \times \operatorname{prox}_{\gamma_n g} \times \mathrm{Id}$. Hence, applying Theorem 7.1 to this decomposition in $\mathbf{X}$, we obtain the following realization of Framework 1.2.

Proposition 7.5 *Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ be such that* $0 \in \operatorname{sri}(L(\operatorname{dom} f) - \operatorname{dom} g)$. *Suppose that the primal problem*

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + g(Lx) \tag{7.24}$$

admits solutions and consider the dual problem

$$\underset{v^* \in \mathcal{G}}{\text{minimize}} \;\; f^*(-L^* v^*) + g^*(v^*). \tag{7.25}$$

Let $(x_0, y_0, v_0^*) \in \mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G}$, let $\varepsilon \in \left]0, 1/(1 + \sqrt{1 + \|L\|^2})\right[$, and let $(\gamma_n)_{n \in \mathbb{N}}$ be a sequence in $[\varepsilon, (1 - \varepsilon)/\sqrt{1 + \|L\|^2}]$. Iterate

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} r_n = \gamma_n (L x_n - y_n) \\ m_{1,n} = \operatorname{prox}_{\gamma_n f} \big(x_n - \gamma_n L^* v_n^*\big) \\ m_{2,n} = \operatorname{prox}_{\gamma_n g} \big(y_n + \gamma_n v_n^*\big) \\ x_{n+1} = m_{1,n} - \gamma_n L^* r_n \\ y_{n+1} = m_{2,n} + \gamma_n r_n \\ v_{n+1}^* = v_n^* + \gamma_n \big(L m_{1,n} - m_{2,n}\big). \end{array} \right. \end{array} \tag{7.26}$$

Then $(x_n)_{n \in \mathbb{N}}$ and $(v_n^*)_{n \in \mathbb{N}}$ converge weakly to solutions to (7.24) and (7.25), respectively.

Remark 7.6 Let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon,(1-\varepsilon)\min\{1,1/\|L\|\}/2]$. Algorithm (7.26) bears a certain resemblance with the iterative scheme

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
p_n = v_n^* + \mu_n(Lx_n - y_n)\\
x_{n+1} = \operatorname{prox}_{\mu_n f}\big(x_n - \mu_n L^* p_n\big)\\
y_{n+1} = \operatorname{prox}_{\mu_n g}\big(y_n + \mu_n p_n\big)\\
v_{n+1}^* = v_n^* + \mu_n\big(Lx_{n+1} - y_{n+1}\big)
\end{array}
\right.
\end{array}
\tag{7.27}
$$

proposed in [118] to solve (7.24)–(7.25) in a finite-dimensional setting.

Remark 7.7 In the finite-dimensional context of [177], the saddle operator (7.22) was split as $\mathcal{S}_{\boldsymbol{F}} = \boldsymbol{M}_1 + \boldsymbol{M}_2$, where

$$
\begin{cases}
\boldsymbol{M}_1\colon (x,y,v^*) \mapsto \big(\partial f(x) + L^*v^*\big) \times \{0\} \times \{-Lx\}\\
\boldsymbol{M}_2\colon (x,y,v^*) \mapsto \{0\} \times \big(\partial g(y) - v^*\big) \times \{y\}.
\end{cases}
\tag{7.28}
$$

Given $\gamma \in\,]0,+\infty[$, $\mu_1 \in \mathbb{R}$, $\mu_2 \in \mathbb{R}$, and $(x_0,y_0,v_0^*) \in \mathcal{H}\oplus\mathcal{G}\oplus\mathcal{G}$, applying the Douglas–Rachford algorithm (6.8) to find a zero of $\boldsymbol{M}_1+\boldsymbol{M}_2$ leads to the algorithm [177]

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
x_{n+1} \in \underset{x\in\mathcal{H}}{\operatorname{Argmin}}\ \left(f(x) + \langle Lx \mid v_n^*\rangle + \dfrac{1}{2\gamma}\|Lx - y_n\|^2 + \dfrac{\gamma\mu_1^2}{2}\|x - x_n\|^2\right)\\
y_{n+1} = \underset{y\in\mathcal{G}}{\operatorname{argmin}}\ \left(g(y) - \langle y \mid v_n^*\rangle + \dfrac{1}{2\gamma}\|Lx_{n+1} - y\|^2 + \dfrac{\gamma\mu_2^2}{2}\|y - y_n\|^2\right)\\
v_{n+1}^* = v_n^* + \gamma^{-1}\big(Lx_{n+1} - y_{n+1}\big).
\end{array}
\right.
\end{array}
\tag{7.29}
$$

When $\mu_1 = \mu_2 = 0$, we recover the alternating direction method of multipliers (ADMM) discussed in Remark 6.9(ii). Just like ADMM, (7.29) necessitates a potentially complex minimization involving f and L jointly to construct x_{n+1}. By contrast, (7.26) achieves full splitting of f, g, and L.

Remark 7.8 In view of Example 3.23, the above saddle operator formalism can be extended to the more general primal-dual inclusion pair of Problem 3.7. As in Proposition 7.5, a zero (x,y,v^*) of the saddle operator $\mathcal{S}$ of (3.25) can be constructed by executing (7.26), where $\operatorname{prox}_{\gamma_n f}$ is replaced with $J_{\gamma_n A}$ and $\operatorname{prox}_{\gamma_n g}$ with $J_{\gamma_n B}$. In this setting, the weak limits x and v^* solve, respectively, the primal inclusion (3.7) and the dual inclusion (3.8).

7.4.3 Mixtures of composite, Lipschitzian, and parallel-sum operators

The Kuhn–Tucker operator of Lemma 3.8 employed in Section 7.4.1 can be expressed in block format as

$$\boldsymbol{\mathcal{K}} = \boldsymbol{M} + \boldsymbol{S} = \underbrace{\begin{bmatrix} A & 0 \\ 0 & B^{-1} \end{bmatrix}}_{\text{monotone}} + \underbrace{\begin{bmatrix} 0 & L^* \\ -L & 0 \end{bmatrix}}_{\text{skew}}. \tag{7.30}$$

A Kuhn–Tucker point was obtained in Proposition 7.3 by applying the forward-backward-forward algorithm (7.2) to $\boldsymbol{M}$ and $\boldsymbol{S}$. In doing so, we did not exploit the linearity and skewness of $\boldsymbol{S}$, but just the fact that it is monotone and Lipschitzian. Let us observe that, if we fill the diagonal of $\boldsymbol{S}$ with monotone Lipschitzian operators $Q\colon \mathcal{H} \to \mathcal{H}$ and $D^{-1}\colon \mathcal{G} \to \mathcal{G}$, we obtain a new monotone and Lipschitzian operator $\boldsymbol{Q}\colon \mathbf{X} \to \mathbf{X}$. In lieu of (7.30), we then consider the decomposition

$$\boldsymbol{\mathcal{K}} = \boldsymbol{M} + \boldsymbol{Q} = \underbrace{\begin{bmatrix} A & 0 \\ 0 & B^{-1} \end{bmatrix}}_{\text{monotone}} + \underbrace{\begin{bmatrix} Q & L^* \\ -L & D^{-1} \end{bmatrix}}_{\text{monotone and Lipschitzian}}. \tag{7.31}$$

Using (2.62), we write

$$\boldsymbol{\mathcal{K}} = \begin{bmatrix} A + Q & L^* \\ -L & (B \,\square\, D)^{-1} \end{bmatrix} \tag{7.32}$$

and interpret it as a variant of the Kuhn–Tucker operator (3.10) associated with Problem 3.7 in which A is replaced with $A + Q$ and B with $B \,\square\, D$. In other words, the primal inclusion is to

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + L^*\big((B \,\square\, D)(Lx)\big) + Qx \tag{7.33}$$

and the dual inclusion is to

$$\text{find } y^* \in \mathcal{G} \text{ such that } 0 \in -L\big((A+Q)^{-1}(-L^*y^*)\big) + B^{-1}y^* + D^{-1}y^* \tag{7.34}$$

or, equivalently,

$$\text{find } y^* \in \mathcal{G} \text{ such that } (\exists x \in \mathcal{H}) \begin{cases} -L^*y^* \in Ax + Qx \\ Lx \in B^{-1}y^* + D^{-1}y^*. \end{cases} \tag{7.35}$$

As in Lemma 3.8, for every $(x, y^*) \in \mathbf{X}$,

$$(x, y^*) \in \operatorname{zer} \boldsymbol{\mathcal{K}} \quad \Rightarrow \quad \begin{cases} x \text{ solves (7.33)} \\ y^* \text{ solves (7.35)} \end{cases} \tag{7.36}$$

and we therefore recover the embedding principle of Framework 1.2.

Example 7.9 In the above setting, set $\mathbf{X} = \mathcal{H} \oplus \mathcal{G}$, let $\mathcal{K}$ be the Kuhn–Tucker operator of (7.32), and let $\mathcal{T}\colon \mathbf{X} \to \mathcal{H}\colon (x, y^*) \mapsto x$. Then $(\mathbf{X}, \mathcal{K}, \mathcal{T})$ is an embedding of (7.33).

The primal-dual inclusion problem (7.33)–(7.34) was first investigated in [145], where it was solved via Tseng's forward-backward-forward algorithm. Here is [145, Theorem 3.1(ii)(c)–(d)], which describes this approach when the operators L, B, and D above are deployed in a product space $\mathcal{G} = \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p$ in the spirit of Problem 3.11 (further analysis of the asymptotic behavior of the method in special cases can be found in [62]).

Proposition 7.10 *Let* $0 < p \in \mathbb{N}$, *let* $\mu \in \left]0, +\infty\right[$, *let* $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ *be maximally monotone, let* $Q\colon \mathcal{H} \to \mathcal{H}$ *be monotone and* μ*-Lipschitzian. For every* $k \in \{1, \ldots, p\}$, *let* $\nu_k \in \left]0, +\infty\right[$, *let* $\mathcal{G}_k$ *be a real Hilbert space, let* $B_k\colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ *be maximally monotone, let* $D_k\colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ *be maximally monotone and such that* $D_k^{-1}\colon \mathcal{G}_k \to \mathcal{G}_k$ *is* ν_k*-Lipschitzian, and assume that* $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$. *Suppose that the set* Z *of solutions to the primal inclusion*

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + \sum_{k=1}^{p} L_k^*\big((B_k \,\square\, D_k)(L_k x)\big) + Qx \tag{7.37}$$

is not empty and let Z^* *be the set of solutions to the dual inclusion*

$$\begin{aligned}&\text{find } y_1^* \in \mathcal{G}_1, \ldots, y_p^* \in \mathcal{G}_p \text{ such that}\\ &\qquad (\exists x \in \mathcal{H}) \begin{cases} -\sum_{k=1}^{p} L_k^* y_k^* \in Ax + Qx \\ \big(\forall k \in \{1, \ldots, p\}\big)\; L_k x \in B_k^{-1} y_k^* + D_k^{-1} y_k^*. \end{cases}\end{aligned} \tag{7.38}$$

Set

$$\beta = \max\{\mu, \nu_1, \ldots, \nu_p\} + \sqrt{\sum_{k=1}^{p} \|L_k\|^2}, \tag{7.39}$$

let $x_0 \in \mathcal{H}$, *let* $(y_{1,0}^*, \ldots, y_{p,0}^*) \in \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p$, *let* $\varepsilon \in \left]0, 1/(\beta+1)\right[$, *and let*

$(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $[\varepsilon,(1-\varepsilon)/\beta]$. *Iterate*

$$
\begin{array}{l}
\textit{for } n=0,1,\ldots\\
\left\lfloor
\begin{array}{l}
y_{1,n}=x_n-\gamma_n\big(Qx_n+\sum_{k=1}^{p}L_k^*y_{k,n}^*\big)\\
m_{1,n}=J_{\gamma_nA}\,y_{1,n}\\
\text{for } k=1,\ldots,p\\
\left\lfloor
\begin{array}{l}
y_{2,k,n}^*=y_{k,n}^*+\gamma_n\big(L_kx_n-D_k^{-1}y_{k,n}^*\big)\\
m_{2,k,n}^*=J_{\gamma_nB_k^{-1}}y_{2,k,n}^*\\
q_{2,k,n}^*=m_{2,k,n}^*+\gamma_n\big(L_km_{1,n}-D_k^{-1}m_{2,k,n}^*\big)\\
y_{k,n+1}^*=y_{k,n}^*-y_{2,k,n}^*+q_{2,k,n}^*
\end{array}
\right.\\
q_{1,n}=m_{1,n}-\gamma_n\big(Qm_{1,n}+\sum_{k=1}^{p}L_k^*m_{2,k,n}^*\big)\\
x_{n+1}=x_n-y_{1,n}+q_{1,n}.
\end{array}
\right.
\end{array}
\tag{7.40}
$$

Then there exist $x\in Z$ *and* $(y_1^*,\ldots,y_p^*)\in Z^*$ *such that* $x_n\rightharpoonup x$, *and, for every* $k\in\{1,\ldots,p\}$, $y_{k,n}^*\rightharpoonup y_k^*$.

Proof. The duality between (7.37) and (7.38) follows as in Problem 3.11, by replacing A with $A+Q$ and $(B_k^{-1})_{1\leqslant k\leqslant p}$ with $(B_k^{-1}+D_k^{-1})_{1\leqslant k\leqslant p}$. Now set

$$
\begin{cases}
\mathcal{G}=\mathcal{G}_1\oplus\cdots\oplus\mathcal{G}_p\\
B\colon\mathcal{G}\to 2^{\mathcal{G}}\colon(y_1,\ldots,y_p)\mapsto B_1y_1\times\cdots\times B_py_p\\
D\colon\mathcal{G}\to 2^{\mathcal{G}}\colon(y_1,\ldots,y_p)\mapsto D_1y_1\times\cdots\times D_py_p\\
L\colon\mathcal{H}\to\mathcal{G}\colon x\mapsto(L_1x,\ldots,L_px),
\end{cases}
\tag{7.41}
$$

define $\boldsymbol{M}$ and $\boldsymbol{Q}$ as in (7.31), and set

$$
(\forall n\in\mathbb{N})\qquad
\begin{cases}
\boldsymbol{x}_n=\big(x_n,y_{1,n}^*,\ldots,y_{p,n}^*\big)\\
\boldsymbol{m}_n=\big(m_{1,n},m_{2,1,n}^*,\ldots,m_{2,p,n}^*\big).
\end{cases}
\tag{7.42}
$$

Then $\boldsymbol{M}$ is maximally monotone and $\boldsymbol{Q}$ is monotone and β-Lipschitzian [145, Equation (3.11)] and, following the same steps as in the proof of Proposition 7.3, we rewrite (7.40) as

$$
\begin{array}{l}
\text{for } n=0,1,\ldots\\
\left\lfloor
\begin{array}{l}
\boldsymbol{b}_n^*=\gamma_n\boldsymbol{Q}\boldsymbol{x}_n\\
\boldsymbol{m}_n=J_{\gamma_n\boldsymbol{M}}(\boldsymbol{x}_n-\boldsymbol{b}_n^*)\\
\boldsymbol{x}_{n+1}=\boldsymbol{m}_n-\gamma_n\boldsymbol{Q}\boldsymbol{m}_n+\boldsymbol{b}_n^*
\end{array}
\right.
\end{array}
\tag{7.43}
$$

and conclude by invoking Theorem 7.1 and (7.36). ☐

Remark 7.11 In (7.37), suppose that $p=1$, $\mathcal{G}_1=\mathcal{H}$, $L_1=\mathrm{Id}$, $B_1=B$, $D_1=\{0\}^{-1}$, and $\mathrm{zer}(A+B+Q)\neq\varnothing$. Let $x_0\in\mathcal{H}$, let $y_0^*\in\mathcal{H}$, let $\varepsilon\in\,]0,1/(\mu+2)[$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (1-\varepsilon)/(\mu+1)]$. Then we deduce from Proposition 7.10 that the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the iterations

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
y_n = x_n - \gamma_n\big(Qx_n + y_n^*\big) \\
p_n = J_{\gamma_n A}\, y_n \\
q_n^* = J_{\gamma_n B^{-1}}\big(y_n^* + \gamma_n x_n\big) \\
x_{n+1} = x_n - y_n + p_n - \gamma_n\big(Qp_n + q_n^*\big) \\
y_{n+1}^* = q_n^* + \gamma_n(p_n - x_n).
\end{array}
\right.
\end{array}
\tag{7.44}
$$

converges weakly to a zero of $A + B + Q$. An alternative method to solve this inclusion is proposed in [349], with constant proximal parameters $(\gamma_n)_{n\in\mathbb{N}}$ and the feature that it coincides with the unrelaxed version of the Douglas–Rachford algorithm when $Q = 0$ (in the spirit of the method of Section 6.4.3 where Q is cocoercive).

Example 7.12 In Proposition 7.10, make the additional assumptions that $Q = 0$ and, for every $k \in \{1, \ldots, p\}$, $\mathcal{G}_k = \mathcal{H}$, $L_k = \mathrm{Id}$, and D_k^{-1} is strictly monotone. Then (7.37) collapses to

$$
\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + \sum_{k=1}^{p} (B_k \,\square\, D_k)(x). \tag{7.45}
$$

It is shown in [130, Proposition 4.2] that (7.45) is an exact relaxation of the (possibly inconsistent) instance of the problem

$$
\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax \text{ and } \big(\forall k \in \{1, \ldots, p\}\big)\ 0 \in B_k x \tag{7.46}
$$

in the sense that the solutions to (7.45) are the same as those to (7.46) when the latter happen to exist.

The specialization of Proposition 7.10 to minimization is as follows. It features the ability to split infimal convolutions (see (2.7)) together with linearly composed functions.

Example 7.13 ([145, Theorem 4.2(ii)(b)–(c)]) Let $0 < p \in \mathbb{N}$, let $\mu \in \left]0, +\infty\right[$, let $f \in \Gamma_0(\mathcal{H})$, and let $h\colon \mathcal{H} \to \mathbb{R}$ be convex, differentiable, and such that ∇h is μ-Lipschitzian. For every $k \in \{1, \ldots, p\}$, let $\nu_k \in \left]0, +\infty\right[$, let $\mathcal{G}_k$ be a real Hilbert space, let $g_k \in \Gamma_0(\mathcal{G}_k)$, let $\ell_k \in \Gamma_0(\mathcal{G}_k)$ be $1/\nu_k$-strongly convex, and suppose that $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$. Let Z be the set of solutions to the primal problem

$$
\underset{x\in\mathcal{H}}{\text{minimize}}\ f(x) + \sum_{k=1}^{p} (g_k \,\square\, \ell_k)(L_k x) + h(x), \tag{7.47}
$$

let Z^* be the set of solutions to the dual problem

$$\underset{y_1^*\in\mathcal{G}_1,\dots,y_p^*\in\mathcal{G}_p}{\text{minimize}} \quad (f^*\,\square\, h^*)\biggl(-\sum_{k=1}^{p} L_k^* y_k^*\biggr) + \sum_{k=1}^{p}\bigl(g_k^*(y_k^*) + \ell_k^*(y_k^*)\bigr), \tag{7.48}$$

and suppose that

$$\operatorname{zer}\biggl(\partial f + \sum_{k=1}^{p} L_k^* \circ \bigl(\partial g_k \,\square\, \partial \ell_k\bigr) \circ L_k + \nabla h\biggr) \neq \varnothing. \tag{7.49}$$

Set

$$\beta = \max\{\mu, \nu_1, \dots, \nu_p\} + \sqrt{\sum_{k=1}^{p} \|L_k\|^2}, \tag{7.50}$$

let $x_0 \in \mathcal{H}$, let $(y_{1,0}^*, \dots, y_{p,0}^*) \in \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p$, let $\varepsilon \in\,]0, 1/(\beta+1)[$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (1-\varepsilon)/\beta]$. Iterate

$$\begin{array}{l}
\text{for } n = 0, 1, \dots \\
\left\lfloor
\begin{array}{l}
y_{1,n} = x_n - \gamma_n\bigl(\nabla h(x_n) + \sum_{k=1}^{p} L_k^* y_{k,n}^*\bigr) \\
m_{1,n} = \operatorname{prox}_{\gamma_n f}\, y_{1,n} \\
\text{for } k = 1, \dots, p \\
\left\lfloor
\begin{array}{l}
y_{2,k,n}^* = y_{k,n}^* + \gamma_n\bigl(L_k x_n - \nabla \ell_k^*(y_{k,n}^*)\bigr) \\
m_{2,k,n}^* = \operatorname{prox}_{\gamma_n g_k^*}\, y_{2,k,n}^* \\
q_{2,k,n}^* = m_{2,k,n}^* + \gamma_n\bigl(L_k m_{1,n} - \nabla \ell_k^*(m_{2,k,n}^*)\bigr) \\
y_{k,n+1}^* = y_{k,n}^* - y_{2,k,n}^* + q_{2,k,n}^*
\end{array}
\right. \\
q_{1,n} = m_{1,n} - \gamma_n\bigl(\nabla h(m_{1,n}) + \sum_{k=1}^{p} L_k^* m_{2,k,n}^*\bigr) \\
x_{n+1} = x_n - y_{1,n} + q_{1,n}.
\end{array}
\right.
\end{array} \tag{7.51}$$

Then there exist $x \in Z$ and $(y_1^*, \dots, y_p^*) \in Z^*$ such that $x_n \rightharpoonup x$, and, for every $k \in \{1, \dots, p\}$, $y_{k,n}^* \rightharpoonup y_k^*$.

Remark 7.14 Conditions under which (7.49) holds are provided in [145, Proposition 4.3].

8 Forward-backward splitting

8.1 Preview

The forward-backward splitting method is a basic algorithm for solving Problem 3.1 when B is cocoercive. At iteration n, given a step size $\gamma_n \in\,]0, +\infty[$, a discrete dynamics associated with the Cauchy problem (5.1) with $M = A + B$ is

$$\frac{x_n - x_{n+1}}{\gamma_n} \in Ax_{n+1} + Bx_n. \tag{8.1}$$

It amounts to performing a forward Euler step relative to the operator B and a backward Euler step relative to the operator A. In view of (2.18), this means that $x_{n+1} = J_{\gamma_n A}(x_n - \gamma_n B x_n)$. This iteration scheme goes back to the gradient-projection method [213, 258] for the constrained minimization of a smooth function (see Example 8.7 below) and its extension to variational inequalities [27, 277].

8.2 Fejérian algorithm

We establish a new, geometric proof of the convergence of a relaxed primal-dual version of the forward-backward algorithm found in [154, Proposition 4.4(iii)] for the primal result and in [37, Theorem 26.14(ii)] for the dual result, where the proximal parameters $(\gamma_n)_{n\in\mathbb{N}}$ are constant. Related primal results and special cases can be found in [197, 254, 255, 278, 374]. The importance of cocoercivity in establishing weak convergence was first identified by Mercier [277] in the context of variational inequalities and, more generally, in [278].

Theorem 8.1 *Let $\alpha \in \left]0, +\infty\right[$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let $B\colon \mathcal{H} \to \mathcal{H}$ be α-cocoercive. Let $\varepsilon \in \left]0, \alpha/(\alpha+1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (2-\varepsilon)\alpha]$, and let*

$$(\forall n \in \mathbb{N}) \quad \varepsilon \leqslant \mu_n \leqslant (1-\varepsilon)\frac{4\alpha - \gamma_n}{2\alpha}. \tag{8.2}$$

Suppose that the set Z of solutions to the problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx \tag{8.3}$$

is not empty and let Z^* be the set of solutions to the dual problem

$$\text{find } x^* \in \mathcal{H} \text{ such that } 0 \in -A^{-1}(-x^*) + B^{-1}x^*. \tag{8.4}$$

Let $x_0 \in \mathcal{H}$ and iterate

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} b_n^* = \gamma_n B x_n \\ w_n = J_{\gamma_n A}(x_n - b_n^*) \\ x_{n+1} = x_n + \mu_n(w_n - x_n). \end{array} \right. \end{array} \tag{8.5}$$

Then the following hold:

(i) *$(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.*

(ii) *Z^* contains a single point $\overline{x}^*$ and $(\forall z \in Z)\ Bz = \overline{x}^*$.*

(iii) *$(Bx_n)_{n\in\mathbb{N}}$ converges strongly to $\overline{x}^*$.*

Proof. The proof hinges on an application of Theorem 4.12 with

$$W = A,\ C = B,\ \text{and}\ (\forall n \in \mathbb{N})\ \ U_n = \gamma_n^{-1}\mathrm{Id} - B \text{ and } q_n = x_n. \tag{8.6}$$

In this setting

$$(\forall n \in \mathbb{N}) \quad J_{W+C}^{U_n} = J_{A+B}^{U_n} = \big(\gamma_n^{-1}\mathrm{Id}+A\big)\circ\big(\gamma_n^{-1}\mathrm{Id}-B\big) = J_{\gamma_n A}\circ(\mathrm{Id}-\gamma_n B) \tag{8.7}$$

and the variables of (4.34) become

$$(\forall n \in \mathbb{N}) \quad \begin{cases} w_n = J_{\gamma_n A}(x_n - \gamma_n B x_n) \\ t_n^* = \dfrac{x_n - w_n}{\gamma_n} \\ \delta_n = \left(\dfrac{1}{\gamma_n} - \dfrac{1}{4\alpha}\right)\|w_n - x_n\|^2. \end{cases} \tag{8.8}$$

Furthermore, we derive from [95, Proposition 3.9] that (7.5) holds. Now set

$$(\forall n \in \mathbb{N}) \quad \lambda_n = \frac{4\alpha\mu_n}{4\alpha - \gamma_n}. \tag{8.9}$$

Then (8.2) yields

$$(\forall n \in \mathbb{N}) \quad \varepsilon \leqslant \frac{4\alpha\varepsilon}{4\alpha-\varepsilon} \leqslant \lambda_n \leqslant \frac{4\alpha(1-\varepsilon)(4\alpha-\gamma_n)}{(4\alpha-\gamma_n)2\alpha} \leqslant 2-\varepsilon. \tag{8.10}$$

We also deduce from (8.8) that

$$(\forall n \in \mathbb{N}) \quad \delta_n \leqslant 0 \ \Leftrightarrow\ w_n = x_n \ \Leftrightarrow\ t_n^* = 0. \tag{8.11}$$

Hence, (4.34) yields

$$(\forall n \in \mathbb{N}) \quad d_n = \frac{\mu_n}{\lambda_n}(x_n - w_n). \tag{8.12}$$

Altogether, we arrive at the conclusion that the sequence $(x_n)_{n\in\mathbb{N}}$ produced by (8.5) coincides with that of (4.34). Hence, by Theorem 4.12(i) and (8.10),

$$\sum_{n\in\mathbb{N}} \|d_n\|^2 < +\infty. \tag{8.13}$$

In turn, upon invoking (8.12), we obtain

$$w_n - x_n \to 0. \tag{8.14}$$

(i): In view of (8.10), condition (ii)(b) in Theorem 4.12(ii) is fulfilled. On the other hand, since Lemma 2.48(iii) asserts that the operators $(\gamma_n U_n)_{n\in\mathbb{N}}$ are nonexpansive, (8.14) implies that $\|U_n w_n - U_n x_n\| \leqslant \|w_n - x_n\|/\varepsilon \to 0$, so that condition (ii)(c) is also fulfilled. Thus, the assertion follows from Theorem 4.12(ii).

(ii): The strong monotonicity of B^{-1} implies that of $-A^{-1}\circ(-\mathrm{Id})+B^{-1}$. Hence, [37, Corollary 23.37(ii)] asserts that (8.4) admits a unique solution $\overline{x}^*$. Now let $z \in Z$. Then $-Bz \in Az$ and therefore $-z \in -A^{-1}(-Bz)$. Thus, $0 = -z + z \in -A^{-1}(-Bz) + B^{-1}(Bz)$, i.e., $Bz \in Z^* = \{\overline{x}^*\}$.

(iii): It follows from (i) and (8.14) that $(x_n)_{n\in\mathbb{N}}$ and $(w_n)_{n\in\mathbb{N}}$ are bounded. Now let $z \in Z$. We retrieve from (4.27) that

$$(\forall n \in \mathbb{N}) \quad \langle z - w_n \mid t_n^* \rangle \leqslant \langle x_n - w_n \mid Bx_n - Bz \rangle - \alpha\|Bx_n - Bz\|^2. \tag{8.15}$$

Hence, the Cauchy–Schwarz inequality, (2.32), (8.8), and (8.14) imply that

$$\begin{aligned}\alpha\|Bx_n - Bz\|^2 &\leqslant \|w_n - x_n\|\,\|Bx_n - Bz\| + \|w_n - z\|\,\|t_n^*\| \\ &\leqslant \frac{1}{\alpha}\|w_n - x_n\|\,\|x_n - z\| + \frac{1}{\gamma_n}\|w_n - z\|\,\|w_n - x_n\| \\ &\to 0. \end{aligned} \tag{8.16}$$

In view of (ii), $Bx_n \to Bz = \overline{x}^*$. ☐

The following examples address Example 3.2 and Example 3.3, respectively.

Example 8.2 Let $\alpha \in \left]0,+\infty\right[$, let $f \in \Gamma_0(\mathcal{H})$, let $B\colon \mathcal{H} \to \mathcal{H}$ be α-cocoercive, suppose that the set Z of solutions to the variational inequality

$$\text{find } x \in \mathcal{H} \text{ such that } (\forall y \in \mathcal{H}) \ \langle x - y \mid Bx \rangle + f(x) \leqslant f(y) \tag{8.17}$$

is not empty, and let Z^* be the set of solutions to the dual problem

$$\text{find } x^* \in \mathcal{H} \text{ such that } 0 \in -\partial f^*(-x^*) + B^{-1}x^*. \tag{8.18}$$

Let $x_0 \in \mathcal{H}$, let $\varepsilon \in \left]0, \alpha/(\alpha+1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (2-\varepsilon)\alpha]$, and suppose that $(\mu_n)_{n\in\mathbb{N}}$ satisfies (8.2). Iterate

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} b_n^* = \gamma_n B x_n \\ w_n = \operatorname{prox}_{\gamma_n f}(x_n - b_n^*) \\ x_{n+1} = x_n + \mu_n(w_n - x_n). \end{array}\right.\end{array} \tag{8.19}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z and $(Bx_n)_{n\in\mathbb{N}}$ converges strongly to the unique point in Z^*.

Proof. Use Example 2.12 and Example 2.35 and set $A = \partial f$ in Theorem 8.1. ☐

Example 8.3 Let $\alpha \in]0, +\infty[$, let C be a nonempty closed convex subset of $\mathcal{H}$, let $B\colon \mathcal{H} \to \mathcal{H}$ be α-cocoercive, suppose that the set Z of solutions to the variational inequality

$$\text{find } x \in C \text{ such that } (\forall y \in C) \ \langle x - y \mid Bx\rangle \leqslant 0 \tag{8.20}$$

is not empty, and let Z^* be the set of solutions to the dual problem

$$\text{find } x^* \in \mathcal{H} \text{ such that } 0 \in -\partial\sigma_C(-x^*) + B^{-1}x^*. \tag{8.21}$$

Let $x_0 \in \mathcal{H}$, let $\varepsilon \in]0, \alpha/(\alpha+1)[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (2-\varepsilon)\alpha]$, and suppose that $(\mu_n)_{n\in\mathbb{N}}$ satisfies (8.2). Iterate

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} b_n^* = \gamma_n B x_n \\ w_n = \operatorname{proj}_C(x_n - b_n^*) \\ x_{n+1} = x_n + \mu_n(w_n - x_n). \end{array} \right. \end{array} \tag{8.22}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z and $(Bx_n)_{n\in\mathbb{N}}$ converges strongly to the unique point in Z^*.

Proof. Use Example 2.36 and (2.2), and set $f = \iota_C$ in Example 8.2. □

The following example focuses on the minimization in the setting of Problem 3.5(ii). This framework has found a multitude of applications, especially in the areas of signal processing and machine learning [15, 45, 115, 149, 152, 164, 232, 382].

Example 8.4 Let $\beta \in]0, +\infty[$, let $f \in \Gamma_0(\mathcal{H})$ and let $g\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable. Suppose that ∇g is β-Lipschitzian and that the set Z of solutions to the problem

$$\underset{x\in\mathcal{H}}{\text{minimize}} \ f(x) + g(x) \tag{8.23}$$

is not empty, and let Z^* be the set of solutions to the dual problem

$$\underset{x^*\in\mathcal{H}}{\text{minimize}} \ f^*(-x^*) + g^*(x^*). \tag{8.24}$$

Let $x_0 \in \mathcal{H}$, let $\varepsilon \in]0, 1/(\beta+1)[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (2-\varepsilon)/\beta]$, and suppose that

$$(\forall n \in \mathbb{N}) \quad \varepsilon \leqslant \mu_n \leqslant (1-\varepsilon)\frac{4-\beta\gamma_n}{2}. \tag{8.25}$$

Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
b_n^* = \gamma_n \nabla g(x_n) \\
w_n = \operatorname{prox}_{\gamma_n f}(x_n - b_n^*) \\
x_{n+1} = x_n + \mu_n (w_n - x_n).
\end{array}
\right.
\end{array}
\tag{8.26}
$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z and $(\nabla g(x_n))_{n\in\mathbb{N}}$ converges strongly to the unique point in Z^*.

Proof. The claim is established by applying Theorem 8.1(i) with $A = \partial f$ (see Example 2.12) and $B = \nabla g$ (see Lemma 2.2). □

Remark 8.5 In some applications, it may be of interest to quantify the asymptotic behavior of the function values $(f(x_n) + g(x_n))_{n\in\mathbb{N}}$ produced by (8.26). This topic has been the focus of a lot of interest since the publication of the influential papers [43, 44, 112]; see [201] and its bibliography for recent results on the unrelaxed implementation of (8.26) with constant proximal parameters.

The following example, taken from [152], models linear inverse problems in which the prior knowledge is modeled by penalizing the coefficients of the decomposition of the ideal solution in an orthonormal basis (see [149, 160, 193] for special cases).

Example 8.6 Suppose that $\mathcal{H}$ is separable, let $(e_k)_{k\in\mathbb{K}\subset\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$, let $y \in \mathcal{G}$, suppose that $0 \neq L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$, and let $(\phi_k)_{k\in\mathbb{K}}$ be functions in $\Gamma_0(\mathbb{R})$ such that $(\forall k \in \mathbb{K})\ \phi_k \geqslant 0 = \phi_k(0)$. Suppose that the set Z of solutions to the problem

$$
\underset{x\in\mathcal{H}}{\text{minimize}} \quad \sum_{k\in\mathbb{K}} \phi_k\big(\langle x \mid e_k\rangle\big) + \frac{1}{2}\|Lx - y\|^2 \tag{8.27}
$$

is not empty. Let $x_0 \in \mathcal{H}$, let $\varepsilon \in \big]0, 1/(\|L\|^2 + 1)\big[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\big[\varepsilon, (2-\varepsilon)/\|L\|^2\big]$, and suppose that

$$
(\forall n \in \mathbb{N}) \quad \varepsilon \leqslant \mu_n \leqslant (1-\varepsilon)\frac{4 - \|L\|^2\gamma_n}{2}. \tag{8.28}
$$

Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
b_n^* = \gamma_n L^*(Lx_n - y) \\
w_n = \sum_{k\in\mathbb{K}} \big(\operatorname{prox}_{\gamma_n \phi_k} \langle x_n - b_n^* \mid e_k\rangle\big) e_k \\
x_{n+1} = x_n + \mu_n (w_n - x_n).
\end{array}
\right.
\end{array}
\tag{8.29}
$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.

Proof. Set $f\colon x \mapsto \sum_{k\in\mathbb{K}} \phi_k(\langle x \mid e_k\rangle)$ and $g\colon x \mapsto \|Lx-y\|^2/2$. Then, as shown in [152, Example 2.19], $f \in \Gamma_0(\mathcal{H})$ and $\operatorname{prox}_{\gamma f}\colon x \mapsto \sum_{k\in\mathbb{K}}(\operatorname{prox}_{\gamma_n\phi_k}\langle x \mid e_k\rangle)e_k$. On the other hand, g is convex and differentiable and $\nabla g\colon x \mapsto L^*(Lx-y)$ is $\|L\|^2$-Lipschitzian. Altogether, the conclusion follows from Example 8.4. ☐

Next, we specialize Example 8.4 to the gradient-projection method, which minimizes a smooth function over a convex set (see Example 3.6) and goes back to [213, 258].

Example 8.7 Let $\beta \in \left]0,+\infty\right[$, let C be a nonempty closed convex subset of $\mathcal{H}$, and let $g\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable. Suppose that ∇g is β-Lipschitzian and that the set Z of solutions to the problem

$$\underset{x\in C}{\text{minimize}}\;\; g(x) \tag{8.30}$$

is not empty, and let Z^* be the set of solutions to the dual problem

$$\underset{x^*\in\mathcal{H}}{\text{minimize}}\;\; \sigma(-x^*) + g^*(x^*). \tag{8.31}$$

Let $x_0 \in \mathcal{H}$, let $\varepsilon \in \left]0, 1/(\beta+1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (2-\varepsilon)/\beta]$, and suppose that $(\mu_n)_{n\in\mathbb{N}}$ satisfies (8.25). Iterate

$$\begin{array}{l}\text{for } n = 0, 1, \ldots\\ \left\lfloor\begin{array}{l} b_n^* = \gamma_n\nabla g(x_n)\\ w_n = \operatorname{proj}_C(x_n - b_n^*)\\ x_{n+1} = x_n + \mu_n(w_n - x_n).\end{array}\right.\end{array} \tag{8.32}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z and $(\nabla g(x_n))_{n\in\mathbb{N}}$ converges strongly to the unique point in Z^*.

Proof. Set $f = \iota_C$ in Example 8.4. Alternatively, set $B = \nabla g$ in Example 8.3. ☐

Remark 8.8 In [22], the backward-forward iterations

$$\begin{array}{l}\text{for } n = 0, 1, \ldots\\ \left\lfloor\begin{array}{l} p_n = J_{\gamma A}x_n\\ q_n = p_n - \gamma Bp_n\\ x_{n+1} = x_n + \mu_n(q_n - x_n)\end{array}\right.\end{array} \tag{8.33}$$

are studied and shown to be related to the forward-backward iterations applied to Yosida envelopes of B and A.

8.3 Haugazeau-like algorithm

As seen in [152, Remark 5.12], the strong convergence of $(x_n)_{n\in\mathbb{N}}$ in Theorem 8.1(i) may fail. Item (i) below on the strong convergence of a best approximation forward-backward algorithm extends [140, Theorem 5.6(i) and Remark 5.5], where $(\forall n \in \mathbb{N})\ \gamma_n = \gamma \in \,]0, 2\alpha[$ and $\mu_n \leqslant 1$.

Theorem 8.9 *Let $\alpha \in \,]0, +\infty[$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let $B\colon \mathcal{H} \to \mathcal{H}$ be α-cocoercive. Let $\varepsilon \in \,]0, \min\{1/2, 2\alpha\}[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 2\alpha]$, and let*

$$(\forall n \in \mathbb{N}) \quad \varepsilon \leqslant \mu_n \leqslant \frac{4\alpha - \gamma_n}{4\alpha}. \tag{8.34}$$

Suppose that the set Z of solutions to the problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Bx \tag{8.35}$$

is not empty and let Z^* be the set of solutions to the dual

$$\text{find } x^* \in \mathcal{H} \text{ such that } 0 \in -A^{-1}(-x^*) + B^{-1}x^*. \tag{8.36}$$

Let $x_0 \in \mathcal{H}$ and iterate

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} b_n^* = \gamma_n B x_n \\ w_n = J_{\gamma_n A}(x_n - b_n^*) \\ x_{n+1} = \mathsf{Q}\big(x_0, x_n, x_n + \mu_n(w_n - x_n)\big), \end{array} \right. \end{array} \tag{8.37}$$

where Q is defined in Lemma 4.6. Then the following hold:

(i) *$(x_n)_{n\in\mathbb{N}}$ converges strongly to $\operatorname{proj}_Z x_0$.*

(ii) *Z^* contains a single point $\overline{x}^*$ and $(Bx_n)_{n\in\mathbb{N}}$ converges strongly to $\overline{x}^*$.*

Proof. We apply Theorem 4.14 in the setting of (8.6), using the same variables as in (8.8) and $(\lambda_n)_{n\in\mathbb{N}}$ defined as in (8.9). Then (8.11) holds and

$$(\forall n \in \mathbb{N}) \quad \varepsilon \leqslant \frac{4\alpha\varepsilon}{4\alpha - \varepsilon} \leqslant \lambda_n \leqslant 1. \tag{8.38}$$

Therefore the sequence $(x_n)_{n\in\mathbb{N}}$ produced by (8.37) coincides with that of (4.44). Hence, by Theorem 4.14(i),

$$w_n - x_n \to 0. \tag{8.39}$$

(i): This follows from Theorem 4.14(ii) since, as in the proof of Theorem 8.1(i), its conditions (ii)(b) and (ii)(d) are fulfilled.

(ii): Since B is continuous, (i) and Theorem 8.1(ii) imply that $Bx_n \to B(\operatorname{proj}_Z x_0) \in Z^*$, where Z^* is a singleton. ☐

8.4 Special cases and variants

8.4.1 Projected Landweber method

In inverse problems, constrained least-squares estimation has a long history [51, 53, 184, 298, 313]. We address the numerical solution of this problem from the viewpoint of the forward-backward algorithm to obtain a relaxed version of the projected Landweber method with iteration-dependent parameters.

Proposition 8.10 *Let $\mathcal{G}$ be a real Hilbert space, suppose that $0 \neq L \in \mathcal{B}(\mathcal{H},\mathcal{G})$, let $y \in \mathcal{G}$, and let C be a closed convex subset of $\mathcal{H}$ such that the set Z of solutions to the problem*

$$\underset{x \in C}{\text{minimize}} \quad \frac{1}{2}\|Lx - y\|^2 \tag{8.40}$$

is not empty. Let $x_0 \in \mathcal{H}$, let $\varepsilon \in \left]0, 1/(\|L\|^2+1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\left[\varepsilon, (2-\varepsilon)/\|L\|^2\right]$, and suppose that $(\mu_n)_{n\in\mathbb{N}}$ satisfies (8.28). Iterate

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
b_n^* = \gamma_n L^*(Lx_n - y) \\
w_n = \operatorname{proj}_C(x_n - b_n^*) \\
x_{n+1} = x_n + \mu_n(w_n - x_n).
\end{array}
\right.
\end{array} \tag{8.41}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.

Proof. Apply Example 8.7 with $g \colon x \mapsto \|Lx - y\|^2/2$. $\square$

Proposition 8.10 was established in [184, Section 3.1] with $(\forall n \in \mathbb{N})$ $\lambda_n = 1$ and $\gamma_n = \gamma \in \left]0, 2/\|L\|^2\right[$. There, it was also conjectured that the convergence was strong, which was disproved in [152, Remark 5.12]. This motivates the following result.

Proposition 8.11 *Let $\mathcal{G}$ be a real Hilbert space, suppose that $0 \neq L \in \mathcal{B}(\mathcal{H},\mathcal{G})$, let $y \in \mathcal{G}$, let C be a closed convex subset of $\mathcal{H}$, and suppose that the set Z of solutions to (8.40) is not empty. Let $x_0 \in \mathcal{H}$, let $\varepsilon \in \left]0, \min\{1/2, 2/\|L\|^2\}\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\left[\varepsilon, 2/\|L\|^2\right]$, and suppose that $(\forall n \in \mathbb{N})$ $\varepsilon \leqslant \mu_n \leqslant 1 - \|L\|^2\gamma_n/4$. Iterate*

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
b_n^* = \gamma_n L^*(Lx_n - y) \\
w_n = \operatorname{proj}_C(x_n - b_n^*) \\
x_{n+1} = \mathsf{Q}\big(x_0, x_n, x_n + \mu_n(w_n - x_n)\big),
\end{array}
\right.
\end{array} \tag{8.42}$$

where Q is defined in Lemma 4.6(ii). Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to $\operatorname{proj}_Z x_0$.

Proof. Follow the pattern of the proof of Proposition 8.10 and use Example 2.36 to apply Theorem 8.9(i) with $A = N_C$ and $B : x \mapsto L^*(Lx - y)$. ☐

Here is an application of Proposition 8.10 to the problem of finding the best approximation to a point from a linearly transformed convex set.

Example 8.12 Consider the setting of Proposition 8.10 with the assumption that $L(C)$ is closed, which guarantees that (8.40) admits solutions. Then $x_n \rightharpoonup x$, where x solves (8.40). Furthermore, if we set $p = Lx$, then $p = \operatorname{proj}_{L(C)} y$. Hence, upon rewriting (8.41) as

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
q_n = Lx_n \\
b_n^* = \gamma_n L^*(q_n - y) \\
w_n = \operatorname{proj}_C(x_n - b_n^*) \\
x_{n+1} = x_n + \mu_n(w_n - x_n)
\end{array}
\right.
\end{array}
\tag{8.43}
$$

and invoking the weak continuity of L, we conclude that $q_n \rightharpoonup \operatorname{proj}_{L(C)} y$.

Example 8.13 Let $\mathcal{G}$ be a real Hilbert space, and suppose that $0 \neq L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$ and that $\operatorname{ran} L$ is closed. Additionally, let $x_0 \in \mathcal{H}$, let $\varepsilon \in \left]0, 1/(\|L\|^2 + 1)\right[$, and let $(\nu_n)_{n\in\mathbb{N}}$ be a sequence in $\left[\varepsilon, (2-\varepsilon)/\|L\|^2\right]$. Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
q_n = Lx_n \\
x_{n+1} = x_n - \nu_n L^* q_n
\end{array}
\right.
\end{array}
\tag{8.44}
$$

and let q be the minimal-norm element of $\operatorname{ran} L$. Then $q_n \rightharpoonup q$.

Proof. Apply Example 8.12 with $C = \mathcal{H}$ and $y = 0$. ☐

The next example is about a composite best approximation problem.

Example 8.14 Let $\mathcal{G}$ be a real Hilbert space, let $y \in \mathcal{G}$, and let $0 < p \in \mathbb{N}$. For every $k \in \{1, \ldots, p\}$, let $\mathcal{H}_k$ be a real Hilbert space, let C_k be a nonempty closed convex subset of $\mathcal{H}_k$, let $0 \neq L_k \in \mathcal{B}(\mathcal{H}_k, \mathcal{G})$, and let $x_{k,0} \in \mathcal{H}_k$. Suppose that $\sum_{k=1}^p L_k(C_k)$ is closed and set $\beta = \sum_{k=1}^p \|L_k\|^2$. Furthermore, let $\varepsilon \in \left]0, 1/(\beta + 1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (2-\varepsilon)/\beta]$, and suppose that $(\mu_n)_{n\in\mathbb{N}}$ satisfies (8.25). Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
q_n = \sum_{k=1}^p L_k x_{k,n} \\
\text{for } k = 1, \ldots, p \\
\left\lfloor
\begin{array}{l}
b_{k,n}^* = \gamma_n L_k^*(q_n - y) \\
w_{k,n} = \operatorname{proj}_{C_k}(x_{k,n} - b_{k,n}^*) \\
x_{k,n+1} = x_{k,n} + \mu_n(w_{k,n} - x_{k,n}).
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{8.45}
$$

Then $q_n \rightharpoonup \mathrm{proj}_{\sum_{k=1}^p L_k(C_k)}\, y$.

Proof. Set $\boldsymbol{\mathcal{H}} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_p$, $\boldsymbol{C} = C_1 \times \cdots \times C_p$, and

$$\boldsymbol{L}\colon \boldsymbol{\mathcal{H}} \to \mathcal{G}\colon (x_k)_{1\leqslant k\leqslant p} \mapsto \sum_{k=1}^p L_k x_k. \tag{8.46}$$

Then $\mathrm{proj}_{\boldsymbol{C}}\colon (x_k)_{1\leqslant k\leqslant p} \mapsto (\mathrm{proj}_{C_k} x_k)_{1\leqslant k\leqslant p}$ (see Examples 2.36 and 2.37), $\|\boldsymbol{L}\|^2 = \beta$, and $\boldsymbol{L}^*\colon \mathcal{G} \to \boldsymbol{\mathcal{H}}\colon y^* \mapsto (L_1^* y^*, \ldots, L_p^* y^*)$. Altogether, the result is an application of Example 8.12 to $\boldsymbol{C}$ and $\boldsymbol{L}$ in $\boldsymbol{\mathcal{H}}$. ☐

As an application of Example 8.14, we address the problem of computing the best approximation from the Minkowski sum of closed convex sets; see [34, 175, 276, 320, 351, 388, 390] for instances of decompositions with respect to such sums.

Example 8.15 Let $z \in \mathcal{H}$ and $0 < p \in \mathbb{N}$. For every $k \in \{1, \ldots, p\}$, let C_k be a nonempty closed convex subset of $\mathcal{H}$ and let $x_{k,0} \in \mathcal{H}$. Suppose that $\sum_{k=1}^p C_k$ is closed, let $\varepsilon \in \left]0, 1/(p+1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (2-\varepsilon)/p]$, and suppose that $(\forall n \in \mathbb{N})\ \varepsilon \leqslant \mu_n \leqslant (1-\varepsilon)(2 - p\gamma_n/2)$. Iterate

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
q_n = \sum_{k=1}^p x_{k,n} \\
b_n^* = \gamma_n (q_n - z) \\
\text{for } k = 1, \ldots, p \\
\left\lfloor \begin{array}{l}
w_{k,n} = \mathrm{proj}_{C_k}(x_{k,n} - b_n^*) \\
x_{k,n+1} = x_{k,n} + \mu_n (w_{k,n} - x_{k,n}).
\end{array} \right.
\end{array} \right.
\end{array} \tag{8.47}$$

Then $q_n \rightharpoonup \mathrm{proj}_{\sum_{k=1}^p C_k}\, z$.

Proof. Apply Example 8.14 with $\mathcal{G} = \mathcal{H}$, $y = z$, and $(\forall k \in \{1, \ldots, p\})\ \mathcal{H}_k = \mathcal{H}$ and $L_k = \mathrm{Id}$. ☐

8.4.2 Partial Yosida approximation to inconsistent common zero problems

We extend a framework proposed in [127, Section 6.3], where no linear transformations were present. We start with the following composite common zero problem (see [100] for a special case).

Problem 8.16 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone and let $0 < p \in \mathbb{N}$. For every $k \in \{1, \ldots, p\}$, let $\mathcal{G}_k$ be a real Hilbert space, let $B_k\colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ be maximally monotone, and suppose that $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$. The objective is to

$$\text{find } x \in \mathrm{zer}\, A \text{ such that } \big(\forall k \in \{1, \ldots, p\}\big)\ L_k x \in \mathrm{zer}\, B_k. \tag{8.48}$$

Example 8.17 Suppose that, in Problem 8.16, $A = N_C$, where C is a nonempty closed convex subset of $\mathcal{H}$, and, for every $k \in \{1, \ldots, p\}$, $B_k = N_{D_k}$, where D_k is a nonempty closed convex subset of $\mathcal{G}_k$. Then (8.48) is the *split feasibility problem* [328]

$$\text{find } x \in C \text{ such that } \big(\forall k \in \{1, \ldots, p\}\big) \quad L_k x \in D_k. \tag{8.49}$$

Example 8.18 Suppose that, in Problem 8.16, $A = \partial f$, where $f \in \Gamma_0(\mathcal{H})$, and, for every $k \in \{1, \ldots, p\}$, $\mathcal{G}_k = \mathcal{H}$, $L_k = \mathrm{Id}$, and $B_k = \partial f_k$, where $f_k \in \Gamma_0(\mathcal{H})$. Then (8.48) becomes

$$\text{find } x \in \big(\operatorname{Argmin} f\big) \cap \bigcap_{k=1}^{p} \operatorname{Argmin} f_k. \tag{8.50}$$

Example 8.19 Suppose that, in Problem 8.16, $A = N_C$, where C is a nonempty closed convex subset of $\mathcal{H}$, and, for every $k \in \{1, \ldots, p\}$, $B_k = (\mathrm{Id} - F_k + r_k)^{-1} - \mathrm{Id}$, where $F_k \colon \mathcal{G}_k \to \mathcal{G}_k$ is firmly nonexpansive and $r_k \in \mathcal{G}_k$. Then (8.48) becomes

$$\text{find } x \in C \text{ such that } \big(\forall k \in \{1, \ldots, p\}\big) \quad F_k(L_k x) = r_k. \tag{8.51}$$

Note that the operators $(\mathrm{Id} - F_k + r_k)_{1 \leqslant k \leqslant p}$ are firmly nonexpansive as well, which makes the operators $(B_k)_{1 \leqslant k \leqslant p}$ maximally monotone by Lemma 2.34(iii). This formulation was investigated in [153] in the context of recovering a signal in C from p nonlinear observations modeled as outputs of Wiener systems (see also Example 5.12).

Our focus here is on situations in which (8.48) is not guaranteed to have solutions (see [105, 133, 138, 212] for concrete illustrations). In such environments, it is natural to approximate it by a more general problem, which exhibits better regularity properties and admits solutions. We propose the following relaxation of Problem 8.16, in which $\operatorname{dom} A$ serves as a hard constraint.

Problem 8.20 Consider the setting of Problem 8.16 and let $(\rho_k)_{1 \leqslant k \leqslant p}$ and $(\omega_k)_{1 \leqslant k \leqslant p}$ be in $]0, +\infty[$. The objective is to solve the *partial Yosida approximation*

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + \sum_{k=1}^{p} \omega_k L_k^* \big({}^{\rho_k}B_k(L_k x)\big) \tag{8.52}$$

to Problem 8.16.

The fact that Problem 8.20 is an appropriate relaxation of Problem 8.16 is supported by the following argument.

Proposition 8.21 *Suppose that the set of solutions to Problem 8.16 is not empty. Then it coincides with the set of solutions to Problem 8.20.*

Proof. Let $\overline{x}$ be a solution to Problem 8.16. Then (2.22) yields

$$0 = -\sum_{k=1}^{p} \omega_k L_k^*\big({}^{\rho_k}B_k(L_k\overline{x})\big) \in A\overline{x}, \tag{8.53}$$

which shows that $\overline{x}$ solves Problem 8.20. Now let x be a solution to Problem 8.20. Then

$$-\sum_{k=1}^{p} \omega_k L_k^*\big({}^{\rho_k}B_k(L_k x)\big) \in Ax. \tag{8.54}$$

It follows from (8.53), (8.54), the monotonicity of A, and the cocoercivity of the operators $({}^{\rho_k}B_k)_{1\leqslant k\leqslant p}$ (see Example 2.7) that

$$\begin{aligned}
0 &\geqslant \left\langle x - \overline{x} \,\middle|\, \sum_{k=1}^{p} \omega_k L_k^*\big({}^{\rho_k}B_k(L_k x)\big) - \sum_{k=1}^{p} \omega_k L_k^*\big({}^{\rho_k}B_k(L_k\overline{x})\big) \right\rangle \\
&= \sum_{k=1}^{p} \omega_k \left\langle L_k x - L_k\overline{x} \,\middle|\, {}^{\rho_k}B_k(L_k x) - {}^{\rho_k}B_k(L_k\overline{x}) \right\rangle \\
&\geqslant \sum_{k=1}^{p} \omega_k \rho_k \big\| {}^{\rho_k}B_k(L_k x) - {}^{\rho_k}B_k(L_k\overline{x}) \big\|^2 \\
&= \sum_{k=1}^{p} \omega_k \rho_k \big\| {}^{\rho_k}B_k(L_k x) \big\|^2.
\end{aligned} \tag{8.55}$$

Hence, we deduce from (2.22) that $(\forall k \in \{1,\ldots,p\})$ $L_k x \in \operatorname{zer} {}^{\rho_k}B_k = \operatorname{zer} B_k$. In view of (8.54), we conclude that x solves Problem 8.16. □

Remark 8.22 It should be emphasized that Problem 8.20 is a relaxation of Problem 8.16, and not of the inclusion

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + \sum_{k=1}^{p} \omega_k L_k^*\big(B_k(L_k x)\big). \tag{8.56}$$

In particular, $\operatorname{zer}(A + {}^{\rho}B) \neq \operatorname{zer}(A + B)$ when $\operatorname{zer}(A + B) \neq \varnothing$. However, the problem of finding a zero of $A + {}^{\rho}B$ can be regarded as a regularization of that of finding a zero of $A + B$ in the sense that solutions to the former approaches a particular solution of the latter as $\rho \to 0$ [270, 278, 295].

Example 8.23 Consider the setting of Example 8.17 and let $(\forall k \in \{1,\ldots,p\})$ $\rho_k = 1$. Then (8.52) relaxes the possibly inconsistent problem (8.49) to the problem

$$\underset{x\in C}{\text{minimize}} \;\; \sum_{k=1}^{p} \omega_k d_{D_k}^2(L_k x). \tag{8.57}$$

(i) Assume that, for every $k \in \{1,\ldots,p\}$, $\mathcal{G}_k = \mathcal{H}$ and $L_k = \mathrm{Id}$. Then (8.57) is the relaxed formulation of [133].

(ii) Assume that $\mathcal{H} = \mathbb{R}^N$, $C = \mathbb{R}^N$, and, for every $k \in \{1,\ldots,p\}$, $\mathcal{G}_k = \mathbb{R}$, $L_k\colon x \mapsto u_k^\top x$ with $u_k \in \mathbb{R}^N$, and $D_k = \{\eta_k\}$ with $\eta_k \in \mathbb{R}$. Let $U \in \mathbb{R}^{p\times N}$ be the matrix with rows $u_1^\top,\ldots,u_p^\top$ and set $y = (\eta_k)_{1\leqslant k\leqslant p}$. Then (8.49) amounts to solving the linear system $Ux = y$ and (8.57) to minimizing $x \mapsto \|Ux - y\|^2$. This least-squares relaxation was proposed by Legendre [252] and rediscovered by Gauss [202].

Example 8.24 Consider the setting of Example 8.18 and recall that $(\forall k \in \{1,\ldots,p\})$ ${}^{\rho_k}(\partial f_k) = \{\nabla({}^{\rho_k} f_k)\}$ [37, Example 23.3]. Thus, (8.52) relaxes the possibly inconsistent problem (8.50) to the problem

$$\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + \sum_{k=1}^{p} \omega_k \big({}^{\rho_k} f_k\big)(x). \tag{8.58}$$

This formulation arises in particular in federated learning [306].

Example 8.25 Consider the setting of Example 8.19 and let $(\forall k \in \{1,\ldots,p\})$ $\rho_k = 1$. Then it follows from Example 2.14 and (2.21) that (8.52) relaxes the possibly inconsistent problem (8.51) to the variational inequality problem (see Problem 3.3)

$$\text{find } x \in C \text{ such that } (\forall y \in C) \;\; \sum_{k=1}^{p} \omega_k \langle L_k(y - x) \mid F_k(L_k x) - r_k\rangle \geqslant 0, \tag{8.59}$$

which is precisely the relaxation of (8.51) studied in [153].

Let us now solve Problem 8.20 with the forward-backward algorithm.

Proposition 8.26 *Consider the setting of Problem 8.20, suppose that its set Z of solutions is not empty, and set*

$$\alpha = \frac{1}{\displaystyle\sum_{k=1}^{p} \frac{\omega_k \|L_k\|^2}{\rho_k}}. \tag{8.60}$$

Let $x_0 \in \mathcal{H}$, let $\varepsilon \in \left]0, \alpha/(\alpha+1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (2-\varepsilon)\alpha]$, and suppose that $(\mu_n)_{n\in\mathbb{N}}$ satisfies (8.2). Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for } k = 1, \ldots, p \\
\left\lfloor
\begin{array}{l}
y_{k,n} = L_k x_n \\
p_{k,n} = \rho_k^{-1}\big(y_{k,n} - J_{\rho_k B_k} y_{k,n}\big)
\end{array}
\right. \\
b_n^* = \gamma_n \displaystyle\sum_{k=1}^{p} \omega_k L_k^* p_{k,n} \\
w_n = J_{\gamma_n A}(x_n - b_n^*) \\
x_{n+1} = x_n + \mu_n(w_n - x_n).
\end{array}
\right.
\end{array}
\tag{8.61}
$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.

Proof. Define

$$
B = \sum_{k=1}^{p} \omega_k L_k^* \circ ({}^{\rho_k}B_k) \circ L_k. \tag{8.62}
$$

Then it follows from [37, Proposition 4.12] and Example 2.7 that B is α-cocoercive. Since (8.61) is a specialization of (8.5), Theorem 8.1(i) furnishes the desired conclusion. ☐

8.4.3 Backward-backward splitting

We focus on the following special case of Problem 8.20.

Problem 8.27 Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let $\rho \in \left]0, +\infty\right[$. The objective is to

$$
\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + {}^{\rho}Bx. \tag{8.63}
$$

Proposition 8.28 *Consider the setting of Problem 8.27 under the assumption that $Z = \operatorname{zer}(A + {}^{\rho}B) \neq \varnothing$. Let $x_0 \in \mathcal{H}$, let $\varepsilon \in \left]0, \rho/(\rho+1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (2-\varepsilon)\rho]$, and suppose that $(\mu_n)_{n\in\mathbb{N}}$ satisfies (8.2) with $\alpha = \rho$. Iterate*

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
p_n = \rho^{-1}\big(x_n - J_{\rho B} x_n\big) \\
w_n = J_{\gamma_n A}(x_n - \gamma_n p_n) \\
x_{n+1} = x_n + \mu_n(w_n - x_n).
\end{array}
\right.
\end{array}
\tag{8.64}
$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.

Proof. Apply Proposition 8.26 with $p = 1$, $\mathcal{G}_1 = \mathcal{H}$, $L_1 = \mathrm{Id}$, $B_1 = B$, $\omega_1 = 1$, and $\rho_1 = \rho$. ☐

Example 8.29 In particular, if we execute (8.64) with, for every $n \in \mathbb{N}$, $\gamma_n = \rho$ and $\mu_n = 1$, then

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = J_{\rho A}\big(J_{\rho B} x_n\big). \tag{8.65}$$

This recursion is known as the *backward-backward algorithm*, as it alternates two backward Euler steps. As derived above, it is a special case of (8.61) and therefore of the forward-backward algorithm (8.5). Its asymptotic behavior has been studied in [38, 278] (see also [264, 305] for ergodic convergence).

Example 8.30 Let f and g be functions in $\Gamma_0(\mathcal{H})$. In Problem 8.27, suppose that $A = \partial f$ and $B = \partial g$. Then, as in Example 8.24, (8.65) becomes

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; f(x) + {}^{\rho}g(x) \tag{8.66}$$

and (8.65) reduces to the *alternating proximal point algorithm*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \mathrm{prox}_{\rho f}\big(\mathrm{prox}_{\rho g}\, x_n\big). \tag{8.67}$$

This method was first investigated in [1], with further developments in [38].

Example 8.31 Let C and D be nonempty closed convex subsets of $\mathcal{H}$. In Example 8.30, suppose that $f = \iota_C$ and $g = \iota_D$. Then (8.67) is the problem of finding a point in C at minimal distance from D and (8.67) yields the *alternating projection method*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \mathrm{proj}_C\big(\mathrm{proj}_D\, x_n\big), \tag{8.68}$$

which was first investigated in [120]. Its weak convergence was established in [220, Theorem 2]

Example 8.32 Let $f \in \Gamma_0(\mathcal{H})$, $h \in \Gamma_0(\mathcal{H})$, $z \in \mathcal{H}$, and $\rho \in \left]0, +\infty\right[$. The problem is to

$$\underset{x \in \mathcal{H}, w \in \mathcal{H}}{\text{minimize}} \;\; f(x) + h(w) + \frac{1}{2\rho}\|x + w - z\|^2. \tag{8.69}$$

Following [152, Section 4.4], set $g\colon y \mapsto h(z - y)$. Then, with the change of variable $y = z - w$, the objective of (8.69) is to

$$\underset{x \in \mathcal{H}, y \in \mathcal{H}}{\text{minimize}} \;\; f(x) + g(y) + \frac{1}{2\rho}\|x - y\|^2, \tag{8.70}$$

which is precisely (8.66) in terms of the variable x. Now let $x_0 \in \mathcal{H}$, let $\varepsilon \in]0, \rho/(\rho+1)[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (2-\varepsilon)\rho]$, and let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 1]$. Applying algorithm (8.64) to $A = \partial f$ and $B = \partial g$, and noting that $J_{\rho B} = \mathrm{prox}_{\rho g} \colon x \mapsto z - \mathrm{prox}_{\rho h}(z - x)$ yields

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
p_n = \rho^{-1}\big(x_n - z + \mathrm{prox}_{\rho h}(z - x_n)\big) \\
w_n = \mathrm{prox}_{\gamma_n f}(x_n - \gamma_n p_n) \\
x_{n+1} = x_n + \mu_n(w_n - x_n).
\end{array}
\right.
\end{array}
\tag{8.71}
$$

It follows from Proposition 8.28 that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point x such that $(x, \mathrm{prox}_{\rho h}(z - x))$ solves (8.69).

Next, we revisit the problem of projecting onto the Minkowski sum of two convex sets (see Example 8.15).

Example 8.33 Let C and D be nonempty closed convex subsets of $\mathcal{H}$ such that $C + D$ is closed, and let $z \in \mathcal{H}$. Upon setting $f = \iota_C$, $h = \iota_D$, and $\rho = 1$ in Example 8.32, (8.69) specializes to the problem of finding the projection of z onto $C + D$. Now let $x_0 \in C$, let $\varepsilon \in]0, 1/2[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 2 - \varepsilon]$, and let $(\mu_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 1]$. Then (8.71) assumes the form

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
p_n = x_n - z + \mathrm{proj}_D(z - x_n) \\
w_n = \mathrm{proj}_C(x_n - \gamma_n p_n) \\
x_{n+1} = x_n + \mu_n(w_n - x_n)
\end{array}
\right.
\end{array}
\tag{8.72}
$$

and it follows from Proposition 8.28 that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point x such that $\mathrm{proj}_{C+D}\, z = x + \mathrm{proj}_D(z - x)$. This best approximation algorithm was first obtained in [351, Theorem 2.1] in the case when $(\forall n \in \mathbb{N})\ \gamma_n = \mu_n = 1$, i.e.,

$$
(\forall n \in \mathbb{N}) \quad x_{n+1} = \mathrm{proj}_C\big(z - \mathrm{proj}_D(z - x_n)\big). \tag{8.73}
$$

8.4.4 Dual implementation

We present a framework for solving strongly monotone composite inclusion problems by applying the forward-backward algorithm to the dual problem. The embedding underlying this approach is that of Example 3.22.

Problem 8.34 Let $\rho \in]0, +\infty[$, let $0 < p \in \mathbb{N}$, let $z \in \mathcal{H}$, and let $A \colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. For every $k \in \{1, \ldots, p\}$, let $B_k \colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ be maximally monotone, let $\nu_k \in]0, +\infty[$, let $D_k \colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ be maximally monotone and

ν_k-strongly monotone, and suppose that $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$. Further, suppose that

$$z \in \operatorname{ran}\left(A + \sum_{k=1}^{p} L_k^* \circ (B_k \,\square\, D_k) \circ L_k + \rho \mathrm{Id}\right). \tag{8.74}$$

The problem is to solve the primal inclusion

$$\text{find} \;\; x \in \mathcal{H} \;\; \text{such that} \;\; z \in Ax + \sum_{k=1}^{p} L_k^*\big((B_k \,\square\, D_k)(L_k x)\big) + \rho x, \tag{8.75}$$

together with the dual inclusion

$$\begin{aligned} &\text{find} \;\; y_1^* \in \mathcal{G}_1, \ldots, y_p^* \in \mathcal{G}_p \;\; \text{such that} \;\; \big(\forall k \in \{1, \ldots, p\}\big) \\ &\qquad 0 \in -L_k\left(J_{A/\rho}\left(\frac{1}{\rho}\left(z - \sum_{j=1}^{p} L_j^* y_j^*\right)\right)\right) + B_k^{-1} y_k^* + D_k^{-1} y_k^*. \end{aligned} \tag{8.76}$$

We refer to [151, Proposition 5.2(iv)] for sufficient conditions that guarantee (8.74). The mechanism to solve (8.75) dually hinges on the following properties.

Proposition 8.35 ([151, Proposition 5.2(ii)–(iii)]) *Consider the setting of Problem 8.34 and set*

$$M = A + \sum_{k=1}^{p} L_k^* \circ (B_k \,\square\, D_k) \circ L_k \quad \textit{and} \quad \overline{x} = J_{M/\rho}(z/\rho). \tag{8.77}$$

Then the following hold:

(i) $\overline{x}$ *is the unique solution to the primal problem* (8.75).

(ii) *The dual problem* (8.76) *admits solutions and, if* $(\overline{y}_k^*)_{1 \leqslant k \leqslant p}$ *solves* (8.76), *then*

$$\overline{x} = J_{A/\rho}\left(\rho^{-1}\left(z - \sum_{k=1}^{p} L_k^* \overline{y}_k^*\right)\right). \tag{8.78}$$

We now apply the forward-backward algorithm of Theorem 8.1 to the dual inclusion (8.76) to construct a sequence $(x_n)_{n \in \mathbb{N}}$ which converges strongly to the solution to primal inclusion (8.75). The following result is an adaptation of [151, Corollary 5.4].

Proposition 8.36 *Consider the setting of Problem 8.34 and set*

$$\nu=\min_{1\leqslant k\leqslant p}\nu_k\quad\textit{and}\quad\alpha=\frac{1}{\dfrac{1}{\nu}+\dfrac{1}{\rho}\displaystyle\sum_{1\leqslant k\leqslant p}\|L_k\|^2}. \tag{8.79}$$

Let $\varepsilon\in\left]0,\alpha/(\alpha+1)\right[$, *let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $[\varepsilon,(2-\varepsilon)\alpha]$, *suppose that* $(\mu_n)_{n\in\mathbb{N}}$ *satisfies* (8.2), *and, for every* $k\in\{1,\ldots,p\}$, *let* $y^*_{k,0}\in\mathcal{G}_k$. *Iterate*

$$\begin{array}{l}\textit{for}\ n=0,1,\ldots\\ \left\lfloor\begin{array}{l}q_n=z-\sum_{k=1}^p L_k^*y^*_{k,n}\\ x_n=J_{A/\rho}(q_n/\rho)\\ \text{for}\ \ k=1,\ldots,p\\ \left\lfloor\begin{array}{l}w_{k,n}=y^*_{k,n}+\gamma_n\big(L_kx_n-D_k^{-1}y^*_{k,n}\big)\\ y^*_{k,n+1}=y^*_{k,n}+\mu_n\big(J_{\gamma_nB_k^{-1}}w_{k,n}-y^*_{k,n}\big).\end{array}\right.\end{array}\right.\end{array} \tag{8.80}$$

Then the following hold for the solution $\overline{x}$ *to* (8.75) *and for some solution* $\overline{\boldsymbol{y}}^*=(\overline{y}^*_1,\ldots,\overline{y}^*_p)$ *to* (8.76)*:*

(i) $(\forall k\in\{1,\ldots,p\})\ y^*_{k,n}\rightharpoonup\overline{y}^*_k$.

(ii) $x_n\to\overline{x}$.

Proof. We deduce from [37, Proposition 22.11(ii)] that, for every $k\in\{1,\ldots,p\}$, D_k^{-1} is ν_k-cocoercive with $\operatorname{dom}D_k^{-1}=\mathcal{G}_k$. Let us set $\boldsymbol{\mathcal{G}}=\mathcal{G}_1\oplus\cdots\oplus\mathcal{G}_p$ and

$$\begin{cases}T\colon\mathcal{H}\to\mathcal{H}\colon x\mapsto J_{\rho^{-1}A}\big(\rho^{-1}(z-x)\big)\\ \boldsymbol{A}\colon\boldsymbol{\mathcal{G}}\to 2^{\boldsymbol{\mathcal{G}}}\colon\boldsymbol{y}^*\mapsto\displaystyle\bigtimes_{1\leqslant k\leqslant p}B_k^{-1}y^*_k\\ \boldsymbol{D}\colon\boldsymbol{\mathcal{G}}\to\boldsymbol{\mathcal{G}}\colon\boldsymbol{y}^*\mapsto\big(D_k^{-1}y^*_k\big)_{1\leqslant k\leqslant p}\\ \boldsymbol{L}\colon\mathcal{H}\to\boldsymbol{\mathcal{G}}\colon x\mapsto\big(L_kx\big)_{1\leqslant k\leqslant p}\\ \boldsymbol{B}=\boldsymbol{D}-\boldsymbol{L}\circ T\circ\boldsymbol{L}^*.\end{cases} \tag{8.81}$$

It follows from Lemmas 2.23 and 2.24 that $\boldsymbol{A}$ is maximally monotone, from (8.79) that $\boldsymbol{D}$ is ν-cocoercive, from Lemma 2.34(iii) that $-T$ is ρ-cocoercive, and hence from [37, Proposition 4.12] that

$$\boldsymbol{B}=\boldsymbol{D}+\boldsymbol{L}\circ(-T)\circ\boldsymbol{L}^*\ \text{is}\ 1/(1/\nu+\|\boldsymbol{L}\|^2/\rho)\text{-cocoercive}. \tag{8.82}$$

Since $\|\boldsymbol{L}\|^2\leqslant\sum_{k=1}^p\|L_k\|^2$, (8.79) implies that $\boldsymbol{B}$ is α-cocoercive. Next, let us define $(\forall n\in\mathbb{N})\ \boldsymbol{y}^*_n=(y^*_{k,n})_{1\leqslant k\leqslant p}$ and $\boldsymbol{w}_n=(w_{k,n})_{1\leqslant k\leqslant p}$. Then, upon combining

(8.81) and Example 2.37, (8.80) can be rewritten as

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{w}_n = \boldsymbol{y}_n^* - \gamma_n \boldsymbol{B}\boldsymbol{y}_n^* \\
\boldsymbol{y}_{n+1}^* = \boldsymbol{y}_n^* + \mu_n\big(J_{\gamma_n \boldsymbol{A}}\, \boldsymbol{w}_n - \boldsymbol{y}_n^*\big),
\end{array}
\right.
\end{array}
\tag{8.83}
$$

and the dual problem (8.76) as

$$
\text{find } \ \boldsymbol{y}^* \in \boldsymbol{\mathcal{G}} \ \text{ such that } \ \boldsymbol{0} \in \boldsymbol{A}\boldsymbol{y}^* + \boldsymbol{B}\boldsymbol{y}^*. \tag{8.84}
$$

(i): In view of the above, the claim follows from Theorem 8.1(i).
(ii): We derive from Proposition 8.35, (8.80), and (8.81) that

$$
\overline{x} = T(\boldsymbol{L}^*\overline{\boldsymbol{y}}^*) \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_n = T(\boldsymbol{L}^*\boldsymbol{y}_n^*). \tag{8.85}
$$

In turn, we deduce from the ρ-cocoercivity of $-T$, (i), the monotonicity of $\boldsymbol{D}$, and the Cauchy–Schwarz inequality that

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad \rho\|x_n - \overline{x}\|^2 &= \rho\|T(\boldsymbol{L}^*\boldsymbol{y}_n^*) - T(\boldsymbol{L}^*\overline{\boldsymbol{y}}^*)\|^2 \\
&\leqslant \langle \boldsymbol{L}^*(\boldsymbol{y}_n^* - \overline{\boldsymbol{y}}^*) \mid T(\boldsymbol{L}^*\overline{\boldsymbol{y}}^*) - T(\boldsymbol{L}^*\boldsymbol{y}_n^*)\rangle \\
&= \langle \boldsymbol{y}_n^* - \overline{\boldsymbol{y}}^* \mid (\boldsymbol{L} \circ T \circ \boldsymbol{L}^*)\overline{\boldsymbol{y}}^* - (\boldsymbol{L} \circ T \circ \boldsymbol{L}^*)\boldsymbol{y}_n^*\rangle \\
&\leqslant \langle \boldsymbol{y}_n^* - \overline{\boldsymbol{y}}^* \mid \boldsymbol{D}\boldsymbol{y}_n^* - \boldsymbol{D}\overline{\boldsymbol{y}}^*\rangle \\
&\quad\; - \langle \boldsymbol{y}_n^* - \overline{\boldsymbol{y}}^* \mid (\boldsymbol{L} \circ T \circ \boldsymbol{L}^*)\boldsymbol{y}_n^* - (\boldsymbol{L} \circ T \circ \boldsymbol{L}^*)\overline{\boldsymbol{y}}^*\rangle \\
&= \langle \boldsymbol{y}_n^* - \overline{\boldsymbol{y}}^* \mid \boldsymbol{B}\boldsymbol{y}_n^* - \boldsymbol{B}\overline{\boldsymbol{y}}^*\rangle \\
&\leqslant \delta\|\boldsymbol{B}\boldsymbol{y}_n^* - \boldsymbol{B}\overline{\boldsymbol{y}}^*\|
\end{aligned}
\tag{8.86}
$$

where, by (i),

$$
\delta = \sup_{n\in\mathbb{N}} \|\boldsymbol{y}_n^* - \overline{\boldsymbol{y}}^*\| < +\infty. \tag{8.87}
$$

Therefore, using (8.83) and Theorem 8.1(ii)–(iii), we conclude that $\|x_n - \overline{x}\| \to 0$. □

Here is an application to strongly convex minimization problems that arise in particular in mechanics [185, 278] and in signal processing [134, 135, 318].

Example 8.37 Let $0 < p \in \mathbb{N}$, let $z \in \mathcal{H}$, let $f \in \Gamma_0(\mathcal{H})$, and let $^1(f^*)$ be the Moreau envelope of f^* (see (2.11)). For every $k \in \{1, \ldots, p\}$, let $g_k \in \Gamma_0(\mathcal{G}_k)$, let $\nu_k \in \left]0, +\infty\right[$, let $h_k \in \Gamma_0(\mathcal{G}_k)$ be ν_k-strongly convex, and suppose that $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$. Define α as in (8.79) and suppose that

$$
z \in \operatorname{ran}\Big(\partial f + \sum_{k=1}^{p} L_k^* \circ (\partial g_k \,\Box\, \partial h_k) \circ L_k + \mathrm{Id}\Big). \tag{8.88}
$$

Then the primal problem

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x)+\sum_{k=1}^{p}(g_k\,\square\,h_k)(L_kx)+\frac{1}{2}\|x-z\|^2 \tag{8.89}$$

admits a unique solution $\overline{x}$, namely

$$\overline{x}=\operatorname{prox}_{f+\sum_{k=1}^{p}(g_k\,\square\,h_k)\circ L_k}z, \tag{8.90}$$

and the dual problem is

$$\underset{y_1^*\in\mathcal{G}_1,\ldots,y_p^*\in\mathcal{G}_p}{\text{minimize}}\quad {}^{1}\!\left(f^*\right)\!\left(z-\sum_{k=1}^{p}L_k^*y_k^*\right)+\sum_{k=1}^{p}\big(g_k^*(y_k^*)+h_k^*(y_k^*)\big). \tag{8.91}$$

Now let $\varepsilon\in\,]0,\alpha/(\alpha+1)[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon,(2-\varepsilon)\alpha]$, suppose that $(\mu_n)_{n\in\mathbb{N}}$ satisfies (8.2), and, for every $k\in\{1,\ldots,p\}$, let $y_{k,0}^*\in\mathcal{G}_k$. Iterate

$$\begin{array}{l}
\text{for } n=0,1,\ldots\\
\left\lfloor\begin{array}{l}
q_n=z-\sum_{k=1}^{p}L_k^*y_{k,n}^*\\
x_n=\operatorname{prox}_f q_n\\
\text{for } k=1,\ldots,p\\
\left\lfloor\begin{array}{l}
w_{k,n}=y_{k,n}^*+\gamma_n\big(L_kx_n-\nabla h_k^*(y_{k,n}^*)\big)\\
y_{k,n+1}^*=y_{k,n}^*+\mu_n\big(\operatorname{prox}_{\gamma_n g_k^*}w_{k,n}-y_{k,n}^*\big).
\end{array}\right.
\end{array}\right.
\end{array} \tag{8.92}$$

Then the following hold:

(i) There exists a solution $(\overline{y}_1^*,\ldots,\overline{y}_p^*)$ to (8.91) such that $(\forall k\in\{1,\ldots,p\})$ $y_{k,n}^*\rightharpoonup\overline{y}_k^*$.

(ii) $x_n\to\overline{x}$.

Proof. Apply Proposition 8.36 with $\rho=1$, $A=\partial f$, and $(\forall k\in\{1,\ldots,p\})$ $B_k=\partial g_k$ and $D_k=\partial h_k$ (see [151, Eample 5.6] for details). $\square$

Remark 8.38 In Example 8.37, suppose that $\mathcal{H}=H_0^1(\Omega)$, where Ω is a bounded open domain in $\mathbb{R}^2$, $p=1$, $\mathcal{G}_1=L^2(\Omega)\oplus L^2(\Omega)$, $L_1=\nabla$, $g_1=\mu\|\cdot\|_{2,1}$ with $\mu\in\,]0,+\infty[$, and $h_1=\iota_{\{0\}}$. Then (8.89) reduces to

$$\underset{x\in H_0^1(\Omega)}{\text{minimize}}\;\; f(x)+\mu\int_\Omega|\nabla x(\omega)|_2d\omega+\frac{1}{2}\|x-z\|^2. \tag{8.93}$$

In mechanics, (8.93) has been studied for certain potentials f [185]. For instance, $f=0$ yields Mossolov's problem and its dual analysis is carried out in [185, Section IV.3.1]. In image processing, Mossolov's problem corresponds to the total variation denoising problem. In 1980, Mercier [278] proposed a dual projection algorithm to solve Mossolov's problem. In image processing, this approach was rediscovered in a discrete setting in [110, 111].

8.4.5 Barycentric Dykstra-like algorithm

Using Proposition 8.36 and, thereby, the forward-backward algorithm, we obtain a method for computing the resolvent of a sum of maximally monotone operators. This result, which generalizes the barycentric Dykstra algorithm of [199] for projecting onto an intersection of closed convex sets, was originally derived in [128, Theorem 3.3] with different techniques.

Proposition 8.39 *Let $0 < p \in \mathbb{N}$, let $z \in \mathcal{H}$, and, for every $k \in \{1,\dots,p\}$, let $A_k\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Suppose that*

$$z \in \operatorname{ran}\left(\sum_{k=1}^{p} A_k + \mathrm{Id}\right) \tag{8.94}$$

and consider the inclusion problem

$$\text{find } x \in \mathcal{H} \text{ such that } z \in \sum_{k=1}^{p} A_k x + x. \tag{8.95}$$

Set $x_0 = z$ and $(\forall k \in \{1,\dots,p\})$ $z_{k,0} = z$. Iterate

$$\begin{array}{l} \text{for } n = 0,1,\dots \\ \left\lfloor \begin{array}{l} \text{for } k = 1,\dots,p \\ \left\lfloor\ r_{k,n} = J_{pA_k} z_{k,n} \right. \\ x_{n+1} = (1/p)\sum_{k=1}^{p} r_{k,n} \\ \text{for } k = 1,\dots,p \\ \left\lfloor\ z_{k,n+1} = z_{k,n} - r_{k,n} + x_{n+1}. \right. \end{array} \right. \end{array} \tag{8.96}$$

Then $x_n \to J_{\sum_{k=1}^{p} A_k} z$.

Proof. First, we observe that (8.94)–(8.95) is the special case of (8.74)–(8.75) in which $A = 0$ and, for every $k \in \{1,\dots,p\}$, $\mathcal{G}_k = \mathcal{H}$, $B_k = A_k$, $L_k = \mathrm{Id}$, and $D_k = \{0\}^{-1}$. Moreover, the cocoercivity constant in (8.79) is $\alpha = 1/p$. With this scenario, implementing (8.80) with, for every $n \in \mathbb{N}$, $\mu_n = 1$ and $\gamma_n = 1/p$, and, for every $k \in \{1,\dots,p\}$, $y^*_{k,0} = 0$ leads to the recursion

$$\begin{array}{l} \text{for } n = 0,1,\dots \\ \left\lfloor \begin{array}{l} x_n = z - \sum_{k=1}^{p} y^*_{k,n} \\ \text{for } k = 1,\dots,p \\ \left\lfloor\ y^*_{k,n+1} = J_{A_k^{-1}/p}\big(y^*_{k,n} + x_n/p\big) \right. \end{array} \right. \end{array} \tag{8.97}$$

and Proposition 8.36(ii) guarantees that $x_n \to J_{\sum_{k=1}^p A_k} z$. Alternatively, with the initialization $x_0 = z$, we rewrite (8.97) as

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for } k = 1, \ldots, p \\
\left\lfloor\; y^*_{k,n+1} = J_{A_k^{-1}/p}\big(y^*_{k,n} + x_n/p\big) \right. \\
x_{n+1} = z - \sum_{k=1}^p y^*_{k,n+1}.
\end{array}
\right.
\end{array}
\tag{8.98}
$$

Let us introduce the variables $(\forall n \in \mathbb{N})(\forall k \in \{1, \ldots, p\})\ z_{k,n} = p y^*_{k,n} + x_n$, where $z_{k,0} = x_0 = z$. Then (8.98) corresponds to the iterations

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
x_{n+1} = z - \sum_{k=1}^p J_{A_k^{-1}/p}\big(z_{k,n}/p\big) \\
\text{for } k = 1, \ldots, p \\
\left\lfloor\; z_{k,n+1} = p\, J_{A_k^{-1}/p}\big(z_{k,n}/p\big) + x_{n+1}. \right.
\end{array}
\right.
\end{array}
\tag{8.99}
$$

By construction,

$$
(\forall n \in \mathbb{N}) \quad \sum_{k=1}^p z_{k,n} = pz. \tag{8.100}
$$

Hence, appealing to (2.21), (8.99) becomes

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
x_{n+1} = (1/p) \sum_{k=1}^p J_{pA_k} z_{k,n} \\
\text{for } k = 1, \ldots, p \\
\left\lfloor\; z_{k,n+1} = z_{k,n} - J_{pA_k} z_{k,n} + x_{n+1}, \right.
\end{array}
\right.
\end{array}
\tag{8.101}
$$

which is precisely (8.96). ☐

Example 8.40 Consider the instantiation of Proposition 8.39 in which, for every $k \in \{1, \ldots, p\}$, $A_k = \partial f_k$, with $f_k \in \Gamma_0(\mathcal{H})$, and execute (8.96), which becomes

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for } k = 1, \ldots, p \\
\left\lfloor\; r_{k,n} = \operatorname{prox}_{pf_k} z_{k,n} \right. \\
x_{n+1} = (1/p) \sum_{k=1}^p r_{k,n} \\
\text{for } k = 1, \ldots, p \\
\left\lfloor\; z_{k,n+1} = z_{k,n} - r_{k,n} + x_{n+1}. \right.
\end{array}
\right.
\end{array}
\tag{8.102}
$$

Then $x_n \to \operatorname{prox}_{\sum_{k=1}^p f_k} z$.

Our last example addresses the barycentric Dykstra algorithm *per se*. The original Dykstra algorithm was devised in [174] to project onto the intersection of closed convex cones (see also [223] for general closed convex sets whose intersection has a nonempty interior) in Euclidean spaces using periodic applications of the projectors onto the individual sets. Convergence of this periodic scheme in the general case of arbitrary closed and convex sets in Hilbert spaces was established in [64] (see [36] for an extension to monotone operators). The barycentric version described below, in which all the projectors are used at each iteration, was devised in [199, Section 6]. Its connection with the forward-backward algorithm is discussed in [134, Remark 3.8] and [135, Remark 2.3], and its asymptotic behavior in the inconsistent case in [32, Theorem 6.1].

Example 8.41 In Example 8.40, suppose that, for every $k \in \{1, \ldots, p\}$, $f_k = \iota_{C_k}$, where C_k is a nonempty closed convex subset of $\mathcal{H}$. Then algorithm (8.102) becomes

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for } k = 1, \ldots, p \\
\left\lfloor\; r_{k,n} = \operatorname{proj}_{C_k} z_{k,n} \right. \\
x_{n+1} = (1/p) \sum_{k=1}^{p} r_{k,n} \\
\text{for } k = 1, \ldots, p \\
\left\lfloor\; z_{k,n+1} = z_{k,n} - r_{k,n} + x_{n+1} \right.
\end{array}
\right.
\end{array}
\tag{8.103}
$$

and $x_n \to \operatorname{proj}_{\bigcap_{k=1}^{p} C_k} z$.

8.4.6 Renorming

We preface our discussion with a renormed version of Theorem 8.1.

Proposition 8.42 *Let $\alpha \in \left]0, +\infty\right[$, let $\beta \in \left]0, +\infty\right[$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $B\colon \mathcal{H} \to \mathcal{H}$ be α-cocoercive, let $U \in \mathcal{B}(\mathcal{H})$ be self-adjoint and β-strongly monotone, and let $\mathcal{X}$ be the real Hilbert space obtained by endowing $\mathcal{H}$ with the scalar product $(x, y) \mapsto \langle Ux \mid y\rangle$. Let $\varepsilon \in \left]0, \alpha\beta/(\alpha\beta + 1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (2 - \varepsilon)\alpha\beta]$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 1]$. Suppose that the set Z of solutions to the problem*

$$\textit{find } x \in \mathcal{H} \textit{ such that } 0 \in Ax + Bx \tag{8.104}$$

is not empty and let Z^* be the set of solutions to the dual problem

$$\textit{find } x^* \in \mathcal{H} \textit{ such that } 0 \in -A^{-1}(-x^*) + B^{-1}x^*. \tag{8.105}$$

Let $x_0 \in \mathcal{H}$ and iterate

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
u_n^* = \gamma_n^{-1} U x_n - B x_n \\
w_n = \big(\gamma_n^{-1} U + A\big)^{-1} u_n^* \\
x_{n+1} = x_n + \lambda_n (w_n - x_n).
\end{array}
\right.
\end{array}
\tag{8.106}
$$

Then the following hold:

(i) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* Z.

(ii) Z^* *contains a single point* $\overline{x}^*$ *and* $(\forall z \in Z)\ Bz = \overline{x}^*$.

(iii) $(Bx_n)_{n\in\mathbb{N}}$ *converges strongly to* $\overline{x}^*$.

Proof. We derive from Lemma 2.25 and Example 2.39 that

$$
(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \Big(J_{\gamma_n U^{-1} \circ A}\big(x_n - \gamma_n U^{-1}(Bx_n)\big) - x_n \Big), \tag{8.107}
$$

where $U^{-1} \circ A \colon \mathcal{X} \to 2^{\mathcal{X}}$ is maximally monotone, $U^{-1} \circ B \colon \mathcal{X} \to \mathcal{X}$ is $\alpha\beta$-cocoercive, and $\operatorname{zer}(A + B) = \operatorname{zer}(U^{-1} \circ (A + B))$. Hence the assertions follow from Theorem 8.1 applied to $U^{-1} \circ A$ and $U^{-1} \circ B$ in $\mathcal{X}$. ☐

Remark 8.43 In terms of the warped resolvents of Section 2.4.3, (8.106) can be condensed into

$$
(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n \Big(J^{U_n}_{\gamma_n (A+B)} x_n - x_n \Big), \quad \text{where} \quad U_n = U - \gamma_n B. \tag{8.108}
$$

We present an approach proposed in [387], which revisited the primal-dual setting of [145] discussed in Proposition 7.10 by replacing the monotone Lipschitz property of the operators C and $(D_k^{-1})_{1\leqslant k\leqslant p}$ with the stronger cocoercivity property.

Proposition 8.44 ([387, Theorem 3.1(i)]) *Let* $0 < p \in \mathbb{N}$, *let* $\alpha \in \left]0, +\infty\right[$, *let* $A \colon \mathcal{H} \to 2^{\mathcal{H}}$ *be maximally monotone, and let* $C \colon \mathcal{H} \to \mathcal{H}$ *be* α*-cocoercive. For every* $k \in \{1, \ldots, p\}$, *let* $\beta_k \in \left]0, +\infty\right[$, *let* $\mathcal{G}_k$ *be a real Hilbert space, let* $B_k \colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ *be maximally monotone, let* $D_k \colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ *be maximally monotone and* β_k*-strongly monotone, and suppose that* $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$. *Additionally, suppose that the set* Z *of solutions to the primal inclusion*

$$
\textit{find } x \in \mathcal{H} \textit{ such that } 0 \in Ax + \sum_{k=1}^{p} L_k^* \big((B_k \,\square\, D_k)(L_k x) \big) + Cx \tag{8.109}
$$

is not empty and let Z^* be the set of solutions to the dual inclusion

$$\text{find}\;\; y_1^* \in \mathcal{G}_1, \ldots, y_p^* \in \mathcal{G}_p \;\;\text{such that}$$
$$(\exists\, x \in \mathcal{H}) \begin{cases} x \in (A+C)^{-1}\Big(-\sum_{k=1}^p L_k^* y_k^*\Big) \\ \big(\forall k \in \{1,\ldots,p\}\big)\; L_k x \in B_k^{-1} y_k^* + D_k^{-1} y_k^*. \end{cases} \tag{8.110}$$

Let $\varepsilon \in \,]0,1[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 1]$, let $x_0 \in \mathcal{H}$, let $(y_{1,0}^*, \ldots, y_{p,0}^*) \in \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p$, let $\tau \in \,]0,+\infty[$, and let $(\sigma_1, \ldots, \sigma_p) \in \,]0,+\infty[^p$. Set

$$\aleph = \min\{\alpha, \beta_1, \ldots, \beta_p\} \quad \text{and} \quad \beta = \frac{1 - \sqrt{\tau \sum_{k=1}^p \sigma_k \|L_k\|^2}}{\max\{\tau, \sigma_1, \ldots, \sigma_p\}} \tag{8.111}$$

and assume that

$$\aleph\beta > \frac{1}{2}. \tag{8.112}$$

Iterate

$$\begin{array}{l}
\text{for}\; n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
x_n^* = \tau\big(\sum_{k=1}^p L_k^* y_{k,n}^* + C x_n\big) \\
p_n = J_{\tau A}(x_n - x_n^*) \\
x_{n+1} = x_n + \lambda_n(p_n - x_n) \\
\text{for}\; k = 1, \ldots, p \\
\left\lfloor \begin{array}{l}
y_{k,n} = \sigma_k\big(L_k(2p_n - x_n) - D_k^{-1} y_{k,n}^*\big) \\
q_{k,n}^* = J_{\sigma_k B_k^{-1}}(y_{k,n}^* + y_{k,n}) \\
y_{k,n+1}^* = y_{k,n}^* + \lambda_n(q_{k,n}^* - y_{k,n}^*).
\end{array}\right.
\end{array}\right.
\end{array} \tag{8.113}$$

Then there exist $x \in Z$ and $(y_1^*, \ldots, y_p^*) \in Z^*$ such that $x_n \rightharpoonup x$, and, for every $k \in \{1, \ldots, p\}$, $y_{k,n}^* \rightharpoonup y_k^*$.

Proof. Set $\mathbf{X} = \mathcal{H} \oplus \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p$ and

$$\begin{cases}
\boldsymbol{M}\colon \mathbf{X} \to 2^{\mathbf{X}}\colon (x, y_1^*, \ldots, y_p^*) \mapsto \\
\qquad\qquad \big(Ax + \sum_{k=1}^p L_k^* y_k^*\big) \times \big(-L_1 x + B_1^{-1} y_1^*\big) \times \cdots \times \big(-L_p x + B_p^{-1} y_p^*\big) \\
\boldsymbol{C}\colon \mathbf{X} \to \mathbf{X}\colon (x, y_1^*, \ldots, y_p^*) \mapsto \big(Cx, D_1^{-1} y_1^*, \ldots, D_p^{-1} y_p^*\big) \\
\boldsymbol{U}\colon \mathbf{X} \to \mathbf{X}\colon (x, y_1^*, \ldots, y_p^*) \mapsto \\
\qquad\qquad \big(\tau^{-1} x - \sum_{k=1}^p L_k^* y_k^*, -L_1 x + \sigma_1^{-1} y_1^*, \ldots, -L_p x + \sigma_p^{-1} y_p^*\big).
\end{cases} \tag{8.114}$$

As in (5.61), $\boldsymbol{M}$ is maximally monotone, while $\boldsymbol{C}$ is $\aleph$-cocoercive. Furthermore, $\boldsymbol{U} \in \mathcal{B}(\mathcal{H})$ is self-adjoint and, as shown in [387, Equation (3.20)], (8.112) implies that it is β-strongly monotone. Now set $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = (x_n, y^*_{1,n}, \ldots, y^*_{p,n})$ and $\boldsymbol{w}_n = (p_n, q^*_{1,n}, \ldots, q^*_{p,n})$. Then, adopting the same pattern as in the proof of Example 5.20, we rewrite (8.113) as

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{u}^*_n = \boldsymbol{U}\boldsymbol{x}_n - \boldsymbol{C}\boldsymbol{x}_n \\
\boldsymbol{w}_n = \big(\boldsymbol{U} + \boldsymbol{M}\big)^{-1}\boldsymbol{u}^*_n \\
\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \lambda_n(\boldsymbol{w}_n - \boldsymbol{x}_n),
\end{array}
\right.
\end{array}
\tag{8.115}
$$

and thus recover (8.106) with $(\forall n \in \mathbb{N})$ $\gamma_n = 1 < 2\aleph\beta$. We therefore appeal to Proposition 8.42(i) to obtain the weak convergence of $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ to a point $(x, y^*_1, \ldots, y^*_p) \in \operatorname{zer}(\boldsymbol{M} + \boldsymbol{C})$. However, replacing A with $A + C$ and $(B_k^{-1})_{1\leqslant k\leqslant p}$ with $(B_k^{-1} + D_k^{-1})_{1\leqslant k\leqslant p}$ in Lemma 3.12(ii) yields $\operatorname{zer}(\boldsymbol{M} + \boldsymbol{C}) \subset Z \times Z^*$. ☐

Remark 8.45 In terms of Framework 1.2, the embedding underlying Proposition 8.44 employs $\mathbf{X} = \mathcal{H} \oplus \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p$, $\mathfrak{M} = \boldsymbol{M} + \boldsymbol{C}$, and $\mathfrak{T} \colon \mathbf{X} \to \mathcal{H} \colon (x, y^*_1, \ldots, y^*_p) \mapsto x$.

The following application to minimization revisits the setting of Example 7.13 and Remark 7.14.

Example 8.46 Let $0 < p \in \mathbb{N}$, let $\alpha \in \,]0, +\infty[$, let $f \in \Gamma_0(\mathcal{H})$, and let $h \colon \mathcal{H} \to \mathbb{R}$ be convex, differentiable, and such that ∇h is $1/\alpha$-Lipschitzian. For every $k \in \{1, \ldots, p\}$, let $\beta_k \in \,]0, +\infty[$, let $\mathcal{G}_k$ be a real Hilbert space, let $g_k \in \Gamma_0(\mathcal{G}_k)$, let $\ell_k \in \Gamma_0(\mathcal{G}_k)$ be β_k-strongly convex, and suppose that $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$. Let Z be the set of solutions to the primal problem

$$
\underset{x\in\mathcal{H}}{\text{minimize}} \;\; f(x) + \sum_{k=1}^{p} (g_k \,\square\, \ell_k)(L_k x) + h(x),
\tag{8.116}
$$

let Z^* be the set of solutions to the dual problem

$$
\underset{y^*_1\in\mathcal{G}_1,\ldots,y^*_p\in\mathcal{G}_p}{\text{minimize}} \;\; (f^* \,\square\, h^*)\Bigg(-\sum_{k=1}^{p} L^*_k y^*_k\Bigg) + \sum_{k=1}^{p} \big(g^*_k(y^*_k) + \ell^*_k(y^*_k)\big),
\tag{8.117}
$$

and suppose that

$$
\operatorname{zer}\Bigg(\partial f + \sum_{k=1}^{p} L^*_k \circ \big(\partial g_k \,\square\, \partial \ell_k\big) \circ L_k + \nabla h\Bigg) \neq \varnothing.
\tag{8.118}
$$

Let $\varepsilon \in \,]0, 1[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 1]$, let $x_0 \in \mathcal{H}$, let $(y^*_{1,0}, \ldots, y^*_{p,0}) \in \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p$, let $\tau \in \,]0, +\infty[$, and let $(\sigma_1, \ldots, \sigma_p) \in \,]0, +\infty[^p$ be such that

(8.111)–(8.112) hold. Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
x_n^* = \tau\big(\sum_{k=1}^{p} L_k^* y_{k,n}^* + \nabla h(x_n)\big) \\
p_n = \mathrm{prox}_{\tau f}(x_n - x_n^*) \\
x_{n+1} = x_n + \lambda_n(p_n - x_n) \\
\text{for } k = 1, \ldots, p \\
\left\lfloor
\begin{array}{l}
y_{k,n} = \sigma_k\big(L_k(2p_n - x_n) - \nabla \ell_k^*(y_{k,n}^*)\big) \\
q_{k,n}^* = \mathrm{prox}_{\sigma_k g_k^*}(y_{k,n}^* + y_{k,n}) \\
y_{k,n+1}^* = y_{k,n}^* + \lambda_n(q_{k,n}^* - y_{k,n}^*).
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{8.119}
$$

Then there exist $x \in Z$ and $(y_1^*, \ldots, y_p^*) \in Z^*$ such that $x_n \rightharpoonup x$, and, for every $k \in \{1, \ldots, p\}$, $y_{k,n}^* \rightharpoonup y_k^*$.

Proof. It follows from the arguments presented in [145, Section 4] that this is an application of Proposition 8.44 with $A = \partial f$, $C = \nabla h$, and $(\forall k \in \{1, \ldots, p\})$ $B_k = \partial g_k$ and $D_k = \partial \ell_k$. $\square$

Remark 8.47 If we make the additional assumptions that, for every $k \in \{1, \ldots, p\}$, $\ell_k = \iota_{\{0\}}$ and $\sigma_k = \sigma_1$, Example 8.46 was independently obtained in [155, Section 5]. For this reason, (8.119) in this particular setting is called the *Condat–Vũ* algorithm.

8.5 Forward-backward-half-forward splitting

Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $C\colon \mathcal{H} \to \mathcal{H}$ be cocoercive, and let $Q\colon \mathcal{H} \to \mathcal{H}$ be monotone and Lipschitzian. Then a zero of $M = A + C + Q$ can be constructed through the forward-backward-forward algorithms of Theorem 7.1 or Theorem 7.2, applied to A and the monotone and Lipschitzian operator $B = C + Q$. These algorithms require two applications of B, i.e., two applications of C and Q, at each iteration. However, the algorithms discussed so far require two applications of a monotone Lipschitzian operator per iteration, as in the Antipin–Korpelevič method of Section 7.1 and the forward-backward-forward methods of Sections 7.2 and 7.3, but only one application of a cocoercive operator, as in the Euler method of Section 5.4.1 and the forward-backward methods of Sections 8.2 and 8.3. It is therefore natural to ask whether one can find a zero of $A + C + Q$ using only one application of C per iteration. A positive answer to this question was given in [79] with the following forward-backward-half-forward splitting algorithm. We provide a simple proof of its convergence using our geometric framework.

Proposition 8.48 ([79, Theorem 2.3.1]) *Let $\alpha \in \,]0, +\infty[$, let $\beta \in \,]0, +\infty[$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $C\colon \mathcal{H} \to \mathcal{H}$ be α-cocoercive, let*

$Q\colon \mathcal{H} \to \mathcal{H}$ be monotone and β-Lipschitzian, and suppose that the set of solutions Z to the inclusion

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Cx + Qx \tag{8.120}$$

is not empty. Let $x_0 \in \mathcal{H}$, set $\chi = 4\alpha/(1+\sqrt{1+16\alpha^2\beta^2})$, let $\varepsilon \in \left]0, \chi/(\chi+1)\right[$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (1-\varepsilon)\chi]$. Iterate

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
c_n^* = \gamma_n C x_n \\
q_n^* = \gamma_n Q x_n \\
w_n = J_{\gamma_n A}(x_n - c_n^* - q_n^*) \\
x_{n+1} = w_n - \gamma_n Q w_n + q_n^*.
\end{array}
\right.
\end{array} \tag{8.121}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.

Proof. The claims will be established as an application of Theorem 4.12 with

$$W = A + Q, \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad U_n = \gamma_n^{-1}\mathrm{Id} - C - Q \quad \text{and} \quad q_n = x_n. \tag{8.122}$$

In this setting, [95, Proposition 3.9] implies that (7.5) is satisfied, we have

$$\begin{aligned}
(\forall n \in \mathbb{N}) \quad J_{W+C}^{U_n} &= \big(\gamma_n^{-1}\mathrm{Id} + A\big) \circ \big(\gamma_n^{-1}\mathrm{Id} - C - Q\big) \\
&= J_{\gamma_n A} \circ \big(\mathrm{Id} - \gamma_n(C+Q)\big),
\end{aligned} \tag{8.123}$$

and the variables of (4.34) become

$$(\forall n \in \mathbb{N}) \quad \begin{cases}
w_n = J_{\gamma_n A}\big(x_n - \gamma_n(Cx_n + Qx_n)\big) \\
t_n^* = \big(\gamma_n^{-1}\mathrm{Id} - Q\big)x_n - \big(\gamma_n^{-1}\mathrm{Id} - Q\big)w_n \\
\delta_n = \left(\dfrac{1}{\gamma_n} - \dfrac{1}{4\alpha}\right)\|w_n - x_n\|^2 - \langle w_n - x_n \mid Qw_n - Qx_n\rangle.
\end{cases} \tag{8.124}$$

Now set

$$(\forall n \in \mathbb{N}) \quad \lambda_n = \begin{cases}
\dfrac{\gamma_n\|t_n^*\|^2}{\delta_n}, & \text{if } \delta_n > 0; \\
\varepsilon, & \text{otherwise}
\end{cases} \tag{8.125}$$

and note that the assumptions yield

$$\inf_{n\in\mathbb{N}} \lambda_n > 0 \quad \text{and} \quad \sup_{n\in\mathbb{N}} \lambda_n < 2. \tag{8.126}$$

As a consequence of (8.124) and the properties of Q, we have

$$(\forall n \in \mathbb{N}) \quad \delta_n \leqslant 0 \Rightarrow \left(\frac{1}{\gamma_n} - \frac{1}{4\alpha} - \beta\right)\|w_n - x_n\|^2 \leqslant 0$$
$$\Leftrightarrow w_n = x_n$$
$$\Leftrightarrow t_n^* = 0. \tag{8.127}$$

Hence, (4.34) yields

$$(\forall n \in \mathbb{N}) \quad d_n = \frac{\gamma_n}{\lambda_n} t_n^* = \frac{1}{\lambda_n}\big(x_n - w_n + \gamma_n(Qw_n - Qx_n)\big). \tag{8.128}$$

As a result, the sequence $(x_n)_{n\in\mathbb{N}}$ produced by (8.121) coincides with that of (4.34). Hence, by Theorem 4.12(i) and (8.126), $\sum_{n\in\mathbb{N}} \|d_n\|^2 < +\infty$ which, in view of (8.128), yields

$$(\mathrm{Id} - \gamma_n Q)w_n - (\mathrm{Id} - \gamma_n Q)x_n \to 0. \tag{8.129}$$

However, since $\chi \leqslant 1/\beta$, $(\gamma_n)_{n\in\mathbb{N}}$ lies in $[\varepsilon, (1-\varepsilon)/\beta]$ and Lemma 2.48(i) implies that the operators $(\mathrm{Id} - \gamma_n Q)_{n\in\mathbb{N}}$ are ε-strongly monotone. Hence,

$$(\forall n \in \mathbb{N}) \quad \varepsilon\|w_n - x_n\|^2 \leqslant \langle w_n - x_n \mid (\mathrm{Id} - \gamma_n Q)w_n - (\mathrm{Id} - \gamma_n Q)x_n\rangle \tag{8.130}$$

and, by the Cauchy–Schwarz inequality and (8.129),

$$\|w_n - x_n\| \leqslant \varepsilon^{-1}\|(\mathrm{Id} - \gamma_n Q)w_n - (\mathrm{Id} - \gamma_n Q)x_n\| \to 0. \tag{8.131}$$

In turn, since C is $1/\alpha$-Lipschitzian, these facts confirm that

$$\|U_n w_n - U_n x_n\| \leqslant \gamma_n^{-1}\|(\mathrm{Id} - \gamma_n Q)w_n - (\mathrm{Id} - \gamma_n Q)x_n\| + \|Cw_n - Cx_n\|$$
$$\leqslant \varepsilon^{-1}\|(\mathrm{Id} - \gamma_n Q)w_n - (\mathrm{Id} - \gamma_n Q)x_n\| + \alpha^{-1}\|w_n - x_n\|$$
$$\to 0. \tag{8.132}$$

Thus, the assertion follows from Theorem 4.12(ii) since its conditions (ii)(b) and (ii)(c) are fulfilled. ☐

Remark 8.49 We complement Proposition 8.48 with a few commentaries.

(i) Suppose that $C = 0$. Then, since α can be arbitrarily large, $\chi = 1/\beta$ and (8.121) reverts to the forward-backward-forward algorithm (7.2).

(ii) Suppose that $Q = 0$. Then, since $\beta = 0$, $\chi = 2\alpha$ and (8.121) becomes an unrelaxed version of forward-backward algorithm (8.5).

(iii) Using the geometric pattern of the proof given above, a strongly convergent version of the forward-backward-half-forward algorithm can be derived from Theorem 4.14.

As an illustration, we extend the Lagrangian approach of Proposition 7.5.

Example 8.50 Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$, and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$ be such that $0 \in \operatorname{sri}(L(\operatorname{dom} f) - \operatorname{dom} g)$. Let $\alpha \in \left]0,+\infty\right[$ and let $h\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable and such that ∇h is $1/\alpha$-Lipschitzian. Suppose that the primal problem

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x) + g(Lx) + h(x) \tag{8.133}$$

admits solutions and consider the dual problem

$$\underset{v^*\in\mathcal{G}}{\text{minimize}}\;\; (f^* \,\square\, h^*)(-L^*v^*) + g^*(v^*). \tag{8.134}$$

Let $(x_0, y_0, v_0^*) \in \mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G}$, set $\chi = 4\alpha/(1 + \sqrt{1 + 16\alpha^2(1 + \|L\|^2)}\,)$, let $\varepsilon \in \left]0, \chi/(\chi + 1)\right[$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, (1 - \varepsilon)\chi]$. Iterate

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots\\
\left\lfloor
\begin{array}{l}
c_n^* = \gamma_n \nabla h(x_n)\\
q_{1,n}^* = \gamma_n L^* v_n^*\\
q_{2,n}^* = -\gamma_n v_n^*\\
q_{3,n}^* = \gamma_n (y_n - Lx_n)\\
a_{1,n} = \operatorname{prox}_{\gamma_n f}\big(x_n - c_n^* - q_{1,n}^*\big)\\
a_{2,n} = \operatorname{prox}_{\gamma_n g}\big(y_n - q_{2,n}^*\big)\\
x_{n+1} = a_{1,n} + \gamma_n L^* q_{3,n}^*\\
y_{n+1} = a_{2,n} - \gamma_n q_{3,n}^*\\
v_{n+1}^* = v_n^* + \gamma_n\big(La_{1,n} - a_{2,n}\big).
\end{array}
\right.
\end{array} \tag{8.135}$$

Then $(x_n)_{n\in\mathbb{N}}$ and $(v_n^*)_{n\in\mathbb{N}}$ converge weakly to solutions to (8.133) and (8.134), respectively.

Proof. We adapt the approach of Section 7.4.2. The saddle operator of (7.22)–(7.23) becomes $\mathcal{S} = \boldsymbol{A} + \boldsymbol{C} + \boldsymbol{Q}$, where

$$\begin{cases}
\boldsymbol{A}\colon (x, y, v^*) \mapsto \partial f(x) \times \partial g(y) \times \{0\}\\
\boldsymbol{C}\colon (x, y, v^*) \mapsto \big(\nabla h(x), 0, 0\big)\\
\boldsymbol{Q}\colon (x, y, v^*) \mapsto \big(L^*v^*, -v^*, -Lx + y\big).
\end{cases} \tag{8.136}$$

As in Section 7.4.2, $\boldsymbol{A}$ is maximally monotone and $\boldsymbol{Q}$ is monotone and $\sqrt{1 + \|L\|^2}$-Lipschitzian. Further, by virtue of Lemma 2.2, $\boldsymbol{C}$ is α-cocoercive. Now set

$(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = (x_n, y_n, v_n^*)$, $\boldsymbol{c}_n^* = (c_n^*, 0, 0)$, $\boldsymbol{q}_n^* = (q_{1,n}^*, q_{2,n}^*, q_{3,n}^*)$, and $\boldsymbol{w}_n = (a_{1,n}, a_{2,n}, v_n^* - q_{3,n}^*)$. Then (8.135) assumes the form

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{c}_n^* = \gamma_n \boldsymbol{C} \boldsymbol{x}_n \\
\boldsymbol{q}_n^* = \gamma_n \boldsymbol{Q} \boldsymbol{x}_n \\
\boldsymbol{w}_n = J_{\gamma_n \boldsymbol{A}}(\boldsymbol{x}_n - \boldsymbol{c}_n^* - \boldsymbol{q}_n^*) \\
\boldsymbol{x}_{n+1} = \boldsymbol{w}_n - \gamma_n \boldsymbol{Q} \boldsymbol{w}_n + \boldsymbol{q}_n^*,
\end{array}
\right.
\end{array}
\tag{8.137}
$$

which is (8.121). Hence, by Proposition 8.48, $(x_n, y_n, v_n^*)_{n\in\mathbb{N}}$ converges weakly to a point $(x, y, v^*) \in \operatorname{zer} \boldsymbol{\mathcal{S}}$. ☐

Remark 8.51 Let $\alpha \in \,]0, +\infty[$, let $A: \mathcal{H} \to 2^{\mathcal{H}}$ and $B: \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone, let $C: \mathcal{H} \to \mathcal{H}$ be α-cocoercive, and let $0 \neq L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. As in Remark 7.8, the saddle approach of Example 8.50 has a natural extension to the problem of finding a zero of $A + L^* \circ B \circ L + C$ and the dual problem of finding a zero of $-L \circ (A + C)^{-1} \circ (-L^*) + B^{-1}$. In this setting, the saddle operator is

$$
\begin{array}{llll}
\boldsymbol{\mathcal{S}}\colon & \mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G} & \to & 2^{\mathcal{H}\oplus\mathcal{G}\oplus\mathcal{G}} \\
& (x, y, v^*) & \mapsto & (Ax + Cx + L^*v^*) \times (By - v^*) \times \{-Lx + y\}.
\end{array}
\tag{8.138}
$$

Accordingly, it suffices to replace ∇h with C, $\operatorname{prox}_{\gamma_n f}$ with $J_{\gamma_n A}$, and $\operatorname{prox}_{\gamma_n g}$ with $J_{\gamma_n B}$ in (8.135) to find primal-dual solutions.

9 Block-iterative Kuhn–Tucker projective splitting

9.1 Preview

Unlike the methods described so far, those described in this section were explicitly designed by employing the geometric principle of Theorem 4.2. The terminology *projective splitting* was coined in [181] in the context of an algorithm to solve Problem 3.1 by choosing points in the graph of A and B to construct half-spaces containing an "extended solution set." In the language of Lemma 3.8, this set is actually the set of zeros of the Kuhn–Tucker operator (3.10), which collapses to

$$
\operatorname{zer} \boldsymbol{\mathcal{K}} = \left\{ (x, x^*) \in \mathcal{H} \oplus \mathcal{H} \mid -x^* \in Ax \ \text{ and } \ x \in B^{-1}x^* \right\}. \tag{9.1}
$$

The paper [181] initiated a fruitful line of work towards more complex monotone inclusions [9, 10, 47, 93, 136, 178, 182, 234, 235, 236, 237, 268, 269, 355]. We use the term *Kuhn–Tucker projective splitting* to describe a method that operates through the principles of Framework 1.2, where $\boldsymbol{\mathcal{M}}$ is a Kuhn-Tucker operator. As we shall see, projective splitting algorithms have features quite different from those of the traditional methods of Sections 5–8 and they display an unprecedented level of flexibility in terms of implementation.

9.2 Primal-dual composite inclusions

Let us go back to the composite Problem 3.7. The sets of primal and dual solutions are, respectively,

$$Z = \operatorname{zer}(A + L^* \circ B \circ L) \quad \text{and} \quad Z^* = \operatorname{zer}\big(-L \circ A^{-1} \circ (-L^*) + B^{-1}\big). \tag{9.2}$$

Moreover, as pointed out in Example 3.20, an embedding of (3.7) is $(\mathsf{X}, \mathcal{K}, \mathcal{T})$, where $\mathsf{X} = \mathcal{H} \oplus \mathcal{G}$, $\mathcal{K}$ is the Kuhn–Tucker operator of (3.10), that is,

$$\mathcal{K} \colon \mathsf{X} \to 2^{\mathsf{X}} \colon (x, y^*) \mapsto \big(Ax + L^* y^*\big) \times \big(B^{-1} y^* - Lx\big), \tag{9.3}$$

and $\mathcal{T} \colon \mathsf{X} \to \mathcal{H} \colon (x, y^*) \mapsto x$. The task is therefore to find a zero of $\mathcal{K}$. This is the path followed in the monotone+skew approach of Section 7.4.1. However, this method requires knowledge of $\|L\|$ (or of a tight upper bound for it), which may be difficult to obtain in certain problems. The renormed algorithms of Example 5.20 and [61], the saddle algorithm of Remark 8.51, or the minimal lifting algorithm of [14] share the same potential limitation. On the other hand, the method of Proposition 5.15, which was derived from the method of partial inverses, requires the inversion of linear operators, a task that may also face implementation issues.

A strategy which circumvents the above shortcomings was proposed in [9], where the approach of [181] for solving Problem 3.1 was extended to Problem 3.7. More precisely, it employs the geometric principle of Proposition 4.10 as follows. Let us assume that, at iteration n, points $(a_n, a_n^*) \in \operatorname{gra} A$ and $(b_n, b_n^*) \in \operatorname{gra} B$ are available and set

$$\boldsymbol{m}_n = (a_n, b_n^*) \quad \text{and} \quad \boldsymbol{m}_n^* = (a_n^* + L^* b_n^*, b_n - L a_n). \tag{9.4}$$

Then it is clear from (9.3) that $(\boldsymbol{m}_n, \boldsymbol{m}_n^*) \in \operatorname{gra} \mathcal{K}$. Hence, given $\lambda_n \in \left]0, 2\right[$, iteration n of algorithm (4.32) updates $(x_n, y_n^*) \in \mathsf{X}$ via the routine

$$\left\lfloor \begin{array}{l} (t_n, t_n^*) = (b_n - L a_n, a_n^* + L^* b_n^*) \\ \tau_n = \|t_n\|^2 + \|t_n^*\|^2 \\ \text{if } \tau_n > 0 \\ \left\lfloor \theta_n = \dfrac{\lambda_n}{\tau_n} \max\Big\{0, \langle x_n \mid t_n^* \rangle + \langle t_n \mid y_n^* \rangle - \langle a_n \mid a_n^* \rangle - \langle b_n \mid b_n^* \rangle\Big\} \right. \\ \text{else } \theta_n = 0 \\ (x_{n+1}, y_{n+1}^*) = (x_n - \theta_n t_n^*, y_n^* - \theta_n t_n). \end{array} \right. \tag{9.5}$$

In view of Proposition 4.10(ii), the task is now to specify $(a_n, a_n^*) \in \operatorname{gra} A$ and $(b_n, b_n^*) \in \operatorname{gra} B$ so as to guarantee that $\boldsymbol{m}_n - (x_n, y_n^*) \rightharpoonup 0$ and $\boldsymbol{m}_n^* \to 0$, that is,

$$a_n - x_n \rightharpoonup 0,\ \ b_n^* - y_n^* \rightharpoonup 0,\ \ b_n - L a_n \to 0,\ \text{ and } a_n^* + L^* b_n^* \to 0. \tag{9.6}$$

Given γ_n and σ_n in $]0,+\infty[$, choosing

$$(a_n, a_n^*) = \Big(J_{\gamma_n A}(x_n - \gamma_n L^* y_n^*), \gamma_n^{-1}\big(x_n - J_{\gamma_n A}(x_n - \gamma_n L^* y_n^*)\big) - L^* y_n^*\Big) \quad (9.7)$$

and

$$(b_n, b_n^*) = \Big(J_{\sigma_n B}(Lx_n + \sigma_n y_n^*), \sigma_n^{-1}\big(Lx_n - J_{\sigma_n B}(Lx_n + \sigma_n y_n^*)\big) + y_n^*\Big) \quad (9.8)$$

satisfies this requirement, which leads to the following result.

Proposition 9.1 ([9, Proposition 3.5]) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$. Suppose that the set Z of solutions to the primal inclusion*

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + L^*\big(B(Lx)\big) \quad (9.9)$$

is not empty and let Z^* be the set of solutions to the dual inclusion

$$\text{find } y^* \in \mathcal{G} \text{ such that } 0 \in -L\big(A^{-1}(-L^*y^*)\big) + B^{-1}y^*. \quad (9.10)$$

Let $\varepsilon \in\,]0,1[$, let $(\gamma_n)_{n\in\mathbb{N}}$ and $(\sigma_n)_{n\in\mathbb{N}}$ be sequences in $[\varepsilon, 1/\varepsilon]$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 2-\varepsilon]$, let $x_0 \in \mathcal{H}$, and let $y_0^* \in \mathcal{G}$. Iterate

$$\begin{array}{l}
\textit{for } n = 0,1,\ldots \\
\left\lfloor
\begin{array}{l}
a_n = J_{\gamma_n A}(x_n - \gamma_n L^* y_n^*) \\
l_n = Lx_n \\
b_n = J_{\sigma_n B}(l_n + \sigma_n y_n^*) \\
t_n = b_n - La_n \\
t_n^* = \gamma_n^{-1}(x_n - a_n) + \sigma_n^{-1}L^*(l_n - b_n) \\
\tau_n = \|t_n\|^2 + \|t_n^*\|^2 \\
\textit{if } \tau_n > 0 \\
\left\lfloor\ \theta_n = \lambda_n\big(\gamma_n^{-1}\|x_n - a_n\|^2 + \sigma_n^{-1}\|l_n - b_n\|^2\big)/\tau_n \right. \\
\textit{else } \theta_n = 0 \\
x_{n+1} = x_n - \theta_n t_n^* \\
y_{n+1}^* = y_n^* - \theta_n t_n.
\end{array}
\right.
\end{array} \quad (9.11)$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point $x \in Z$ and $(y_n^*)_{n\in\mathbb{N}}$ converges weakly to a point $y^* \in Z^*$.

Remark 9.2 Here are notable instantiations of Proposition 9.1.

(i) The first instance of (9.11) in the literature seems to be that of [167], where $\mathcal{H}$ and $\mathcal{G}$ are Euclidean spaces, $A = 0$, and $(\forall n \in \mathbb{N})\ \gamma_n = \sigma_n = 1$ and $\lambda_n = \lambda \in\,]0,2[$. Convergence of the primal sequence $(x_n)_{n\in\mathbb{N}}$ was established by different means.

(ii) In the setting of Problem 3.1 (i.e., $\mathcal{G} = \mathcal{H}$ and $L = \mathrm{Id}$), (9.11) was studied in [181]. Under the additional assumptions that $A + B$ is maximally monotone or that $\mathcal{H}$ is finite-dimensional, weak convergence was established in [181, Proposition 3] for a version of (9.11) which allows for an additional relaxation parameter in the definition of a_n.

Remark 9.3 So far, we have presented several methods to solve Problem 3.7; see Proposition 5.15, Example 5.20, Proposition 7.3, and Remark 8.51. Some features that distinguish the splitting algorithm (9.11) from them are as follows.

(i) At each iteration of (9.11), different proximal parameters γ_n and σ_n can be used for the operators A and B and, since ε is chosen by the user, their values can be arbitrarily large.

(ii) The execution of (9.11) does not require that $\|L\|$ or an approximation thereof be known, or the inversion of linear operators.

(iii) A variant of (9.11) exploiting the cocoercivity of some of the operators and activating them via Euler steps is discussed in [236].

(iv) The complexity of certain special cases and variants of (9.11) is investigated in [234, 269].

The following strongly convergent projective splitting algorithm results from Proposition 4.11.

Proposition 9.4 ([10, Proposition 3.5]) *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone, and let $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$. Suppose that the set Z of solutions to the primal inclusion*

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + L^*\big(B(Lx)\big) \tag{9.12}$$

is not empty and let Z^* be the set of solutions to the dual inclusion

$$\text{find } y^* \in \mathcal{G} \text{ such that } 0 \in -L\big(A^{-1}(-L^*y^*)\big) + B^{-1}y^*. \tag{9.13}$$

Let $\varepsilon \in \left]0,1\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ and $(\sigma_n)_{n\in\mathbb{N}}$ be sequences in $[\varepsilon, 1/\varepsilon]$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a *sequence in* $[\varepsilon, 1]$, *let* $x_0 \in \mathcal{H}$, *and let* $y_0^* \in \mathcal{G}$. *Iterate*

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
a_n = J_{\gamma_n A}(x_n - \gamma_n L^* y_n^*) \\
l_n = L x_n \\
b_n = J_{\sigma_n B}(l_n + \sigma_n y_n^*) \\
t_n = b_n - L a_n \\
t_n^* = \gamma_n^{-1}(x_n - a_n) + \sigma_n^{-1} L^*(l_n - b_n) \\
\tau_n = \|t_n\|^2 + \|t_n^*\|^2 \\
\text{if } \tau_n > 0 \\
\left\lfloor\ \theta_n = \lambda_n\big(\gamma_n^{-1}\|x_n - a_n\|^2 + \sigma_n^{-1}\|l_n - b_n\|^2\big)/\tau_n \right. \\
\text{else } \theta_n = 0 \\
r_n = x_n - \theta_n t_n^* \\
r_n^* = y_n^* - \theta_n t_n \\
\chi_n = \theta_n\big(\langle x_0 - x_n \mid t_n^*\rangle + \langle t_n \mid y_0^* - y_n^*\rangle\big) \\
\mu_n = \|x_0 - x_n\|^2 + \|y_0^* - y_n^*\|^2 \\
\nu_n = \theta_n^2 \tau_n \\
\rho_n = \mu_n \nu_n - \chi_n^2 \\
\text{if } \rho_n = 0 \text{ and } \chi_n \geqslant 0 \\
\left\lfloor
\begin{array}{l}
x_{n+1} = r_n \\
y_{n+1}^* = r_n^*
\end{array}
\right. \\
\text{if } \rho_n > 0 \text{ and } \chi_n \nu_n \geqslant \rho_n \\
\left\lfloor
\begin{array}{l}
x_{n+1} = x_0 - \theta_n(1 + \chi_n/\nu_n) t_n^* \\
y_{n+1}^* = y_0^* - \theta_n(1 + \chi_n/\nu_n) t_n
\end{array}
\right. \\
\text{if } \rho_n > 0 \text{ and } \chi_n \nu_n < \rho_n \\
\left\lfloor
\begin{array}{l}
x_{n+1} = x_n + (\nu_n/\rho_n)\big(\chi_n(x_0 - x_n) - \mu_n \theta_n t_n^*\big) \\
y_{n+1}^* = y_n^* + (\nu_n/\rho_n)\big(\chi_n(y_0^* - y_n^*) - \mu_n \theta_n t_n\big).
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{9.14}
$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges strongly to a point* $x \in Z$ *and* $(y_n^*)_{n\in\mathbb{N}}$ *converges strongly to a point* $y^* \in Z^*$.

9.3 Block-iterative asynchronous method

We consider a refinement of Problem 3.11 in which the primal variable is specified in terms of finitely many coordinates, say $\boldsymbol{x} = (x_1, \ldots, x_m)$, where each x_i lies in a Hilbert space $\mathcal{H}_i$. Such coupled systems of inclusions arise in particular in multivariate optimization [1, 16, 17, 130], domain decomposition methods [6, 18, 21], image processing [25, 78, 117, 384], game theory [48, 57, 77, 98], network flow problems [54, 92, 341, 342], machine learning [74, 232, 279, 385], signal processing [75], mean field games [81], statistics [141, 394], tensor completion [200, 288], and semi-definite programming [229, 301].

Problem 9.5 Let $I = \{1, \ldots, m\}$ and $K = \{1, \ldots, p\}$ be nonempty finite sets. For every $i \in I$ and every $k \in K$, let $\mathcal{H}_i$ and $\mathcal{G}_k$ be real Hilbert spaces, let $A_i : \mathcal{H}_i \to 2^{\mathcal{H}_i}$ and $B_k : \mathcal{G}_k \to 2^{\mathcal{G}_k}$ be maximally monotone, and let $L_{ki} \in \mathcal{B}(\mathcal{H}_i, \mathcal{G}_k)$. Set

$$\boldsymbol{\mathcal{H}} = \bigoplus_{i \in I} \mathcal{H}_i \quad \text{and} \quad \boldsymbol{\mathcal{G}} = \bigoplus_{k \in K} \mathcal{G}_k. \tag{9.15}$$

The objective is to solve the primal inclusion

$$\text{find } \boldsymbol{x} \in \boldsymbol{\mathcal{H}} \text{ such that}$$
$$(\forall i \in I) \quad 0 \in A_i x_i + \sum_{k \in K} L_{ki}^* \Bigg(B_k \Bigg(\sum_{j \in I} L_{kj} x_j \Bigg) \Bigg) \tag{9.16}$$

together with the dual inclusion

$$\text{find } \boldsymbol{y}^* \in \boldsymbol{\mathcal{G}} \text{ such that}$$
$$(\exists \boldsymbol{x} \in \boldsymbol{\mathcal{H}}) \quad \begin{cases} (\forall i \in I) \;\; x_i \in A_i^{-1} \Bigg(-\sum_{k \in K} L_{ki}^* y_k^* \Bigg) \\ (\forall k \in K) \;\; \displaystyle\sum_{i \in I} L_{ki} x_i \in B_k^{-1} y_k^*. \end{cases} \tag{9.17}$$

Remark 9.6 There is an oversight in the dual problem given in [136, Problem 1], the correct formulation of the dual inclusion is (9.17).

The counterpart of Lemma 3.12 for Problem 9.5 is as follows.

Lemma 9.7 *In the setting of Problem 9.5, set* $\mathbf{X} = \boldsymbol{\mathcal{H}} \oplus \boldsymbol{\mathcal{G}}$*, and let* $\boldsymbol{Z}$ *and* $\boldsymbol{Z}^*$ *be the sets of solutions to* (9.16) *and* (9.17)*, respectively. Define the Kuhn–Tucker operator of Problem* 9.5 *as*

$$\begin{aligned} \boldsymbol{\mathcal{K}} \colon \mathbf{X} \to 2^{\mathbf{X}} \colon (\boldsymbol{x}, \boldsymbol{y}^*) \mapsto & \\ & \Bigg(A_1 x_1 + \sum_{k \in K} L_{k1}^* y_k^* \Bigg) \times \cdots \times \Bigg(A_m x_m + \sum_{k \in K} L_{km}^* y_k^* \Bigg) \\ & \times \Bigg(-\sum_{i \in I} L_{1i} x_i + B_1^{-1} y_1^* \Bigg) \times \cdots \times \Bigg(-\sum_{i \in I} L_{pi} x_i + B_p^{-1} y_p^* \Bigg) \end{aligned} \tag{9.18}$$

and the set of Kuhn–Tucker points as $\operatorname{zer} \boldsymbol{\mathcal{K}}$*. Then the following hold:*

(i) $\boldsymbol{\mathcal{K}}$ *is maximally monotone.*

(ii) $\operatorname{zer} \boldsymbol{\mathcal{K}}$ *is a closed convex subset of* $\boldsymbol{Z} \times \boldsymbol{Z}^*$.

(iii) $\boldsymbol{Z}^* \neq \varnothing \Leftrightarrow \operatorname{zer} \boldsymbol{\mathcal{K}} \neq \varnothing \Rightarrow \boldsymbol{Z} \neq \varnothing$.

Example 9.8 In the setting of Problem 9.5, set $\mathsf{X} = \mathcal{H} \oplus \mathcal{G}$, let $\mathcal{K}$ be the Kuhn–Tucker operator of (9.18), and let $\mathcal{T}\colon \mathsf{X} \to \mathcal{H}\colon (\boldsymbol{x}, \boldsymbol{y}^*) \mapsto \boldsymbol{x}$. Then it follows from Lemma 9.7(ii) that $(\mathsf{X}, \mathcal{K}, \mathcal{T})$ is an embedding of (9.16).

When the monotone operators $(A_i)_{1 \leqslant i \leqslant m}$ and $(B_k)_{1 \leqslant k \leqslant p}$ are taken to be subdifferentials, Problem 9.5 specializes to a multivariate minimization problem under a suitable qualification condition.

Example 9.9 Define $\mathcal{H}$ and $\mathcal{G}$ as in Problem 9.5. For every $i \in I$ and every $k \in K$, let $f_i \in \Gamma_0(\mathcal{H}_i)$, let $g_k \in \Gamma_0(\mathcal{G}_k)$, and let $L_{ki} \in \mathcal{B}(\mathcal{H}_i, \mathcal{G}_k)$. Suppose that (existence of a Kuhn–Tucker point)

$$(\exists\, \boldsymbol{x} \in \mathcal{H})(\exists\, \boldsymbol{y}^* \in \mathcal{G}) \quad \begin{cases} (\forall i \in I) \quad -\displaystyle\sum_{k\in K} L_{ki}^* y_k^* \in \partial f_i(x_i) \\ (\forall k \in K) \quad \displaystyle\sum_{i\in I} L_{ki} x_i \in \partial g_k^*(y_k^*). \end{cases} \tag{9.19}$$

The objective is to solve the primal minimization problem

$$\underset{\boldsymbol{x}\in\mathcal{H}}{\text{minimize}} \quad \sum_{i\in I} f_i(x_i) + \sum_{k\in K} g_k\biggl(\sum_{i\in I} L_{ki} x_i\biggr) \tag{9.20}$$

together with its dual problem

$$\underset{\boldsymbol{y}^*\in\mathcal{G}}{\text{minimize}} \quad \sum_{i\in I} f_i^*\biggl(-\sum_{k\in K} L_{ki}^* y_k^*\biggr) + \sum_{k\in K} g_k^*(y_k^*). \tag{9.21}$$

In an attempt to recast Problem 9.5 as a realization of Problem 3.7, let us define

$$\begin{cases} A\colon \mathcal{H} \to 2^{\mathcal{H}}\colon \boldsymbol{x} \mapsto A_1 x_1 \times \cdots \times A_m x_m \\ B\colon \mathcal{G} \to 2^{\mathcal{G}}\colon \boldsymbol{y} \mapsto B_1 y_1 \times \cdots \times B_p y_p \\ L\colon \mathcal{H} \to \mathcal{G}\colon \boldsymbol{x} \mapsto \bigl(\sum_{i\in I} L_{1i} x_i, \ldots, \sum_{i\in I} L_{pi} x_i\bigr). \end{cases} \tag{9.22}$$

Upon injecting these operators into (9.11) and invoking Example 2.37, we obtain an algorithm that requires that $m+p$ resolvents be evaluated at each iteration. In large-scale problems, m and/or p can be huge and this requirement poses implementation issues as the only information flow within an iteration is from the m operators $(A_i)_{i\in I}$ calculations to the p operators $(B_k)_{k\in K}$ calculations. This results in an algorithm in which large blocks of calculations must be performed before any information is exchanged between subsystems. Thus, if some small subset of the subsystems represented by the operators $(A_i)_{i\in I}$ or $(B_k)_{k\in K}$ are more computation-intensive than others, load balancing can become problematic: most processors may have to sit idle while the remaining few complete their tasks. More generally, none of the methods discussed so far can handle block-processing or asynchronicity.

The algorithm we present now was conceived in [136] around combined objectives which were beyond the reach of the existing splitting algorithms:

- **Block iterations:** At iteration n, it necessitates calculation of new points in the graphs of only some of the operators, say $(A_i)_{i\in I_n}$ and $(B_k)_{k\in K_n}$ with $I_n \subset I$ and $K_n \subset K$. The deterministic control sequences $(I_n)_{n\in\mathbb{N}}$ and $(K_n)_{n\in\mathbb{N}}$ dictate how frequently the various operators are used.
- **Asynchronicity:** A new point $(a_{i,n}, a^*_{i,n}) \in \operatorname{gra} A_i$ being incorporated into the calculations at iteration n may be based on data $x_{i,\pi_i(n)}$ and $(y^*_{k,\pi_i(n)})_{k\in K}$ available at some possibly earlier iteration $\pi_i(n) \leqslant n$. Therefore, the calculation of $(a_{i,n}, a^*_{i,n})$ could have been initiated at iteration $\pi_i(n)$, with its results becoming available only at iteration n. Likewise, for every $k \in K_n$, the computation of $(b_{k,n}, b^*_{k,n}) \in \operatorname{gra} B_k$ can be initiated at some iteration $\omega_k(n) \leqslant n$, based on $(x_{i,\omega_k(n)})_{i\in I}$ and $y^*_{k,\omega_k(n)}$.
- **Convergence:** It guarantees (weak or strong) convergence of the iterates to primal and dual solutions.

Remark 9.10 Regarding block iterations for Problem 9.5, a product space version of the Douglas–Rachford algorithm was introduced in [146], which features random activation of the blocks. A random block-iterative version of the forward-backward algorithm was also proposed in [146], which led in [310] to algorithms for Problem 9.5 via the renorming techniques presented in Section 8.4.6 (for specialized block-iterative forward-backward algorithms tailored for instances of Example 9.9, see [74, 266, 350, 372]). These methods differ from the deterministic ones presented below in that they operate under stochastic assumptions on the underlying processes, have a less predictable computational load over the iterations, have less freedom in the choice of the proximal parameters, and offer only almost sure convergence guarantees (see also [99] for numerical comparisons).

Going back to (9.5) in the setting of (9.22) and Lemma 9.7, what is actually needed at iteration n to create the half-space containing $\operatorname{zer}\boldsymbol{\mathcal{K}}$ are points

$$\begin{cases}(a_{i,n}, a^*_{i,n}) \in \operatorname{gra} A_i, & \text{for } i \in I;\\ (b_{k,n}, b^*_{k,n}) \in \operatorname{gra} B_k, & \text{for } k \in K.\end{cases} \tag{9.23}$$

The key observation is that not all of these points have to be new in order to obtain a new half-space. In other words, we can update only some of them while keeping old ones and still create a new half-space onto which the current primal-dual iterate $(\boldsymbol{x}_n, \boldsymbol{y}^*_n) = (x_{1,n}, \dots, x_{m,n}, y^*_{1,n}, \dots, y^*_{p,n})$ will be projected. How often the points in the individual graphs should be updated, and in which fashion, will be regulated by the following rules.

Assumption 9.11 *Given* $0 < R \in \mathbb{N}$, $(I_n)_{n\in\mathbb{N}}$ *is a sequence of nonempty subsets of* I, *and* $(K_n)_{n\in\mathbb{N}}$ *is a sequence of nonempty subsets of* K *such that*

$$I_0 = I,\ K_0 = K,\ \ \textit{and}\ \ (\forall n \in \mathbb{N})\ \begin{cases} \displaystyle\bigcup_{j=n}^{n+R-1} I_j = I \\ \displaystyle\bigcup_{j=n}^{n+R-1} K_j = K. \end{cases} \tag{9.24}$$

Assumption 9.12 $T \in \mathbb{N}$ *and, for every* $i \in I$ *and every* $k \in K$, $(\pi_i(n))_{n\in\mathbb{N}}$ *and* $(\omega_k(n))_{n\in\mathbb{N}}$ *are sequences in* $\mathbb{N}$ *such that* $(\forall n \in \mathbb{N})$ $n - T \leqslant \pi_i(n) \leqslant n$ *and* $n - T \leqslant \omega_k(n) \leqslant n$.

With these considerations and by making selections for the updated points $(a_{i,n}, a^*_{i,n})_{i\in I_n}$ and $(b^*_{k,n}, b^*_{k,n})_{k\in K_n}$ akin to those of (9.7) and (9.8), we arrive at the following realization of (9.5).

Algorithm 9.13 Consider the setting of Problem 9.5, suppose that Assumptions 9.11 and 9.12 are in force, let $\varepsilon \in\,]0,1[$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 2-\varepsilon]$. For every $i \in I$, let $(\gamma_{i,n})_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 1/\varepsilon]$ and let $x_{i,0} \in \mathcal{H}_i$. For every $k \in K$, let $(\sigma_{k,n})_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 1/\varepsilon]$ and let

$y^*_{k,0} \in \mathcal{G}_k$. Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for every } i \in I_n \\
\left\lfloor
\begin{array}{l}
l^*_{i,n} = \sum_{k\in K} L^*_{ki} y^*_{k,\pi_i(n)} \\
a_{i,n} = J_{\gamma_{i,\pi_i(n)} A_i}\big(x_{i,\pi_i(n)} - \gamma_{i,\pi_i(n)} l^*_{i,n}\big) \\
a^*_{i,n} = \gamma^{-1}_{i,\pi_i(n)}(x_{i,\pi_i(n)} - a_{i,n}) - l^*_{i,n}
\end{array}
\right. \\
\text{for every } i \in I \smallsetminus I_n \\
\lfloor\ \big(a_{i,n}, a^*_{i,n}\big) = \big(a_{i,n-1}, a^*_{i,n-1}\big) \\
\text{for every } k \in K_n \\
\left\lfloor
\begin{array}{l}
l_{k,n} = \sum_{i\in I} L_{ki} x_{i,\omega_k(n)} \\
b_{k,n} = J_{\sigma_{k,\omega_k(n)} B_k}\big(l_{k,n} + \sigma_{k,\omega_k(n)} y^*_{k,\omega_k(n)}\big) \\
b^*_{k,n} = y^*_{k,\omega_k(n)} + \sigma^{-1}_{k,\omega_k(n)}(l_{k,n} - b_{k,n})
\end{array}
\right. \\
\text{for every } k \in K \smallsetminus K_n \\
\lfloor\ \big(b_{k,n}, b^*_{k,n}\big) = \big(b_{k,n-1}, b^*_{k,n-1}\big) \\
\text{for every } i \in I \\
\lfloor\ t^*_{i,n} = a^*_{i,n} + \sum_{k\in K} L^*_{ki} b^*_{k,n} \\
\text{for every } k \in K \\
\lfloor\ t_{k,n} = b_{k,n} - \sum_{i\in I} L_{ki} a_{i,n} \\
\tau_n = \sum_{i\in I} \|t^*_{i,n}\|^2 + \sum_{k\in K} \|t_{k,n}\|^2 \\
\text{if } \tau_n > 0 \\
\left\lfloor
\begin{array}{l}
\theta_n = \dfrac{\lambda_n}{\tau_n} \max\Big\{0, \sum_{i\in I}\big(\langle x_{i,n} \mid t^*_{i,n}\rangle - \langle a_{i,n} \mid a^*_{i,n}\rangle\big) \\
\qquad\qquad + \sum_{k\in K}\big(\langle t_{k,n} \mid y^*_{k,n}\rangle - \langle b_{k,n} \mid b^*_{k,n}\rangle\big)\Big\}
\end{array}
\right. \\
\text{else } \theta_n = 0 \\
\text{for every } i \in I \\
\lfloor\ x_{i,n+1} = x_{i,n} - \theta_n t^*_{i,n} \\
\text{for every } k \in K \\
\lfloor\ y^*_{k,n+1} = y^*_{k,n} - \theta_n t_{k,n}.
\end{array}
\right.
\end{array}
\tag{9.25}
$$

Weak convergence is obtained by applying the principles of Proposition 4.10(ii).

Theorem 9.14 ([136, Theorem 13]) *Consider the setting of Problem 9.5 and Algorithm 9.13, and suppose that the Kuhn–Tucker operator $\mathcal{K}$ of (9.18) has zeros. Then, for every $i \in I$, $(x_{i,n})_{n\in\mathbb{N}}$ converges weakly to a point $x_i \in \mathcal{H}_i$ and, for every $k \in K$, $(y^*_{k,n})_{n\in\mathbb{N}}$ converges weakly to a point $y^*_k \in \mathcal{G}_k$. In addition, $(x_i)_{i\in I}$ solves the primal problem (9.16) and $(y^*_k)_{k\in K}$ solves the dual problem (9.17).*

Remark 9.15 Here are a few comments on algorithm (9.13).

(i) The synchronous implementation is obtained by taking, for every $n \in \mathbb{N}$, every $i \in I_n$, and every $k \in K_n$, $\pi_i(n) = \omega_k(n) = n$.

(ii) We recover [9, Theorem 4.3] (and in particular Proposition 9.4 when $m = p = 1$) in the special case when the implementation is synchronous, and at every iteration n, every operator is used (i.e., $I_n = I$ and $K_n = K$), with $\gamma_{i,n} = \gamma_n$ for every $i \in I$ and $\sigma_{k,n} = \sigma_n$ for every $k \in K$.

(iii) The specialization of Theorem 9.14 to the minimization setting of Example 9.9 is obtained by replacing each $J_{\gamma_{i,\pi_i(n)} A_i}$ with $\operatorname{prox}_{\gamma_{i,\pi_i(n)} f_i}$ and each $J_{\sigma_{k,\omega_k(n)} B_k}$ with $\operatorname{prox}_{\sigma_{k,\omega_k(n)} g_k}$. Numerical experiments are presented in [99] in the context of signal recovery and machine learning, and in [183] in the context of stochastic programming.

(iv) For the strongly convergent variant of Theorem 9.14 based on Proposition 4.11, see [136, Theorem 15].

(v) When $m = 1$ and $A = 0$, a variant that takes into account the fact that some of the operators $(B_k)_{k \in K}$ may be monotone and Lipschitzian, and which activate them via Euler steps is presented in [237] (see also [235]).

10 Block-iterative saddle projective splitting

10.1 Preview

In all the algorithms discussed so far, each monotone operator has one of three properties: it is set-valued, single-valued and cocoercive, or single-valued and Lipschitzian. In addition, at each iteration, a set-valued operator is used once via its resolvents, a cocoercive operator once via a Euler step, and a Lipschitzian operator twice via Euler steps. This is particularly the case in the forward-backward-half-forward algorithm of Section 8.5, the objective of which is to find a zero of

$$M = A + C + Q, \quad \text{where} \quad \begin{cases} A\colon \mathcal{H} \to 2^{\mathcal{H}} & \text{is maximally monotone} \\ C\colon \mathcal{H} \to \mathcal{H} & \text{is cocoercive} \\ Q\colon \mathcal{H} \to \mathcal{H} & \text{is monotone and Lipschitzian.} \end{cases} \tag{10.1}$$

On the other hand, the Kuhn–Tucker projective splitting techniques of Section 9 activate all the operators via their resolvents (exceptions were noted in Remarks 9.3(iii) and 9.15(v), but they concern special cases of Problem 9.5). Furthermore, they are not designed to handle problems such as (7.37) or (8.109), which incorporate parallel sums.

In this section, following [97], we unify all the problem formulations encountered in Sections 5–9 by including parallel sums in the system of monotone inclusions of Problem 9.5, and decomposing each operator in the resulting problem as in (10.1). In addition, nonlinear coupling operators $(R_i)_{i \in I}$ are incorporated.

Problem 10.1 Let $(\mathcal{H}_i)_{i\in I}$ and $(\mathcal{G}_k)_{k\in K}$ be finite families of real Hilbert spaces, and set

$$\boldsymbol{\mathcal{H}}=\bigoplus_{i\in I}\mathcal{H}_i \quad\text{and}\quad \boldsymbol{\mathcal{G}}=\bigoplus_{k\in K}\mathcal{G}_k. \tag{10.2}$$

For every $i\in I$ and every $k\in K$, suppose that the following are satisfied:

[a] $A_i\colon\mathcal{H}_i\to 2^{\mathcal{H}_i}$ is maximally monotone, $C_i\colon\mathcal{H}_i\to\mathcal{H}_i$ is cocoercive with constant $\alpha_i^c\in\,]0,+\infty[$, $Q_i\colon\mathcal{H}_i\to\mathcal{H}_i$ is monotone and Lipschitzian with constant $\alpha_i^\ell\in[0,+\infty[$, and $R_i\colon\boldsymbol{\mathcal{H}}\to\mathcal{H}_i$.

[b] $B_k^m\colon\mathcal{G}_k\to 2^{\mathcal{G}_k}$ is maximally monotone, $B_k^c\colon\mathcal{G}_k\to\mathcal{G}_k$ is cocoercive with constant $\beta_k^c\in\,]0,+\infty[$, and $B_k^\ell\colon\mathcal{G}_k\to\mathcal{G}_k$ is monotone and Lipschitzian with constant $\beta_k^\ell\in[0,+\infty[$.

[c] $D_k^m\colon\mathcal{G}_k\to 2^{\mathcal{G}_k}$ is maximally monotone, $D_k^c\colon\mathcal{G}_k\to\mathcal{G}_k$ is cocoercive with constant $\delta_k^c\in\,]0,+\infty[$, and $D_k^\ell\colon\mathcal{G}_k\to\mathcal{G}_k$ is monotone and Lipschitzian with constant $\delta_k^\ell\in[0,+\infty[$.

[d] $L_{ki}\in\mathcal{B}(\mathcal{H}_i,\mathcal{G}_k)$.

In addition,

[e] $\boldsymbol{R}\colon\boldsymbol{\mathcal{H}}\to\boldsymbol{\mathcal{H}}\colon\boldsymbol{x}\mapsto(R_i\boldsymbol{x})_{i\in I}$ is monotone and Lipschitzian with constant $\chi\in[0,+\infty[$.

The objective is to solve the primal problem

$$\begin{aligned}&\text{find }\ \boldsymbol{x}=(x_i)_{i\in I}\in\boldsymbol{\mathcal{H}}\ \text{ such that }\ (\forall i\in I)\ \ 0\in A_ix_i+C_ix_i+Q_ix_i+R_i\boldsymbol{x}\\ &\qquad+\sum_{k\in K}L_{ki}^*\Bigg(\Big(\big(B_k^m+B_k^c+B_k^\ell\big)\,\square\,\big(D_k^m+D_k^c+D_k^\ell\big)\Big)\Bigg(\sum_{j\in I}L_{kj}x_j\Bigg)\Bigg)\end{aligned} \tag{10.3}$$

and the associated dual problem

$$\begin{aligned}&\text{find }\ \boldsymbol{y}^*=(y_k^*)_{k\in K}\in\boldsymbol{\mathcal{G}}\ \text{ such that }\ (\exists\,\boldsymbol{x}\in\boldsymbol{\mathcal{H}})\\ &\begin{cases}(\forall i\in I)\ \ -\displaystyle\sum_{k\in K}L_{ki}^*y_k^*\in A_ix_i+C_ix_i+Q_ix_i+R_i\boldsymbol{x}\\ (\forall k\in K)\ \ y_k^*\in\Big(\big(B_k^m+B_k^c+B_k^\ell\big)\,\square\,\big(D_k^m+D_k^c+D_k^\ell\big)\Big)\Bigg(\displaystyle\sum_{i\in I}L_{ki}x_i\Bigg).\end{cases}\end{aligned} \tag{10.4}$$

Here is an instance of Problem 10.1 which is not captured by previous monotone inclusion models.

Example 10.2 We consider a game theoretic minimax problem. Let I be a finite set and suppose that $\varnothing \neq J \subset I$. For every $i \in I$, the strategy x_i of player i belongs to a real Hilbert space $\mathcal{H}_i$. A strategy profile is a point

$$\boldsymbol{x} = (x_i)_{i\in I} \in \bigoplus_{i\in I} \mathcal{H}_i, \tag{10.5}$$

and the associated profile of the players other than $i \in I$ is $\boldsymbol{x}_{\smallsetminus i} = (x_j)_{j\in I\smallsetminus\{i\}}$. For every $i \in I$ and every

$$(x_i, \boldsymbol{y}) \in \mathcal{H}_i \oplus \bigoplus_{j\in I} \mathcal{H}_j, \tag{10.6}$$

we set $(x_i; \boldsymbol{y}_{\smallsetminus i}) = (y_1, \ldots, y_{i-1}, x_i, y_{i+1}, \ldots, y_p)$. Now set

$$\boldsymbol{\mathcal{U}} = \bigoplus_{i\in I\smallsetminus J} \mathcal{H}_i, \quad \boldsymbol{\mathcal{V}} = \bigoplus_{j\in J} \mathcal{H}_j, \quad \text{and} \quad \boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{U}} \oplus \boldsymbol{\mathcal{V}}, \tag{10.7}$$

and, for every $i \in I$, let $f_i \in \Gamma_0(\mathcal{H}_i)$. Further, let $\boldsymbol{F}\colon \boldsymbol{\mathcal{H}} \to \mathbb{R}$ be differentiable with a Lipschitzian gradient and such that, for every $\boldsymbol{u} \in \boldsymbol{\mathcal{U}}$ and every $\boldsymbol{v} \in \boldsymbol{\mathcal{V}}$, the functions $-\boldsymbol{F}(\boldsymbol{u}, \cdot)$ and $\boldsymbol{F}(\cdot, \boldsymbol{v})$ are convex. We consider the multivariate minimax problem

$$\underset{\boldsymbol{u}\in\boldsymbol{\mathcal{U}}}{\text{minimize}}\;\; \underset{\boldsymbol{v}\in\boldsymbol{\mathcal{V}}}{\text{maximize}} \;\; \sum_{i\in I\smallsetminus J} f_i(u_i) + \boldsymbol{F}(\boldsymbol{u}, \boldsymbol{v}) - \sum_{j\in J} f_j(v_j). \tag{10.8}$$

Now define

$$(\forall i \in I) \quad \boldsymbol{h}_i\colon \boldsymbol{\mathcal{H}} \to \mathbb{R}\colon (\boldsymbol{u}, \boldsymbol{v}) \mapsto \begin{cases} \boldsymbol{F}(\boldsymbol{u}, \boldsymbol{v}), & \text{if } i \in I \smallsetminus J; \\ -\boldsymbol{F}(\boldsymbol{u}, \boldsymbol{v}), & \text{if } i \in J. \end{cases} \tag{10.9}$$

Then (10.8) can be put in the form

$$\text{find } \boldsymbol{x} \in \boldsymbol{\mathcal{H}} \text{ such that } (\forall i \in I) \;\; x_i \in \operatorname{Argmin} f_i + \boldsymbol{h}_i(\cdot; \boldsymbol{x}_{\smallsetminus i}). \tag{10.10}$$

Since

$$(\forall i \in I)(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}}) \quad \nabla_i \boldsymbol{h}_i(\boldsymbol{x}) = \begin{cases} \nabla_i \boldsymbol{F}(\boldsymbol{x}), & \text{if } i \in I \smallsetminus J; \\ -\nabla_i \boldsymbol{F}(\boldsymbol{x}), & \text{if } i \in J, \end{cases} \tag{10.11}$$

the operator

$$\boldsymbol{R}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon \boldsymbol{x} \mapsto \big(\nabla_i \boldsymbol{h}_i(\boldsymbol{x})\big)_{i\in I} = \Big(\big(\nabla_i \boldsymbol{F}(\boldsymbol{x})\big)_{i\in I\smallsetminus J}, \big(-\nabla_j \boldsymbol{F}(\boldsymbol{x})\big)_{j\in J}\Big) \tag{10.12}$$

is monotone [335, 336] and Lipschitzian. Now, for every $i \in I$, set $A_i = \partial f_i$. Then, by Fermat's rule, (10.10) is equivalent to

$$\text{find } \boldsymbol{x} \in \boldsymbol{\mathcal{H}} \text{ such that } (\forall i \in I) \;\; 0 \in A_i x_i + R_i \boldsymbol{x}, \tag{10.13}$$

which shows that (10.8) is an instantiation of (10.3). Special cases of (10.8) under the above assumptions arise in [17, 148, 226, 297, 342, 371, 396].

Our objective is to solve Problem 10.1 with the same level of flexibility and the same primal-dual convergence guarantees as in Theorem 9.14, i.e., to achieve full splitting of all the operators using an asynchronous block-iterative algorithm without knowledge of the norms of the linear operators or inversion of linear operators. In addition, all the single-valued operators should be activated via Euler steps.

10.2 Saddle operator formulation

The approach adopted in Section 9 to break Problem 9.5 into manageable pieces hinged on the Kuhn–Tucker operator of Lemma 9.7 to obtain the embedding of Framework 1.2. This strategy does not appear to lead to a full splitting of Problem 10.1, as it contains a larger number of operators. We therefore require an embedding in a space $\mathbf{X}$ which is bigger than the primal-dual space $\mathcal{H}_1 \oplus \cdots \mathcal{H}_m \oplus \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_p$ of Theorem 9.14. As discussed in Remark 8.51, saddle operators are defined on a bigger space than Kuhn–Tucker operators (for instance, $\mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G}$ versus $\mathcal{H} \oplus \mathcal{G}$ in (8.138)) and their zeros still provide primal-dual solutions. Following Framework 1.2, as we did in Example 3.23, the methodology of *saddle projective splitting* is to introduce a saddle operator for Problem 10.1. We shall then devise asynchronous block-iterative splitting algorithms based on the geometric principles of Theorems 4.8 and 4.9 to find a zero of it, from which solutions to Problem 10.1 will be extracted. This is outlined in the following lemma.

Lemma 10.3 ([97, Proposition 1]) *Define $\mathcal{H}$ and $\mathcal{G}$ as in (10.2), set $\mathbf{X} = \mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G} \oplus \mathcal{G}$, and define the saddle operator of Problem 10.1 as*

$$
\begin{aligned}
\boldsymbol{\mathcal{S}} \colon \mathbf{X} \to 2^{\mathbf{X}} \colon (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{v}^*) \mapsto \\
&\Bigg(\bigtimes_{i\in I} \bigg(A_i x_i + C_i x_i + Q_i x_i + R_i \boldsymbol{x} + \sum_{k\in K} L_{ki}^* v_k^* \bigg), \\
&\bigtimes_{k\in K} \big(B_k^m y_k + B_k^c y_k + B_k^\ell y_k - v_k^* \big), \\
&\bigtimes_{k\in K} \big(D_k^m z_k + D_k^c z_k + D_k^\ell z_k - v_k^* \big), \\
&\bigtimes_{k\in K} \bigg\{ y_k + z_k - \sum_{i\in I} L_{ki} x_i \bigg\} \Bigg),
\end{aligned}
\tag{10.14}
$$

let $\boldsymbol{Z}$ be the set of solutions to (10.3) and let $\boldsymbol{Z}^*$ be the set of solutions to (10.4). Then the following hold:

(i) $\boldsymbol{\mathcal{S}}$ *is maximally monotone.*

(ii) $\operatorname{zer}\mathcal{S}$ *is closed and convex.*

(iii) *Suppose that* $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{v}^*)\in\operatorname{zer}\mathcal{S}$. *Then* $(\boldsymbol{x},\boldsymbol{v}^*)\in\boldsymbol{Z}\times\boldsymbol{Z}^*$.

(iv) $\boldsymbol{Z}^*\neq\varnothing \Leftrightarrow \operatorname{zer}\mathcal{S}\neq\varnothing \Rightarrow \boldsymbol{Z}\neq\varnothing$.

We thus obtain the following generalization of Example 3.23.

Example 10.4 In the setting of Problem 10.1, set

$$\mathbf{X}=\mathcal{H}\oplus\mathcal{G}\oplus\mathcal{G}\oplus\mathcal{G}, \tag{10.15}$$

let $\mathcal{S}$ be the saddle operator of (10.14), and let

$$\mathcal{T}\colon\mathbf{X}\to\mathcal{H}\colon(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{v}^*)\mapsto\boldsymbol{x}. \tag{10.16}$$

Then it follows from Lemma 10.3(iii) that $(\mathbf{X},\mathcal{S},\mathcal{T})$ is an embedding of (10.3).

Thus, to solve Problem 10.1 via Theorem 4.8, we need a decomposition of the saddle operator (10.14) as $\mathcal{S}=\mathbf{W}+\mathbf{C}$, where $\mathbf{W}\colon\mathbf{X}\to 2^{\mathbf{X}}$ is maximally monotone and $\mathbf{C}\colon\mathbf{X}\to\mathbf{X}$ is α-cocoercive. This will be achieved with

$$\mathbf{C}\colon\mathbf{X}\to\mathbf{X}\colon(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{v}^*)\mapsto\Big(\big(C_ix_i\big)_{i\in I},\big(B_k^{c}y_k\big)_{k\in K},\big(D_k^{c}z_k\big)_{k\in K},\mathbf{0}\Big) \tag{10.17}$$

and $\alpha=\min\{\alpha_i^c,\beta_k^c,\delta_k^c\}_{i\in I,k\in K}$. These considerations lead to the following implementation of (4.23).

Algorithm 10.5 In the setting of Problem 10.1, set

$$\alpha=\min\big\{\alpha_i^c,\beta_k^c,\delta_k^c\big\}_{\substack{i\in I\\ k\in K}}, \tag{10.18}$$

let $\sigma\in\left]1/(4\alpha),+\infty\right[$ and $\varepsilon\in\left]0,1\right[$ be such that

$$\frac{1}{\varepsilon}>\sigma+\max\Big\{\alpha_i^{\ell}+\chi,\beta_k^{\ell},\delta_k^{\ell}\Big\}_{\substack{i\in I\\ k\in K}}, \tag{10.19}$$

and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon,2-\varepsilon]$. For every $i\in I$, let $(\gamma_{i,n})_{n\in\mathbb{N}}$ be a sequence in $\big[\varepsilon,1/(\alpha_i^{\ell}+\chi+\sigma)\big]$ and let $x_{i,0}\in\mathcal{H}_i$. For every $k\in K$, let $(\mu_{k,n})_{n\in\mathbb{N}}$ be a sequence in $\big[\varepsilon,1/(\beta_k^{\ell}+\sigma)\big]$, let $(\rho_{k,n})_{n\in\mathbb{N}}$ be a sequence in $\big[\varepsilon,1/(\delta_k^{\ell}+\sigma)\big]$, let $(\sigma_{k,n})_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon,1/\varepsilon]$, and let $\{y_{k,0},z_{k,0},v_{k,0}^*\}\subset\mathcal{G}_k$. Suppose that Assumptions 9.11 and 9.12 are in force and iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots\\
\quad \text{for every } i \in I_n\\
\quad\quad l^*_{i,n} = Q_i x_{i,\pi_i(n)} + R_i \boldsymbol{x}_{\pi_i(n)} + \sum_{k\in K} L^*_{ki} v^*_{k,\pi_i(n)};\\
\quad\quad a_{i,n} = J_{\gamma_{i,\pi_i(n)} A_i}\big(x_{i,\pi_i(n)} - \gamma_{i,\pi_i(n)}(l^*_{i,n} + C_i x_{i,\pi_i(n)})\big);\\
\quad\quad a^*_{i,n} = \gamma^{-1}_{i,\pi_i(n)}(x_{i,\pi_i(n)} - a_{i,n}) - l^*_{i,n} + Q_i a_{i,n};\\
\quad\quad \xi_{i,n} = \|a_{i,n} - x_{i,\pi_i(n)}\|^2;\\
\quad \text{for every } i \in I \smallsetminus I_n\\
\quad\quad a_{i,n} = a_{i,n-1};\ a^*_{i,n} = a^*_{i,n-1};\ \xi_{i,n} = \xi_{i,n-1};\\
\quad \text{for every } k \in K_n\\
\quad\quad u^*_{k,n} = v^*_{k,\omega_k(n)} - B^\ell_k y_{k,\omega_k(n)};\, w^*_{k,n} = v^*_{k,\omega_k(n)} - D^\ell_k z_{k,\omega_k(n)};\\
\quad\quad b_{k,n} = J_{\mu_{k,\omega_k(n)} B^m_k}\big(y_{k,\omega_k(n)} + \mu_{k,\omega_k(n)}(u^*_{k,n} - B^c_k y_{k,\omega_k(n)})\big);\\
\quad\quad d_{k,n} = J_{\rho_{k,\omega_k(n)} D^m_k}\big(z_{k,\omega_k(n)} + \rho_{k,\omega_k(n)}(w^*_{k,n} - D^c_k z_{k,\omega_k(n)})\big);\\
\quad\quad e^*_{k,n} = \sigma_{k,\omega_k(n)}\big(\sum_{i\in I} L_{ki} x_{i,\omega_k(n)} - y_{k,\omega_k(n)} - z_{k,\omega_k(n)}\big)\\
\quad\quad\qquad\quad + v^*_{k,\omega_k(n)};\\
\quad\quad q^*_{k,n} = \mu^{-1}_{k,\omega_k(n)}(y_{k,\omega_k(n)} - b_{k,n}) + u^*_{k,n} + B^\ell_k b_{k,n} - e^*_{k,n};\\
\quad\quad t^*_{k,n} = \rho^{-1}_{k,\omega_k(n)}(z_{k,\omega_k(n)} - d_{k,n}) + w^*_{k,n} + D^\ell_k d_{k,n} - e^*_{k,n};\\
\quad\quad \eta_{k,n} = \|b_{k,n} - y_{k,\omega_k(n)}\|^2 + \|d_{k,n} - z_{k,\omega_k(n)}\|^2;\\
\quad\quad e_{k,n} = b_{k,n} + d_{k,n} - \sum_{i\in I} L_{ki} a_{i,n};\\
\quad \text{for every } k \in K \smallsetminus K_n\\
\quad\quad b_{k,n} = b_{k,n-1};\ d_{k,n} = d_{k,n-1};\ e^*_{k,n} = e^*_{k,n-1};\\
\quad\quad q^*_{k,n} = q^*_{k,n-1};\ t^*_{k,n} = t^*_{k,n-1};\ \eta_{k,n} = \eta_{k,n-1};\\
\quad\quad e_{k,n} = b_{k,n} + d_{k,n} - \sum_{i\in I} L_{ki} a_{i,n};\\
\quad \text{for every } i \in I\\
\quad\quad p^*_{i,n} = a^*_{i,n} + R_i \boldsymbol{a}_n + \sum_{k\in K} L^*_{ki} e^*_{k,n};\\
\quad \Delta_n = -(4\alpha)^{-1}\big(\sum_{i\in I}\xi_{i,n} + \sum_{k\in K}\eta_{k,n}\big) + \sum_{i\in I}\langle x_{i,n} - a_{i,n} \mid p^*_{i,n}\rangle\\
\quad\qquad + \sum_{k\in K}\big(\langle y_{k,n} - b_{k,n} \mid q^*_{k,n}\rangle + \langle z_{k,n} - d_{k,n} \mid t^*_{k,n}\rangle\\
\quad\qquad\qquad + \langle e_{k,n} \mid v^*_{k,n} - e^*_{k,n}\rangle\big);\\
\quad \text{if } \Delta_n > 0\\
\quad\quad \theta_n = \lambda_n\Delta_n/\big(\sum_{i\in I}\|p^*_{i,n}\|^2 + \sum_{k\in K}\big(\|q^*_{k,n}\|^2 + \|t^*_{k,n}\|^2 + \|e_{k,n}\|^2\big)\big);\\
\quad\quad \text{for every } i \in I\\
\quad\quad\quad x_{i,n+1} = x_{i,n} - \theta_n p^*_{i,n};\\
\quad\quad \text{for every } k \in K\\
\quad\quad\quad y_{k,n+1} = y_{k,n} - \theta_n q^*_{k,n};\ z_{k,n+1} = z_{k,n} - \theta_n t^*_{k,n};\\
\quad\quad\quad v^*_{k,n+1} = v^*_{k,n} - \theta_n e_{k,n};\\
\quad \text{else}\\
\quad\quad \text{for every } i \in I\\
\quad\quad\quad x_{i,n+1} = x_{i,n};\\
\quad\quad \text{for every } k \in K\\
\quad\quad\quad y_{k,n+1} = y_{k,n};\ z_{k,n+1} = z_{k,n};\ v^*_{k,n+1} = v^*_{k,n}.
\end{array}
\tag{10.20}
$$

10.3 Convergence

The convergence properties of Algorithm 10.5 are laid out in the following theorem.

Theorem 10.6 ([97, Theorem 1(iv)]) *Consider the setting of Problem 10.1 and Algorithm 10.5, and suppose that the saddle operator $\mathcal{S}$ of (10.14) has zeros. Then, for every $i \in I$, $(x_{i,n})_{n\in\mathbb{N}}$ converges weakly to a point $x_i \in \mathcal{H}_i$ and, for every $k \in K$, $(v_{k,n}^*)_{n\in\mathbb{N}}$ converges weakly to a point $v_k^* \in \mathcal{G}_k$. In addition, $(x_i)_{i\in I}$ solves the primal problem (10.3) and $(v_k^*)_{k\in K}$ solves the dual problem (10.4).*

Remark 10.7 The strongly convergent variant of Theorem 10.6 based on Theorem 4.9 is proposed in [97, Theorem 2(iv)].

Remark 10.8 A fact that has not be appreciated previously is that Theorem 10.6 contains as special cases various weak convergence results of Sections 7–8. Thus, suppose that

$$I = K = \{1\}, \;\; R_1 = 0, \;\text{ and }\; L_{11} = 0. \tag{10.21}$$

Then Problem 10.1 reduces to finding a zero of $A_1+C_1+Q_1$ (see (8.120)), (10.20) reduces to the forward-backward-half-forward algorithm (8.121), and Theorem 10.6 reduces to Proposition 8.48. This covers both the forward-backward-forward algorithm (7.2) for $C_1 = 0$ (Theorem 7.1) and the unrelaxed forward-backward algorithm (8.5) for $Q_1 = 0$ (Theorem 8.1). In a similar fashion, we can recover the multivariate forward-backward-forward algorithm of [130] by choosing

$$(\forall i \in I)(\forall k \in K) \;\; C_i = R_i = 0 \;\text{ and }\; B_k^c = B_k^\ell = D_k^c = D_k^\ell = 0. \tag{10.22}$$

Going back to the simple inclusion problem (8.120), Theorem 10.6 offers several other possibilities, for instance by implementing it with

$$\begin{aligned} &I = K = \{1\}, A_1 = A, \;\; R_1 = C_1 = Q_1 = 0, \;\; L_{11} = \mathrm{Id}, \\ &\qquad B_1^m = 0\,, B_1^c = C, \;\; B_1^\ell = Q, \;\text{ and } D_1^m = D_1^c = D_1^\ell = \{0\}^{-1}. \end{aligned} \tag{10.23}$$

As mentioned earlier, Problem 10.1 encompasses all the problems discussed earlier. Theorem 10.6 can therefore be used to provide alternative algorithms to solve them in an asynchronous and block-iterative manner, and with operator-dependent proximal parameters (these features are absent from the algorithms of Sections 5–8). Here is an example.

Example 10.9 In Problem 10.1, suppose that

$$\begin{aligned} &I = \{1\}, \;\; K = \{1, \ldots, p\}, \;\; A_1 = A, C_1 = R_1 = 0, Q_1 = Q, \;\text{ and } (\forall k \in K) \\ &L_{k1} = L_k, B_k^m = B_k, B_k^c = B_k^\ell = 0, D_k^m = D_k, \text{and } D_k^c = D_k^\ell = 0. \end{aligned} \tag{10.24}$$

Then we obtain the primal-dual inclusions (7.37)–(7.38) of Proposition 7.10, and Theorem 10.6 furnishes a flexible alternative to Proposition 7.10 which, in addition, places no restriction on the operators $(D_k)_{k\in K}$, with the algorithm

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots\\
\left\lfloor
\begin{array}{l}
l_n^* = Qx_{\pi(n)} + \sum_{k\in K} L_k^* v_{k,\pi(n)}^*;\\
a_n = J_{\gamma_{\pi(n)}A}\big(x_{\pi(n)} - \gamma_{\pi(n)} l_n^*\big);\\
a_n^* = \gamma_{\pi(n)}^{-1}(x_{\pi(n)} - a_n) - l_n^* + Qa_n;\\
\text{for every } k \in K_n\\
\left\lfloor
\begin{array}{l}
b_{k,n} = J_{\mu_{k,\omega_k(n)}B_k}\big(y_{k,\omega_k(n)} + \mu_{k,\omega_k(n)} v_{k,\omega_k(n)}^*\big);\\
d_{k,n} = J_{\rho_{k,\omega_k(n)}D_k}\big(z_{k,\omega_k(n)} + \rho_{k,\omega_k(n)} v_{k,\omega_k(n)}^*\big);\\
e_{k,n}^* = \sigma_{k,\omega_k(n)}\big(L_k x_{\omega_k(n)} - y_{k,\omega_k(n)} - z_{k,\omega_k(n)}\big) + v_{k,\omega_k(n)}^*;\\
q_{k,n}^* = \mu_{k,\omega_k(n)}^{-1}(y_{k,\omega_k(n)} - b_{k,n}) + v_{k,\omega_k(n)}^* - e_{k,n}^*;\\
t_{k,n}^* = \rho_{k,\omega_k(n)}^{-1}(z_{k,\omega_k(n)} - d_{k,n}) + v_{k,\omega_k(n)}^* - e_{k,n}^*;\\
\eta_{k,n} = \|b_{k,n} - y_{k,\omega_k(n)}\|^2 + \|d_{k,n} - z_{k,\omega_k(n)}\|^2;\\
e_{k,n} = b_{k,n} + d_{k,n} - L_k a_n;
\end{array}
\right.\\
\text{for every } k \in K \smallsetminus K_n\\
\left\lfloor
\begin{array}{l}
b_{k,n} = b_{k,n-1};\ d_{k,n} = d_{k,n-1};\ e_{k,n}^* = e_{k,n-1}^*;\\
q_{k,n}^* = q_{k,n-1}^*;\ t_{k,n}^* = t_{k,n-1}^*;\ \eta_{k,n} = \eta_{k,n-1};\\
e_{k,n} = b_{k,n} + d_{k,n} - L_k a_n;
\end{array}
\right.\\
p_n^* = a_n^* + \sum_{k\in K} L_k^* e_{k,n}^*;\\
\Delta_n = -(4\alpha)^{-1}\big(\|a_n - x_{\pi(n)}\|^2 + \sum_{k\in K}\eta_{k,n}\big) + \langle x_n - a_n \mid p_n^*\rangle\\
\qquad\quad + \sum_{k\in K}\big(\langle y_{k,n} - b_{k,n} \mid q_{k,n}^*\rangle + \langle z_{k,n} - d_{k,n} \mid t_{k,n}^*\rangle\\
\qquad\qquad\qquad + \langle e_{k,n} \mid v_{k,n}^* - e_{k,n}^*\rangle\big);\\
\text{if } \Delta_n > 0\\
\left\lfloor
\begin{array}{l}
\theta_n = \lambda_n\Delta_n/\big(\|p_n^*\|^2 + \sum_{k\in K}\big(\|q_{k,n}^*\|^2 + \|t_{k,n}^*\|^2 + \|e_{k,n}\|^2\big)\big);\\
x_{n+1} = x_n - \theta_n p_n^*;\\
\text{for every } k \in K\\
\left\lfloor
\begin{array}{l}
y_{k,n+1} = y_{k,n} - \theta_n q_{k,n}^*;\ z_{k,n+1} = z_{k,n} - \theta_n t_{k,n}^*;\\
v_{k,n+1}^* = v_{k,n}^* - \theta_n e_{k,n};
\end{array}
\right.
\end{array}
\right.\\
\text{else}\\
\left\lfloor
\begin{array}{l}
x_{n+1} = x_n;\\
\text{for every } k \in K\\
\lfloor\ y_{k,n+1} = y_{k,n};\ z_{k,n+1} = z_{k,n};\ v_{k,n+1}^* = v_{k,n}^*.
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{10.25}
$$

Remark 10.10 In the same vein as Example 10.9, we can solve the primal-dual inclusions (8.109)–(8.110) of Proposition 8.44 via Theorem 10.6 by making the modifications $C_1 = C$ and $Q_1 = 0$ in (10.24).

11 Extensions and variants

The following flowchart summarizes the articulation of the main splitting methods presented in the previous sections (a similar flowchart can be drawn for the chain of strong convergence results starting with the Haugazeau principle of Theorem 4.7, then Theorem 4.9, etc.).

- Cutting plane Fejér principle (Theorem 4.2)
  $\Downarrow$
  - Graph-based cuts (Theorem 4.8)
    - Section 9 (Block-iterative Kuhn–Tucker projective splitting)
    - Section 10 (Block-iterative saddle projective splitting)
    - Warped resolvent splitting (Theorem 4.12)
      $\Downarrow$
      - Section 5 (Proximal point algorithm)
      - Section 6 (Douglas–Rachford splitting)
      - Section 7 (Forward-backward-forward splitting)
      - Section 8 (Forward-backward splitting). (11.1)

This flowchart suggests that any extension or variant of the main theorems of Section 4 (Theorems 4.2, 4.8, and 4.12) will lead to further splitting methods or, at least, different implementations of them. We discuss some of the possible variations on the basic geometric principles we have employed.

The basic operating principle of Theorem 4.2 is Fejér-monotonicity, i.e., its property (i). There are extensions of this notion which preserve the main weak convergence conclusions. For instance the notion of quasi-Fejér monotonicity, introduced in [188] and studied in detail in [126], requires that there exist a summable sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ in $[0,+\infty[$ such that

$$(\forall z \in Z)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - z\|^2 \leqslant \|x_n - z\|^2 + \varepsilon_n. \tag{11.2}$$

It follows from [126, Section 3] that Theorem 4.2 remains valid if, for some sequence $(e_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}} \lambda_n\|e_n\| < +\infty$, we use an approximate projection $p_n = \operatorname{proj}_{H_n} x_n + e_n$ in (4.1) (see also [146] for a stochastic version of this result that allows for random iteration modeling). This summable error framework can be propagated in (11.1) to recover approximate implementation results from [61, 127, 130, 145, 155, 339, 387]. Variable metric quasi-Fejér-monotonicity is an extension of (11.2) described by

$$(\forall z \in Z)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - z\|^2_{U_{n+1}} \leqslant \|x_n - z\|^2_{U_n} + \varepsilon_n, \tag{11.3}$$

where $(U_n)_{n\in\mathbb{N}}$ is a sequence of strongly monotone operators in $\mathcal{B}(\mathcal{H})$ satisfying certain properties [150]. It follows from [150, Theorem 3.3] that the conclusions of Theorem 4.2 remain valid in this setting, which amounts to changing the metric of $\mathcal{H}$ at each iteration. See [119, 151] for applications to forward-backward splitting, [343] for applications to multiplier methods, and [323] for considerations on the choice of the variable metrics. All the results derived from Theorem 4.2 can be revisited in this variable-metric context. Another extension of (11.2) of interest is the multi-step quasi-Fejér-monotonicity notion

$$(\forall z \in Z)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\|^2 \leqslant \sum_{j=0}^{n} \mu_{n,j} \|x_j - x\|^2 + \varepsilon_n \tag{11.4}$$

of [139, Lemma 2.2], where $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ is an array in $[0,+\infty[$ satisfying certain properties. This setting led to deterministic block-iterative implementations of the forward-backward algorithm [139, Proposition 4.9] in the spirit of methods found in [287, 289] in the minimization case.

The hybrid proximal-extragradient/projection methods of [357, 358, 359, 361] revolve around a variant of Proposition 4.10 in which, at iteration n, (m_n, m_n^*) is merely required to be in the graph of a perturbed version of M, which permits us to recover certain iterative methods beyond the proximal point algorithm. See also [367] for more recent work along these lines, where approximate resolvents are used to recover an instance of the forward-backward algorithm.

As is apparent from (11.1), many convergence results we have discussed follow from Theorem 4.12. We now present a perturbed extension of it in which, at iteration n, the warped resolvent is applied at a point $\widetilde{x}_n$ and not necessarily at the current iterate x_n. The special case when $C = 0$, $(q_n)_{n\in\mathbb{N}} = (w_n)_{n\in\mathbb{N}}$, and conditions (ii)(b) and (ii)(c) of Theorem 4.12 are fulfilled appears in [95, Theorem 4.2].

Theorem 11.1 *Let $\alpha \in\,]0,+\infty[$, let $W\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $C\colon \mathcal{H} \to \mathcal{H}$ be α-cocoercive and such that $Z = \operatorname{zer}(W + C) \neq \varnothing$, let $x_0 \in \mathcal{H}$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$. Further, for every $n \in \mathbb{N}$, let $\widetilde{x}_n \in \mathcal{H}$ and let $U_n\colon \mathcal{H} \to \mathcal{H}$ be an operator such that* $\operatorname{ran} U_n \subset \operatorname{ran}(U_n + W + C)$ *and* $U_n + W + C$

is injective. Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
w_n = J_{W+C}^{U_n} \widetilde{x}_n \\
w_n^* = U_n \widetilde{x}_n - U_n w_n - C w_n \\
q_n \in \mathcal{H} \\
t_n^* = w_n^* + C q_n \\
\delta_n = \langle x_n - w_n \mid t_n^* \rangle - \|w_n - q_n\|^2/(4\alpha) \\
d_n = \begin{cases} \dfrac{\delta_n}{\|t_n^*\|^2} t_n^*, & \text{if } \delta_n > 0; \\ 0, & \text{otherwise} \end{cases} \\
x_{n+1} = x_n - \lambda_n d_n.
\end{array}
\right.
\end{array}
\tag{11.5}
$$

Suppose that $\widetilde{x}_n - x_n \to 0$. Then the conclusions of Theorem 4.12 remain valid if the condition $U_n w_n - U_n x_n \to 0$ in (ii)(c) is replaced by $U_n w_n - U_n \widetilde{x}_n \to 0$.

Proof. Adapt the pattern of the proof of Theorem 4.12. □

Remark 11.2 The auxiliary sequence $(\widetilde{x}_n)_{n\in\mathbb{N}}$ in Theorem 11.1 adds considerable breadth to the scope of the algorithm, compared to that of Theorem 4.12. Here are some illustrations of the condition $\widetilde{x}_n - x_n \to 0$, where we assume that $\inf_{n\in\mathbb{N}} \lambda_n > 0$ and $\sup_{n\in\mathbb{N}} \lambda_n < 2$.

(i) At iteration n, $\widetilde{x}_n$ can model an additive perturbation of x_n, say $\widetilde{x}_n = x_n + e_n$. Here, the error sequence $(e_n)_{n\in\mathbb{N}}$ need only satisfy $\|e_n\| \to 0$ and not the usual summability condition $\sum_{n\in\mathbb{N}} \|e_n\| < +\infty$ required in the quasi-Fejérian splitting methods of [61, 126, 127, 130, 145, 387].

(ii) In the spirit of inertial methods [20, 43, 112, 137, 317], let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ and set $(\forall n \in \mathbb{N} \smallsetminus \{0\})$ $\widetilde{x}_n = x_n + \alpha_n(x_n - x_{n-1})$. In these methods, $\alpha_n(x_n - x_{n-1}) \to 0$, which guarantees that $\|\widetilde{x}_n - x_n\| \to 0$, as required.

(iii) More generally, weak convergence results can be derived from Theorem 11.1 for iterations with memory, that is,

$$
(\forall n \in \mathbb{N}) \quad \widetilde{x}_n = \sum_{j=0}^{n} \mu_{n,j} x_j, \quad \text{where}
$$

$$
(\mu_{n,j})_{0\leqslant j\leqslant n} \in \mathbb{R}^{n+1} \quad \text{and} \quad \sum_{j=0}^{n} \mu_{n,j} = 1. \tag{11.6}
$$

Here we have $\widetilde{x}_n - x_n \to 0$ if $(1 - \mu_{n,n})x_n - \sum_{j=0}^{n-1} \mu_{n,j} x_j \to 0$. In the case of standard inertial methods, weak convergence requires more stringent conditions on the weights $(\mu_{n,j})_{n\in\mathbb{N}, 0\leqslant j\leqslant n}$ [137].

(iv) As indicated in (11.1), Theorem 9.14 on the Kuhn–Tucker projective splitting algorithm was derived from Proposition 4.10, hence from Theorem 4.8, and it does not appear possible to derive it from Theorem 4.12. However, as shown in [93, Corollary 4], Theorem 9.14 follows from Theorem 11.1 (implemented with $C = 0$ and $q_n = w_n$) through a suitable choice of the auxiliary sequence $(\widetilde{x}_n)_{n\in\mathbb{N}}$. This last example provides further confirmation of the effectiveness of warped resolvents.

Acknowledgment. The author thanks Minh N. Bùi for his careful proofreading of the paper and his suggestions.

References

[1] F. Acker and M. A. Prestel, Convergence d'un schéma de minimisation alternée, *Ann. Fac. Sci. Toulouse V. Sér. Math.* vol. 2, pp. 1–9, 1980.

[2] S. Adly, A. Hantoute, and B. K. Le, Maximal monotonicity and cyclic monotonicity arising in nonsmooth Lur'e dynamical systems, *J. Math. Anal. Appl.*, vol. 448, pp. 691–706, 2017.

[3] S. Agmon, The relaxation method for linear inequalities, *Canad. J. Math.*, vol. 6, pp. 382–392, 1954.

[4] Ya. Alber and I. Ryazantseva, *Nonlinear Ill-Posed Problems of Monotone Type*. Springer, New York, 2006.

[5] G. Alduncin, Composition duality principles for mixed variational inequalities, *Math. Comput. Modelling*, vol. 41, pp. 639–654, 2005.

[6] G. Alduncin, Multidomain optimal control of variational subpotential mixed evolution inclusions, *Appl. Math. Optim.*, vol. 88, art. 35, 2023.

[7] 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.

[8] M. Alimohammady, M. Ramazannejad, and M. Roohi, Notes on the difference of two monotone operators, *Optim. Lett.*, vol. 8, pp. 81–84, 2014.

[9] A. Alotaibi, P. L. Combettes, and N. Shahzad, Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn–Tucker set, *SIAM J. Optim.*, vol. 24, pp. 2076–2095, 2014.

[10] A. Alotaibi, P. L. Combettes, and N. Shahzad, Best approximation from the Kuhn–Tucker set of composite monotone inclusions, *Numer. Funct. Anal. Optim.*, vol. 36, pp. 1513–1532, 2015.

[11] W. N. Anderson, Jr. and G. E. Trapp, A class of monotone operator functions related to electrical network theory, *Linear Algebra Appl.*, vol. 15, pp. 53–67, 1976.

[12] A. S. Antipin, On a method for convex programs using a symmetrical modification of the Lagrange function, *Èkonom. i Mat. Metody*, vol. 12, pp. 1164–1173, 1976.

[13] K. Aoyama, Y. Kimura, and W. Takahashi, Maximal monotone operators and maximal monotone functions for equilibrium problems, *J. Convex Anal.*, vol. 15, pp. 395–409, 2008.

[14] F. J. Aragón-Artacho, R. I. Boţ, and D. Torregrosa-Belén, A primal-dual splitting algorithm for composite monotone inclusions with minimal lifting, *Numer. Algorithms*, vol. 93, pp. 103–130, 2023.

[15] A. Argyriou, R. Foygel, and N. Srebro, Sparse prediction with the k-support norm, *Proc. Adv. Neural Inform. Process. Syst. Conf.*, vol. 25, pp. 1457–1465, 2012.

[16] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, Alternating proximal algorithms for weakly coupled convex minimization problems. Applications to dynamical games and PDE's, *J. Convex Anal.*, vol. 15, pp. 485–506, 2008.

[17] 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.

[18] H. Attouch, L. M. Briceño–Arias, and P. L. Combettes, A strongly convergent primal-dual method for nonoverlapping domain decomposition, *Numer. Math.*, vol. 133, pp. 433–470, 2016.

[19] H. Attouch, G. Buttazzo, and G. Michaille, *Variational Analysis in Sobolev and BV Spaces*, 2nd ed. SIAM, Philadelphia, PA, 2014.

[20] H. Attouch and A. Cabot, Convergence of a relaxed inertial proximal algorithm for maximally monotone operators, *Math. Program.*, vol. A184, pp. 243–287, 2020.

[21] H. Attouch, A. Cabot, P. Frankel, and J. Peypouquet, Alternating proximal algorithms for linearly constrained variational inequalities: application to domain decomposition for PDE's, *Nonlinear Anal.*, vol. 74, pp. 7455–7473, 2011.

[22] 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.

[23] H. Attouch and M. Théra, A general duality principle for the sum of two operators, *J. Convex Anal.*, vol. 3, pp. 1–24, 1996.

[24] J.-P. Aubin and A. Cellina, *Differential Inclusions: Set-Valued Maps and Viability Theory.* Springer, New York, 1984.

[25] J.-F. Aujol and A. Chambolle, Dual norms and image decomposition models, *Int. J. Comput. Vision*, vol. 63, pp. 85–104, 2005.

[26] 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.

[27] A. B. Bakušinskiĭ and B. T. Polyak, The solution of variational inequalities, *Soviet Math. Dokl.*, vol. 15, pp. 1705–1710, 1974.

[28] S. Banert, R. I. Boţ, and E. R. Csetnek, Fixing and extending some recent results on the ADMM algorithm, *Numer. Algorithms*, vol. 86, pp. 1303–1325, 2021.

[29] V. Barbu, *Nonlinear Differential Equations of Monotone Types in Banach Spaces*. Springer, New York, 2010.

[30] S. Bartz, H. H. Bauschke, S. M. Moffat, and X. Wang, The resolvent average of monotone operators: Dominant and recessive properties, *SIAM J. Optim.*, vol. 26, pp. 602–634, 2016.

[31] 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.

[32] H. H. Bauschke and J. M. Borwein, Dykstra's alternating projection algorithm for two sets, *J. Approx. Theory*, vol. 79, pp. 418–443, 1994.

[33] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.*, vol. 42, pp. 596–636, 2003.

[34] H. H. Bauschke, M. N. Bùi, and X. Wang, On sums and convex combinations of projectors onto convex sets, *J. Approx. Theory*, vol. 242, pp. 31–57, 2019.

[35] H. H. Bauschke and P. L. Combettes, A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces, *Math. Oper. Res.*, vol. 26, pp. 248–264, 2001.

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

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

[38] H. H. Bauschke, P. L. Combettes, and S. Reich, The asymptotic behavior of the composition of two resolvents, *Nonlinear Anal.*, vol. 60, pp. 283–301, 2005.

[39] H. H. Bauschke, F. Deutsch, and H. Hundal, Characterizing arbitrarily slow convergence in the method of alternating projections, *Int. Trans. Oper. Res.*, vol. 16, pp. 413–425, 2009.

[40] H. H. Bauschke, V. R. Koch, and H. M. Phan, Stadium norm and Douglas–Rachford splitting: A new approach to road design optimization, *Oper. Res.*, vol. 64, pp. 201–218, 2016.

[41] 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.

[42] H. H. Bauschke and W. M. Moursi, On the Douglas–Rachford algorithm, *Math. Program.*, vol. A164, pp. 263–284, 2017.

[43] 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.

[44] A. Beck and M. Teboulle, Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems, *IEEE Trans. Image Process.*, vol. 18, pp. 2419–2434, 2009.

[45] A. Beck and M. Teboulle, Gradient-based algorithms with applications to signal recovery problems, in: D. P. Palomar and Y. C. Eldar (eds.), *Convex Optimization in Signal Processing and Communications*, pp. 42–88. Cambridge University Press, Cambridge, UK, 2010.

[46] S. R. Becker and P. L. Combettes, An algorithm for splitting parallel sums of linearly composed monotone operators, with applications to signal recovery, *J. Nonlinear Convex Anal.*, vol. 15, pp. 137–159, 2014.

[47] E. M. Bednarczuk, A. Jezierska, and K. E. Rutkowski, Proximal primal-dual best approximation algorithm with memory, *Comput. Optim. Appl.*, vol. 71, pp. 767–794, 2018.

[48] G. Belgioioso, A. Nedich, and S. Grammatico, Distributed generalized Nash equilibrium seeking in aggregative games on time-varying networks, *IEEE Trans. Automat. Control*, vol. 66, pp. 2061–2075, 2021.

[49] 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.

[50] E. Beltrami, A note regarding abstract operators and passive networks, *Quart. Appl. Math.*, vol. 30, pp. 369–370, 1972.

[51] M. Benning and M. Burger, Modern regularization methods for inverse problems, *Acta Numer.*, vol. 27, pp 1–111, 2018.

[52] C. Berge and A. Ghouila–Houri, *Programmes, Jeux, et Réseaux de Transport*. Dunod, Paris, 1962. English translation: *Programming, Games and Transportation Networks*. Wiley, New York, 1965.

[53] M. Bertero, D. Bindi, P. Boccacci, M. Cattaneo, C. Eva, and V. Lanza, Application of the projected Landweber method to the estimation of the source time function in seismology, *Inverse Problems*, vol. 13, pp. 465–486, 1997.

[54] D. P. Bertsekas, *Network Optimization: Continuous and Discrete Models*. Athena Scientific, Belmont, MA, 1998.

[55] G. Birkhoff and R. S. Varga, Implicit alternating direction methods, *Trans. Amer. Math. Soc.*, vol. 92, pp. 13–24, 1959.

[56] E. Blum and W. Oettli, From optimization and variational inequalities to equilibrium problems, *Math. Student*, vol. 63, pp. 123–145, 1994.

[57] E. Börgens and C. Kanzow, ADMM-Type methods for generalized Nash equilibrium problems in Hilbert spaces, *SIAM J. Optim.*, vol. 31, pp. 377–403, 2021.

[58] J. M. Borwein, Fifty years of maximal monotonicity, *Optim. Lett.*, vol. 4, pp. 473–490, 2010.

[59] R. I. Boţ, *Conjugate Duality in Convex Optimization*. Springer, Berlin, 2010.

[60] R. I. Boţ and E. R. Csetnek, ADMM for monotone operators: Convergence analysis and rates, *Adv. Comput. Math.*, vol. 45, pp. 327–359, 2019.

[61] R. I. Boţ and C. Hendrich, A Douglas–Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators, *SIAM J. Optim.*, vol. 23, pp. 2541–2565, 2013.

[62] 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 Vis.*. vol. 49, pp. 551–568, 2014.

[63] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Machine Learn.*, vol. 3, pp. 1–122, 2010.

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

[65] K. Bredies, E. Chenchene, D. A. Lorenz, and E. Naldi, Degenerate preconditioned proximal point algorithms, *SIAM J. Optim.*, vol. 32, pp. 2376–2401, 2022.

[66] L. M. Brègman, The method of successive projection for finding a common point of convex sets, *Soviet Math. – Dokl.*, vol. 6, pp. 688–692, 1965.

[67] L. M. Bregman, The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming, *USSR Comput. Math. Math. Phys.*, vol. 7, pp. 200–217, 1967.

[68] H. Brézis, Les opérateurs monotones, *Séminaire Choquet – Initiation à l'Analyse*, tome 5, exp. no. 10, pp. 1–33, 1966.

[69] H. Brézis, Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations, in [405], pp. 101–156.

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

[71] H. Brézis and F. Browder, Partial differential equations in the 20th century, *Adv. Math.*, vol. 135, pp. 76–144, 1998.

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

[73] L. M. Briceño–Arias, A Douglas–Rachford splitting method for solving equilibrium problems, *Nonlinear Anal.*, vol. 75, pp. 6053–6059, 2012.

[74] L. M. Briceño–Arias, G. Chierchia, E. Chouzenoux, and J.-C. Pesquet, A random block-coordinate Douglas–Rachford splitting method with low computational complexity for binary logistic regression, *Comput. Optim. Appl.*, vol. 72, pp. 707–726, 2019.

[75] 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.

[76] 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.

[77] L. M. Briceño–Arias and P. L. Combettes, Monotone operator methods for Nash equilibria in non-potential games, in: *Computational and Analytical Mathematics*, (D. Bailey et al., eds.), pp. 143–159. Springer, New York, 2013.

[78] L. M. Briceño–Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik, Proximal algorithms for multicomponent image recovery problems, *J. Math. Imaging Vision*, vol. 41, pp. 3–22, 2011.

[79] L. M. Briceño-Arias and D. Davis, Forward-backward-half forward algorithm for solving monotone inclusions, *SIAM J. Optim.*, vol. 28, pp. 2839–2871, 2018.

[80] L. M. Briceño–Arias, J. Deride, S. López–Rivera, and F. J. Silva, A primal-dual partial inverse algorithm for constrained monotone inclusions: Applications to stochastic programming and mean field games, *Appl. Math. Optim.*, vol. 87, art. 21, 2023.

[81] L. M. Briceño–Arias, D. Kalise, and F. J. Silva, Proximal methods for stationary mean field games with local couplings, *SIAM J. Control Optim.*, vol. 56, pp. 801–836, 2018.

[82] L. M. Briceño–Arias and F. Roldán, Primal-dual splittings as fixed point iterations in the range of linear operators, *J. Global Optim.*, vol. 85, pp. 847–866, 2023.

[83] B. Brogliato, R. Lozano, B. Maschke, and O. Egeland, *Dissipative Systems Analysis and Control – Theory and Applications*, 2nd ed. Springer, New York, 2007.

[84] B. Brogliato and A. Tanwani, Dynamical systems coupled with monotone set-valued operators: Formalisms, applications, well-posedness, and stability, *SIAM Rev.*, vol. 62, pp. 3–129, 2020.

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

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

[87] F. E. Browder, Nonlinear operators and nonlinear equations of evolution in Banach spaces, *Proc. Symp. Pure Math.*, vol. 18, pp. 1–308, 1968/1976.

[88] R. E. Bruck, The iterative solution of the equation $y \in x + Tx$ for a monotone operator T in Hilbert space, *Bull. Amer. Math. Soc.*, vol. 79, pp. 1258–1261, 1973.

[89] R. E. Bruck, A strongly convergent iterative solution of $0 \in U(x)$ for a maximal monotone operator U in Hilbert space, *J. Math. Anal. Appl.*, vol. 48, pp. 114–126, 1974.

[90] R. E. Bruck, An iterative solution of a variational inequality for certain monotone operators in Hilbert space, *Bull. Amer. Math. Soc.*, vol. 81, pp. 890–892, 1975. Corrigendum: vol. 82, p. 353, 1976.

[91] R. C. Buck, *Advanced Calculus*, 1st ed. McGraw-Hill, New York, 1956.

[92] M. N. Bùi, A decomposition method for solving multicommodity network equilibria, *Oper. Res. Lett.*, vol. 50, pp. 40–44, 2022.

[93] M. N. Bùi, Projective splitting as a warped proximal algorithm, *Appl. Math. Optim.*, vol. 85, art. 4, 2022.

[94] M. N. Bùi and P. L. Combettes, The Douglas–Rachford algorithm converges only weakly, *SIAM J. Control Optim.*, vol. 58, pp. 1118–1120, 2020.

[95] M. N. Bùi and P. L. Combettes, Warped proximal iterations for monotone inclusions, *J. Math. Anal. Appl.*, vol. 491, art. 124315, 2020.

[96] M. N. Bùi and P. L. Combettes, Bregman forward-backward operator splitting, *Set-Valued Var. Anal.*, vol. 29, pp. 583–603, 2021.

[97] M. N. Bùi and P. L. Combettes, Multivariate monotone inclusions in saddle form, *Math. Oper. Res.*, vol. 47, pp. 1082–1109, 2022.

[98] M. N. Bùi and P. L. Combettes, Analysis and numerical solution of a modular convex Nash equilibrium problem, *J. Convex Anal.*, vol. 29, pp. 1007–1021, 2022.

[99] M. N. Bùi, P. L. Combettes, and Z. C. Woodstock, Block-activated algorithms for multicomponent fully nonsmooth minimization, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 5428–5432, 2022.

[100] C. Byrne, Y. Censor, A. Gibali, and S. Reich, The split common null point problem, *J. Nonlinear Convex Anal.*, vol. 13, pp. 759–775, 2012.

[101] M. K. Camlibel and J. M. Schumacher, Linear passive systems and maximal monotone mappings, *Math. Program.*, vol. B157, pp. 397–420, 2016.

[102] A. Cauchy, Méthode générale pour la résolution des systèmes d'équations simmultanées, *C. R. Acad. Sci. Paris*, vol. 25, pp. 536–538, 1847.

[103] I. Cederbaum, On optimal operation of communication nets, *J. Franklin Inst.*, vol. 274, pp. 130–141, 1962.

[104] A. Cegielski, *Iterative Methods for Fixed Point Problems in Hilbert Spaces*, Lecture Notes in Math., vol. 2057. Springer, Heidelberg, 2012.

[105] Y. Censor and M. Zaknoon, Algorithms and convergence results of projection methods for inconsistent feasibility problems: A review, *Pure Appl. Funct. Anal.*, vol. 3, pp. 565–586, 2018.

[106] Y. Censor and S. A. Zenios, Proximal minimization algorithm with D-functions, *J. Optim. Theory Appl.*, vol. 73, pp. 451–464, 1992.

[107] T. Chaffey, S. Banert, P. Giselsson, and R. Pates, Circuit analysis using monotone+skew splitting, *Eur. J. Control*, vol. 74, art. 100854, 2023.

[108] T. Chaffey, F. Forni, and R. Sepulchre, Graphical nonlinear system analysis, *IEEE Trans. Automat. Control*, vol. 68, pp. 6067–6081, 2023.

[109] T. Chaffey and R. Sepulchre, Monotone one-port circuits, *IEEE Trans. Autom. Control*, vol. 69, pp. 783–796, 2024.

[110] A. Chambolle, An algorithm for total variation minimization and applications, *J. Math. Imaging Vision*, vol. 20, pp. 89–97, 2004.

[111] A. Chambolle, Total variation minimization and a class of binary MRF model, *Lecture Notes in Comput. Sci.*, vol. 3757, pp 136–152, 2005.

[112] 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.

[113] 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.

[114] A. Chambolle and T. Pock, An introduction to continuous optimization for imaging, *Acta Numer.*, vol. 25, pp. 161–319, 2016.

[115] R. H. Chan, S. Setzer, and G. Steidl, Inpainting by flexible Haar-wavelet shrinkage, *SIAM J. Imaging Sci.*, vol. 1, pp. 273–293, 2008.

[116] S. H. Chan, X. Wang, and O. A. Elgendy, Plug-and-play ADMM for image restoration: Fixed-point convergence and applications, *IEEE Trans. Comput. Imaging*, vol. 3, pp. 84–98, 2017.

[117] 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.

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

[119] G. H.-G. Chen and R. T. Rockafellar, Convergence rates in forward-backward splitting, *SIAM J. Optim.*, vol. 7, pp. 421–444, 1997.

[120] W. Cheney and A. A. Goldstein, Proximity maps for convex sets, *Proc. Amer. Math. Soc.*, vol. 10, pp. 448–450, 1959.

[121] E. W. Cheney and A. A. Goldstein, Newton's method for convex programming and Tchebycheff approximation, *Numer. Math.*, vol. 1, pp. 253–268, 1959.

[122] G. Chierchia, E. Chouzenoux, P. L. Combettes, and J.-C. Pesquet, The proximity operator repository. http://proximity-operator.net/

[123] C. Clason and T. Valkonen, Primal-dual extragradient methods for nonlinear nonsmooth PDE-constrained optimization, *SIAM J. Optim.*, vol. 27, pp. 1314–1339, 2017.

[124] G. Cohen, Nash equilibria: Gradient and decomposition algorithms, *Large Scale Syst.*, vol. 12, pp. 173–184, 1987.

[125] P. L. Combettes, Fejér-monotonicity in convex optimization, in: *Encyclopedia of Optimization,* (C. A. Floudas and P. M. Pardalos, Eds.), vol. 2, Springer-Verlag, New York, 2001, pp. 106–114. (Also available in 2nd ed., pp. 1016–1024, 2009.)

[126] P. L. Combettes, Quasi-Fejérian analysis of some optimization algorithms, in: *Inherently Parallel Algorithms for Feasibility and Optimization*, (D. Butnariu, Y. Censor, and S. Reich, eds.), pp. 115–152. Elsevier, New York, 2001.

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

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

[129] P. L. Combettes, Can one genuinely split $m > 2$ monotone operators? *Workshop on Algorithms and Dynamics for Games and Optimization*, Playa Blanca, Tongoy, Chile, October 14-18, 2013.
https://pcombet.math.ncsu.edu/2013open-pbs1.pdf

[130] P. L. Combettes, Systems of structured monotone inclusions: Duality, algorithms, and applications, *SIAM J. Optim.*, vol. 23, pp. 2420–2447, 2013.

[131] P. L. Combettes, Monotone operator theory in convex optimization, *Math. Program.*, vol. B170, pp. 177–206, 2018.

[132] P. L. Combettes, Resolvent and proximal compositions, *Set-Valued Var. Anal.*, vol. 31, art. 22, 2023.

[133] P. L. Combettes and P. Bondon, Hard-constrained inconsistent signal feasibility problems, *IEEE Trans. Signal Process.*, vol. 47, pp. 2460–2468, 1999.

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

[135] P. L. Combettes, Đinh Dũng, and B. C. Vũ, Proximity for sums of composite functions, *J. Math. Anal. Appl.*, vol. 380, pp. 680–688, 2011.

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

[137] P. L. Combettes and L. E. Glaudin, Quasinonexpansive iterations on the affine hull of orbits: From Mann's mean value algorithm to inertial methods, *SIAM J. Optim.*, vol. 27, pp. 2356–2380, 2017.

[138] P. L. Combettes and L. E. Glaudin, Proximal activation of smooth functions in splitting algorithms for convex image recovery, *SIAM J. Imaging Sci.*, vol. 12, pp. 1905–1935, 2019.

[139] P. L. Combettes and L. E. Glaudin, Solving composite fixed point problems with block updates, *Adv. Nonlinear Anal.*, vol. 10, pp. 1154–1177, 2021.

[140] P. L. Combettes and S. A. Hirstoaga, Equilibrium programming in Hilbert spaces, *J. Nonlinear Convex Anal.*, vol. 6, pp. 117–136, 2005.

[141] P. L. Combettes and C. L. Müller, Perspective maximum likelihood-type estimation via proximal decomposition, *Electron. J. Stat.*, vol. 14, pp. 207–238, 2020.

[142] P. L. Combettes and C. L. Müller, Regression models for compositional data: General log-contrast formulations, proximal optimization, and microbiome data applications, *Stat. Biosciences*, vol. 13, pp. 217–242, 2021.

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

[144] 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.

[145] 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.

[146] P. L. Combettes and J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping, *SIAM J. Optim.*, vol. 25, pp. 1221–1248, 2015.

[147] P. L. Combettes and J.-C. Pesquet, Deep neural network structures solving variational inequalities, *Set-Valued Var. Anal.*, vol. 28, pp. 491–518, 2020.

[148] P. L. Combettes and J.-C. Pesquet, Fixed point strategies in data science, *IEEE Trans. Signal Process.*, vol. 69, pp. 3878–3905, 2021.

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

[150] P. L. Combettes and B. C. Vũ, Variable metric quasi-Fejér monotonicity, *Nonlinear Anal.*, vol. 78, pp. 17–31, 2013.

[151] 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.

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

[153] P. L. Combettes and Z. C. Woodstock, A variational inequality model for the construction of signals from inconsistent nonlinear equations, *SIAM J. Imaging Sci.*, vol. 15, pp. 84–109, 2022.

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

[155] 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.

[156] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi, Proximal splitting algorithms for convex optimization: A tour of recent advances, with new twists, *SIAM Rev.*, vol. 65, pp. 375–435, 2023.

[157] H. B. Curry, The method of steepest descent for non-linear minimization problems, *Quart. Appl. Math.*, vol. 2, pp. 258–261, 1944.

[158] S. Dafermos, Traffic equilibrium and variational inequalities, *Transport. Sci.*, vol. 14, pp. 42–54, 1980.

[159] G. Darboux, Mémoire sur les fonctions discontinues, *Ann. Sci. École Normale Sup.*, Sér. 2, vo. 4, pp. 57–112, 1875.

[160] I. Daubechies, M. Defrise, and C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Comm. Pure Appl. Math.*, vol. 57, pp. 1413–1457, 2004.

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

[162] C. A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*. Academic, New York, 1975.

[163] C. A. Desoer and F. F. Wu, Nonlinear monotone networks, *SIAM J. Appl. Math.*, vol. 26, pp. 315–333, 1974.

[164] N. Dexter, H. Tran, and C. G. Webster, On the strong convergence of forward-backward splitting in reconstructing jointly sparse signals, *Set-Valued Var. Anal.*, vol. 30, pp. 543–557, 2022.

[165] V. Doležal, Feedback systems described by monotone operators, *SIAM J. Control Optim.*, vol. 17, pp. 339–364, 1979.

[166] V. Doležal, *Monotone Operators and Applications in Control and Network Theory*. Elsevier, New York, 1979.

[167] Y. Dong, An LS-free splitting method for composite mappings, *Appl. Math. Lett.*, vol. 18, pp. 843–848, 2005.

[168] Q.-L. Dong, Y. J. Cho, S. He, P. M. Pardalos, and T. M. Rassias, *The Krasnosel'skiĭ–Mann Iterative Method – Recent Progress and Applications*. Springer, New York, 2022.

[169] J. Douglas, On the numerical integration of $\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2 = \partial u/\partial t$ by implicit methods, *J. Soc. Indust. Appl. Math.*, vol. 3, pp. 42–65, 1955.

[170] J. Douglas and H. H. Rachford, On the numerical solution of heat conduction problems in two or three space variables, *Trans. Amer. Math. Soc.*, vol. 82, pp. 421–439, 1956.

[171] R. J. Duffin, Nonlinear networks I, *Bull. Amer. Math. Soc.*, vol. 52, pp. 833–838, 1946.

[172] R. J. Duffin, Nonlinear networks IIa, *Bull. Amer. Math. Soc.*, vol. 53, pp. 963–971, 1947.

[173] R. J. Duffin, Nonlinear networks IIb, *Bull. Amer. Math. Soc.*, vol. 54, pp. 119–127, 1948.

[174] R. L. Dykstra, An algorithm for restricted least squares regression, *J. Amer. Stat. Assoc.*, vol. 78, pp. 837–842, 1983.

[175] B. C. Eaves, Subdivisions from primal and dual cones and polytopes, *Linear Algebra and Its Applications*, vol. 62, pp. 277–285, 1984.

[176] J. Eckstein, Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Math. Oper. Res.*, vol. 18, pp. 202–226, 1993.

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

[178] J. Eckstein, A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers, *J. Optim. Theory Appl.*, vol. 173, pp. 155–182, 2017.

[179] 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.

[180] J. Eckstein and M. C. Ferris, Smooth methods of multipliers for complementarity problems, *Math. Program.*, vol. A86, pp. 65–90, 1999.

[181] J. Eckstein and B. F. Svaiter, A family of projective splitting methods for the sum of two maximal monotone operators, *Math. Program.*, vol. 111, pp. 173–199, 2008.

[182] J. Eckstein and B. F. Svaiter, General projective splitting methods for sums of maximal monotone operators, *SIAM J. Control Optim.*, vol. 48, pp. 787–811, 2009.

[183] J. Eckstein, J.-P. Watson, and D. L. Woodruff, Projective hedging algorithms for multistage stochastic programming, supporting distributed and asynchronous implementation, *Oper. Res.*, published online 2023-07-17.

[184] B. Eicke, Iteration methods for convexly constrained ill-posed problems in Hilbert space, *Numer. Funct. Anal. Optim.*, vol. 13, pp. 413–429, 1992.

[185] I. Ekeland and R. Temam, *Analyse Convexe et Problèmes Variationnels.* Dunod, Paris, 1974. English translation: *Convex Analysis and Variational Problems.* SIAM, Philadelphia, PA, 1999.

[186] I. I. Eremin, Methods of Fejér approximations in convex programming, *Mat. Zametki*, vol. 3, pp. 217–234, 1968.

[187] I. I. Eremin, On the speed of convergence in the method of Fejér approximations, *Mat. Zametki*, vol. 4, pp. 53–62, 1968.

[188] Yu. M. Ermol'ev and A. D. Tuniev, Random Fejér and quasi-Fejér sequences, *Theory of Optimal Solutions – Akad. Nauk Ukrainskoĭ SSR Kiev,* vol. 2, pp. 76–83, 1968.

[189] F. Facchinei, A. Fischer, and V. Piccialli, On generalized Nash games and variational inequalities, *Oper. Res. Lett.*, vol. 35, pp. 159–164, 2007.

[190] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems.* Springer, New York, 2003.

[191] L. Fejér, Über die Lage der Nullstellen von Polynomen, die aus Minimumforderungen gewisser Art entspringen, *Math. Ann.*, vol. 85, pp. 41–48, 1922.

[192] G. Fichera, Sul problema elastostatico di Signorini con ambigue condizioni al contorno, *Atti Accad. Naz. Lincei Rend. Ser. VIII*, vol. 34, pp. 138–142, 1963.

[193] M. A. T. Figueiredo and R. D. Nowak, An EM algorithm for wavelet-based image restoration, *IEEE Trans. Image Process.*, vol. 12, pp. 906–916, 2003.

[194] M. Fortin and R. Glowinski (eds.), *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems.* North-Holland, Amsterdam, 1983.

[195] A. Froda, *Sur la Distribution des Propriétés de Voisinage des Fonctions de Variables Réelles.* Hermann, Paris, 1929.

[196] M. Fukushima, The primal Douglas–Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem, *Math. Program.*, vol. 72, pp. 1–15, 1996.

[197] D. Gabay, Applications of the method of multipliers to variational inequalities, in: [194], pp. 299–331. North-Holland, Amsterdam, 1983.

[198] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite elements approximations, *Comput. Math. Appl.*, vol. 2, pp. 17–40, 1976.

[199] N. Gaffke and R. Mathar, A cyclic projection algorithm via duality, *Metrika*, vol. 36, pp. 29–54, 1989.

[200] S. Gandy, B. Recht, and I. Yamada, Tensor completion and low-n-rank tensor recovery via convex optimization, *Inverse Problems*, vol. 27, art. 025010, 2011.

[201] G. Garrigos, L. Rosasco, and S. Villa, Convergence of the forward-backward algorithm: Beyond the worst-case with the help of geometry, *Math. Program.*, vol. A198, pp. 937–996, 2023.

[202] C. F. Gauss, *Theoria Motus Corporum Coelestium*. Perthes and Besser, Hamburg, 1809.

[203] P. Gautam, D. R. Sahu, A. Dixit, and T. Som, Forward-backward-half forward dynamical systems for monotone inclusion problems with application to v-GNE, *J. Optim. Theory Appl.*, vol. 190, pp. 491–523, 2021.

[204] A. Genel and J. Lindenstrauss, An example concerning fixed points, *Israel J. Math.*, vol. 22, pp. 81–86, 1975.

[205] A. Ghizzetti (ed.), *Theory and Applications of Monotone Operators*, Proceedings of a NATO Advanced Study Institute held in Venice, Italy, June 17–30, 1968. Edizioni Oderisi, Gubbio, 1969.

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

[207] P. Giselsson, Nonlinear forward-backward splitting with projection correction, *SIAM J. Optim.*, vol. 31, pp. 2199–2226, 2021.

[208] 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, *C. R. Acad. Sci. Paris*, vol. A278, 1649–1652, 1974; see also *RAIRO Anal. Numer.*, vol. 9, pp. 41–76, 1975.

[209] R. Glowinski and P. Le Tallec (eds.), *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. SIAM, Philadelphia, PA, 1989.

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

[211] D. Goeleven, *Complementarity and Variational Inequalities in Electronics*. Academic, London, 2017.

[212] M. Goldburg and R. J. Marks II, Signal synthesis in the presence of an inconsistent set of constraints, *IEEE Trans. Circuits Syst.*, vol. 32, pp. 647–663, 1985.

[213] A. A. Goldstein, Convex programming in Hilbert space, *Bull. Amer. Math. Soc.*, vol. 70, pp. 709–710, 1964.

[214] E. G. Gol'shtein and N. V. Tret'yakov, Modified Lagrangians in convex programming and their generalizations, *Math. Program. Studies*, vol. 10, pp. 86–97, 1979.

[215] E. G. Golshtein and N. V. Tretyakov, *Modified Lagrangians and Monotone Maps in Optimization.* Wiley, New York, 1996.

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

[217] M. Golomb, Über Systeme von nichtlinearen Integralgleichungen, *Publ. Math. Univ. Belgrade*, vol. 5, pp. 52–83, 1936.

[218] E. G. Gol'shtein, A general approach to decomposition of optimization systems, *Sov. J. Comput. Syst. Sci.*, vol. 25, pp. 105–114, 1987.

[219] C. W. Groetsch, A note on segmenting Mann iterates, *J. Math. Anal. Appl.*, vol. 40, pp. 369–372, 1972.

[220] L. G. Gubin, B. T. Polyak, and E. V. Raik, The method of projections for finding the common point of convex sets, *Comput. Math. Math. Phys.*, vol. 7, pp. 1–24, 1967.

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

[222] H. Hahn, *Theorie der reellen Funktionen*. Springer, Berlin, 1921.

[223] S. P. Han, A successive projection method, *Math. Program.*, vol. 40, pp. 1–14, 1988.

[224] Y. Haugazeau, Sur la minimisation de formes quadratiques avec contraintes, *C. R. Acad. Sci. Paris*, vol. A264, pp. 322–324, 1967.

[225] Y. Haugazeau, *Sur les Inéquations Variationnelles et la Minimisation de Fonctionnelles Convexes*. Thèse, Université de Paris, Paris, France, 1968.

[226] Y. He and R. D. C. Monteiro, Accelerating block-decomposition first-order methods for solving composite saddle-point and two-player Nash equilibrium problems, *SIAM J. Optim.*, vol. 25, pp. 2182–2211, 2015.

[227] 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.

[228] M. R. Hestenes, Multiplier and gradient methods, *J. Optim. Theory Appl.*, vol. 4, pp. 303–320, 1969.

[229] H. Hu, R. Sotirov, and H. Wolkowicz, Facial reduction for symmetry reduced semidefinite and doubly nonnegative programs, *Math. Program.*, vol. A200, pp. 475–529, 2023.

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

[231] 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.

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

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

[234] P. R. Johnstone and J. Eckstein, Convergence rates for projective splitting, *SIAM J. Optim.*, vol. 29, pp. 1931–1957, 2019.

[235] P. R. Johnstone and J. Eckstein, Projective splitting with forward steps only requires continuity, *Optim. Lett.*, vol. 14, pp. 229–247, 2020.

[236] P. R. Johnstone and J. Eckstein, Single-forward-step projective splitting: Exploiting cocoercivity, *Comput. Optim. Appl.*, vol. 78, pp. 125–166, 2021.

[237] P. R. Johnstone and J. Eckstein, Projective splitting with forward steps, *Math. Program.*, vol. A191, pp. 631–670, 2022.

[238] J. L. Joly and P. J. Laurent, Stability and duality in convex minimization problems, *Rev. Française Informat. Recherche Opérationnelle*, sér. R2, vol. 5, pp. 3–42, 1971.

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

[240] R. I. Kačurovskiĭ, Non-linear monotone operators in Banach spaces, *Russian Math. Surveys*, vol. 23, pp. 117–165, 1968.

[241] T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed. Springer, New York, 1980.

[242] R. B. Kellogg, A nonlinear alternating direction method, *Math. Comp.*, vol. 23, pp. 23–27, 1969.

[243] J. E. Kelley, The cutting-plane method for solving convex programs, *J. SIAM*, vol. 8, pp. 703–712, 1960.

[244] D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications.* Academic, New York, 1980.

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

[246] G. M. Korpelevič, The extragradient method for finding saddle points and other problems, *Èkonom. i Mat. Metody*, vol. 12, pp. 747–756, 1976.

[247] M. A. Krasnosel'skiĭ, Two remarks on the method of successive approximations, *Uspekhi Mat. Nauk*, vol. 10, pp. 123–127, 1955.

[248] A. V. Kryanev, The solution of incorrectly posed problems by methods of successive approximations, *Soviet Math. Dokl.*, vol. 14, pp. 673–676, 1973.

[249] P. Latafat and P. Patrinos, Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators, *Comput. Optim. Appl.*, vol. 68, pp. 57–93, 2017.

[250] P. J. Laurent and B. Martinet, Méthodes duales pour le calcul du minimum d'une fonction convexe sur une intersection de convexes, *Lecture Notes in Math.*, vol. 132, pp. 159–180, 1970.

[251] 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.

[252] A. M. Legendre, *Nouvelles Méthodes pour la Détermination des Orbites des Comètes.* Firmin Didot, Paris, 1805.

[253] 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.

[254] B. Lemaire, Stability of the iteration method for nonexpansive mappings, *Serdica Math. J.*, vol. 22, pp. 331–340, 1996.

[255] B. Lemaire, Which fixed point does the iteration method select?, *Lecture Notes in Econom. and Math. Systems*, vol. 452, pp. 154–167, 1997.

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

[257] 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.

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

[259] J. Lieutaud, Approximations d'opérateurs monotones par des méthodes de splitting, in: [205], pp. 259–264.

[260] J. Lieutaud, *Approximation d'Opérateurs par des Méthodes de Décomposition*. Thèse, Université de Paris, 1969.

[261] S. Lindstrom and B. Sims, Survey: Sixty years of Douglas–Rachford, *J. Aust. Math. Soc.*, vol. 110, pp. 333–370, 2021.

[262] J.-L. Lions (ed.), *Numerical Analysis of Partial Differential Equations*, Lectures given at CIME summer school held in Ispra (Varese), Italy, July 3–11, 1967. Reprint: Springer, New York, 2010.

[263] J.-L. Lions, *Quelques Méthodes de Résolution des Problèmes aux Limites Non Linéaires.* Dunod, Paris, 1969.

[264] P.-L. Lions, Une méthode itérative de résolution d'une inéquation variationnelle, *Israel J. Math.*, vol. 31, pp. 204–208, 1978.

[265] 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.

[266] J. Liu and S. J. Wright, Asynchronous stochastic coordinate descent: Parallelism and convergence properties, *SIAM J. Optim.*, vol. 25, pp. 351–376, 2015.

[267] H. Lu, R. M. Freund, and Yu. Nesterov, Relatively smooth convex optimization by first-order methods, and applications, *SIAM J. Optim.*, vol. 28, pp. 333–354, 2018.

[268] M. P. Machado, On the complexity of the projective splitting and Spingarn's methods for the sum of two maximal monotone operators, *J. Optim. Theory Appl.*, vol. 178, pp. 153–190, 2018.

[269] M. P. Machado and M. R. Sicre, A projective splitting method for monotone inclusions: Iteration-complexity and application to composite optimization, *J. Optim. Theory Appl.*, vol. 198, pp. 552–587, 2023.

[270] Ph. Mahey and D. T. Pham, Partial regularization of the sum of two maximal monotone operators, *RAIRO Modélisation Math. Analyse Numér.*, vol. 27, pp. 375–392, 1993.

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

[272] Yu. Malitsky and M. K. Tam, Resolvent splitting for sums of monotone operators with minimal lifting, *Math. Program.*, vol. A201, pp. 231–262, 2023.

[273] W. R. Mann, Mean value methods in iteration, *Proc. Amer. Math. Soc.*, vol. 4, pp. 506–510, 1953.

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

[275] 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.

[276] J. E. Martínez-Legaz and A. Seeger, A general cone decomposition theory based on efficiency, *Math. Program.*, vol. 65, pp. 1–20, 1994.

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

[278] B. Mercier, *Inéquations Variationnelles de la Mécanique* (Publications Mathématiques d'Orsay, no. 80.01). Université de Paris-XI, Orsay, France, 1980.

[279] C. A. Micchelli, J. M. Morales, and M. Pontil, Regularizers for structured sparsity, *Adv. Comput. Math.*, vol. 38, pp. 455–489, 2013.

[280] W. Millar, Some general theorems for non-linear systems possessing resistance, *London, Edinburgh, Dublin Phil. Mag. J. Sci.*, vol. 42, pp. 1150–1160, 1951.

[281] G. J. Minty, Monotone networks, *Proc. R. Soc. Lond. A*, vol. 57, pp. 194–212, 1960.

[282] G. J. Minty, Solving steady-state nonlinear networks of "monotone" elements, *IRE Trans. Circuit Theory*, vol. 8, pp. 99–104, 1961.

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

[284] 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.

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

[286] G. J. Minty, On some aspects of the theory of monotone operators, in: [205], pp. 67–82.

[287] K. Mishchenko, F. Iutzeler, and J. Malick, A distributed flexible delay-tolerant proximal gradient algorithm, *SIAM J. Optim.*, vol. 30, pp. 933–959, 2020.

[288] T. Mizoguchi and I. Yamada, Hypercomplex tensor completion via convex optimization, *IEEE Trans. Signal Process.*, vol. 67, pp. 4078–4092, 2019.

[289] A. Mokhtari, M. Gürbüzbalaban, and A. Ribeiro, Surpassing gradient descent provably: A cyclic incremental method with linear convergence rate, *SIAM J. Optim.*, vol. 28, pp. 1420–1447, 2018.

[290] J. J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris Sér. A Math.*, vol. 255, pp. 2897–2899, 1962.

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

[292] J. J. Moreau, *Fonctionnelles Convexes* (Séminaire Jean Leray sur les Équations aux Dérivées Partielles, no. 2). Collège de France, Paris, 1966–1967.

[293] U. Mosco, Dual variational inequalities, *J. Math. Anal. Appl.*, vol. 40, pp. 202–206, 1972.

[294] T. S. Motzkin and I. J. Schoenberg, The relaxation method for linear inequalities, *Canad. J. Math.*, vol. 6, pp. 393–404, 1954.

[295] A. Moudafi, On the regularization of the sum of two maximal monotone operators, *Nonlinear Anal.*, vol. 42, pp. 1203–1208, 2000.

[296] A. Moudafi and M. Théra, Proximal and dynamical approaches to equilibrium problems, *Lecture Notes in Econom. and Math. Systems*, vol. 477, pp. 187–201, 1999.

[297] A. Nemirovski, Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems, *SIAM J. Optim.*, vol. 15, pp. 229–251, 2004.

[298] A. Neubauer, Tikhonov-regularization of ill-posed linear operator equations on closed convex sets, *J. Approx. Theory*, vol. 53, pp. 304–320, 1988.

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

[300] D. O'Connor and L. Vandenberghe, Primal-dual decomposition by operator splitting and applications to image deblurring, *SIAM J. Imaging Sci.*, vol. 7, pp. 1724–1754, 2014.

[301] D. E. Oliveira, H. Wolkowicz, and Y. Xu, ADMM for the SDP relaxation of the QAP, *Math. Program. Comput.*, vol. 10, pp. 631–658, 2018.

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

[303] J. M. Papakonstantinou and R. A. Tapia, Origin and evolution of the secant method in one dimension, *Amer. Math. Monthly*, vol. 120, pp. 500–518, 2013.

[304] D. Pascali and S. Sburlan, *Nonlinear Mappings of Monotone Type*. Editura Academiei, Bucuresti, Romania, 1978.

[305] G. B. Passty, Ergodic convergence to a zero of the sum of monotone operators in Hilbert space, *J. Math. Anal. Appl.*, vol. 72, pp. 383–390, 1979.

[306] R. Pathak and M. J. Wainwright, FedSplit: An algorithmic framework for fast federated optimization, *Proc. Adv. Neural Inform. Process. Syst. Conf.*, vol. 33. pp. 7057–7066, 2020.

[307] D. W. Peaceman and H. H. Rachford, The numerical solution of parabolic and elliptic differential equations, *J. Soc. Indust. Appld Math.*, vol. 3, pp. 28–41, 1955.

[308] T. Pennanen, Dualization of generalized equations of maximal monotone type, *SIAM J. Optim.*, vol. 10, pp. 809–835, 2000.

[309] T. Pennanen, A splitting method for composite mappings, *Numer. Funct. Anal. Optim.*, vol. 23, pp. 875–890, 2002.

[310] J.-C. Pesquet and A. Repetti, A class of randomized primal-dual algorithms for distributed optimization, *J. Nonlinear Convex Anal.*, vol. 16, pp. 2453–2490, 2015.

[311] J.-C. Pesquet, A. Repetti, M. Terris, and Y. Wiaux, Learning maximally monotone operators for image recovery, *SIAM J. Imaging Sci.*, vol. 14, pp. 1206–1237, 2021.

[312] W. V. Petryshyn, On the extension and solution of nonlinear operator equations, *Illinois J. Math.*, vol. 10, pp. 255–274, 1966.

[313] D. L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, *J. Assoc. Comput. Mach.*, vol. 9, pp. 84–97, 1962.

[314] R. S. Phillips, Dissipative operators and hyperbolic systems of partial differential equations, *Trans. Amer. Math. Soc.*, vol. 90, pp. 193–254, 1959.

[315] G. Pierra, Éclatement de contraintes en parallèle pour la minimisation d'une forme quadratique, *Lecture Notes in Comput. Sci.*, vol. 41, pp. 200–218, 1976.

[316] G. Pierra, Decomposition through formalization in a product space, *Math. Program.*, vol. 28, pp. 96–115, 1984.

[317] B. T. Polyak, Some methods of speeding up the convergence of iteration methods, *USSR Comput. Math. Math. Phys.*, vol. 4, pp. 1–17, 1964.

[318] L. C. Potter and K. S. Arun, A dual approach to linear inverse problems with convex constraints, *SIAM J. Control Optim.*, vol. 31, pp. 1080–1092, 1993.

[319] M. J. D. Powell, A method for nonlinear constraints in minimization problems, in: *Optimization*, R. Fletcher (ed.), Academic, pp. 283–298, 1969.

[320] X. Qin and N. T. An, Smoothing algorithms for computing the projection onto a Minkowski sum of convex sets, *Comput. Optim. Appl.*, vol. 74, pp. 821–850, 2019.

[321] H. Raguet, A note on the forward-Douglas-Rachford splitting for monotone inclusion and convex optimization, *Optim. Lett.*, vol. 13, pp. 717–740, 2019.

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

[323] H. Raguet and L. Landrieu, Preconditioning of a generalized forward-backward splitting and application to optimization on graphs, *SIAM J. Imaging Sci.*, vol. 8, pp. 2706–2739, 2015.

[324] E. Raik, Fejér type methods in Hilbert space, *Eesti NSV Tead. Akad. Toimetised Füüs.-Mat.*, vol. 16, pp. 286–293, 1967.

[325] E. Raik, A class of iterative methods with Fejér-monotone sequences, *Eesti NSV Tead. Akad. Toimetised Füüs.-Mat.*, vol. 18, pp. 22–26, 1969.

[326] A. Renaud and G. Cohen, An extension of the auxiliary problem principle to nonsymmetric auxiliary operators, *ESAIM Control Optim. Calc. Var.*, vol. 2, pp. 281–306, 1997.

[327] H. J. Reich, *Functional Circuits and Oscillators.* Van Nostrand, New York, 1961.

[328] S. Reich, M. T. Truong, and T. N. H. Mai, The split feasibility problem with multiple output sets in Hilbert spaces, *Optim. Lett.*, vol. 14, pp. 2335–2353, 2020.

[329] S. M. Robinson, A reduction method for variational inequalities, *Math. Program.*, vol. 80, pp. 161–169, 1998.

[330] S. M. Robinson, Composition duality and maximal monotonicity, *Math. Program.*, vol. 85, pp. 1–13, 1999.

[331] S. M. Robinson, Generalized duality in variational analysis, in: N. Hadjisavvas and P. M. Pardalos (eds.), *Advances in Convex Analysis and Global Optimization*. Dordrecht, The Netherlands, Kluwer, 2001, pp. 205–219.

[332] R. T. Rockafellar, Duality and stability in extremum problems involving convex functions, *Pacific J. Math.*, vol. 21, pp. 167–187, 1967.

[333] R. T. Rockafellar, Convex functions and duality in optimization problems and dynamics, in: *Mathematical Systems Theory and Economics I*, (H. W. Kuhn and G. P. Szegö, eds.), pp. 117–141. Springer, New York, 1969.

[334] R. T. Rockafellar, *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[335] 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.

[336] 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, Amsterdam, 1971.

[337] R. T. Rockafellar, The multiplier method of Hestenes and Powell applied to convex programming, *J. Optim. Theory Appl.*, vol. 12, pp. 555–562, 1973.

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

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

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

[341] R. T. Rockafellar, *Network Flows and Monotropic Optimization*. Wiley, New York, 1984.

[342] R. T. Rockafellar, Monotone relations and network equilibrium, in: *Variational Inequalities and Network Equilibrium Problems*, (F. Giannessi and A. Maugeri, eds.), pp. 271–288. Plenum Press, New York, 1995.

[343] R. T. Rockafellar, Generalizations of the proximal method of multipliers in convex optimization, *Comput. Optim. Appl.*, vol. 87, pp. 219–247, 2024.

[344] R. T. Rockafellar and J. Sun, Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging, *Math. Program.*, vol. A174, pp. 453–471, 2019.

[345] 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.

[346] E. K. Ryu, Uniqueness of DRS as the 2 operator resolvent-splitting and impossibility of 3 operator resolvent-splitting, *Math. Program.*, vol. A182, pp. 233–273, 2020.

[347] E. K. Ryu, Y. Liu, and W. Yin, Douglas–Rachford splitting and ADMM for pathological convex optimization, *Comput. Optim. Appl.*, vol. 74, pp. 747–778, 2019.

[348] E. K. Ryu, A. B. Taylor, C. Bergeling, and P. Giselsson, Operator splitting performance estimation: Tight contraction factors and optimal parameter selection, *SIAM J. Optim.*, vol. 30, pp. 2251–2271, 2020.

[349] E. K. Ryu and B. C. Vũ, Finding the forward-Douglas-Rachford-forward method, *J. Optim. Theory Appl.*, vol. 184, pp. 858–876, 2020.

[350] S. Salzo and S. Villa, Parallel random block-coordinate forward-backward algorithm: A unified convergence analysis, *Math. Program.*, vol. A193, pp. 225–269, 2022.

[351] A. Seeger, Alternating projection and decomposition with respect to two convex sets, *Math. Japon.*, vol. 47, pp. 273–280, 1998.

[352] R. Shefi and M. Teboulle, Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization, *SIAM J. Optim.*, vol. 24, pp. 269–297, 2014.

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

[354] M. Sibony, Méthodes itératives pour les équations et inéquations aux dérivées partielles non linéaires de type monotone, *Calcolo*, vol. 7, pp. 65–183, 1970.

[355] M. R. Sicre, On the complexity of a hybrid proximal extragradient projective method for solving monotone inclusion problems, *Comput. Optim. Appl.*, vol. 76, pp. 991–1019, 2020.

[356] S. Singh, G. Weiss, and M. Tucsnak, A class of incrementally scattering-passive nonlinear systems, *Automatica*, vol. 142, art. 110369, 2022.

[357] M. V. Solodov, A class of decomposition methods for convex optimization and monotone variational inclusions via the hybrid inexact proximal point framework, *Optim. Methods Softw.*, vol. 19, pp. 557–575, 2004.

[358] M. V. Solodov and B. F. Svaiter, A hybrid projection-proximal point algorithm, *J. Convex Anal.*, vol. 6, pp. 59–70, 1999.

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

[360] M. V. Solodov and B. F. Svaiter, Forcing strong convergence of proximal point iterations in a Hilbert space, *Math. Program.*, vol. A87, pp. 189–202, 2000.

[361] M. V. Solodov and B. F. Svaiter, A unified framework for some inexact proximal point algorithms, *Numer. Funct. Anal. Optim.*, vol. 22, pp. 1013–1035, 2001.

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

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

[364] 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.

[365] G. Steidl and T. Teuber, Removing multiplicative noise by Douglas–Rachford splitting methods, *J. Math. Imaging Vis.*, vol. 36, pp. 168–184, 2010.

[366] B. F. Svaiter, On weak convergence of the Douglas–Rachford method, *SIAM J. Control Optim.*, vol. 49, pp. 280–287, 2011.

[367] B. F. Svaiter, A class of Fejér convergent algorithms, approximate resolvents and the hybrid proximal-extragradient method, *J. Optim. Theory Appl.*, vol. 162, pp. 133–153, 2014.

[368] M. Teboulle, Entropic proximal mappings with applications to nonlinear programming, *Math. Oper. Res.*, vol. 17, pp. 670–690, 1992.

[369] M. Teboulle, A simplified view of first order methods for optimization, *Math. Program.*, vol. B170, pp. 67–96, 2018.

[370] B. D. H. Tellegen, The gyrator, a new electric network element, *Philips Res. Rept.*, vol. 3, pp. 81–101, 1948.

[371] K. K. Thekumparampil, P. Jain, P. Netrapalli, and S. Oh, Efficient algorithms for smooth minimax optimization, *Proc. Adv. Neural Inform. Process. Syst. Conf.*, vol. 32, 2019.

[372] C. Traoré, S. Salzo, and S. Villa, Convergence of an asynchronous block-coordinate forward-backward algorithm for convex composite optimization, *Comput. Optim. Appl.*, vol. 86, pp. 303–344, 2023.

[373] P. Tseng, Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming, *Math. Program.*, vol. 48, pp. 249–263, 1990.

[374] 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.

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

[376] M. M. Vaĭnberg, *Variatsionnye Metody Issledovaniya Nelineinykh Operatorov*. Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow, 1956. English translation: *Variational Methods for the Study of Non-Linear Operators*. Holden-Day, San Francisco, 1964.

[377] M. M. Vaĭnberg, New theorems for non-linear operators and equations, *Dokl. Akad. Nauk SSSR*, vol. 129, pp. 1199–1202, 1959.

[378] M. M. Vaĭnberg, On the convergence of the method of steepest descent for nonlinear equations, *Dokl. Akad. Nauk SSSR*, vol. 130. pp. 9–12, 1960.

[379] M. M. Vaĭnberg, On the convergence of the process of steepest descent for nonlinear equations, *Sibirsk. Mat. Zh.*, vol. 2. pp. 201–220, 1961.

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

[381] 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.

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

[383] A. F. Veinott, The supporting hyperplane method for unimodal programming, *Oper. Res.*, vol. 15, pp. 147–152, 1967.

[384] L. A. Vese and S. J. Osher, Image denoising and decomposition with total variation minimization and oscillatory functions, *J. Math. Imaging Vision*, vol. 20, pp. 7–18, 2004.

[385] S. Villa, L. Rosasco, S. Mosci, and A. Verri, Proximal methods for the latent group lasso penalty, *Comput. Optim. Appl.*, vol. 58, pp. 381–407, 2014.

[386] M. I. Vishik, Boundary-value problems for quasilinear strongly elliptic systems of equations having divergence form, *Soviet Math. Dokl.*, vol. 2, pp. 643–647, 1961.

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

[388] X. Wang, J. Zhang, and W. Zhang, The distance between convex sets with Minkowski sum structure: Application to collision detection, *Comput. Optim. Appl.*, vol. 77, pp. 465–490, 2020.

[389] E. Winston and J. Z. Kolter, Monotone operator equilibrium networks, *Proc. Conf. Adv. Neural Inform. Process. Syst.*, vol. 33, pp. 10718–10728, 2020.

[390] J.-H. Won, J. Xu, and K. Lange, Projection onto Minkowski sums with application to constrained learning, *Proc. 36th Int. Conf. Machine Learn.*, pp. 3642–3651, 2019.

[391] S. J. Wright and B. Recht, *Optimization for Data Analysis*. Cambridge University Press, Cambridge, UK, 2022.

[392] F. Xue, A generalized forward-backward splitting operator: Degenerate analysis and applications, *Comput. Appl. Math.*, vol. 42, art. 9, 2023.

[393] F. Xue, Equivalent resolvents of Douglas–Rachford splitting and other operator splitting algorithms: A unified degenerate proximal point analysis, *Optimization*, published online 2023-07-03.

[394] X. Yan and J. Bien, Rare feature selection in high dimensions, *J. Amer. Statist. Assoc.*, vol. 116, pp. 887–900, 2021.

[395] P. Yi and S. Ching, Synthesis of recurrent neural dynamics for monotone inclusion with application to Bayesian inference, *Neural Networks*, vol. 131, pp. 231–241, 2020.

[396] T. Yoon and E. K. Ryu, Accelerated algorithms for smooth convex-concave minimax problems with $O(1/k^2)$ rate on squared gradient norm, *Proc. 38th Int. Conf. Machine Learn.*, pp. 12098–12109, 2021.

[397] D. C. Youla, Mathematical theory of image restoration by the method of convex projections, in: H. Stark (ed.) *Image Recovery: Theory and Application*, pp. 29–77. Academic Press, San Diego, CA, 1987.

[398] Y. Yu, J. Peng, X. Han, and A. Cui, A primal Douglas–Rachford splitting method for the constrained minimization problem in compressive sensing, *Circuits Syst. Signal Process.*, vol. 36, pp. 4022–4049, 2017.

[399] G. Zames, On the input-output stability of time-varying nonlinear feedback systems part I: Conditions derived using concepts of loop gain, conicity, and positivity, *IEEE Trans. Autom. Control*, vol. 11, pp. 228–238, 1966.

[400] G. Zames, On the input-output stability of time-varying nonlinear feedback systems part II: Conditions involving circles in the frequency plane and sector nonlinearities, *IEEE Trans. Autom. Control*, vol. 11, pp. 465–476, 1966.

[401] G. Zames and P. L. Falb, Stability conditions for systems with monotone and slope-restricted nonlinearities, *SIAM J. Control*, vol. 6, pp. 89–108, 1968.

[402] W. I. Zangwill, *Nonlinear Programming – A Unified Approach.* Prentice-Hall, Englewood Cliffs, NJ, 1969.

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

[404] E. H. Zarantonello, The closure of the numerical range contains the spectrum, *Bull. Amer. Math. Soc.*, vol. 70, pp. 781–787, 1964.

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

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