ABSTRACT

BÙI, MINH NHỰT. The Warped Resolvent of a Set-Valued Operator: Theory and Applications. (Under the direction of Patrick L. Combettes.)

This dissertation develops novel nonlinear analysis tools and methodologies to advance the field of monotone operator theory and its applications. First, we show that the Douglas–Rachford algorithm and Spingarn's method of partial inverses can fail to converge strongly. Second, we introduce the notion of a warped resolvent as an extension of the classical resolvent, study its properties, and propose weakly and strongly convergent warped proximal iteration principles. This framework unifies and extends several so-far unrelated ones, such as projective splitting and Tseng's forward-backward-forward method. In addition, the warped resolvent framework is shown to be an effective device to produce new and flexible splitting methods for complex monotone inclusion problems. Next, we introduce a Bregman forward-backward method for solving monotone inclusions, establish its weak convergence, and show that it captures and extends several iterative methods. Another contribution is to propose a new saddle presentation to study and solve highly structured systems of monotone inclusions. This leads to highly flexible asynchronous block-iterative algorithms. Applications of warped resolvents in the context of variational inequalities, convex optimization, Nash equilibrium, and network flows are discussed. Finally, we analyze the merits and performance of block-activated algorithms for solving multicomponent fully nonsmooth minimization problems with applications to machine learning and image recovery.

This dissertation was supported by the National Science Foundation under grants CCF-1715671 and DMS-1818946.

The Warped Resolvent of a Set-Valued Operator: Theory and Applications

by
Minh Nhựt Bùi

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Mathematics

Raleigh, North Carolina

2021

APPROVED BY:

Kazufumi Ito

Ryan Murray

Tien Khai Nguyen

Reha Uzsoy

Patrick L. Combettes
Chair of Advisory Committee

DEDICATION

Dành tặng gia đình yêu thương: Mẹ Hương, Cha Hóa, em gái Nguyệt Minh.
Cảm ơn gia đình, vì tất cả.

BIOGRAPHY

Minh Nhựt Bùi received his Bachelor of Science from Hồ Chí Minh City University of Education in 2016 and his Master of Science in Mathematics from the University of British Columbia in July 2018. He became a Ph.D. student in the Department of Mathematics at North Carolina State University under the supervision of Professor Patrick L. Combettes in August 2018.

ACKNOWLEDGMENTS

To Professor Combettes: I would like to express my most sincere gratitude for everything you have done for me. Your professionalism, high research standards, broad knowledge, extraordinary mentorship, and encouragement have shaped me into who I am today as a researcher. I am thankful for the opportunity to be your doctoral student at NC State, and I deeply appreciate the countless hours you have spent to work with me and to prepare me for the academic career ahead. My three years of doing research at NC State were absolutely wonderful and I could not have dreamed of any better experience. From the bottom of my heart, thank you!

I thank Prof. Ito, Prof. Murray, Prof. Nguyễn, and Prof. Uzsoy for agreeing to serve on my advisory committee.

To Thế Cường, Minh Đức, Tiến Kha, Thanh Minh, Thanh Nhàn, Văn Thành, Tú Trinh, and Zev: Thank you for being parts of my graduate school journey. I am grateful for the experience and memories that we have had together.

Thương gửi gia đình của con: Cảm ơn cả nhà vì đã luôn là điểm tựa vững chắc cho con trên cả hành trình dài này. Con có được ngày hôm nay đều nhờ có sự ủng hộ và động viên của Mẹ và Cha ngay từ những giây phút đầu tiên. Cảm ơn gia đình vì đã luôn tin tưởng và sát cánh cùng con.

This dissertation was supported by the National Science Foundation under grants CCF-1715671 and DMS-1818946.

TABLE OF CONTENTS

LIST OF FIGURES

NOTATION AND DEFINITIONS

The following notation is used throughout this dissertation.

General notation

- $\mathcal{H}$, $\mathcal{H}_i$, $\mathcal{G}$, $\mathcal{G}_k$, $\mathcal{K}$, $\mathcal{K}_i$: Real Hilbert spaces.
- $\langle \cdot \mid \cdot \rangle$: Scalar product of a real Hilbert space.
- $\|\cdot\|$: Norm.
- $\bigoplus_{i\in I}\mathcal{H}_i$: Hilbert direct sum of a family $(\mathcal{H}_i)_{i\in I}$ of real Hilbert spaces, that is,

$$\bigoplus_{i\in I}\mathcal{H}_i = \left\{ \boldsymbol{x} = (x_i)_{i\in I} \in \bigtimes_{i\in I}\mathcal{H}_i \;\middle|\; \sum_{i\in I}\|x_i\|_i^2 < +\infty \right\}$$

  equipped with the scalar product $(\boldsymbol{x},\boldsymbol{y}) \mapsto \sum_{i\in I}\langle x_i \mid y_i\rangle_i$.
- Id: Identity operator.
- $2^{\mathcal{H}}$: Power set of $\mathcal{H}$.
- L^*: Adjoint of a bounded linear operator $L\colon \mathcal{H}\to\mathcal{G}$.
- $\to$: Strong convergence.
- $\rightharpoonup$: Weak convergence.
- $\|L\| = \sup\big\{\|Lx\| \mid x\in\mathcal{H},\ \|x\|\leqslant 1\big\}$: Norm of a bounded linear operator $L\colon\mathcal{H}\to\mathcal{G}$.

Notation and definitions relative to a function $f\colon\mathcal{H}\to[-\infty,+\infty]$

- $\operatorname{dom} f = \big\{x\in\mathcal{H} \mid f(x) < +\infty\big\}$: Domain of f.
- $\operatorname{epi} f = \big\{(x,\xi)\in\mathcal{H}\times\mathbb{R} \mid f(x)\leqslant\xi\big\}$: Epigraph of f.
- f is proper if $-\infty\notin f(\mathcal{H})$ and $\operatorname{dom} f\neq\varnothing$.
- Suppose that f is proper. Then $\operatorname{Argmin} f = \big\{x\in\mathcal{H} \mid f(x) = \inf f(\mathcal{H})\big\}$ is the set of minimizers of f over $\mathcal{H}$.
- f is convex if $\operatorname{epi} f$ is a convex subset of $\mathcal{H}\oplus\mathbb{R}$.
- f is lower semicontinuous if $\operatorname{epi} f$ is a closed subset of $\mathcal{H}\oplus\mathbb{R}$.
- $\Gamma_0(\mathcal{H})$: Set of proper lower semicontinuous convex functions from $\mathcal{H}$ to $]-\infty,+\infty]$.
- Let $(\mathcal{H}_i)_{i\in I}$ be a finite family of real Hilbert spaces and, for every $i\in I$, let $f_i\colon\mathcal{H}_i\to\,]-\infty,+\infty]$. Then

$$\bigoplus_{i\in I} f_i\colon \bigoplus_{i\in I}\mathcal{H}_i \to\,]-\infty,+\infty] : (x_i)_{i\in I}\mapsto \sum_{i\in I} f_i(x_i). \tag{1}$$

- $f^*\colon\mathcal{H}\to[-\infty,+\infty] : x^*\mapsto \sup_{x\in\mathcal{H}}(\langle x\mid x^*\rangle - f(x))$: Conjugate of f.

- Suppose that f is proper. Then

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

  is the subdifferential of f.
- Suppose that $f \in \Gamma_0(\mathcal{H})$. Then, for every $x \in \mathcal{H}$, $\operatorname{prox}_f x$ denotes the unique minimizer of the function $f + (1/2)\|\cdot - x\|^2$. The proximity operator of f is $\operatorname{prox}_f\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \operatorname{prox}_f x$.

Notation and definitions relative to a subset C of $\mathcal{H}$

- ι_C: Indicator function of C, that is,

$$\iota_C\colon \mathcal{H} \to [0, +\infty] : x \mapsto \begin{cases} 0, & \text{if } x \in C; \\ +\infty, & \text{if } x \notin C. \end{cases} \tag{3}$$

- $\operatorname{int} C$: Interior of C.
- $\overline{C}$: Closure of C.
- $\operatorname{sri} C$: Strong relative interior of C, that is,

$$\operatorname{sri} C = \left\{ x \in C \;\middle|\; \bigcup_{\lambda \in]0,+\infty[} \lambda(C - x) \text{ is a closed vector subspace of } \mathcal{H} \right\}. \tag{4}$$

- $\operatorname{ri} C$: Relative interior of C, that is,

$$\operatorname{ri} C = \left\{ x \in C \;\middle|\; \bigcup_{\lambda \in]0,+\infty[} \lambda(C - x) \text{ is a vector subspace of } \mathcal{H} \right\}. \tag{5}$$

- $d_C\colon \mathcal{H} \to [0, +\infty] : x \mapsto \inf \|C - x\|$: Distance function to C.
- Suppose that C is nonempty, closed, and convex. Then $\operatorname{proj}_C = \operatorname{prox}_{\iota_C}$.

Notation and definitions relative to an operator $T\colon \mathcal{H} \to \mathcal{H}$

- $\operatorname{Fix} T = \big\{x \in \mathcal{H} \mid Tx = x\big\}$: Set of fixed points of T.
- T is cocoercive with constant $\beta \in \left]0, +\infty\right[$ if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Tx - Ty\rangle \geqslant \beta\|Tx - Ty\|^2. \tag{6}$$

- T is Lipschitzian with constant $\beta \in \left[0, +\infty\right[$ if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\| \leqslant \beta\|x - y\|. \tag{7}$$

- T is nonexpansive if it is Lipschitzian with constant 1.

Notation and definitions relative to a set-valued operator $M\colon \mathcal{H} \to 2^{\mathcal{H}}$

- $\operatorname{dom} M = \big\{x \in \mathcal{H} \mid Mx \neq \varnothing\big\}$: Domain of M.
- $\operatorname{ran} M = \bigcup_{x\in\mathcal{H}} Mx$: Range of M.
- $\operatorname{zer} M = \big\{x \in \mathcal{H} \mid 0 \in Mx\big\}$: Set of zeros of M.
- $\operatorname{gra} M = \big\{(x, x^*) \in \mathcal{H} \times \mathcal{H} \mid x^* \in Mx\big\}$: Graph of M.
- M^{-1}: Inverse of M, that is,

$$M^{-1}\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x^* \mapsto \big\{x \in \mathcal{H} \mid x^* \in Mx\big\}. \tag{8}$$

- $J_M = (\mathrm{Id} + M)^{-1}$: Resolvent of M.
- M is monotone if

$$\big(\forall (x, x^*) \in \operatorname{gra} M\big)\big(\forall (y, y^*) \in \operatorname{gra} M\big) \quad \langle x - y \mid x^* - y^*\rangle \geqslant 0. \tag{9}$$

- M is maximally monotone if M is monotone and, for every monotone operator $\widetilde{M}\colon \mathcal{H} \to 2^{\mathcal{H}}$, $\operatorname{gra} M \subset \operatorname{gra} \widetilde{M} \Rightarrow M = \widetilde{M}$.

Chapter 1

INTRODUCTION

1.1 Overview

Throughout this chapter, $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces. A fundamental problem in nonlinear analysis is the following.

Problem 1.1 Let $M\colon \mathcal{H} \to 2^{\mathcal{H}}$ be monotone. The objective is to

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } 0 \in M\overline{x}. \tag{1.1}$$

From a theoretical viewpoint, Problem 1.1 is a powerful modeling framework that captures concrete scenarios in fields as diverse as partial differential equations [22, 25, 28, 29, 37, 83, 88, 92, 95, 118, 126], mechanics [84, 97], image recovery [1, 2, 10, 35, 49, 51, 58, 85, 89, 102], game theory [6, 17, 34, 52, 66, 91, 125], evolution inclusions [5, 7, 23, 24, 98], integral equations [26, 27], network flows [7, 74, 80, 115, 116], systems theory [46], domain decomposition [6, 8], machine learning [9, 30], optimal control [13], signal processing [11, 31, 56, 62, 66, 68], optimization [14, 55, 71, 72, 93, 111, 112], matrix estimation [16, 47], mean field games [36, 78, 87], statistics [50, 59, 60, 90, 108, 124], neural networks [65], variational inequalities [77, 121], tensor completion [82, 99], location problem [100, 103], and optimal transportation [104, 105]. From a practical viewpoint, the most elementary method for solving Problem 1.1 in the case where M is maximally monotone is the proximal point algorithm [28, 113]

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = J_{\gamma_n M} x_n, \quad \text{where } \gamma_n \in \left]0, +\infty\right[\text{ and } J_{\gamma_n M} = (\mathrm{Id} + \gamma_n M)^{-1}. \tag{1.2}$$

The implementation of this method may be hindered by the difficulty of evaluating the resolvents $(J_{\gamma_n M})_{n\in\mathbb{N}}$; for example, there is no closed-form expression for the resolvent of the maximally monotone operator

$$\mathcal{H} \oplus \mathcal{H} \to 2^{\mathcal{H}\oplus\mathcal{H}}\colon (x, v^*) \mapsto \big(\partial f(x) + v^*\big) \times \big(-x + \partial g^*(v^*)\big), \tag{1.3}$$

which arises in Rockafellar's saddle formalism [111, 112] for the problem of minimizing $f+g$, where $f\in\Gamma_0(\mathcal{H})$ and $g\in\Gamma_0(\mathcal{H})$. To circumvent this issue, it is often assumed that the operator M in Problem 1.1 can be expressed as the sum of maximally monotone operators $A\colon\mathcal{H}\to 2^{\mathcal{H}}$ and $B\colon\mathcal{H}\to 2^{\mathcal{H}}$, where the resolvents $(J_{\gamma A})_{\gamma\in]0,+\infty[}$ are easy to compute and B satisfies one of the following: the resolvents $(J_{\gamma B})_{\gamma\in]0,+\infty[}$ are easy to compute, $B\colon\mathcal{H}\to\mathcal{H}$ is cocoercive, or $B\colon\mathcal{H}\to\mathcal{H}$ is Lipschitzian.

Problem 1.2 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

$$\text{find } \overline{x}\in\mathcal{H} \text{ such that } 0\in A\overline{x}+B\overline{x}. \tag{1.4}$$

The three fundamental algorithms for solving Problem 1.2 are:

- **(Douglas–Rachford algorithm [96])** Let $x_0\in\mathcal{H}$ and $\gamma\in\,]0,+\infty[$. Iterate

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l} x_n=J_{\gamma B}y_n\\ z_n=J_{\gamma A}(2x_n-y_n)\\ y_{n+1}=y_n+z_n-x_n.\end{array}\right.\end{array} \tag{1.5}$$

- **(Forward-backward algorithm [97])** In Problem 1.2, suppose that $B\colon\mathcal{H}\to\mathcal{H}$ and that B is β-cocoercive for some $\beta\in\,]0,+\infty[$. Let $x_0\in\mathcal{H}$ and $\gamma\in\,]0,2\beta[$. Iterate

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l} y_n=x_n-\gamma Bx_n\\ x_{n+1}=J_{\gamma A}y_n.\end{array}\right.\end{array} \tag{1.6}$$

- **(Tseng's forward-backward-forward algorithm [122])** In Problem 1.2, suppose that $B\colon\mathcal{H}\to\mathcal{H}$ and that B is β-Lipschitzian for some $\beta\in\,]0,+\infty[$. Let $x_0\in\mathcal{H}$, let $\varepsilon\in\,]0,1/(\beta+1)[$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be in $[\varepsilon,(1-\varepsilon)/\beta]$. Iterate

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l} y_n=x_n-\gamma_n Bx_n\\ p_n=J_{\gamma_n A}y_n\\ q_n=p_n-\gamma_n Bp_n\\ x_{n+1}=x_n-y_n+q_n.\end{array}\right.\end{array} \tag{1.7}$$

Many complex splitting algorithms in the literature (e.g., [7, 21, 32, 53, 54, 56, 63, 69, 73, 81, 86, 94, 114, 119, 123] and the references therein) are reformulations of (1.5)–(1.7) in suitable spaces. Let us illustrate this by considering the state-of-the-art model [54].

Example 1.3 Let $(\mathcal{H}_i)_{i\in I}$ and $(\mathcal{G}_k)_{k\in K}$ be finite families of real Hilbert spaces and let $(\mu_i)_{i\in I}$ and $(\nu_k)_{k\in K}$ be in $[0,+\infty[$. For every $i\in I$ and every $k\in K$, let $A_i\colon\mathcal{H}_i\to 2^{\mathcal{H}_i}$ and $B_k\colon\mathcal{G}_k\to 2^{\mathcal{G}_k}$

be maximally monotone, let $C_i\colon \mathcal{H}_i \to \mathcal{H}_i$ be monotone and μ_i-Lipschitzian, 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, let $z_i \in \mathcal{H}_i$, let $r_k \in \mathcal{G}_k$, and let $L_{ki}\colon \mathcal{H}_i \to \mathcal{G}_k$ be linear and bounded. The objective is to solve the primal system of monotone inclusions

$$\text{find } (\overline{x}_i)_{i\in I} \in \bigoplus_{i\in I} \mathcal{H}_i \text{ such that}$$
$$(\forall i \in I)\ \ z_i \in A_i\overline{x}_i + \sum_{k\in K} L_{ki}^*\Bigg((B_k \,\square\, D_k)\Bigg(\sum_{j\in I} L_{kj}\overline{x}_j - r_k\Bigg)\Bigg) + C_i\overline{x}_i, \tag{1.8}$$

together with the dual system

$$\text{find } (\overline{v}_k)_{k\in K} \in \bigoplus_{k\in K} \mathcal{G}_k \text{ such that}$$
$$-r_k \in -\sum_{i\in I} L_{ki}(A_i + C_i)^{-1}\Bigg(z_i - \sum_{j\in K} L_{ji}^*\overline{v}_k\Bigg) + B_k^{-1}\overline{v}_k + D_k^{-1}\overline{v}_k. \tag{1.9}$$

By applying Tseng's forward-backward-forward algorithm (1.7) to the primal-dual setting

$$\begin{cases} \boldsymbol{\mathcal{H}} = \big(\bigoplus_{i\in I}\mathcal{H}_i\big) \oplus \big(\bigoplus_{k\in K}\mathcal{G}_k\big) \\ \boldsymbol{A}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}\colon \big((x_i)_{i\in I}, (v_k)_{k\in K}\big) \mapsto \big(\times_{i\in I}(-z_i + A_i x_i)\big) \times \big(\times_{k\in K}(r_k + B_k^{-1}v_k)\big) \\ \boldsymbol{B}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon \big((x_i)_{i\in I}, (v_k)_{k\in K}\big) \mapsto \Big(\big(C_i x_i + \sum_{k\in K} L_{ki}^* v_k\big)_{i\in I}, \big(D_k^{-1}v_k - \sum_{i\in I} L_{ki}x_i\big)_{k\in K}\Big), \end{cases} \tag{1.10}$$

we obtain the splitting scheme [54, Eq. (2.4)] for solving Problem 1.2. This method achieves full splitting of Problem 1.2 in the sense that the set-valued operators $(A_i)_{i\in I}$ and $(B_k)_{k\in K}$ are activated independently via backward resolvent steps and the single-valued operators $(D_k^{-1})_{k\in K}$ and $(L_{ki})_{i\in I,k\in K}$ are activated via forward steps.

Another approach to monotone inclusions is the projective splitting framework, which was first proposed in [70] for the monotone inclusion

$$\text{find } \overline{x} \in \mathbb{R}^N \text{ such that } 0 \in L^*\big(B(L\overline{x})\big), \quad \text{where } \begin{cases} B\colon \mathbb{R}^P \to 2^{\mathbb{R}^P} \text{ is maximally monotone} \\ L\colon \mathbb{R}^N \to \mathbb{R}^P \text{ is linear,} \end{cases} \tag{1.11}$$

in [75] for Problem 1.2, and then in [76] for the problem of finding a zero of a sum of maximally monotone operators. These approaches were unified and extended in [3], where the special case of (1.8) where, for every $i \in I$ and every $k \in K$, $C_i = 0$ and $D_k^{-1} = 0$, was considered, i.e.,

$$\text{find } (\overline{x}_i)_{i\in I} \in \bigoplus_{i\in I}\mathcal{H}_i \text{ such that } (\forall i \in I)\ \ z_i \in A_i\overline{x}_i + \sum_{k\in K} L_{ki}^*\Bigg(B_k\Bigg(\sum_{j\in I} L_{kj}\overline{x}_j - r_k\Bigg)\Bigg). \tag{1.12}$$

The state of the art in projective splitting is [57], which is the first method where (1.12) is solved by an algorithm which has the ability to incorporate the result of calculations initiated at earlier iterations, and requires to activate only a subgroup of operators at every iteration.

Despite the significant developments in the last decade, there remain many important open questions in the area of monotone operator splitting. We list below the ones that will be addressed in this dissertation:

(Q1) The weak convergence of the sequence $(x_n)_{n\in\mathbb{N}}$ generated by (1.5) was established in [120]. While various additional conditions on the operators A and B in Problem 1.2 have been proposed to ensure the strong convergence of $(x_n)_{n\in\mathbb{N}}$ [14, 53, 96], it remains an open question whether $(x_n)_{n\in\mathbb{N}}$ can fail to converge strongly in the general setting of Problem 1.2.

(Q2) As mentioned above, the execution of (1.2) depends on the ease of evaluating $(J_{\gamma_n M})_{n\in\mathbb{N}} = ((\mathrm{Id} + \gamma_n M)^{-1})_{n\in\mathbb{N}}$, which is not the case in many situations. For example, in [10, 14, 18–20, 33, 54, 57, 63, 69, 107, 123], M is a composite operator assembled from several elementary blocks that can be linear operators and monotone operators and there is no convenient way to express $J_{\gamma M}$ in terms of these elementary blocks. A question is therefore to seek an extension of the classical resolvent that can be tailored to the structure of M and is thus potentially easier to compute than the classical resolvent. Further, can we explore a new path for solving Problem 1.1 with this new resolvent at the core of our analysis?

(Q3) Several forward-backward methods based on Bregman distances have been proposed with various assumptions [12, 67, 101, 109] for solving some special cases of finding a zero of the sum of two maximally monotone operators acting on a Banach space. Is there a Bregman forward-backward framework that unifies these algorithms and can solve problems beyond their reach? The motivation for this is that standard splitting algorithms are not applicable beyond Hilbert spaces. In addition, there has been a significant body of work (see, e.g., [11, 12, 15, 48, 61, 79, 101, 104, 117]) showing the benefits of replacing standard distances by Bregman distances, even in Euclidean spaces.

(Q4) What is the connection between the approach of [32, 54, 63] and the projective splitting method of [3, 70, 75, 76]? Beyond the fact that they both rely on the Fejér monotonicity principle, connections remain elusive.

(Q5) Splitting methods have been applied to solving Nash equilibrium under convexity assumption, but with simple settings (see, e.g., [6, 17, 34, 52, 66, 91, 125] and the references therein). Can we develop flexible splitting algorithms for more general Nash equilibrium models?

(Q6) Rockafellar proposed in [116] a multicommodity network equilibrium model and studied some of its properties. The pertinence of this model is demonstrated in [115, Chapter 8] and [116], where it is shown to capture formulations arising in areas such as traffic assignment, hydraulic networks, and price equilibrium. Thus far, the numerical aspect of this model has not been considered. Can we develop an efficient algorithm for this model?

(Q7) There seem to be only two primary methods for solving multicomponent fully nonsmooth minimization that are block-activated in the sense that they require to activate only a subgroup of functions at each iteration [57, 64]. Nevertheless, little effort has been devoted to assessing them. Can we shed more light on the implementation, the features, and the behavior of these algorithms, compare their merits, and provide numerical experiments illustrating their performance?

(Q8) Rockafellar's saddle formalism plays a central role in analyzing and solving primal-dual convex optimization problems [111, 112]. What is the extension of this notion to monotone inclusions in duality in the sense of [106, 110]? This is motivated by the fact that there are monotone inclusions arising in applications which involve non-subdifferential operators (see, e.g., [7, 55, 66, 77, 92] and the references therein).

1.2 Contributions and organization

This dissertation, which produced the articles [38–45], provides answers to the open questions **(Q1)**–**(Q8)**. More precisely, the main contributions of this dissertation are the following:

- We answer **(Q1)** in Chapter 2 by providing counterexamples showing that the Douglas–Rachford algorithm [96] and the method of partial inverses [119] can fail to converge strongly.
- To address **(Q2)**, we introduce in Chapter 3 the novel notion of a warped resolvent and investigate its properties. Moreover, we provide weak and strong warped proximal iteration principles and establish their convergence in Theorems 3.16 and 3.22. In Section 3.2.5, the warped proximal iterations are shown to capture [3, 4, 122] and to provide new and flexible splitting algorithms for complex monotone inclusions.
- Chapter 4 is devoted to **(Q3)**. There, a Bregman forward-backward splitting algorithm for monotone inclusions in Banach spaces is introduced, and its weak convergence is established. This scheme is shown to unifies [12, 67, 101, 109]. We also establish rates of convergence in the minimization setting.
- An answer to **(Q4)** is provided in Chapter 5, where we show that [57, Algorithm 12] can be viewed as an instantiation of the warped proximal iterations of Theorem 3.16. Thus, the frameworks of [3, 32, 54, 63, 70, 75, 76] are special cases of the warped resolvent iterations.
- We address **(Q5)** in Chapter 6. We propose an asynchronous block-iterative algorithm for solving a highly modular Nash equilibrium problem. Our methodology relies on the warped resolvent algorithm of Theorem 3.16.

- Chapter 7 focuses on **(Q6)**. We present a flexible decomposition method based on [57, Algorithm 12] for solving the multicommodity network equilibrium model proposed by Rockafellar in [116].
- To address **(Q7)**, we devote Chapter 8 to assess the block-activated algorithms of [57,64]. Our numerical experiments are in the areas of machine learning and image recovery.
- In Chapter 9, we focus on **(Q8)**. A saddle formulation for studying and solving highly structured systems of monotone inclusions is proposed. Various applications are discussed, and instantiations of the proposed framework in the context of variational inequalities and minimization problems are presented.
- We conclude this dissertation in Chapter 10 with future research directions.

References

[1] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, Fast image recovery using variable splitting and constrained optimization, *IEEE Trans. Image Process.*, vol. 19, pp. 2345–2356, 2010.

[2] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems, *IEEE Trans. Image Process.*, vol. 20, pp. 681–695, 2011.

[3] 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.

[4] 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.

[5] H. Attouch, *Variational Convergence for Functions and Operators*. Pitman, Boston, MA, 1984.

[6] 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.

[7] 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.

[8] 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.

[9] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, Optimization with sparsity-inducing penalties, *Found. Trends Machine Learn.*, vol. 4, pp. 1–106, 2012.

[10] S. Banert, A. Ringh, J. Adler, J. Karlsson, and O. Öktem, Data-driven nonsmooth optimization, *SIAM J. Optim.*, vol. 30, pp. 102–131, 2020.

[11] 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.

[12] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.*, vol. 42, pp. 596–636, 2003.

[13] H. Bauschke, R. S. Burachik, and C. Y. Kaya, Constraint splitting and projection methods for optimal control of double integrator, in: *Splitting Algorithms, Modern Operator Theory, and Applications*, (H. H. Bauschke et al, eds.), pp. 45–68. Springer, New York, 2019.

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

[15] H. H. Bauschke, M. N. Dao, and S. B. Lindstrom, Regularizing with Bregman–Moreau envelopes, *SIAM J. Optim.*, vol. 28, pp. 3208–3228, 2018.

[16] A. Benfenati, E. Chouzenoux, and J.-C. Pesquet, Proximal approaches for matrix optimization problems: Application to robust precision matrix estimation, *Signal Process.*, vol. 169, art. 107417, 2020.

[17] 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.

[18] R. I. Boţ and E. R. Csetnek, ADMM for monotone operators: Convergence analysis and rates, *Adv. Comput. Math.*, vol. 45, pp. 327–359, 2019.

[19] R. I. Boţ, E. R. Csetnek, and A. Heinrich, A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators, *SIAM J. Optim.*, vol. 23, pp. 2011–2036, 2013.

[20] 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.

[21] 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.

[22] H. Brézis, Monotonicity methods in Hilbert spaces and some applications to nonlinear partial differential equations, in: *Contributions to Nonlinear Functional Analysis*, (E. Zarantonello, ed.), pp. 101–156. Academic Press, New York, 1971.

[23] H. Brézis, Équations d'évolution du second ordre associées à des opérateurs monotones, *Israel J. Math.*, vol. 12, pp. 51–60, 1972.

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

[25] H. Brézis, Monotone operators, nonlinear semigroups, and applications, in: *Proceedings of the International Congress of Mathematicians*, (R. D. James, ed.), pp. 249–255. Canadian Mathematical Congress, Vancouver, BC, 1975.

[26] H. Brézis and F. E. Browder, Nonlinear integral equations and systems of Hammerstein type, *Adv. Math.*, vol. 18, pp. 115–147, 1975.

[27] H. Brézis and F. E. Browder, Maximal monotone operators in nonreflexive Banach spaces and nonlinear integral equations of Hammerstein type, *Bull. Amer. Math. Soc.*, vol. 81, pp. 82–88, 1975.

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

[29] H. Brézis et M. Sibony, Méthodes d'approximation et d'itération pour les opérateurs monotones, *Arch. Ration. Mech. Anal.*, vol. 28, pp. 59–82, 1967/1968.

[30] 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.

[31] 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.

[32] 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.

[33] L. M. Briceño-Arias, Forward-partial inverse-forward splitting for solving monotone inclusions, *J. Optim. Theory Appl.*, vol. 166, pp. 391–413, 2015.

[34] 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.

[35] 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.

[36] 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.

[37] F. E. Browder and C. P. Gupta, Monotone operators and nonlinear integral equations of Hammerstein type, *Bull. Amer. Math. Soc.*, vol. 75, pp. 1347–1353, 1969.

[38] M. N. Bùi, Projective splitting as a warped proximal algorithm, submitted.

[39] M. N. Bùi, A decomposition method for solving multicommodity network equilibrium, submitted.

[40] M. N. Bùi and P. L. Combettes, The Douglas–Rachford algorithm converges only weakly, *SIAM J. Control Optim.*, vol. 58, pp. 1118–1120, 2020.

[41] M. N. Bùi and P. L. Combettes, Warped proximal iterations for monotone inclusions, *J. Math. Anal. Appl.*, vol. 491, art. 124315, 21 pp., 2020.

[42] M. N. Bùi and P. L. Combettes, Bregman forward-backward operator splitting, *Set-Valued Var. Anal.*, vol. 29, pp. 583–603, 2021.

[43] M. N. Bùi and P. L. Combettes, Multivariate monotone inclusions in saddle form, *Math. Oper. Res.*, to appear.

[44] M. N. Bùi and P. L. Combettes, A warped resolvent algorithm to construct Nash equilibria, submitted.

[45] M. N. Bùi, P. L. Combettes, and Z. C. Woodstock, Block-activated algorithms for multicomponent fully nonsmooth minimization, submitted.

[46] M. K. Camlibel and J. M. Schumacher, Linear passive systems and maximal monotone mappings, *Math. Program.*, vol. B157, pp. 397–420, 2016.

[47] E. J. Candès and B. Recht, Exact matrix completion via convex optimization, *Found. Comput. Math.*, vol. 9, pp. 717–772, 2009.

[48] Y. Censor and S. A. Zenios, *Parallel Optimization – Theory, Algorithms and Applications.* Oxford University Press, New York, 1997.

[49] A. Chambolle and T. Pock, An introduction to continuous optimization for imaging, *Acta Numer.*, vol. 25, pp. 161–319, 2016.

[50] E. C. Chi and K. Lange, Splitting methods for convex clustering, *J. Comput. Graph. Statist.*, vol. 24, pp. 994–1013, 2015.

[51] G. Chierchia, N. Pustelnik, J.-C. Pesquet, and B. Pesquet-Popescu, Epigraphical splitting for solving constrained convex optimization problems with proximal tools, *Signal Image Video Process.*, vol. 9, pp. 1737–1749, 2015.

[52] G. Cohen, Nash equilibria: Gradient and decomposition algorithms, *Large Scale Syst.*, vol. 12, pp. 173–184, 1987.

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

[54] P. L. Combettes, Systems of structured monotone inclusions: Duality, algorithms, and applications, *SIAM J. Optim.*, vol. 23, pp. 2420–2447, 2013.

[55] P. L. Combettes, Monotone operator theory in convex optimization, *Math. Program.*, vol. B170, pp. 177–206, 2018.

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

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

[58] 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.

[59] P. L. Combettes and C. L. Müller, Perspective maximum likelihood-type estimation via proximal decomposition, *Electron. J. Stat.*, vol. 14, pp. 207–238, 2020.

[60] P. L. Combettes and C. L. Müller, Regression models for compositional data: General log-contrast formulations, proximal optimization, and microbiome data applications, *Stat. Biosci.*, vol. 13, pp. 217–242, 2021.

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

[62] P. L. Combettes and J.-C. Pesquet, Proximal splitting methods in signal processing, in: *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, (H. H. Bauschke et al., eds.), pp. 185–212. Springer, New York, 2011.

[63] 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.

[64] 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.

[65] P. L. Combettes and J.-C. Pesquet, Deep neural network structures solving variational inequalities, *Set-Valued Var. Anal.*, vol. 28, pp. 491–518, 2020.

[66] P. L. Combettes and J.-C. Pesquet, Fixed point strategies in data science, *IEEE Trans. Signal Process.*, vol. 69, pp. 3878–3905, 2021.

[67] 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.

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

[69] 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.

[70] Y. Dong, An LS-free splitting method for composite mappings, *Appl. Math. Lett.*, vol. 18, pp. 843–848, 2005.

[71] J. Eckstein, Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming, *Math. Oper. Res.*, vol. 18, pp. 202–226, 1993.

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

[73] 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.

[74] J. Eckstein and M. Fukushima, Some reformulations and applications of the alternating direction method of multipliers, in: *Large Scale Optimization: State of the Art*, (W. W.

Hager, D. W. Hearn, and P. M. Pardalos, eds.), pp. 115–134. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1994.

[75] 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.

[76] 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.

[77] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer-Verlag, New York, 2003.

[78] R. Ferreira and D. Gomes, Existence of weak solutions to stationary mean-field games through variational inequalities, *SIAM J. Math. Anal.*, vol. 50, pp. 5969–6006, 2018.

[79] J. Frecon, S. Salzo, and M. Pontil, Bilevel learning of the group lasso structure, *Adv. Neural Inform. Process. Syst.*, vol. 31, pp. 8301–8311, 2018.

[80] M. Fukushima, The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with applications to the traffic equilibrium problem, *Math. Program.*, vol. 72, pp. 1–15, 1996.

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

[82] 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.

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

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

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

[86] E. G. Gol'shtein, A general approach to decomposition of optimization systems, *Sov. J. Comput. Syst. Sci.*, vol. 25, pp. 105–114, 1987.

[87] D. A. Gomes and J. Saúde, Numerical methods for finite-state mean-field games satisfying a monotonicity condition, *Appl. Math. Optim.*, vol. 83, pp. 51–82, 2021.

[88] M. Hintermüller and K. Kunisch, Path-following methods for a class of constrained minimization problems in function space, *SIAM J. Optim.*, vol. 17, pp. 159–187, 2006.

[89] M. Hintermüller and G. Stadler, An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration, *SIAM J. Sci. Comput.*, vol. 28, pp. 1–23, 2006.

[90] P. R. Johnstone and J. Eckstein, Projective splitting with forward steps, *Math. Program.*, published online 2020-09-30.

[91] A. Kannan and U. V. Shanbhag, Distributed computation of equilibria in monotone Nash games via iterative regularization techniques, *SIAM J. Optim.*, vol. 22, pp. 1177–1205, 2012.

[92] A. Kaplan and R. Tichatschke, Proximal point method and elliptic regularization, *Nonlinear Anal.*, vol. 71, pp. 4525–4543, 2009.

[93] N. Komodakis and J.-C. Pesquet, Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems, *IEEE Signal Process. Mag.*, vol. 32, pp. 31–54, 2015.

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

[95] J.-L. Lions, *Quelques Méthodes de Résolution des Problèmes aux Limites Non Linéaires*. Dunod, Paris, 1969.

[96] 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.

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

[98] S. Migórski, A. Ochal, and M. Sofonea, Evolutionary inclusions and hemivariational inequalities, in: *Advances in Variational and Hemivariational Inequalities: Theory, Numerical Analysis, and Applications*, (W. Han, S. Migórski, and M. Sofonea, eds.), pp. 39–64. Springer, New York, 2015.

[99] T. Mizoguchi and I. Yamada, Hypercomplex tensor completion via convex optimization, *IEEE Trans. Signal Process.*, vol. 67, pp. 4078–4092, 2019.

[100] N. M. Nam, T. A. Nguyen, R. B. Rector, and J. Sun, Nonsmooth algorithms and Nesterov's smoothing technique for generalized Fermat-Torricelli problems, *SIAM J. Optim.*, vol. 24, pp. 1815–1839, 2014.

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

[102] 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.

[103] 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.

[104] G. Ortiz-Jiménez, M. El Gheche, E. Simou, H. Petric Maretić, and P. Frossard, Forward-backward splitting for optimal transport based problems, *Proc. Intl. Conf. Acoust., Speech, Signal Process.*, pp. 5405–5409, 2020.

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

[106] T. Pennanen, Dualization of generalized equations of maximal monotone type, *SIAM J. Optim.*, vol. 10, pp. 809–835, 2000.

[107] H. Raguet, A note on the forward-Douglas–Rachford splitting for monotone inclusion and convex optimization, *Optim. Lett.*, vol. 13, pp. 717–740, 2019.

[108] P. Ravikumar, M. J. Wainwright, G. Raskutti, and B. Yu, High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence, *Electron. J. Statist.*, vol. 5, pp. 935–980, 2011.

[109] 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.

[110] S. M. Robinson, Composition duality and maximal monotonicity, *Math. Program.*, vol. 85, pp. 1–13, 1999.

[111] 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.

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

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

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

[115] R. T. Rockafellar, *Network Flows and Monotropic Optimization*. Wiley, New York, 1984.

[116] 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.

[117] S. Salzo, The variable metric forward-backward splitting algorithm under mild differentiability assumptions, *SIAM J. Optim.*, vol. 27, pp. 2153–2181, 2017.

[118] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*. AMS, Providence, RI, 1997.

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

[120] B. F. Svaiter, On weak convergence of the Douglas–Rachford method, *SIAM J. Control Optim.*, vol. 49, pp. 280–287, 2011.

[121] 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.

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

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

[124] X. Yan and J. Bien, Rare feature selection in high dimensions, *J. Amer. Statist. Assoc.*, vol. 116, pp. 887–900, 2021.

[125] P. Yi and L. Pavel, An operator splitting approach for distributed generalized Nash equilibria computation, *Automatica*, vol. 102, pp. 111–121, 2019.

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

Chapter 2

THE DOUGLAS–RACHFORD ALGORITHM CONVERGES ONLY WEAKLY

2.1 Introduction and context

We provide a complete answer to question **(Q1)** of Chapter 1 in Counterexample 2.2. In addition, Counterexample 2.4 shows that the method of partial inverses can fail to converge strongly.

This chapter presents the following article:

M. N. Bùi and P. L. Combettes, The Douglas–Rachford algorithm converges only weakly, *SIAM Journal on Control and Optimization*, vol. 58, no. 2, pp. 1118–1120, 2020.

2.2 Article: The Douglas–Rachford algorithm converges only weakly

Abstract. We show that the weak convergence of the Douglas–Rachford algorithm for finding a zero of the sum of two maximally monotone operators cannot be improved to strong convergence. Likewise, we show that strong convergence can fail for the method of partial inverses.

The original Douglas–Rachford splitting algorithm was designed to decompose positive systems of linear equations [3]. It evolved in [5] into a powerful method for finding a zero of the sum of two maximally monotone operators in Hilbert spaces, a problem which is ubiquitous in applied mathematics (see [1] for background on monotone operators). In this context, the Douglas–Rachford algorithm constitutes a prime decomposition method in areas such as control, partial differential equations, optimization, statistics, variational inequalities, mechanics, optimal transportation, machine learning, and signal processing. Its asymptotic behavior is described next.

Theorem 2.1 *Let $\mathcal{H}$ be a real Hilbert space, and let A and B be set-valued maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$ with resolvents $J_A = (\mathrm{Id} + A)^{-1}$ and $J_B = (\mathrm{Id} + B)^{-1}$. Suppose that $\mathrm{zer}(A+B) = \big\{x \in \mathcal{H} \mid 0 \in Ax + Bx\big\} \neq \varnothing$, let $y_0 \in \mathcal{H}$, and iterate*

$$(\forall n \in \mathbb{N}) \quad x_n = J_B y_n \quad \textit{and} \quad y_{n+1} = y_n + J_A(2x_n - y_n) - x_n. \tag{2.1}$$

Then the following hold for some $(y, x) \in \mathrm{gra}\, J_B$:

(i) *$x = J_A(2x - y)$, $y_n \rightharpoonup y$, and $x \in \mathrm{zer}(A+B)$.*

(ii) *$x_n \rightharpoonup x$.*

Property (i) was established in [5]. Let us note that, since J_B is not weakly sequentially continuous in general, the weak convergence of $(y_n)_{n\in\mathbb{N}}$ in (i) does not imply (ii). The latter was first established in [7] (see also [1, Theorem 26.11(iii)] for an alternate proof). While various additional conditions on A and B have been proposed to ensure the strong convergence of the sequence $(x_n)_{n\in\mathbb{N}}$ in (2.1) [1, 2, 5], it remains an open question whether it can fail in the general setting of Theorem 2.1. We show that this is indeed the case. Our argument relies on a result of Hundal [4] concerning the method of alternating projections.

Counterexample 2.2 In Theorem 2.1, suppose that $\mathcal{H}$ is infinite-dimensional and separable. Let $(e_k)_{k\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$, let $V = \{e_0\}^{\perp}$, let $y_0 = e_2$, and let K be the smallest closed convex cone containing the set

$$\big\{\exp\big(-100\xi^3\big)e_0 + \cos\big(\pi(\xi - \lfloor\xi\rfloor)/2\big)e_{\lfloor\xi\rfloor+1} + \sin\big(\pi(\xi - \lfloor\xi\rfloor)/2\big)e_{\lfloor\xi\rfloor+2} \;\big|\; \xi \in \left[0, +\infty\right[\big\}, \tag{2.2}$$

where $\lfloor\xi\rfloor$ denotes the integer part of $\xi \in \left[0, +\infty\right[$. Let proj_V and proj_K be the projection operators onto V and K, and set

$$A\colon x \mapsto \begin{cases} V^{\perp}, & \text{if } x \in V; \\ \varnothing, & \text{if } x \notin V \end{cases} \quad \text{and} \quad B = \big(\mathrm{proj}_V \circ \mathrm{proj}_K \circ \mathrm{proj}_V\big)^{-1} - \mathrm{Id}. \tag{2.3}$$

Then A and B are maximally monotone, and the sequence $(x_n)_{\in\mathbb{N}}$ constructed in Theorem 2.1 converges weakly, but not strongly, to a zero of $A + B$.

Proof. We first note that A is maximally monotone by virtue of [1, Examples 6.43 and and 20.26]. Now set $T = \mathrm{proj}_V \circ \mathrm{proj}_K \circ \mathrm{proj}_V$. Then it follows from [1, Example 4.14] that T is firmly nonexpansive, that is,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Tx - Ty \rangle \geqslant \|Tx - Ty\|^2. \tag{2.4}$$

In turn, we derive from [1, Proposition 23.10] that $B = T^{-1} - \mathrm{Id}$ is maximally monotone. Next, we observe that $0 \in \operatorname{zer} A$ and that, since K is a closed cone, $0 \in K$. Thus, $0 = (\operatorname{proj}_V \circ \operatorname{proj}_K \circ \operatorname{proj}_V)0$, which implies that $0 \in \operatorname{zer} B$. Hence,

$$0 \in \operatorname{zer}(A + B). \tag{2.5}$$

Now set

$$z_0 = \exp(-100)e_0 + e_2 \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad z_{n+1} = \operatorname{proj}_K\big(\operatorname{proj}_V z_n\big). \tag{2.6}$$

Then, by nonexpansiveness of proj_K,

$$(\forall n \in \mathbb{N}) \quad \|z_{n+1}\|^2 = \|\operatorname{proj}_K\big(\operatorname{proj}_V z_n\big) - \operatorname{proj}_K 0\|^2 \leqslant \|\operatorname{proj}_V z_n\|^2 = \|z_n\|^2 - \|\operatorname{proj}_V z_n - z_n\|^2 \tag{2.7}$$

and, therefore,

$$\operatorname{proj}_V z_n - z_n \to 0. \tag{2.8}$$

As shown in [4], we also have

$$z_n \rightharpoonup 0 \quad \text{and} \quad z_n \not\to 0. \tag{2.9}$$

On the other hand, we derive from (2.3) that

$$J_A = \operatorname{proj}_V \quad \text{and} \quad J_B = \operatorname{proj}_V \circ \operatorname{proj}_K \circ \operatorname{proj}_V, \tag{2.10}$$

and from (2.6) that $\operatorname{proj}_V z_0 = e_2 = y_0$. It thus follows from (2.1) and (2.6) that $x_0 = \operatorname{proj}_V(\operatorname{proj}_K(\operatorname{proj}_V y_0)) = \operatorname{proj}_V(\operatorname{proj}_K(\operatorname{proj}_V z_0)) = \operatorname{proj}_V z_1$. Now, assume that, for some $n \in \mathbb{N}$, $y_n = \operatorname{proj}_V z_n$ and $x_n = \operatorname{proj}_V z_{n+1}$. Since x_n and y_n lie in V, we derive from (2.1) and (2.10) that

$$y_{n+1} = y_n + \operatorname{proj}_V(2x_n - y_n) - x_n = x_n = \operatorname{proj}_V z_{n+1} \tag{2.11}$$

and hence that

$$x_{n+1} = \big(\operatorname{proj}_V \circ \operatorname{proj}_K \circ \operatorname{proj}_V\big)\big(\operatorname{proj}_V z_{n+1}\big) = \operatorname{proj}_V\big(\operatorname{proj}_K\big(\operatorname{proj}_V z_{n+1}\big)\big) = \operatorname{proj}_V z_{n+2}. \tag{2.12}$$

We have thus proven by induction that

$$(\forall n \in \mathbb{N}) \quad x_n = \operatorname{proj}_V z_{n+1}. \tag{2.13}$$

In view of (2.8), we obtain $x_n - z_{n+1} \to 0$ and therefore derive from (2.9) and (2.5) that $x_n \rightharpoonup 0 \in \operatorname{zer}(A + B)$ and $x_n \not\to 0$. ☐

Next, we settle a similar open question for Spingarn's method of partial inverses [6] by showing that its strong convergence can fail.

Theorem 2.3 ([6]) *Let $\mathcal{H}$ be a real Hilbert space, let $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let V be a closed vector subspace of $\mathcal{H}$. Suppose that the problem*

$$\text{find } x \in V \text{ and } u \in V^{\perp} \text{ such that } u \in Bx \tag{2.14}$$

has at least one solution. Let $x_0 \in V$, let $u_0 \in V^{\perp}$, and iterate

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \mathrm{proj}_V\big(J_B(x_n+u_n)\big) \quad \text{and} \quad u_{n+1} = \mathrm{proj}_{V^\perp}\big(J_{B^{-1}}(x_n+u_n)\big). \tag{2.15}$$

Then $(x_n,u_n)_{n\in\mathbb{N}}$ converges weakly to a solution to (2.14).

Counterexample 2.4 Define $\mathcal{H}$, V, K, and B as in Counterexample 2.2, and set $x_0 = e_2$ and $u_0 = 0$. Then $(0,0)$ solves (2.14) and the sequence $(x_n,u_n)_{n\in\mathbb{N}}$ constructed in Theorem 2.3 converges weakly, but not strongly, to $(0,0)$.

Proof. Since $J_B = \mathrm{proj}_V \circ \mathrm{proj}_K \circ \mathrm{proj}_V$ and $J_{B^{-1}} = \mathrm{Id} - J_B$, (2.15) implies that

$$(\forall n \in \mathbb{N}) \quad \begin{cases} x_{n+1} = \big(\mathrm{proj}_V \circ \mathrm{proj}_K \circ \mathrm{proj}_V\big)(x_n+u_n) \\ u_{n+1} = \mathrm{proj}_{V^\perp}\big(x_n+u_n - \big(\mathrm{proj}_V \circ \mathrm{proj}_K \circ \mathrm{proj}_V\big)(x_n+u_n)\big). \end{cases} \tag{2.16}$$

We therefore obtain inductively that

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \mathrm{proj}_V\big(\mathrm{proj}_K x_n\big) \quad \text{and} \quad u_n = 0. \tag{2.17}$$

Now define $(z_n)_{n\in\mathbb{N}}$ as in (2.6). Then, by induction, $(\forall n \in \mathbb{N})$ $x_n = \mathrm{proj}_V z_n$. Hence, in view of (2.8) and (2.9), we conclude that $0 \not\leftarrow x_n \rightharpoonup 0$. ☐

References

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

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

[3] 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.

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

[5] 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.

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

[7] B. F. Svaiter, On weak convergence of the Douglas–Rachford method, *SIAM J. Control Optim.,* vol. 49, pp. 280–287, 2011.

Chapter 3

WARPED PROXIMAL ITERATIONS FOR MONOTONE INCLUSIONS

3.1 Introduction and context

To address question **(Q2)** of Chapter 1, we introduce the notion of a warped resolvent as a generalization of the classical resolvent, study its properties, and devise warped proximal iteration principles. The pertinence of this warped resolvent framework is illustrated.

This chapter presents the following article:

> M. N. Bùi and P. L. Combettes, Warped proximal iterations for monotone inclusions, *Journal of Mathematical Analysis and Applications*, vol. 491, no. 1, art. 124315, 21 pp., 2020.

3.2 Article: Warped proximal iterations for monotone inclusions

Abstract. Resolvents of set-valued operators play a central role in various branches of mathematics and in particular in the design and the analysis of splitting algorithms for solving monotone inclusions. We propose a generalization of this notion, called warped resolvent, which is constructed with the help of an auxiliary operator. The properties of warped resolvents are investigated and connections are made with existing notions. Abstract weak and strong convergence principles based on warped resolvents are proposed and shown to not only provide a synthetic view of splitting algorithms but to also constitute an effective device to produce new solution methods for challenging inclusion problems.

3.2.1 Introduction

A generic problem in nonlinear analysis and optimization is to find a zero of a maximally monotone operator $M\colon \mathcal{X} \to 2^{\mathcal{X}}$, where $\mathcal{X}$ is a real Hilbert space. The most elementary method designed for this task is the proximal point algorithm [34]

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = J_{\gamma_n M} x_n, \quad \text{where} \quad \gamma_n \in \left]0, +\infty\right[\quad \text{and} \quad J_{\gamma_n M} = (\mathrm{Id} + \gamma_n M)^{-1}. \tag{3.1}$$

In practice, the execution of (3.1) may be hindered by the difficulty of evaluating the resolvents $(J_{\gamma_n M})_{n\in\mathbb{N}}$. Thus, even in the simple case when M is the sum of two monotone operators A and B, there is no mechanism to express conveniently the resolvent of M in terms of operators involving A and B separately. To address this issue, various splitting strategies have been proposed to handle increasingly complex formulations in which M is a composite operator assembled from several elementary blocks that can be linear operators and monotone operators [5, 7, 9–12, 17, 18, 21, 22, 30, 37]. In the present paper, we explore a different path by placing at the core of our analysis the following extension of the classical notion of a resolvent.

Definition 3.1 (Warped resolvent) Let $\mathcal{X}$ be a reflexive real Banach space with topological dual $\mathcal{X}^*$, let D be a nonempty subset of $\mathcal{X}$, let $K\colon D \to \mathcal{X}^*$, and let $M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be such that $\operatorname{ran} K \subset \operatorname{ran}(K + M)$ and $K + M$ is injective (see Definition 3.2). The warped resolvent of M with kernel K is $J_M^K = (K + M)^{-1} \circ K$.

A main motivation for introducing warped resolvents is that, through judicious choices of a kernel K tailored to the structure of an inclusion problem, one can create simple patterns to design and analyze new, flexible, and modular splitting algorithms. At the same time, the theory required to analyze the static properties of warped resolvents as nonlinear operators, as well as the dynamics of algorithms using them, needs to be developed as it cannot be extrapolated from the classical case, where K is simply the identity operator. In the present paper, this task is undertaken and we illustrate the pertinence of warped iteration methods through applications to challenging monotone inclusion problems.

The paper is organized as follows. Section 3.2.2 is dedicated to notation and background. In Section 3.2.3, we provide important illustrations of Definition 3.1 and make connections with constructions found in the literature. The properties of warped resolvents are also discussed in that section. Weakly and strongly convergent warped proximal iteration methods are introduced and analyzed in Section 3.2.4. Besides the use of kernels varying at each iteration, our framework also features evaluations of warped resolvents at points that may not be the current iterate, which adds considerable flexibility and models in particular inertial phenomena and other perturbations. New splitting algorithms resulting from the proposed warped iteration constructs are devised in Section 3.2.5 to solve monotone inclusions.

3.2.2 Notation and background

Throughout the paper, $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ are reflexive real Banach spaces. We denote the canonical pairing between $\mathcal{X}$ and its topological dual $\mathcal{X}^*$ by $\langle\cdot,\cdot\rangle$, and by Id the identity operator. The weak convergence of a sequence $(x_n)_{n\in\mathbb{N}}$ to x is denoted by $x_n \rightharpoonup x$, while $x_n \to x$ denotes its strong convergence. The space of bounded linear operators from $\mathcal{X}$ to $\mathcal{Y}$ is denoted by $\mathcal{B}(\mathcal{X},\mathcal{Y})$, and we set $\mathcal{B}(\mathcal{X}) = \mathcal{B}(\mathcal{X},\mathcal{X})$.

Let $M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$. We denote by $\operatorname{gra} M = \big\{(x,x^*) \in \mathcal{X}\times\mathcal{X}^* \mid x^* \in Mx\big\}$ the graph of M, by $\operatorname{dom} M = \big\{x \in \mathcal{X} \mid Mx \neq \varnothing\big\}$ the domain of M, by $\operatorname{ran} M = \big\{x^* \in \mathcal{X}^* \mid (\exists\, x \in \mathcal{X})\ x^* \in Mx\big\}$ the range of M, by $\operatorname{zer} M = \big\{x \in \mathcal{X} \mid 0 \in Mx\big\}$ the set of zeros of M, and by M^{-1} the inverse of M, i.e., $\operatorname{gra} M^{-1} = \big\{(x^*,x) \in \mathcal{X}^*\times\mathcal{X} \mid x^* \in Mx\big\}$. Further, M is monotone if

$$\big(\forall (x,x^*) \in \operatorname{gra} M\big)\big(\forall (y,y^*) \in \operatorname{gra} M\big) \quad \langle x-y, x^*-y^*\rangle \geqslant 0, \tag{3.2}$$

and maximally monotone if, in addition, there exists no monotone operator $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ such that $\operatorname{gra} M \subset \operatorname{gra} A \neq \operatorname{gra} M$. We say that M is uniformly monotone with modulus $\phi\colon\ [0,+\infty[\to [0,+\infty]$ if ϕ is increasing, vanishes only at 0, and

$$\big(\forall (x,x^*) \in \operatorname{gra} M\big)\big(\forall (y,y^*) \in \operatorname{gra} M\big) \quad \langle x-y, x^*-y^*\rangle \geqslant \phi\big(\|x-y\|\big). \tag{3.3}$$

In particular, M is strongly monotone with constant $\alpha \in\,]0,+\infty[$ if it is uniformly monotone with modulus $\phi = \alpha|\cdot|^2$.

Definition 3.2 An operator $M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ is injective if $(\forall x \in \mathcal{X})(\forall y \in \mathcal{X})\ Mx \cap My \neq \varnothing \Rightarrow x = y$.

The following lemma, which concerns a type of duality for monotone inclusions studied in [20, 29, 32], will be instrumental.

Lemma 3.3 *Let $A\colon \mathcal{Y} \to 2^{\mathcal{Y}^*}$ and $B\colon \mathcal{Z} \to 2^{\mathcal{Z}^*}$ be maximally monotone, let $L \in \mathcal{B}(\mathcal{Y},\mathcal{Z})$, let $s^* \in \mathcal{Y}^*$, and let $r \in \mathcal{Z}$. Suppose that $\mathcal{X} = \mathcal{Y}\times\mathcal{Z}\times\mathcal{Z}^*$ (hence $\mathcal{X}^* = \mathcal{Y}^*\times\mathcal{Z}^*\times\mathcal{Z}$), define*

$$M\colon \mathcal{X} \to 2^{\mathcal{X}^*}\colon (x,y,v^*) \mapsto (-s^* + Ax + L^*v^*) \times (By - v^*) \times \{r - Lx + y\}, \tag{3.4}$$

and set $Z = \big\{(x,v^*) \in \mathcal{Y}\times\mathcal{Z}^* \mid s^* - L^*v^* \in Ax \text{ and } Lx - r \in B^{-1}v^*\big\}$. In addition, denote by $\mathscr{P}$ the set of solutions to the primal problem

$$\text{find } x \in \mathcal{Y} \text{ such that } s^* \in Ax + L^*\big(B(Lx-r)\big), \tag{3.5}$$

and by $\mathscr{D}$ the set of solutions to the dual problem

$$\text{find } v^* \in \mathcal{Z}^* \text{ such that } -r \in -L\big(A^{-1}(s^* - L^*v^*)\big) + B^{-1}v^*. \tag{3.6}$$

Then the following hold:

(i) *Z is a closed convex subset of $\mathscr{P} \times \mathscr{D}$.*

(ii) *M is maximally monotone.*

(iii) *Suppose that $(x, y, v^*) \in \operatorname{zer} M$. Then $(x, v^*) \in Z$, $x \in \mathscr{P}$, and $v^* \in \mathscr{D}$.*

(iv) *$\mathscr{P} \neq \varnothing \Leftrightarrow \mathscr{D} \neq \varnothing \Leftrightarrow Z \neq \varnothing \Leftrightarrow \operatorname{zer} M \neq \varnothing$.*

Proof. (i): [20, Proposition 2.1(i)(a)].

(ii): Define

$$\begin{cases} C\colon \mathcal{X} \to 2^{\mathcal{X}^*}\colon (x, y, v^*) \mapsto (-s^* + Ax) \times By \times \{r\} \\ S\colon \mathcal{X} \to \mathcal{X}^*\colon (x, y, v^*) \mapsto (L^*v^*, -v^*, -Lx + y). \end{cases} \tag{3.7}$$

It follows from the maximal monotonicity of A and B that C is maximally monotone. On the other hand, S is linear and bounded, and

$$\big(\forall (x, y, v^*) \in \mathcal{X}\big) \quad \big\langle (x, y, v^*), S(x, y, v^*) \big\rangle = \langle x, L^*v^* \rangle - \langle y, v^* \rangle + \langle y - Lx, v^* \rangle = 0. \tag{3.8}$$

Thus, we derive from [35, Section 17] that S is maximally monotone with $\operatorname{dom} S = \mathcal{X}$. In turn, [35, Theorem 24.1(a)] asserts that $M = C + S$ is maximally monotone.

(iii): We deduce from (3.4) that $s^* \in Ax + L^*v^*$, $v^* \in By$, and $y = Lx - r$; hence $v^* \in B(Lx - r)$. Consequently, $s^* - L^*v^* \in Ax$ and $Lx - r \in B^{-1}v^*$, which yields $(x, v^*) \in Z$. Finally, (i) entails that $x \in \mathscr{P}$ and $v^* \in \mathscr{D}$.

(iv): By [20, Proposition 2.1(i)(c)], $\mathscr{P} \neq \varnothing \Leftrightarrow \mathscr{D} \neq \varnothing \Leftrightarrow Z \neq \varnothing$. In addition, in view of (iii), $\operatorname{zer} M \neq \varnothing \Rightarrow Z \neq \varnothing$. Suppose that $(x, v^*) \in Z$ and set $y = Lx - r$. Then $y = Lx - r \in B^{-1}v^*$ and $s^* \in Ax + L^*v^*$. Hence $0 \in By - v^*$ and $0 \in -s^* + Ax + L^*v^*$. Altogether, $0 \in (-s^* + Ax + L^*v^*) \times (By - v^*) \times \{r - Lx + y\} = M(x, y, v^*)$, i.e., $(x, y, v^*) \in \operatorname{zer} M$. ☐

Now suppose that $\mathcal{X}$ is a real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle$. An operator $T\colon \mathcal{X} \to \mathcal{X}$ is nonexpansive if it is 1-Lipschitzian, α-averaged with $\alpha \in \left]0, 1\right[$ if $\mathrm{Id} + (1/\alpha)(T - \mathrm{Id})$ is nonexpansive, firmly nonexpansive if it is $1/2$-averaged, and β-cocoercive with $\beta \in \left]0, +\infty\right[$ if βT is firmly nonexpansive. Averaged operators were introduced in [4]. The projection operator onto a nonempty closed convex subset C of $\mathcal{X}$ is denoted by proj_C. The resolvent of $M\colon \mathcal{X} \to 2^{\mathcal{X}}$ is $J_M = (\mathrm{Id} + M)^{-1}$.

3.2.3 Warped resolvents

We provide illustrations of Definition 3.1 and then study the properties of warped resolvents.

Our first example is the warped resolvent of a subdifferential. This leads to the following notion, which extends Moreau's classical proximity operator in Hilbert spaces [28].

Example 3.4 (Warped proximity operator) Let D be a nonempty subset of $\mathcal{X}$, let $K\colon D \to \mathcal{X}^*$, and let $\varphi\colon \mathcal{X} \to \left]-\infty, +\infty\right]$ be a proper lower semicontinuous convex function such that ran $K \subset \operatorname{ran}(K + \partial\varphi)$ and $K + \partial\varphi$ is injective. The warped proximity operator of φ with kernel K is $\operatorname{prox}_{\varphi}^{K} = (K + \partial\varphi)^{-1} \circ K$. It is characterized by the variational inequality

$$\big(\forall (x,p) \in \mathcal{X} \times \mathcal{X}\big) \quad p = \operatorname{prox}_{\varphi}^{K} x \quad \Leftrightarrow \quad (\forall y \in \mathcal{X}) \quad \langle y - p, Kx - Kp \rangle + \varphi(p) \leqslant \varphi(y). \tag{3.9}$$

In particular, in the case of normal cones, we arrive at the following definition (see Figure 3.1).

Example 3.5 (Warped projection operator) Let D be a nonempty subset of $\mathcal{X}$, let $K\colon D \to \mathcal{X}^*$, and let C be a nonempty closed convex subset of $\mathcal{X}$ with normal cone operator N_C such that $\operatorname{ran} K \subset \operatorname{ran}(K + N_C)$ and $K + N_C$ is injective. The warped projection operator onto C with kernel K is $\operatorname{proj}_{C}^{K} = (K + N_C)^{-1} \circ K$. It is characterized by

$$\big(\forall (x,p) \in \mathcal{X} \times \mathcal{X}\big) \quad p = \operatorname{proj}_{C}^{K} x \quad \Leftrightarrow \quad \big[p \in C \quad \text{and} \quad (\forall y \in C) \quad \langle y - p, Kx - Kp \rangle \leqslant 0\big]. \tag{3.10}$$

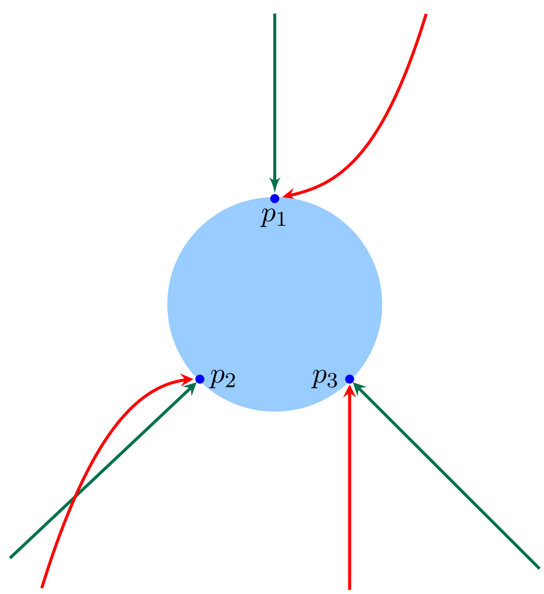

Figure 3.1 Warped projections onto the closed unit ball C centered at the origin in the Euclidean plane. Sets of points projecting onto p_1, p_2, and p_3 for the kernels $K_1 = \operatorname{Id}$ (in green) and $K_2\colon (\xi_1, \xi_2) \mapsto (\xi_1^3/2 + \xi_1/5 - \xi_2, \xi_1 + \xi_2)$ (in red). Note that K_2 is not a gradient.

Example 3.6 Suppose that $\mathcal{X}$ is strictly convex, let $M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be maximally monotone, and let K be the normalized duality mapping of $\mathcal{X}$. Then J_{M}^{K} is a well-defined warped resolvent which was introduced in [26].

Example 3.7 Let $M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be maximally monotone such that $\operatorname{zer} M \neq \varnothing$, let $f\colon \mathcal{X} \to$

$]-\infty,+\infty]$ be a Legendre function [6] such that $\operatorname{dom} M \subset \operatorname{int}\operatorname{dom} f$, and set $K = \nabla f$. Then it follows from [6, Corollary 3.14(ii)] that J_M^K is a well-defined warped resolvent, called the D-resolvent of M in [6].

Example 3.8 Let $M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be maximally monotone and let $K\colon \mathcal{X} \to \mathcal{X}^*$ be strictly monotone, surjective, and 3^* monotone in the sense that [39, Definition 32.40(c)]

$$(\forall x \in \operatorname{dom} M)(\forall x^* \in \operatorname{ran} M) \quad \sup_{(y,y^*)\in\operatorname{gra} M} \langle x-y, y^*-x^*\rangle < +\infty. \tag{3.11}$$

Then it follows from [8, Theorem 2.3] that J_M^K is a well-defined warped resolvent, called the K-resolvent of M in [8].

Example 3.9 Let $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ and $B\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be maximally monotone, and let $f\colon \mathcal{X} \to]-\infty,+\infty]$ be a proper lower semicontinuous convex function which is essentially smooth [6]. 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 $K\colon D \to \mathcal{X}^*\colon x \mapsto \nabla f(x) - Bx$. Then the warped resolvent J_M^K is well defined and coincides with the Bregman forward-backward operator $(\nabla f + A)^{-1} \circ (\nabla f - B)$ investigated in [13], where it is shown to capture a construction found in [31].

Example 3.10 Consider the setting of Lemma 3.3. For simplicity (more general kernels can be considered), take $s^* = 0$, $r = 0$, and assume that $\mathcal{Y}$ and $\mathcal{Z}^*$ are strictly convex with normalized duality mapping $K_{\mathcal{Y}}$ and $K_{\mathcal{Z}^*}$. As seen in Lemma 3.3(i), finding a zero of the Kuhn–Tucker operator $U\colon \mathcal{Y}\times\mathcal{Z}^* \to 2^{\mathcal{Y}^*\times\mathcal{Z}}\colon (x,v^*) \mapsto (Ax + L^*v^*) \times (B^{-1}v^* - Lx)$ provides a solution to the primal-dual problem (3.5)–(3.6). Now set $K\colon (x,v^*) \mapsto (K_{\mathcal{Y}}x - L^*v^*, Lx + K_{\mathcal{Z}^*}v^*)$. Then the warped resolvent J_U^K is well defined and

$$J_U^K\colon (x,v^*) \mapsto \big((K_{\mathcal{Y}} + A)^{-1}(K_{\mathcal{Y}}x - L^*v^*), (K_{\mathcal{Z}^*} + B^{-1})^{-1}(Lx + K_{\mathcal{Z}^*}v^*)\big). \tag{3.12}$$

For instance, in a Hilbertian setting, $J_U^K\colon (x,v^*) \mapsto (J_A(x - L^*v^*), J_{B^{-1}}(Lx + v^*))$, whereas J_U is intractable; note also that the kernel K is a non-Hermitian bounded linear operator.

Further examples will appear in Section 3.2.5. Let us turn our attention to the properties of warped resolvents.

Proposition 3.11 (viability) *Let D be a nonempty subset of $\mathcal{X}$, let $K\colon D \to \mathcal{X}^*$, and let $M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be such that $\operatorname{ran} K \subset \operatorname{ran}(K+M)$ and $K+M$ is injective. Then $J_M^K\colon D \to D$.*

Proof. By assumption, $\operatorname{dom} J_M^K = \operatorname{dom}((K+M)^{-1}\circ K) = \big\{x \in \operatorname{dom} K \mid Kx \in \operatorname{dom}(K+M)^{-1}\big\} = \big\{x \in D \mid Kx \in \operatorname{ran}(K+M)\big\} = D$. Next, observe that

$$\operatorname{ran} J_M^K = \operatorname{ran}\big((K+M)^{-1} \circ K\big) \subset \operatorname{ran}(K+M)^{-1} = \operatorname{dom}(K+M) \subset \operatorname{dom} K = D. \tag{3.13}$$

Finally, to show that $(K+M)^{-1}$ is at most single-valued, suppose that $(x^*,x_1)\in \operatorname{gra}(K+M)^{-1}$ and $(x^*,x_2)\in\operatorname{gra}(K+M)^{-1}$. Then $\{x^*\}\subset(K+M)x_1\cap(K+M)x_2$ and, since $K+M$ is injective, it follows that $x_1=x_2$. ☐

Sufficient conditions that guarantee that warped resolvents are well defined are made explicit below.

Proposition 3.12 *Let D be a nonempty subset of $\mathcal{X}$, let $K\colon D\to\mathcal{X}^*$, and let $M\colon\mathcal{X}\to 2^{\mathcal{X}^*}$. Then the following hold:*

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

[a] *$K+M$ is surjective.*

[b] *$K+M$ is maximally monotone and $D\cap\operatorname{dom}M$ is bounded.*

[c] *$K+M$ is maximally monotone, $K+M$ is uniformly monotone with modulus ϕ, and $\phi(t)/t\to+\infty$ as $t\to+\infty$.*

[d] *$K+M$ is maximally monotone and strongly monotone.*

[e] *M is maximally monotone, $D=\mathcal{X}$, and K is maximally monotone, strictly monotone, 3^* monotone, and surjective.*

[f] *K is maximally monotone and there exists a lower semicontinuous coercive convex function $\varphi\colon\mathcal{X}\to\mathbb{R}$ such that $M=\partial\varphi$.*

Then $\operatorname{ran}K\subset\operatorname{ran}(K+M)$.

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

[a] *$K+M$ is strictly monotone.*

[b] *M is monotone and K is strictly monotone on* $\operatorname{dom}M$.

[c] *K is monotone and M is strictly monotone.*

[d] *$-(K+M)$ is strictly monotone.*

Then $K+M$ is injective.

Proof. Set $A=K+M$.

(i): Item [a] is clear. We prove the remaining ones as follows.

[b]: It follows from [39, Theorem 32.G] that $\operatorname{ran}A=\mathcal{X}\supset\operatorname{ran}K$.

[c]&[d]: Since [20, Lemma 2.7(ii)] and [39, Corollary 32.35] assert that A is surjective, the claim follows from (i)[a].

[e]: See [8, Theorem 2.3].

[f]: Take $z\in D$ and set $B=A(\cdot+z)-Kz$. By coercivity of φ, there exists $\rho\in\,]0,+\infty[$ such that

$$(\forall x\in\mathcal{X})\quad \|x\|\geqslant\rho\quad\Rightarrow\quad \inf\langle x,\partial\varphi(x+z)\rangle\geqslant\varphi(x+z)-\varphi(z)\geqslant 0. \tag{3.14}$$

Now take $(x,x^*)\in\operatorname{gra}B$ and suppose that $\|x\|\geqslant\rho$. Then $x^*+Kz-K(x+z)\in\partial\varphi(x+z)$ and it follows from (3.14) and the monotonicity of K that

$$0\leqslant\langle x,x^*+Kz-K(x+z)\rangle=\langle x,x^*\rangle-\langle(x+z)-z,K(x+z)-Kz\rangle\leqslant\langle x,x^*\rangle. \tag{3.15}$$

On the other hand, since $\operatorname{dom}\partial\varphi = \mathcal{X}$ [38, Theorems 2.2.20(b) and 2.4.12], A is maximally monotone [35, Theorem 24.1(a)], and so is B. Altogether, [33, Proposition 2] asserts that there exists $\overline{x} \in \mathcal{X}$ such that $0 \in B\overline{x}$. Consequently, $Kz \in A(\overline{x}+z) \subset \operatorname{ran}(K+M)$.

(ii): We need to prove only [a] since [b] and [c] are special cases of it, and [d] is similar. To this end, let $(x_1, x_2) \in \mathcal{X}^2$ and suppose that $Ax_1 \cap Ax_2 \neq \varnothing$. We must show that $x_1 = x_2$. Take $x^* \in Ax_1 \cap Ax_2$. Then (x_1, x^*) and (x_2, x^*) lie in $\operatorname{gra} A$. In turn, since A is strictly monotone and $\langle x_1 - x_2, x^* - x^*\rangle = 0$, we obtain $x_1 = x_2$. ☐

Proposition 3.13 *Let $M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$, let $\gamma \in \left]0,+\infty\right[$, and let $K\colon \mathcal{X} \to \mathcal{X}^*$ be such that $\operatorname{ran} K \subset \operatorname{ran}(K+\gamma M)$ and $K+\gamma M$ is injective. Then the following hold:*

(i) $\operatorname{Fix} J^K_{\gamma M} = \operatorname{zer} M$.

(ii) *Let $x \in \mathcal{X}$ and $p \in \mathcal{X}$. Then $p = J^K_{\gamma M}x \Leftrightarrow (p, \gamma^{-1}(Kx - Kp)) \in \operatorname{gra} M$.*

(iii) *Suppose that M is monotone. Let $x \in \mathcal{X}$ and $y \in \mathcal{X}$, and set $p = J^K_{\gamma M}x$ and $q = J^K_{\gamma M}y$. Then $\langle p-q, Kx-Ky\rangle \geqslant \langle p-q, Kp-Kq\rangle$.*

(iv) *Suppose that M is monotone, that K is uniformly continuous and ϕ-uniformly monotone, and that $\psi\colon t \mapsto \phi(t)/t$ is real-valued on $\left]0,\xi\right[$ for some $\xi \in \left]0,+\infty\right[$ and strictly increasing. Then $J^K_{\gamma M}$ is uniformly continuous.*

(v) *Suppose that M is monotone and that K is β-Lipschitzian and α-strongly monotone for some $\alpha \in \left]0,+\infty\right[$ and $\beta \in \left]0,+\infty\right[$. Then $J^K_{\gamma M}$ is (β/α)-Lipschitzian.*

(vi) *Suppose that M is monotone. Let $x \in \mathcal{X}$, and set $y = J^K_{\gamma M}x$ and $y^* = \gamma^{-1}(Kx - Ky)$. Then $\operatorname{zer} M \subset \big\{z \in \mathcal{X} \mid \langle z-y, y^*\rangle \leqslant 0\big\}$.*

Proof. (i): We derive from Proposition 3.11 that $(\forall x \in \mathcal{X})\ x \in \operatorname{zer} M \Leftrightarrow Kx \in Kx + \gamma Mx \Leftrightarrow x = J^K_{\gamma M}x \Leftrightarrow x \in \operatorname{Fix} J^K_{\gamma M}$.

(ii): We have $p = J^K_{\gamma M}x \Leftrightarrow p = (K+\gamma M)^{-1}(Kx) \Leftrightarrow Kx \in Kp + \gamma Mp \Leftrightarrow Kx - Kp \in \gamma Mp \Leftrightarrow (p, \gamma^{-1}(Kx-Kp)) \in \operatorname{gra} M$.

(iii): This follows from (ii) and the monotonicity of M.

(iv): Let x and y be in $\mathcal{X}$, and set $p = J^K_{\gamma M}x$ and $q = J^K_{\gamma M}y$. Then we deduce from (iii) that

$$\phi(\|p-q\|) \leqslant \langle p-q, Kp-Kq\rangle \leqslant \langle p-q, Kx-Ky\rangle \leqslant \|p-q\|\,\|Kx-Ky\|. \tag{3.16}$$

Now fix $\varepsilon \in \left]0,\xi\right[$ and let $\eta \in \left]0,\psi(\varepsilon)\right]$. By uniform continuity of K, there exists $\delta \in \left]0,+\infty\right[$ such that $\|x-y\| \leqslant \delta \Rightarrow \|Kx-Ky\| \leqslant \eta$. Without loss of generality, suppose that $p \neq q$. Then, if $\|x-y\| \leqslant \delta$, we derive from (3.16) that $\psi(\|p-q\|) \leqslant \|Kx-Ky\| \leqslant \eta \leqslant \psi(\varepsilon)$. Consequently, since ψ is strictly increasing, $\|p-q\| \leqslant \varepsilon$.

(v): Let x and y be in $\mathcal{X}$ and set $p = J^K_{\gamma M}x$ and $q = J^K_{\gamma M}y$. Then we deduce from (iii) that

$$\alpha\|p-q\|^2 \leqslant \langle p-q, Kp-Kq\rangle \leqslant \langle p-q, Kx-Ky\rangle \leqslant \|p-q\|\,\|Kx-Ky\| \leqslant \beta\|p-q\|\,\|x-y\|. \tag{3.17}$$

In turn, $\|p-q\| \leqslant (\beta/\alpha)\|x-y\|$.

(vi): Suppose that $z \in \operatorname{zer} M$. Then $(z, 0) \in \operatorname{gra} M$. On the other hand, we derive from (ii) that $(y, y^*) \in \operatorname{gra} M$. Hence, by monotonicity of M, $\langle y - z, y^* \rangle \geqslant 0$. ☐

In Hilbert spaces, standard resolvents are firmly nonexpansive, hence $1/2$-averaged. A related property for warped resolvents is the following.

Proposition 3.14 *Suppose that $\mathcal{X}$ is a Hilbert space. Let $M \colon \mathcal{X} \to 2^{\mathcal{X}}$ be maximally monotone and let $K \colon \mathcal{X} \to \mathcal{X}$ be averaged with constant $\alpha \in \left]0,1\right[$. Suppose that $K + M$ is 1-strongly monotone. Then J_M^K is averaged with constant $1/(2-\alpha)$.*

Proof. Since K is nonexpansive by virtue of [7, Remark 4.34(i)], it follows from the Cauchy–Schwarz inequality that

$$
\begin{aligned}
(\forall x \in \mathcal{X})(\forall y \in \mathcal{X}) \quad \langle x - y \mid (2\mathrm{Id} + K)x - (2\mathrm{Id} + K)y \rangle &= 2\|x - y\|^2 + \langle x - y \mid Kx - Ky \rangle \\
&\geqslant 2\|x - y\|^2 - \|x - y\|^2 \\
&= \|x - y\|^2
\end{aligned}
\tag{3.18}
$$

and therefore, by continuity of $2\mathrm{Id} + K$, that $2\mathrm{Id} + K$ is maximally monotone [7, Corollary 20.28]. Thus, in the light of [7, Corollary 25.5(i)], $2\mathrm{Id} + K + M$ is maximally monotone. In turn, since $2\mathrm{Id} + K + M$ is strongly monotone by (3.18), [7, Proposition 22.11(ii)] entails that $\operatorname{ran}(3\mathrm{Id} + K + M - \mathrm{Id}) = \operatorname{ran}(2\mathrm{Id} + K + M) = \mathcal{X}$, which yields $\operatorname{ran}(\mathrm{Id} + (K + M - \mathrm{Id})/3) = \mathcal{X}$. Hence, by monotonicity of $K + M - \mathrm{Id}$ and Minty's theorem [7, Theorem 21.1], we infer that $K + M - \mathrm{Id}$ is maximally monotone. Thus, in view of [7, Corollary 23.9], $(K + M)^{-1} = (\mathrm{Id} + K + M - \mathrm{Id})^{-1}$ is averaged with constant $1/2$. Consequently, we infer from [7, Proposition 4.44] that $J_M^K = (K + M)^{-1} \circ K$ is averaged with constant $1/(2 - \alpha)$. ☐

3.2.4 Warped proximal iterations

Throughout this section, $\mathcal{X}$ is a real Hilbert space identified with its dual. We start with an abstract principle for the basic problem of finding a zero of a maximally monotone operator.

Proposition 3.15 *Let $M \colon \mathcal{X} \to 2^{\mathcal{X}}$ be a maximally monotone operator such that $Z = \operatorname{zer} M \neq \varnothing$, let $x_0 \in \mathcal{X}$, let $\varepsilon \in \left]0,1\right[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 2-\varepsilon]$, and let $(y_n, y_n^*)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} M$. Set*

$$
(\forall n \in \mathbb{N}) \quad x_{n+1} =
\begin{cases}
x_n + \dfrac{\lambda_n \langle y_n - x_n \mid y_n^* \rangle}{\|y_n^*\|^2}\, y_n^*, & \text{if } \langle y_n - x_n \mid y_n^* \rangle < 0; \\
x_n, & \text{otherwise.}
\end{cases}
\tag{3.19}
$$

Then the following hold:

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

(ii) *Suppose that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ is in Z. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.*

Proof. By [7, Proposition 23.39], Z is a nonempty closed convex subset of $\mathcal{X}$. Set $(\forall n \in \mathbb{N})$ $H_n = \big\{z \in \mathcal{X} \mid \langle z - y_n \mid y_n^* \rangle \leqslant 0\big\}$. For every $z \in Z$ and every $n \in \mathbb{N}$, since $(z, 0)$ and (y_n, y_n^*) lie in gra M, the monotonicity of M forces $\langle y_n - z \mid y_n^* \rangle \geqslant 0$. Thus $Z \subset \bigcap_{n\in\mathbb{N}} H_n$. In addition, [7, Example 29.20] asserts that

$$(\forall n \in \mathbb{N}) \quad \text{proj}_{H_n} x_n = \begin{cases} x_n + \dfrac{\langle y_n - x_n \mid y_n^* \rangle}{\|y_n^*\|^2}\, y_n^*, & \text{if } \langle y_n - x_n \mid y_n^* \rangle < 0; \\ x_n, & \text{otherwise.} \end{cases} \tag{3.20}$$

Hence, we derive from (3.19) that

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n(\text{proj}_{H_n} x_n - x_n). \tag{3.21}$$

Therefore (i) follows from [16, Equation (10)] and (ii) follows from [16, Proposition 6i)]. ☐

To implement the conceptual principle outlined in Proposition 3.15, one is required to construct points in the graph of the underlying monotone operator. Towards this end, our strategy is to use Proposition 3.13(ii). We shall then seamlessly obtain in Section 3.2.5 a broad class of algorithms to solve a variety of monotone inclusions. It will be convenient to use the notation

$$(\forall y^* \in \mathcal{Y}^*) \quad (y^*)^\sharp = \begin{cases} \dfrac{y^*}{\|y^*\|}, & \text{if } y^* \neq 0; \\ 0, & \text{if } y^* = 0. \end{cases} \tag{3.22}$$

Our first method employs, at iteration n, a warped resolvent based on a different kernel, and this warped resolvent is applied at a point $\widetilde{x}_n$ that may not be the current iterate x_n.

Theorem 3.16 *Let $M \colon \mathcal{X} \to 2^{\mathcal{X}}$ be a maximally monotone operator such that $Z = \text{zer}\, M \neq \varnothing$, let $x_0 \in \mathcal{X}$, let $\varepsilon \in \left]0, 1\right[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 2 - \varepsilon]$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\left[\varepsilon, +\infty\right[$. Further, for every $n \in \mathbb{N}$, let $\widetilde{x}_n \in \mathcal{X}$ and let $K_n \colon \mathcal{X} \to \mathcal{X}$ be a monotone operator such that $\text{ran}\, K_n \subset \text{ran}(K_n + \gamma_n M)$ and $K_n + \gamma_n M$ is injective. Iterate*

$$\begin{array}{l} \textit{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_n = J^{K_n}_{\gamma_n M} \widetilde{x}_n \\ y_n^* = \gamma_n^{-1}(K_n \widetilde{x}_n - K_n y_n) \\ \textit{if } \langle y_n - x_n \mid y_n^* \rangle < 0 \\ \left\lfloor \; x_{n+1} = x_n + \dfrac{\lambda_n \langle y_n - x_n \mid y_n^* \rangle}{\|y_n^*\|^2}\, y_n^* \right. \\ \textit{else} \\ \left\lfloor \; x_{n+1} = x_n. \right. \end{array} \right. \end{array} \tag{3.23}$$

Then the following hold:

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

(ii) *Suppose that the following are satisfied:*

[a] $\widetilde{x}_n - x_n \to 0$.

[b] $\big\langle \widetilde{x}_n - y_n \mid (K_n\widetilde{x}_n - K_n y_n)^\sharp \big\rangle \to 0 \quad \Rightarrow \quad \begin{cases} \widetilde{x}_n - y_n \rightharpoonup 0 \\ K_n\widetilde{x}_n - K_n y_n \to 0. \end{cases}$

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

Proof. By Proposition 3.13(ii),

$$(\forall n \in \mathbb{N}) \quad (y_n, y_n^*) \in \operatorname{gra} M. \tag{3.24}$$

Therefore, (i) follows from Proposition 3.15(i). It remains to prove (ii). To this end, take a strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ and a point $x \in \mathcal{X}$ such that $x_{k_n} \rightharpoonup x$. In view of Proposition 3.15(ii), we must show that $x \in Z$. We infer from (ii)[a] that

$$\widetilde{x}_{k_n} \rightharpoonup x. \tag{3.25}$$

Next, by (3.22) and (3.23), for every $n \in \mathbb{N}$, if $\langle x_n - y_n \mid y_n^* \rangle > 0$, then $y_n^* \neq 0$ and

$$\big\langle x_n - y_n \mid (y_n^*)^\sharp \big\rangle = \frac{\langle x_n - y_n \mid y_n^* \rangle}{\|y_n^*\|} = \lambda_n^{-1}\|x_{n+1} - x_n\| \leqslant \varepsilon^{-1}\|x_{n+1} - x_n\|; \tag{3.26}$$

otherwise, $\langle x_n - y_n \mid y_n^* \rangle \leqslant 0$ and it thus results from (3.22) that

$$\begin{aligned} \big\langle x_n - y_n \mid (y_n^*)^\sharp \big\rangle &= \begin{cases} 0, & \text{if } y_n^* = 0; \\ \dfrac{\langle x_n - y_n \mid y_n^* \rangle}{\|y_n^*\|}, & \text{otherwise} \end{cases} \\ &\leqslant 0 \\ &= \varepsilon^{-1}\|x_{n+1} - x_n\|. \end{aligned} \tag{3.27}$$

Therefore, using (i) and the monotonicity of $(K_n)_{n\in\mathbb{N}}$, we obtain

$$\begin{aligned} 0 \leftarrow \varepsilon^{-1}&\|x_{n+1} - x_n\| \\ &\geqslant \langle x_n - y_n \mid (y_n^*)^\sharp \rangle \\ &= \big\langle x_n - \widetilde{x}_n \mid (K_n\widetilde{x}_n - K_n y_n)^\sharp \big\rangle + \big\langle \widetilde{x}_n - y_n \mid (K_n\widetilde{x}_n - K_n y_n)^\sharp \big\rangle \\ &\geqslant \big\langle x_n - \widetilde{x}_n \mid (K_n\widetilde{x}_n - K_n y_n)^\sharp \big\rangle. \end{aligned} \tag{3.28}$$

However, by the Cauchy–Schwarz inequality and (ii)[a],

$$\big|\big\langle x_n - \widetilde{x}_n \mid (K_n\widetilde{x}_n - K_n y_n)^\sharp \big\rangle\big| \leqslant \|x_n - \widetilde{x}_n\| \to 0. \tag{3.29}$$

Hence, (3.28) implies that $\langle \widetilde{x}_n - y_n \mid (K_n\widetilde{x}_n - K_n y_n)^\sharp \rangle \to 0$. In turn, we deduce from (ii)[b] that $\widetilde{x}_n - y_n \rightharpoonup 0$ and $K_n\widetilde{x}_n - K_n y_n \to 0$. Altogether, since $\sup_{n\in\mathbb{N}} \gamma_n^{-1} \leqslant \varepsilon^{-1}$, it follows from

(3.24) and (3.25) that

$$y_{k_n} = \widetilde{x}_{k_n} + (y_{k_n} - \widetilde{x}_{k_n}) \rightharpoonup x \tag{3.30}$$

and

$$My_{k_n} \ni y^*_{k_n} = \gamma_{k_n}^{-1}(K_{k_n}\widetilde{x}_{k_n} - K_{k_n}y_{k_n}) \to 0. \tag{3.31}$$

Appealing to the maximal monotonicity of M, [7, Proposition 20.38(ii)] allows us to conclude that $x \in Z$. ☐

Remark 3.17 Condition (ii)[b] in Theorem 3.16 is satisfied in particular when there exist α and β in $]0,+\infty[$ such that the kernels $(K_n)_{n\in\mathbb{N}}$ are α-strongly monotone and β-Lipschitzian.

Remark 3.18 The auxiliary sequence $(\widetilde{x}_n)_{n\in\mathbb{N}}$ in Theorem 3.16 can serve several purposes. In general, it provides the flexibility of not applying the warped resolvent to the current iterate. Here are some noteworthy candidates.

(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 many methods, e.g., [11, 17, 21, 37].

(ii) Mimicking the behavior of so-called inertial methods [3, 19], let $(\alpha_n)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathbb{R}$ and set $(\forall n \in \mathbb{N}\smallsetminus\{0\})\ \widetilde{x}_n = x_n + \alpha_n(x_n - x_{n-1})$. Then Theorem 3.16(i) yields $\|\widetilde{x}_n - x_n\| = |\alpha_n|\,\|x_n - x_{n-1}\| \to 0$ and therefore assumption (ii)[a] holds in Theorem 3.16. More generally, weak convergence results can be derived from Theorem 3.16 for iterations with memory, that is,

$$(\forall n \in \mathbb{N}) \quad \widetilde{x}_n = \sum_{j=0}^{n} \mu_{n,j} x_j, \quad \text{where} \quad (\mu_{n,j})_{0\leqslant j\leqslant n} \in \mathbb{R}^{n+1} \quad \text{and} \quad \sum_{j=0}^{n} \mu_{n,j} = 1. \tag{3.32}$$

Here condition (ii)[a] holds 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 $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ [19].

(iii) Nonlinear perturbations can also be considered. For instance, at iteration n, $\widetilde{x}_n = \operatorname{proj}_{C_n} x_n$ is an approximation to x_n from some suitable closed convex set $C_n \subset \mathcal{X}$.

Remark 3.19 The independent work [23] was posted on arXiv at the same time as the report [14] from which our paper is derived. The former uses a notion of resolvents subsumed by Definition 3.1 to explore the application of an algorithm similar to (3.23) with no perturbation (i.e., for every $n \in \mathbb{N}$, $\widetilde{x}_n = x_n$). The work [23] nicely complements ours in the sense that it proposes applications to splitting schemes not discussed here, which further attests to the versatility and effectiveness of the notion of warped proximal iterations.

We now turn our attention to a variant of Theorem 3.16 that guarantees strong convergence of the iterates to a best approximation. In the spirit of Haugazeau's algorithm (see [24, Théorème 3-2] and [7, Corollary 30.15]), it involves outer approximations consisting of the intersection of two half-spaces. For convenience, given $(x,y,z)\in\mathcal{X}^3$, we set

$$H(x,y)=\big\{u\in\mathcal{X}\mid\langle u-y\mid x-y\rangle\leqslant 0\big\} \tag{3.33}$$

and, if $R=H(x,y)\cap H(y,z)\neq\varnothing$, $Q(x,y,z)=\operatorname{proj}_R x$. The latter can be computed explicitly as follows (see [24, Théorème 3-1] or [7, Corollary 29.25]).

Lemma 3.20 *Let $(x,y,z)\in\mathcal{X}^3$. Set $R=H(x,y)\cap H(y,z)$, $\chi=\langle x-y\mid y-z\rangle$, $\mu=\|x-y\|^2$, $\nu=\|y-z\|^2$, and $\rho=\mu\nu-\chi^2$. Then exactly one of the following holds:*

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

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

$$Q(x,y,z)=\begin{cases} z, & \text{if } \rho=0 \text{ and } \chi\geqslant 0;\\ x+(1+\chi/\nu)(z-y), & \text{if } \rho>0 \text{ and } \chi\nu\geqslant\rho;\\ y+(\nu/\rho)\big(\chi(x-y)+\mu(z-y)\big), & \text{if } \rho>0 \text{ and } \chi\nu<\rho. \end{cases} \tag{3.34}$$

Our second abstract convergence principle can now be stated.

Proposition 3.21 *Let $M\colon\mathcal{X}\to 2^{\mathcal{X}}$ be a maximally monotone operator such that $Z=\operatorname{zer}M\neq\varnothing$, let $x_0\in\mathcal{X}$, and let $(y_n,y_n^*)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra}M$. For every $n\in\mathbb{N}$, set*

$$x_{n+1/2}=\begin{cases} x_n+\dfrac{\langle y_n-x_n\mid y_n^*\rangle}{\|y_n^*\|^2}\,y_n^*, & \text{if } \langle y_n-x_n\mid y_n^*\rangle<0;\\ x_n, & \text{otherwise} \end{cases} \quad\text{and}\quad x_{n+1}=Q\big(x_0,x_n,x_{n+1/2}\big). \tag{3.35}$$

Then the following hold:

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

(ii) *Suppose that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ is in Z. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to $\operatorname{proj}_Z x_0$.*

Proof. Set $(\forall n\in\mathbb{N})$ $H_n=\big\{z\in\mathcal{X}\mid\langle z-y_n\mid y_n^*\rangle\leqslant 0\big\}$. Then, as in the proof of Proposition 3.15, Z is a nonempty closed convex subset of $\mathcal{X}$ and $Z\subset\bigcap_{n\in\mathbb{N}}H_n$. On the one hand,

$$(\forall n\in\mathbb{N})\quad x_{n+1/2}=\operatorname{proj}_{H_n}x_n \quad\text{and}\quad x_{n+1}=Q\big(x_0,x_n,x_{n+1/2}\big). \tag{3.36}$$

On the other hand, by (3.33),

$$\begin{aligned}(\forall n\in\mathbb{N})\quad H\big(x_n,x_{n+1/2}\big)&=\begin{cases}\mathcal{X}, & \text{if } x\in H_n;\\ H_n, & \text{otherwise}\end{cases}\\ &\supset Z.\end{aligned} \tag{3.37}$$

The claims therefore follow from [2, Proposition 2.1]. ☐

Theorem 3.22 *Let $M\colon \mathcal{X} \to 2^{\mathcal{X}}$ be a maximally monotone operator such that $Z = \operatorname{zer} M \neq \varnothing$, let $x_0 \in \mathcal{X}$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$ such that $\inf_{n\in\mathbb{N}} \gamma_n > 0$. For every $n \in \mathbb{N}$, let $\widetilde{x}_n \in \mathcal{X}$ and let $K_n\colon \mathcal{X} \to \mathcal{X}$ be a monotone operator such that $\operatorname{ran} K_n \subset \operatorname{ran}(K_n + \gamma_n M)$ and $K_n + \gamma_n M$ is injective. Iterate*

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
y_n = J^{K_n}_{\gamma_n M}\widetilde{x}_n\\
y_n^* = \gamma_n^{-1}(K_n\widetilde{x}_n - K_n y_n)\\
\text{if } \langle y_n - x_n \mid y_n^*\rangle < 0\\
\left\lfloor
\begin{array}{l}
x_{n+1/2} = x_n + \dfrac{\langle y_n - x_n \mid y_n^*\rangle}{\|y_n^*\|^2}\, y_n^*
\end{array}
\right.\\
\text{else}\\
\left\lfloor
\begin{array}{l}
x_{n+1/2} = x_n
\end{array}
\right.\\
x_{n+1} = Q(x_0, x_n, x_{n+1/2}).
\end{array}
\right.
\end{array}
\tag{3.38}
$$

Then the following hold:

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

(ii) *Suppose that the following are satisfied:*

[a] $\widetilde{x}_n - x_n \to 0$.

[b] $\left\langle \widetilde{x}_n - y_n \mid (K_n\widetilde{x}_n - K_n y_n)^\sharp \right\rangle \to 0 \quad \Rightarrow \quad \begin{cases} \widetilde{x}_n - y_n \rightharpoonup 0\\ K_n\widetilde{x}_n - K_n y_n \to 0. \end{cases}$

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

Proof. Proposition 3.13(ii) asserts that $(\forall n \in \mathbb{N})$ $(y_n, y_n^*) \in \operatorname{gra} M$. Thus, we obtain (i) from Proposition 3.21(i). In the light of Proposition 3.21(ii), to establish (ii), we need to show that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ is a zero of M. Since (i) asserts that $x_{n+1/2} - x_n \to 0$, this is done as in the proof of Theorem 3.16(ii). ☐

We complete this section with the following remarks.

Remark 3.23 Suppose that $\mathcal{Y}$ and $\mathcal{Z}$ are real Hilbert spaces and that $\mathcal{X} = \mathcal{Y}\times\mathcal{Z}$. Let $A\colon \mathcal{Y} \to 2^{\mathcal{Y}}$ and $B\colon \mathcal{Z} \to 2^{\mathcal{Z}}$ be maximally monotone, and let $L \in \mathcal{B}(\mathcal{Y},\mathcal{Z})$. Define

$$
M\colon \mathcal{X} \to 2^{\mathcal{X}}\colon (x, v^*) \mapsto (Ax + L^*v^*) \times (-Lx + B^{-1}v^*). \tag{3.39}
$$

In [1,2,18] the problem of finding a zero of M (and hence a solution to the monotone inclusion $0 \in Ax + L^*(B(Lx))$) is approached by generating, at each iteration n, points $(a_n, a_n^*) \in \operatorname{gra} A$ and $(b_n, b_n^*) \in \operatorname{gra} B$. This does provide a point $(y_n, y_n^*) = ((a_n, b_n^*), (a_n^* + L^*b_n^*, -La_n + b_n)) \in \operatorname{gra} M$, which shows that the algorithms proposed in [1,2,18] are actually instances of the conceptual principles laid out in Propositions 3.15 and 3.21. In particular, the primal-dual framework of [1] corresponds to applying Theorem 3.16 to the operator M of (3.39) with kernels

$$(\forall n \in \mathbb{N}) \quad K_n \colon \mathcal{X} \to \mathcal{X} \colon (x, v^*) \mapsto \big(\gamma_n^{-1} x - L^* v^*, Lx + \mu_n v^*\big). \tag{3.40}$$

Likewise, that of [2] corresponds to the application of Theorem 3.22 to this setting.

Remark 3.24 In Theorems 3.16 and 3.22, the algorithms operate by using a single point (y_n, y_n^*) in $\operatorname{gra} M$ at iteration n. It may be advantageous to use a finite family $(y_{i,n}, y_{i,n}^*)_{i \in I_n}$ of points in $\operatorname{gra} M$, say

$$(\forall i \in I_n) \quad (y_{i,n}, y_{i,n}^*) = \Big(J_{\gamma_{i,n} M}^{K_{i,n}} \widetilde{x}_{i,n}, \gamma_{i,n}^{-1}(K_{i,n}\widetilde{x}_{i,n} - K_{i,n} y_{i,n})\Big). \tag{3.41}$$

By monotonicity of M, $(\forall i \in I_n)(\forall z \in \operatorname{zer} M)$ $\langle z \mid y_{i,n}^* \rangle \leqslant \langle y_{i,n} \mid y_{i,n}^* \rangle$. Therefore, using ideas found in the area of convex feasibility algorithms [15, 27], at every iteration n, given strictly positive weights $(\omega_{i,n})_{i \in I_n}$ adding up to 1, we average these inequalities to create a new half-space H_n containing $\operatorname{zer} M$, namely

$$\operatorname{zer} M \subset H_n = \big\{z \in \mathcal{X} \mid \langle z \mid y_n^* \rangle \leqslant \eta_n\big\}, \quad \text{where} \quad \begin{cases} y_n^* = \sum_{i \in I_n} \omega_{i,n} y_{i,n}^* \\ \eta_n = \sum_{i \in I_n} \omega_{i,n} \langle y_{i,n} \mid y_{i,n}^* \rangle. \end{cases} \tag{3.42}$$

Now set

$$\Lambda_n = \begin{cases} \dfrac{\sum_{i \in I_n} \omega_{i,n} \langle y_{i,n} - x_n \mid y_{i,n}^* \rangle}{\big\| \sum_{i \in I_n} \omega_{i,n} y_{i,n}^* \big\|^2}, & \text{if } \sum_{i \in I_n} \omega_{i,n} \langle x_n - y_{i,n} \mid y_{i,n}^* \rangle > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{3.43}$$

Then, employing $\operatorname{proj}_{H_n} x_n = x_n + \Lambda_n \sum_{i \in I_n} \omega_{i,n} y_{i,n}^*$ as the point x_{n+1} in (3.23) and as the point $x_{n+1/2}$ in (3.38) results in multi-point extensions of Theorems 3.16 and 3.22.

3.2.5 Applications

We apply Theorem 3.16 to design new algorithms to solve complex monotone inclusion problems in a real Hilbert space $\mathcal{X}$. We do not mention explicitly minimization problems as they follow, with usual constraint qualification conditions, by considering monotone inclusions involving subdifferentials as maximally monotone operators [7, 17]. For brevity, we do not mention either the strongly convergent counterparts of each of the corollaries below that can be systematically obtained using Theorem 3.22.

Let us note that the most basic instantiation of Theorem 3.16 is obtained by setting $(\forall n \in \mathbb{N})$ $K_n = \mathrm{Id}$, $\widetilde{x}_n = x_n$, and $\lambda_n = 1$. In this case, the warped proximal algorithm (3.23) reduces to the basic proximal point algorithm (3.1).

In connection with Remark 3.18, let us first investigate the convergence of a novel perturbed forward-backward-forward algorithm with memory. This will require the following fact.

Lemma 3.25 *Let $B\colon \mathcal{X} \to \mathcal{X}$ be Lipschitzian with constant $\beta \in \left]0,+\infty\right[$, let $W\colon \mathcal{X} \to \mathcal{X}$ be strongly monotone with constant $\alpha \in \left]0,+\infty\right[$, let $\varepsilon \in \left]0,\alpha\right[$, let $\gamma \in \left]0,(\alpha-\varepsilon)/\beta\right]$, and set $K = W - \gamma B$. Then the following hold:*

(i) *K is ε-strongly monotone.*

(ii) *Suppose that $\alpha = 1$ and $W = \mathrm{Id}$. Then K is cocoercive with constant $1/(2-\varepsilon)$.*

Proof. (i): By the Cauchy–Schwarz inequality,

$$(\forall x \in \mathcal{X})(\forall y \in \mathcal{X}) \quad \begin{aligned}[t] \langle x-y \mid Kx-Ky\rangle &= \langle x-y \mid Wx-Wy\rangle - \gamma\langle x-y \mid Bx-By\rangle \\ &\geqslant \alpha\|x-y\|^2 - \gamma\|x-y\|\,\|Bx-By\| \\ &\geqslant \alpha\|x-y\|^2 - \gamma\beta\|x-y\|^2 \\ &\geqslant \varepsilon\|x-y\|^2. \end{aligned} \tag{3.44}$$

(ii): Since γB is $(1-\varepsilon)$-Lipschitzian, [7, Proposition 4.38] entails that γB is averaged with constant $(2-\varepsilon)/2$. Hence, since $\gamma B = \mathrm{Id} - K$, [7, Proposition 4.39] implies that K is cocoercive with constant $1/(2-\varepsilon)$. ☐

Corollary 3.26 *Let $A\colon \mathcal{X} \to 2^{\mathcal{X}}$ be maximally monotone, let $B\colon \mathcal{X} \to \mathcal{X}$ be monotone and β-Lipschitzian for some $\beta \in \left]0,+\infty\right[$, let $(\alpha,\chi) \in \left]0,+\infty\right[^2$, and let $\varepsilon \in \left]0,\alpha/(\beta+1)\right[$. For every $n \in \mathbb{N}$, let $W_n\colon \mathcal{X} \to \mathcal{X}$ be α-strongly monotone and χ-Lipschitzian, and let $\gamma_n \in [\varepsilon,(\alpha-\varepsilon)/\beta]$. Take $x_0 \in \mathcal{X}$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,2\right[$ such that $0 < \inf_{n\in\mathbb{N}} \lambda_n \leqslant \sup_{n\in\mathbb{N}} \lambda_n < 2$, and let $(e_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{X}$ such that $e_n \to 0$. Furthermore, let $m \in \mathbb{N}\smallsetminus\{0\}$ and let $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ be a real array that satisfies the following:*

[a] *For every integer $n > m$ and every integer $j \in [0, n-m-1]$, $\mu_{n,j} = 0$.*

[b] *For every $n \in \mathbb{N}$, $\sum_{j=0}^{n} \mu_{n,j} = 1$.*

[c] $\sup_{n\in\mathbb{N}} \max_{0\leqslant j\leqslant n} |\mu_{n,j}| < +\infty$.

Iterate

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} \widetilde{x}_n = e_n + \sum_{j=0}^{n} \mu_{n,j} x_j \\ v_n^* = W_n\widetilde{x}_n - \gamma_n B\widetilde{x}_n \\ y_n = (W_n + \gamma_n A)^{-1} v_n^* \\ y_n^* = \gamma_n^{-1}(v_n^* - W_n y_n) + By_n \\ \text{if } \langle y_n - x_n \mid y_n^*\rangle < 0 \\ \left\lfloor \begin{array}{l} x_{n+1} = x_n + \dfrac{\lambda_n\langle y_n - x_n \mid y_n^*\rangle}{\|y_n^*\|^2}\, y_n^* \end{array} \right. \\ \text{else} \\ \left\lfloor \begin{array}{l} x_{n+1} = x_n. \end{array} \right. \end{array} \right. \end{array} \tag{3.45}$$

Suppose that $\operatorname{zer}(A+B) \neq \varnothing$. *Then the following hold:*

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

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

Proof. We apply Theorem 3.16 with $M = A + B$ and $(\forall n \in \mathbb{N})\ K_n = W_n - \gamma_n B$. First, [7, Corollary 20.28] asserts that B is maximally monotone. Therefore, M is maximally monotone by virtue of [7, Corollary 25.5(i)]. Next, in view of Lemma 3.25(i), the kernels $(K_n)_{n\in\mathbb{N}}$ are ε-strongly monotone. Furthermore, the kernels $(K_n)_{n\in\mathbb{N}}$ are Lipschitzian with constant $\alpha + \chi$ since

$$
\begin{aligned}
(\forall x \in \mathcal{X})(\forall y \in \mathcal{X}) \quad \|K_n x - K_n y\| &\leqslant \|W_n x - W_n y\| + \gamma_n \|Bx - By\| \\
&\leqslant \chi\|x - y\| + \frac{\alpha - \varepsilon}{\beta}\beta\|x - y\| \\
&\leqslant (\alpha + \chi)\|x - y\|.
\end{aligned}
\tag{3.46}
$$

Therefore, for every $n \in \mathbb{N}$, since $K_n + \gamma_n M$ is maximally monotone, Proposition 3.12(i)[d]&(ii)[b] entail that $\operatorname{ran} K_n \subset \operatorname{ran}(K_n + \gamma_n M)$ and $K_n + \gamma_n M$ is injective. Let us also observe that (3.45) is a special case of (3.23).

(i): This follows from Theorem 3.16(i).

(ii): Set $\mu = \sup_{n\in\mathbb{N}} \max_{0\leqslant j\leqslant n} |\mu_{n,j}|$. For every integer $n > m$, it results from [a] and [b] that

$$
\begin{aligned}
\|\widetilde{x}_n - x_n\| &= \left\| e_n + \sum_{j=n-m}^{n} \mu_{n,j}(x_j - x_n) \right\| \\
&\leqslant \|e_n\| + \sum_{j=n-m}^{n} |\mu_{n,j}|\|x_j - x_n\| \\
&\leqslant \|e_n\| + \mu \sum_{j=n-m}^{n} \|x_j - x_n\| \\
&= \|e_n\| + \mu \sum_{j=0}^{m} \|x_n - x_{n-j}\|.
\end{aligned}
\tag{3.47}
$$

Therefore, (i) and [c] imply that $\widetilde{x}_n - x_n \to 0$. On the other hand, it follows from Remark 3.17 that condition (ii)[b] in Theorem 3.16 is satisfied. Hence, the conclusion follows from Theorem 3.16(ii). ☐

Next, we recover Tseng's forward-backward-forward algorithm [7, 36].

Corollary 3.27 *Let* $A\colon \mathcal{X} \to 2^{\mathcal{X}}$ *be maximally monotone, let* $B\colon \mathcal{X} \to \mathcal{X}$ *be monotone and* β*-Lipschitzian for some* $\beta \in\,]0, +\infty[$. *Suppose that* $\operatorname{zer}(A + B) \neq \varnothing$, *take* $x_0 \in \mathcal{X}$, *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*

$$\begin{array}{l} \textit{for}\ n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} v_n^* = \gamma_n B x_n \\ y_n = J_{\gamma_n A}(x_n - v_n^*) \\ x_{n+1} = y_n - \gamma_n B y_n + v_n^*. \end{array} \right. \end{array} \tag{3.48}$$

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

Proof. We apply Theorem 3.16 with $M = A+B$ and $(\forall n \in \mathbb{N})$ $K_n = \mathrm{Id} - \gamma_n B$ and $\widetilde{x}_n = x_n$. Note that the kernels $(K_n)_{n\in\mathbb{N}}$ are cocoercive with constant $1/(2-\varepsilon)$ by virtue of Lemma 3.25(ii). Moreover, using Lemma 3.25(i), we deduce that the kernels $(K_n)_{n\in\mathbb{N}}$ are strongly monotone with constant ε. Thus, for every $n \in \mathbb{N}$, since $K_n + \gamma_n M = \mathrm{Id} + \gamma_n A$ is maximally monotone, Proposition 3.12(i)[d]&(ii)[b] assert that $\operatorname{ran} K_n \subset \operatorname{ran}(K_n + \gamma_n M)$ and $K_n + \gamma_n M$ is injective. Now set

$$(\forall n \in \mathbb{N}) \quad y_n^* = \gamma_n^{-1}(K_n x_n - K_n y_n) \quad \text{and} \quad \lambda_n = \begin{cases} \dfrac{\gamma_n \|y_n^*\|^2}{\langle x_n - y_n \mid y_n^* \rangle}, & \text{if } \langle x_n - y_n \mid y_n^* \rangle > 0; \\ \varepsilon, & \text{otherwise.} \end{cases} \tag{3.49}$$

Fix $n \in \mathbb{N}$. Then, by strong monotonicity of K_n and the Cauchy–Schwarz inequality,

$$\varepsilon \|x_n - y_n\|^2 \leqslant \langle x_n - y_n \mid K_n x_n - K_n y_n \rangle \leqslant \|x_n - y_n\|\, \|K_n x_n - K_n y_n\|. \tag{3.50}$$

This implies that $\langle x_n - y_n \mid y_n^* \rangle = \gamma_n^{-1} \langle x_n - y_n \mid K_n x_n - K_n y_n \rangle \leqslant \gamma_n^{-1} \|x_n - y_n\|\, \|K_n x_n - K_n y_n\| \leqslant (\varepsilon\gamma_n)^{-1} \|K_n x_n - K_n y_n\|^2 = \varepsilon^{-1}\gamma_n \|y_n^*\|^2$ and therefore that $\lambda_n \geqslant \varepsilon$. In addition, by cocoercivity of K_n, $\gamma_n \|y_n^*\|^2 = \gamma_n^{-1} \|K_n x_n - K_n y_n\|^2 \leqslant (2-\varepsilon)\gamma_n^{-1} \langle x_n - y_n \mid K_n x_n - K_n y_n \rangle = (2-\varepsilon) \langle x_n - y_n \mid y_n^* \rangle$ and thus $\lambda_n \leqslant 2 - \varepsilon$. Next, we derive from (3.48) that $y_n = J_{\gamma_n M}^{K_n} x_n$. If $\langle x_n - y_n \mid y_n^* \rangle > 0$, then (3.48) and (3.49) yield $x_{n+1} = x_n - \gamma_n y_n^* = x_n + \lambda_n \langle y_n - x_n \mid y_n^* \rangle y_n^* / \|y_n^*\|^2$. Otherwise, $\langle x_n - y_n \mid y_n^* \rangle \leqslant 0$ and the cocoercivity of K_n yields $\|y_n^*\|^2 = \gamma_n^{-2} \|K_n x_n - K_n y_n\|^2 \leqslant (2-\varepsilon)\gamma_n^{-2} \langle x_n - y_n \mid K_n x_n - K_n y_n \rangle \leqslant 0$. Hence, $y_n^* = 0$ and we therefore deduce from (3.48) that $x_{n+1} = x_n$. Thus, (3.48) is an instance of (3.23). Next, condition (ii)[a] in Theorem 3.16 is trivially satisfied and, in view of Remark 3.17, condition (ii)[b] in Theorem 3.16 is also fulfilled. ☐

We conclude this section by further illustrating the effectiveness of warped resolvent iterations by designing a new method to solve an intricate system of monotone inclusions and its dual. We are not aware of a splitting method that could handle such a formulation with a comparable level of flexibility. Special cases of this system appear in [1, 10, 18, 25].

Problem 3.28 Let $(\mathcal{Y}_i)_{i\in I}$ and $(\mathcal{Z}_j)_{j\in J}$ be finite families of real Hilbert spaces. For every $i \in I$ and $j \in J$, let $A_i\colon \mathcal{Y}_i \to 2^{\mathcal{Y}_i}$ and $B_j\colon \mathcal{Z}_j \to 2^{\mathcal{Z}_j}$ be maximally monotone, let $C_i\colon \mathcal{Y}_i \to \mathcal{Y}_i$ be monotone and μ_i-Lipschitzian for some $\mu_i \in \left]0, +\infty\right[$, let $D_j\colon \mathcal{Z}_j \to \mathcal{Z}_j$ be monotone and ν_j-

Lipschitzian for some $\nu_j \in]0,+\infty[$, let $L_{ji} \in \mathcal{B}(\mathcal{Y}_i, \mathcal{Z}_j)$, let $s_i^* \in \mathcal{Y}_i$, and let $r_j \in \mathcal{Z}_j$. Consider the system of coupled inclusions

$$\text{find } (x_i)_{i\in I} \in \bigtimes_{i\in I} \mathcal{Y}_i \text{ such that}$$
$$(\forall i \in I) \quad s_i^* \in A_i x_i + \sum_{j\in J} L_{ji}^*\Big((B_j + D_j)\Big(\sum_{k\in I} L_{jk}x_k - r_j\Big)\Big) + C_i x_i, \tag{3.51}$$

its dual problem

$$\text{find } (v_j^*)_{j\in J} \in \bigtimes_{j\in J} \mathcal{Z}_j \text{ such that}$$
$$\Big(\exists\, (x_i)_{i\in I} \in \bigtimes_{i\in I} \mathcal{Y}_i\Big)(\forall i \in I)(\forall j \in J) \quad \begin{cases} s_i^* - \sum_{k\in J} L_{ki}^* v_k^* \in A_i x_i + C_i x_i \\ v_j^* \in (B_j + D_j)\Big(\sum_{k\in I} L_{jk}x_k - r_j\Big), \end{cases} \tag{3.52}$$

and the associated Kuhn–Tucker set

$$Z = \Big\{\big((x_i)_{i\in I}, (v_j^*)_{j\in J}\big) \;\Big|\; (\forall i \in I)\;\; x_i \in \mathcal{Y}_i \;\text{ and }\; s_i^* - \sum_{k\in J} L_{ki}^* v_k^* \in A_i x_i + C_i x_i,$$
$$\text{and } (\forall j \in J)\;\; v_j^* \in \mathcal{Z}_j \;\text{ and }\; \sum_{k\in I} L_{jk}x_k - r_j \in (B_j + D_j)^{-1} v_j^*\Big\}. \tag{3.53}$$

We denote by $\mathscr{P}$ and $\mathscr{D}$ the sets of solutions to (3.51) and (3.52), respectively. The problem is to find a point in Z.

Corollary 3.29 *Consider the setting of Problem 3.28. For every $i \in I$ and every $j \in J$, let $(\alpha_i, \chi_i, \beta_j, \kappa_j) \in]0,+\infty[^4$, let $\varepsilon_i \in]0, \alpha_i/(\mu_i+1)[$, let $\delta_j \in]0, \beta_j/(\nu_j+1)[$, let $(F_{i,n})_{n\in\mathbb{N}}$ be operators from $\mathcal{Y}_i$ to $\mathcal{Y}_i$ that are α_i-strongly monotone and χ_i-Lipschitzian, let $(W_{j,n})_{n\in\mathbb{N}}$ be operators from $\mathcal{Z}_j$ to $\mathcal{Z}_j$ that are β_j-strongly monotone and κ_j-Lipschitzian; in addition, let $(\gamma_{i,n})_{n\in\mathbb{N}}$ and $(\tau_{j,n})_{n\in\mathbb{N}}$ be sequences in $[\varepsilon_i, (\alpha_i - \varepsilon_i)/\mu_i]$ and $[\delta_j, (\beta_j - \delta_j)/\nu_j]$, respectively. Suppose that $Z \neq \varnothing$ and that*

$$\mathcal{Y} = \bigtimes_{i\in I} \mathcal{Y}_i, \quad \mathcal{Z} = \bigtimes_{j\in J} \mathcal{Z}_j, \quad \text{and} \quad \mathcal{X} = \mathcal{Y} \times \mathcal{Z} \times \mathcal{Z}. \tag{3.54}$$

Let $((x_{i,0})_{i\in I}, (y_{j,0})_{j\in J}, (v_{j,0}^*)_{j\in J})$ and $((\widetilde{x}_{i,n})_{i\in I}, (\widetilde{y}_{j,n})_{j\in J}, (\widetilde{v}_{j,n}^*)_{j\in J})_{n\in\mathbb{N}}$ be in $\mathcal{X}$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$ such that $0 < \inf_{n\in\mathbb{N}} \lambda_n \leqslant \sup_{n\in\mathbb{N}} \lambda_n < 2$. Iterate

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\textit{for every } i \in I \\
\left\lfloor
\begin{array}{l}
l^*_{i,n} = F_{i,n}\widetilde{x}_{i,n} - \gamma_{i,n} C_i \widetilde{x}_{i,n} - \gamma_{i,n} \sum_{j\in J} L^*_{ji} \widetilde{v}^*_{j,n} \\
a_{i,n} = \big(F_{i,n} + \gamma_{i,n} A_i\big)^{-1} (l^*_{i,n} + \gamma_{i,n} s^*_i) \\
o^*_{i,n} = \gamma_{i,n}^{-1} (l^*_{i,n} - F_{i,n} a_{i,n}) + C_i a_{i,n}
\end{array}
\right. \\
\textit{for every } j \in J \\
\left\lfloor
\begin{array}{l}
t^*_{j,n} = W_{j,n}\widetilde{y}_{j,n} - \tau_{j,n} D_j \widetilde{y}_{j,n} + \tau_{j,n} \widetilde{v}^*_{j,n} \\
b_{j,n} = \big(W_{j,n} + \tau_{j,n} B_j\big)^{-1} t^*_{j,n} \\
f^*_{j,n} = \tau_{j,n}^{-1} (t^*_{j,n} - W_{j,n} b_{j,n}) + D_j b_{j,n} \\
c_{j,n} = \sum_{i\in I} L_{ji} \widetilde{x}_{i,n} - \widetilde{y}_{j,n} + \widetilde{v}^*_{j,n} - r_j
\end{array}
\right. \\
\textit{for every } i \in I \\
\left\lfloor
\begin{array}{l}
a^*_{i,n} = o^*_{i,n} + \sum_{j\in J} L^*_{ji} c_{j,n}
\end{array}
\right. \\
\textit{for every } j \in J \\
\left\lfloor
\begin{array}{l}
b^*_{j,n} = f^*_{j,n} - c_{j,n} \\
c^*_{j,n} = r_j + b_{j,n} - \sum_{i\in I} L_{ji} a_{i,n}
\end{array}
\right. \\
\sigma_n = \sum_{i\in I} \|a^*_{i,n}\|^2 + \sum_{j\in J} \big(\|b^*_{j,n}\|^2 + \|c^*_{j,n}\|^2\big) \\
\theta_n = \sum_{i\in I} \langle a_{i,n} - x_{i,n} \mid a^*_{i,n} \rangle + \sum_{j\in J} \big(\langle b_{j,n} - y_{j,n} \mid b^*_{j,n}\rangle + \langle c_{j,n} - v^*_{j,n} \mid c^*_{j,n}\rangle\big) \\
\textit{if } \theta_n < 0 \\
\left\lfloor
\begin{array}{l}
\rho_n = \lambda_n \theta_n / \sigma_n
\end{array}
\right. \\
\textit{else} \\
\left\lfloor
\begin{array}{l}
\rho_n = 0
\end{array}
\right. \\
\textit{for every } i \in I \\
\left\lfloor
\begin{array}{l}
x_{i,n+1} = x_{i,n} + \rho_n a^*_{i,n}
\end{array}
\right. \\
\textit{for every } j \in J \\
\left\lfloor
\begin{array}{l}
y_{j,n+1} = y_{j,n} + \rho_n b^*_{j,n} \\
v^*_{j,n+1} = v^*_{j,n} + \rho_n c^*_{j,n}.
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{3.55}
$$

Suppose that

$$
(\forall i \in I)(\forall j \in J) \quad \widetilde{x}_{i,n} - x_{i,n} \to 0, \quad \widetilde{y}_{j,n} - y_{j,n} \to 0, \quad \textit{and} \quad \widetilde{v}^*_{j,n} - v^*_{j,n} \to 0. \tag{3.56}
$$

Set $(\forall n \in \mathbb{N})$ $x_n = (x_{i,n})_{i\in I}$ *and* $v^*_n = (v^*_{j,n})_{j\in J}$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point* $\overline{x} \in \mathscr{P}$, $(v^*_n)_{n\in\mathbb{N}}$ *converges weakly to a point* $\overline{v}^* \in \mathscr{D}$, *and* $(\overline{x}, \overline{v}^*) \in Z$.

Proof. Define

$$\begin{cases} A\colon \mathcal{Y} \to 2^{\mathcal{Y}}\colon (x_i)_{i\in I} \mapsto \bigtimes_{i\in I} (A_i x_i + C_i x_i) \\ B\colon \mathcal{Z} \to 2^{\mathcal{Z}}\colon (y_j)_{j\in J} \mapsto \bigtimes_{j\in J} (B_j y_j + D_j y_j) \\ L\colon \mathcal{Y} \to \mathcal{Z}\colon (x_i)_{i\in I} \mapsto \Bigg(\sum_{i\in I} L_{ji} x_i\Bigg)_{j\in J} \\ s^* = (s_i^*)_{i\in I} \quad \text{and} \quad r = (r_j)_{j\in J}. \end{cases} \tag{3.57}$$

We observe that

$$L^*\colon \mathcal{Z} \to \mathcal{Y}\colon (v_j^*)_{j\in J} \mapsto \Bigg(\sum_{j\in J} L_{ji}^* v_j^*\Bigg)_{i\in I}. \tag{3.58}$$

In the light of [7, Proposition 20.23], A and B are maximally monotone. On the other hand, we deduce from (3.53), (3.57), and (3.58) that

$$Z = \big\{(x, v^*) \in \mathcal{Y} \times \mathcal{Z} \mid s^* - L^* v^* \in Ax \text{ and } Lx - r \in B^{-1} v^*\big\}. \tag{3.59}$$

Define

$$M\colon \mathcal{X} \to 2^{\mathcal{X}}\colon (x, y, v^*) \mapsto (-s^* + Ax + L^* v^*) \times (By - v^*) \times \{r - Lx + y\}. \tag{3.60}$$

Lemma 3.3(ii) entails that M is maximally monotone. Furthermore, since $Z \neq \varnothing$, Lemma 3.3(iv) yields $\operatorname{zer} M \neq \varnothing$. Next, set

$$S\colon \mathcal{X} \to \mathcal{X}\colon (x, y, v^*) \mapsto (-L^* v^*, v^*, Lx - y) \tag{3.61}$$

and, for every $n \in \mathbb{N}$,

$$K_n\colon \mathcal{X} \to \mathcal{X}\colon (x, y, v^*) \mapsto \Big(\big(\gamma_{i,n}^{-1} F_{i,n} x_i - C_i x_i\big)_{i\in I} - L^* v^*, \big(\tau_{j,n}^{-1} W_{j,n} y_j - D_j y_j\big)_{j\in J} + v^*, Lx - y + v^*\Big) \tag{3.62}$$

and

$$T_n\colon \mathcal{X} \to \mathcal{X}\colon (x, y, v^*) \mapsto \Big(\big(\gamma_{i,n}^{-1} F_{i,n} x_i - C_i x_i\big)_{i\in I}, \big(\tau_{j,n}^{-1} W_{j,n} y_j - D_j y_j\big)_{j\in J}, v^*\Big). \tag{3.63}$$

For every $i \in I$ and every $n \in \mathbb{N}$, using the facts that C_i is μ_i-Lipschitzian, that $F_{i,n}$ is α_i-strongly monotone, and that $\gamma_{i,n} \in [\varepsilon_i, (\alpha_i - \varepsilon_i)/\mu_i]$, Lemma 3.25(i) implies that $F_{i,n} - \gamma_{i,n} C_i$ is ε_i-strongly monotone and therefore, since $\gamma_{i,n}^{-1} \geqslant \mu_i/(\alpha_i - \varepsilon_i)$, it follows that $\gamma_{i,n}^{-1} F_{i,n} - C_i$ is strongly monotone with constant $\varepsilon_i \mu_i/(\alpha_i - \varepsilon_i)$. Likewise, for every $j \in J$ and every $n \in \mathbb{N}$, $\tau_{j,n}^{-1} W_{j,n} - D_j$ is strongly monotone with constant $\delta_j \nu_j/(\beta_j - \delta_j)$. Thus, upon setting

$$\vartheta = \min\left\{\min_{i\in I} \frac{\varepsilon_i \mu_i}{\alpha_i - \varepsilon_i}, \min_{j\in J} \frac{\delta_j \nu_j}{\beta_j - \delta_j}, 1\right\}, \tag{3.64}$$

we get

$$
\begin{aligned}
&(\forall n \in \mathbb{N})\big(\forall (x,y,v^*) \in \mathcal{X}\big)\big(\forall (a,b,c^*) \in \mathcal{X}\big)\\
&\quad \big\langle (x,y,v^*) - (a,b,c^*) \mid T_n(x,y,v^*) - T_n(a,b,c^*)\big\rangle\\
&\qquad = \sum_{i\in I} \big\langle x_i - a_i \mid \big(\gamma_{i,n}^{-1}F_{i,n}x_i - C_i x_i\big) - \big(\gamma_{i,n}^{-1}F_{i,n}a_i - C_i a_i\big)\big\rangle\\
&\qquad\quad + \sum_{j\in J} \big\langle y_j - b_j \mid \big(\tau_{j,n}^{-1}W_{j,n}y_j - D_j y_j\big) - \big(\tau_{j,n}^{-1}W_{j,n}b_j - D_j b_j\big)\big\rangle + \|v^* - c^*\|^2\\
&\qquad \geqslant \vartheta \sum_{i\in I}\|x_i - a_i\|^2 + \vartheta \sum_{j\in J}\|y_j - b_j\|^2 + \vartheta\|v^* - c^*\|^2\\
&\qquad = \vartheta\|(x,y,v^*) - (a,b,c^*)\|^2.
\end{aligned}
\tag{3.65}
$$

Hence, the operators $(T_n)_{n\in\mathbb{N}}$ are ϑ-strongly monotone. However, S is linear, bounded, and $S^* = -S$. It follows that the kernels $(K_n)_{n\in\mathbb{N}} = (T_n + S)_{n\in\mathbb{N}}$ are ϑ-strongly monotone. Now, for every $i \in I$ and every $n \in \mathbb{N}$, since $\gamma_{i,n}^{-1}F_{i,n}$ is Lipschitzian with constant χ_i/ε_i, we deduce that $\gamma_{i,n}^{-1}F_{i,n} - C_i$ is Lipschitzian with constant $\chi_i/\varepsilon_i + \mu_i$. Likewise, for every $j \in J$ and every $n \in \mathbb{N}$, $\tau_{j,n}^{-1}W_{j,n} - D_j$ is Lipschitzian with constant $\kappa_j/\delta_j + \nu_j$. Hence, upon setting

$$
\eta = \max\Big\{\max_{i\in I}\{\chi_i/\varepsilon_i + \mu_i\}, \max_{j\in J}\{\kappa_j/\delta_j + \nu_j\}, 1\Big\},
\tag{3.66}
$$

we obtain

$$
\begin{aligned}
(\forall n \in \mathbb{N})\big(\forall (x,y,v^*) \in \mathcal{X}\big)\big(\forall (a,b,c^*) \in \mathcal{X}\big) \quad & \|T_n(x,y,v^*) - T_n(a,b,c^*)\|^2\\
&= \sum_{i\in I}\big\|\big(\gamma_{i,n}^{-1}F_{i,n}x_i - C_i x_i\big) - \big(\gamma_{i,n}^{-1}F_{i,n}a_i - C_i a_i\big)\big\|^2\\
&\quad + \sum_{j\in J}\big\|\big(\tau_{j,n}^{-1}W_{j,n}y_j - D_j y_j\big) - \big(\tau_{j,n}^{-1}W_{j,n}b_j - D_j b_j\big)\big\|^2 + \|v^* - c^*\|^2\\
&\leqslant \eta^2 \sum_{i\in I}\|x_i - a_i\|^2 + \eta^2\sum_{j\in J}\|y_j - b_j\|^2 + \eta^2\|v^* - c^*\|^2\\
&= \eta^2\|(x,y,v^*) - (a,b,c^*)\|^2.
\end{aligned}
\tag{3.67}
$$

This implies that the operators $(T_n)_{n\in\mathbb{N}}$ are η-Lipschitzian. On the other hand, S is Lipschitzian with constant $\|S\|$. Altogether, the kernels $(K_n)_{n\in\mathbb{N}}$ are Lipschitzian with constant $\eta + \|S\|$. In turn, using Proposition 3.12(i)[d]&(ii)[b], we infer that, for every $n \in \mathbb{N}$, $\operatorname{ran} K_n \subset \operatorname{ran}(K_n + M)$ and $K_n + M$ is injective. Now set

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad & p_n = \big((x_{i,n})_{i\in I}, (y_{j,n})_{j\in J}, (v_{j,n}^*)_{j\in J}\big), \quad \widetilde{p}_n = \big((\widetilde{x}_{i,n})_{i\in I}, (\widetilde{y}_{j,n})_{j\in J}, (\widetilde{v}_{j,n}^*)_{j\in J}\big),\\
& q_n = \big((a_{i,n})_{i\in I}, (b_{j,n})_{j\in J}, (c_{j,n})_{j\in J}\big), \quad \text{and} \quad q_n^* = \big((a_{i,n}^*)_{i\in I}, (b_{j,n}^*)_{j\in J}, (c_{j,n}^*)_{j\in J}\big).
\end{aligned}
\tag{3.68}
$$

In view of (3.62), (3.60), (3.57), and (3.58), we deduce that (3.55) assumes the form

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
q_n = J_M^{K_n} \widetilde{p}_n \\
q_n^* = K_n \widetilde{p}_n - K_n q_n \\
\text{if } \langle q_n - p_n \mid q_n^* \rangle < 0 \\
\left\lfloor\; p_{n+1} = p_n + \dfrac{\lambda_n \langle q_n - p_n \mid q_n^* \rangle}{\|q_n^*\|^2}\, q_n^* \right. \\
\text{else} \\
\left\lfloor\; p_{n+1} = p_n. \right.
\end{array}
\right.
\end{array}
\tag{3.69}
$$

In addition, (3.56) implies that $\widetilde{p}_n - p_n \to 0$. Altogether, in the light of Theorem 3.16 and Remark 3.17, there exists $(\overline{x}, \overline{y}, \overline{v}^*) \in \operatorname{zer} M$ such that $p_n \rightharpoonup (\overline{x}, \overline{y}, \overline{v}^*)$. It follows that $x_n \rightharpoonup \overline{x}$ and $v_n^* \rightharpoonup \overline{v}^*$. Further, we conclude by using Lemma 3.3(iii) that $\overline{x} \in \mathscr{P}$, $\overline{v}^* \in \mathscr{D}$, and $(\overline{x}, \overline{v}^*) \in Z$. ☐

References

[1] 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.

[2] 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.

[3] H. Attouch and A. Cabot, Convergence of a relaxed inertial proximal algorithm for maximally monotone operators, *Math. Program. A*, published online 2019-06-29.

[4] J.-B. Baillon, R. E. Bruck, and S. Reich, On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces, *Houston J. Math.*, vol. 4, pp. 1–9, 1978.

[5] S. Banert, A. Ringh, J. Adler, J. Karlsson, and O. Öktem, Data-driven nonsmooth optimization, *SIAM J. Optim.,* vol. 30, pp. 102–131, 2020.

[6] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.*, vol. 42, pp. 596–636, 2003.

[7] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed., correct. printing. Springer, New York, 2019.

[8] H. H. Bauschke, X. Wang, and L. Yao, General resolvents for monotone operators: Characterization and extension, in: *Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning and Inverse Problems*, (Y. Censor, M. Jiang, and G. Wang, eds.), pp. 57–74. Medical Physics Publishing, Madison, WI, 2010.

[9] R. I. Boţ and E. R. Csetnek, ADMM for monotone operators: Convergence analysis and rates, *Adv. Comput. Math.*, vol. 45, pp. 327–359, 2019.

[10] R. I. Boţ, E. R. Csetnek, and A. Heinrich, A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators, *SIAM J. Optim.*, vol. 23, pp. 2011–2036, 2013.

[11] 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.

[12] L. M. Briceño-Arias, Forward-partial inverse-forward splitting for solving monotone inclusions, *J. Optim. Theory Appl.*, vol. 166, pp. 391–413, 2015.

[13] M. N. Bùi and P. L. Combettes, Bregman forward-backward operator splitting, 2019-09-13. `https://arxiv.org/pdf/1908.03878`

[14] M. N. Bùi and P. L. Combettes, Warped proximal iterations for monotone inclusions, 2019-08-19. `https://arxiv.org/pdf/1908.07077v1`

[15] P. L. Combettes, Construction d'un point fixe commun à une famille de contractions fermes, *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 320, pp. 1385–1390, 1995.

[16] 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.)

[17] P. L. Combettes, Systems of structured monotone inclusions: Duality, algorithms, and applications, *SIAM J. Optim.*, vol. 23, pp. 2420–2447, 2013.

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

[19] 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.

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

[21] 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.

[22] 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.

[23] P. Giselsson, Nonlinear forward-backward splitting with projection correction, 2019-08-20. `https://arxiv.org/pdf/1908.07449v1`

[24] Y. Haugazeau, *Sur les Inéquations Variationnelles et la Minimisation de Fonctionnelles Convexes.* Thèse, Université de Paris, Paris, France, 1968.

[25] P. R. Johnstone and J. Eckstein, Projective splitting with forward steps: Asynchronous and block-iterative operator splitting. `https://arxiv.org/pdf/1803.07043.pdf`

[26] G. Kassay, The proximal points algorithm for reflexive Banach spaces, *Studia Univ. Babeş-Bolyai Math.*, vol. 30, pp. 9–17, 1985.

[27] K. C. Kiwiel and B. Łopuch, Surrogate projection methods for finding fixed points of firmly nonexpansive mappings, *SIAM J. Optim.*, vol. 7, pp. 1084–1102, 1997.

[28] J. J. Moreau, Fonctions convexes duales et points proximaux dans un espace hilbertien, *C. R. Acad. Sci. Paris Sér. A*, vol. 255, pp. 2897–2899, 1962.

[29] T. Pennanen, Dualization of generalized equations of maximal monotone type, *SIAM J. Optim.*, vol. 10, pp. 809–835, 2000.

[30] H. Raguet, A note on the forward-Douglas–Rachford splitting for monotone inclusion and convex optimization, *Optim. Lett.*, vol. 13, pp. 717–740, 2019.

[31] 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.

[32] S. M. Robinson, Composition duality and maximal monotonicity, *Math. Program.*, vol. 85, pp. 1–13, 1999.

[33] R. T. Rockafellar, On the maximality of sums of nonlinear monotone operators, *Trans. Amer. Math. Soc.*, vol. 149, no. 1, pp. 75–88, 1970.

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

[35] S. Simons, *From Hahn-Banach to Monotonicity*, Lecture Notes in Math. 1693, Springer-Verlag, New York, 2008.

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

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

[38] C. Zălinescu, *Convex Analysis in General Vector Spaces.* World Scientific Publishing, River Edge, NJ, 2002.

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

Chapter 4

BREGMAN FORWARD-BACKWARD OPERATOR SPLITTING

4.1 Introduction and context

We devote this chapter to question **(Q3)** of Chapter 1. We propose a novel forward-backward splitting algorithm based on Bregman distances, which is shown to bring together and extend several Bregman iterative methods scattered in the literature. Its weak convergence is established and, in the minimization setting, rates of convergence are obtained.

This chapter presents the following article:

> M. N. Bùi and P. L. Combettes, Bregman forward-backward operator splitting, *Set-Valued and Variational Analysis*, vol. 29, no. 3, pp. 583–603, September 2021.

4.2 Article: Bregman forward-backward operator splitting

Dedicated to Terry Rockafellar on the occasion of his 85th birthday

Abstract. We establish the convergence of the forward-backward splitting algorithm based on Bregman distances for the sum of two monotone operators in reflexive Banach spaces. Even in Euclidean spaces, the convergence of this algorithm has so far been proved only in the case of minimization problems. The proposed framework features Bregman distances that vary over the iterations and a novel assumption on the single-valued operator that captures various properties scattered in the literature. In the minimization setting, we obtain rates that are sharper than existing ones.

4.2.1 Introduction

Throughout, $\mathcal{X}$ is a reflexive real Banach space with topological dual $\mathcal{X}^*$. We are concerned with the following monotone inclusion problem (see Section 4.2.2.1 for notation and definitions).

Problem 4.1 Let $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ and $B\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be maximally monotone, let $f \in \Gamma_0(\mathcal{X})$ be essentially smooth, and let D_f be the Bregman distance associated with f. Set $C = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A$ and $\mathscr{S} = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{zer}(A+B)$. Suppose that $C \subset \operatorname{int}\operatorname{dom} B$, $\mathscr{S} \neq \varnothing$, B is single-valued on $\operatorname{int}\operatorname{dom} B$, and there exist $\delta_1 \in [0,1[$, $\delta_2 \in [0,1]$, and $\kappa \in [0,+\infty[$ such that

$$
\begin{aligned}
&(\forall x \in C)(\forall y \in C)(\forall z \in \mathscr{S})(\forall y^* \in Ay)(\forall z^* \in Az)\\
&\qquad \big\langle y - x, By - Bz\big\rangle \leqslant \kappa D_f(x,y) + \big\langle y - z, \delta_1(y^* - z^*) + \delta_2\big(By - Bz\big)\big\rangle. \qquad (4.1)
\end{aligned}
$$

The objective is to

$$
\text{find } \ x \in \operatorname{int}\operatorname{dom} f \ \text{ such that } \ 0 \in Ax + Bx. \tag{4.2}
$$

The central problem (4.2) has extensive connections with various areas of mathematics and its applications. In Hilbert spaces, if B is cocoercive, a standard method for solving (4.2) is the forward-backward algorithm, which operates with the update $x_{n+1} = (\mathrm{Id} + \gamma A)^{-1}(x_n - \gamma B x_n)$ [17]. This iteration is not applicable beyond Hilbert spaces since A maps to $\mathcal{X}^* \neq \mathcal{X}$. In addition, there has been a significant body of work (see, e.g., [3, 6, 8, 12, 13, 16, 18, 19, 23]) showing the benefits of replacing standard distances by Bregman distances, even in Euclidean spaces. Given a sequence $(\gamma_n)_{n\in\mathbb{N}}$ in $]0,+\infty[$ and a suitable sequence of differentiable convex functions $(f_n)_{n\in\mathbb{N}}$, we propose to solve (4.2) via the iterative scheme

$$
(\forall n \in \mathbb{N}) \quad x_{n+1} = \big(\nabla f_n + \gamma_n A\big)^{-1}\big(\nabla f_n(x_n) - \gamma_n B x_n\big), \tag{4.3}
$$

which consists of first applying a forward (explicit) step involving B and then a backward (implicit) step involving A. Let us note that the convergence of such an iterative process has not yet been established, even in finite-dimensional spaces with a single function $f_n = f$ and constant parameters $\gamma_n = \gamma$. Furthermore, the novel scheme (4.3) will be shown to unify and extend several iterative methods which have thus far not been brought together:

- The Bregman monotone proximal point algorithm

$$
(\forall n \in \mathbb{N}) \quad x_{n+1} = \big(\nabla f + \gamma_n A\big)^{-1}\big(\nabla f(x_n)\big) \tag{4.4}
$$

  of [6] for finding a zero of A in $\operatorname{int}\operatorname{dom} f$, where f is a Legendre function.

- The variable metric forward-backward splitting method

$$
(\forall n \in \mathbb{N}) \quad x_{n+1} = \big(U_n + \gamma_n A\big)^{-1}\big(U_n x_n - \gamma_n B x_n\big) \tag{4.5}
$$

  of [15] for finding a zero of $A + B$ in a Hilbert space, where $(U_n)_{n\in\mathbb{N}}$ is a sequence of strongly positive self-adjoint bounded linear operators.

- The splitting method

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \big(\nabla f_n + \gamma_n \partial\varphi\big)^{-1}\big(\nabla f_n(x_n) - \gamma_n \nabla\psi(x_n)\big) \tag{4.6}$$

  of [18] for finding a minimizer of the sum of the convex functions φ and ψ in $\operatorname{int}\operatorname{dom} f$.

- The Renaud–Cohen algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \big(\nabla f + \gamma A\big)^{-1}\big(\nabla f(x_n) - \gamma B x_n\big) \tag{4.7}$$

  of [20] for finding a zero of $A + B$ in a Hilbert space, where f is real-valued and strongly convex.

Problems which cannot be solved by algorithms (4.4)–(4.7) will be presented in Example 4.10 as well as in Sections 4.2.3.2 and 4.2.3.4. New results on the minimization setting will be presented in Section 4.2.3.3.

The goal of the present paper is to investigate the asymptotic behavior of (4.3) under mild conditions on A, B, and $(f_n)_{n\in\mathbb{N}}$. Let us note that the convergence proof techniques used in the above four frameworks do not extend to (4.3). For instance, the tools of [18] rely heavily on functional inequalities involving φ and ψ. On the other hand, the approach of [15] exploits specific properties of quadratic kernels in Hilbert spaces, while [6] relies on Bregman monotonicity properties of the iterates that will no longer hold in the presence of B. Finally, the proofs of [20] depend on the strong convexity of f, the underlying Hilbertian structure, and the fact that the updating equation is governed by a fixed operator. Our analysis will not only capture these frameworks but also provide new methods to solve problems beyond their reach. It hinges on the theory of Legendre functions and the new condition (4.1), which will be seen to cover in particular various properties such as the cocoercivity assumption used in the standard forward-backward method in Hilbert spaces [7, 17], as well as the seemingly unrelated assumptions used in [6, 15, 18, 20] to study (4.4)–(4.7).

The main result on the convergence of (4.3) is established in Section 4.2.2 for the general scenario described in Problem 4.1. Section 4.2.3 is dedicated to special cases and applications. In the context of minimization problems, convergence rates on the worst behavior of the method are obtained.

4.2.2 Main results

4.2.2.1 Notation and definitions

The norm of $\mathcal{X}$ is denoted by $\|\cdot\|$ and the canonical pairing between $\mathcal{X}$ and $\mathcal{X}^*$ by $\langle\cdot,\cdot\rangle$. If $\mathcal{X}$ is Hilbertian, its scalar product is denoted by $\langle\cdot \mid \cdot\rangle$. The symbols $\rightharpoonup$ and $\to$ denote respectively weak and strong convergence. The set of weak sequential cluster points of a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ is denoted by $\mathfrak{W}(x_n)_{n\in\mathbb{N}}$.

Let $M\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be a set-valued operator. Then $\operatorname{gra} M = \big\{(x,x^*) \in \mathcal{X}\times\mathcal{X}^* \mid x^* \in Mx\big\}$ is the graph of M, $\operatorname{dom} M = \big\{x\in\mathcal{X} \mid Mx \neq \varnothing\big\}$ the domain of M, $\operatorname{ran} M = \big\{x^*\in\mathcal{X}^* \mid (\exists\, x\in\mathcal{X})\; x^*\in Mx\big\}$ the range of M, and $\operatorname{zer} M = \big\{x\in\mathcal{X} \mid 0\in Mx\big\}$ the set of zeros of M. Moreover, M is monotone if

$$\big(\forall (x_1,x_1^*)\in\operatorname{gra} M\big)\big(\forall (x_2,x_2^*)\in\operatorname{gra} M\big)\quad \langle x_1-x_2, x_1^*-x_2^*\rangle \geqslant 0, \tag{4.8}$$

and maximally monotone if, furthermore, there exists no monotone operator from $\mathcal{X}$ to $2^{\mathcal{X}^*}$ the graph of which properly contains $\operatorname{gra} M$.

A function $f\colon \mathcal{X}\to\left]-\infty,+\infty\right]$ is coercive if $\lim_{\|x\|\to+\infty} f(x) = +\infty$ and supercoercive if $\lim_{\|x\|\to+\infty} f(x)/\|x\| = +\infty$. $\Gamma_0(\mathcal{X})$ is the class of lower semicontinuous convex functions $f\colon\mathcal{X}\to\left]-\infty,+\infty\right]$ such that $\operatorname{dom} f = \big\{x\in\mathcal{X} \mid f(x)<+\infty\big\} \neq \varnothing$. Now let $f\in\Gamma_0(\mathcal{X})$. The conjugate of f is the function $f^*\in\Gamma_0(\mathcal{X}^*)$ defined by $f^*\colon \mathcal{X}^*\to\left]-\infty,+\infty\right] : x^*\mapsto \sup_{x\in\mathcal{X}}(\langle x,x^*\rangle - f(x))$, and the subdifferential of f is the maximally monotone operator

$$\partial f\colon \mathcal{X}\to 2^{\mathcal{X}^*} : x\mapsto \big\{x^*\in\mathcal{X}^* \mid (\forall y\in\mathcal{X})\; \langle y-x, x^*\rangle + f(x) \leqslant f(y)\big\}. \tag{4.9}$$

In addition, f is a Legendre function if it is essentially smooth in the sense that ∂f is both locally bounded and single-valued on its domain, and essentially strictly convex in the sense that ∂f^* is locally bounded on its domain and f is strictly convex on every convex subset of $\operatorname{dom}\partial f$ [5]. Suppose that f is Gâteaux differentiable on $\operatorname{int}\operatorname{dom} f \neq \varnothing$. The Bregman distance associated with f is

$$\begin{aligned} D_f\colon \mathcal{X}\times\mathcal{X} &\to [0,+\infty] \\ (x,y) &\mapsto \begin{cases} f(x)-f(y)-\langle x-y, \nabla f(y)\rangle, & \text{if } \; y\in\operatorname{int}\operatorname{dom} f; \\ +\infty, & \text{otherwise.} \end{cases} \end{aligned} \tag{4.10}$$

Given $\alpha\in\left]0,+\infty\right[$, we define

$$\mathcal{C}_\alpha(f) = \big\{g\in\Gamma_0(\mathcal{X}) \mid \operatorname{dom} g = \operatorname{dom} f,\; g \text{ is Gâteaux differentiable on } \operatorname{int}\operatorname{dom} f,\; D_g \geqslant \alpha D_f\big\}. \tag{4.11}$$

4.2.2.2 On condition (4.1)

The following proposition provides several key illustrations of the pertinence of (4.1) in terms of capturing concrete scenarios.

Proposition 4.2 *Consider the setting of Problem 4.1. Then (4.1) holds in each of the following cases:*

(i) $\delta_1\in\left[0,1\right[$, $\delta_2 = 1$, *and* $(\forall x\in C)(\forall y\in C)(\forall z\in\mathscr{S})\; \langle z-x, By-Bz\rangle \leqslant \kappa D_f(x,y)$.

(ii) $\delta_1 = 0$, $\delta_2 = 1$, *and* $B = \partial\psi$, *where* $\psi \in \Gamma_0(\mathcal{X})$ *satisfies*

$$(\forall x \in C)(\forall y \in C)(\forall z \in \mathscr{S}) \quad D_\psi(x,y) \leqslant \kappa D_f(x,y) + D_\psi(x,z) + D_\psi(z,y). \tag{4.12}$$

(iii) $\delta_1 = 0$, $\delta_2 = 1$, *and there exists* $\psi \in \Gamma_0(\mathcal{X})$ *such that* $B = \partial\psi$ *and* $(\forall x \in C)(\forall y \in C)$ $D_\psi(x,y) \leqslant \kappa D_f(x,y)$.

(iv) $\operatorname{dom} B = \mathcal{X}$, *there exists* $\beta \in \left]0,+\infty\right[$ *such that*

$$\big(\forall (x,x^*) \in \operatorname{gra}(A+B)\big)\big(\forall (y,y^*) \in \operatorname{gra}(A+B)\big) \quad \langle x-y, x^*-y^*\rangle \geqslant \beta\|Bx-By\|^2, \tag{4.13}$$

f is Fréchet differentiable on $\mathcal{X}$, ∇f is α-strongly monotone on $\operatorname{dom} A$ *for some* $\alpha \in \left]0,+\infty\right[$, $\varepsilon \in \left]0,2\beta\right[$, $\kappa = 1/(\alpha(2\beta-\varepsilon))$, *and* $\delta_1 = \delta_2 = (2\beta-\varepsilon)/(2\beta)$.

(v) *$A+B$ is strongly monotone with constant $\mu \in \left]0,+\infty\right[$, B is Lipschitzian on* $\operatorname{dom} B = \mathcal{X}$ *with constant $\nu \in \left]0,+\infty\right[$, f is Fréchet differentiable on $\mathcal{X}$, ∇f is α-strongly monotone on* $\operatorname{dom} A$ *for some* $\alpha \in \left]0,+\infty\right[$, $\varepsilon \in \left]0, 2\mu/\nu^2\right[$, $\kappa = \nu^2/(\alpha(2\mu-\varepsilon\nu^2))$, *and* $\delta_1 = \delta_2 = (2\mu-\varepsilon\nu^2)/(2\mu)$.

(vi) $\operatorname{dom} B = \mathcal{X}$, $\beta \in \left]0,+\infty\right[$, *$f$ is Fréchet differentiable on $\mathcal{X}$, ∇f is α-strongly monotone on* $\operatorname{dom} A$ *for some* $\alpha \in \left]0,+\infty\right[$, $\varepsilon \in \left]0,2\beta\right[$, $\kappa = 1/(\alpha(2\beta-\varepsilon))$, $\delta_1 = 0$, $\delta_2 = (2\beta-\varepsilon)/(2\beta)$, *and one of the following is satisfied:*

[a] *B is β-cocoercive, i.e.,*

$$(\forall x \in \mathcal{X})(\forall y \in \mathcal{X}) \quad \langle x-y, Bx-By\rangle \geqslant \beta\|Bx-By\|^2. \tag{4.14}$$

[b] *B is ν-Lipschitzian for some $\nu \in \left]0,+\infty\right[$, and angle bounded with constant $1/(4\beta\nu)$, i.e.,*

$$(\forall x \in \mathcal{X})(\forall y \in \mathcal{X})(\forall z \in \mathcal{X}) \quad \langle y-z, Bz-Bx\rangle \leqslant \frac{1}{4\beta\nu}\langle x-y, Bx-By\rangle. \tag{4.15}$$

[c] *B is $(1/\beta)$-Lipschitzian and there exists $\psi \in \Gamma_0(\mathcal{X})$ such that $B = \nabla\psi$.*

Proof. (i): Let $x \in C$, $y \in C$, and $z \in \mathscr{S}$. Then $\langle y-x, By-Bz\rangle = \langle z-x, By-Bz\rangle + \langle y-z, By-Bz\rangle \leqslant \kappa D_f(x,y) + \langle y-z, \delta_2(By-Bz)\rangle$. In view of the monotonicity of A, we obtain (4.1).

(ii)⇒(i): In the light of [9, Proposition 4.1.5 and Corollary 4.2.5], ψ is Gâteaux differentiable on $\operatorname{int}\operatorname{dom}\psi$ and $B = \nabla\psi$ on $\operatorname{int}\operatorname{dom}\psi = \operatorname{int}\operatorname{dom} B \supset C$. Hence, we derive from (4.12), (4.10), and [6, Proposition 2.3(ii)] that

$$(\forall x \in C)(\forall y \in C)(\forall z \in \mathscr{S}) \quad \kappa D_f(x,y) \geqslant D_\psi(x,y) - D_\psi(x,z) - D_\psi(z,y) = \langle z-x, By-Bz\rangle. \tag{4.16}$$

(iii)⇒(ii): Clear.

(iv): It results from [9, Theorem 4.2.10] that ∇f is continuous. Thus, using the strong monotonicity of ∇f on $\operatorname{dom} A$, we obtain

$$(\forall x \in \overline{\operatorname{dom}}\,A)(\forall y \in \overline{\operatorname{dom}}\,A) \quad \langle x-y, \nabla f(x) - \nabla f(y)\rangle \geqslant \alpha\|x-y\|^2. \tag{4.17}$$

Given x and y in $\overline{\operatorname{dom}}\,A$, define $\phi\colon \mathbb{R} \to \mathbb{R}\colon t \mapsto f(y+t(x-y))$, and observe that, since $\overline{\operatorname{dom}}\,A$ is convex [24, Theorem 3.11.12], $[x,y] \subset \overline{\operatorname{dom}}\,A$ and therefore (4.17) yields

$$\begin{aligned} D_f(x,y) &= \int_0^1 \phi'(t)dt - \langle x-y, \nabla f(y)\rangle \\ &= \int_0^1 \langle x-y, \nabla f(y+t(x-y)) - \nabla f(y)\rangle dt \\ &\geqslant \int_0^1 t\alpha\|x-y\|^2 dt \\ &= \frac{\alpha}{2}\|x-y\|^2. \end{aligned} \tag{4.18}$$

In turn, using (4.13) and (4.18), we deduce that

$$\begin{aligned} &(\forall x \in C)\big(\forall (y,y^*) \in \operatorname{gra} A\big)\big(\forall (z,z^*) \in \operatorname{gra} A\big) \\ &\quad \langle y-x, By-Bz\rangle \leqslant \left\|\frac{y-x}{\sqrt{2\beta-\varepsilon}}\right\| \left\|\sqrt{2\beta-\varepsilon}(By-Bz)\right\| \\ &\qquad\qquad \leqslant \frac{\|y-x\|^2}{2(2\beta-\varepsilon)} + \frac{2\beta-\varepsilon}{2}\|By-Bz\|^2 \qquad (4.19) \\ &\qquad\qquad \leqslant \kappa D_f(x,y) + \big\langle y-z, \delta_1(y^*-z^*) + \delta_2(By-Bz)\big\rangle. \qquad (4.20) \end{aligned}$$

(v)⇒(iv): Set $\beta = \mu/\nu^2$. Then

$$\begin{aligned} &\big(\forall (x,x^*) \in \operatorname{gra}(A+B)\big)\big(\forall (y,y^*) \in \operatorname{gra}(A+B)\big) \\ &\qquad \langle x-y, x^*-y^*\rangle \geqslant \mu\|x-y\|^2 \geqslant \beta\|Bx-By\|^2. \end{aligned} \tag{4.21}$$

(vi): We consider each case separately.

[a]: By arguing as in (4.18), we obtain $(\forall x \in \operatorname{dom} A)(\forall y \in \operatorname{dom} A)\ D_f(x,y) \geqslant (\alpha/2)\|x-y\|^2$. It thus follows from (4.19) and (4.14) that

$$\begin{aligned} &(\forall x \in C)\big(\forall (y,y^*) \in \operatorname{gra} A\big)\big(\forall (z,z^*) \in \operatorname{gra} A\big) \\ &\quad \langle y-x, By-Bz\rangle \leqslant \frac{\|y-x\|^2}{2(2\beta-\varepsilon)} + \frac{2\beta-\varepsilon}{2}\|By-Bz\|^2 \\ &\qquad\qquad \leqslant \kappa D_f(x,y) + \big\langle y-z, \delta_2(By-Bz)\big\rangle. \end{aligned} \tag{4.22}$$

[b]⇒[a]: We derive from [1, Proposition 4] that B is cocoercive with constant β.

[c]⇒[a]: This follows from [1, Corollaire 10]. ☐

Remark 4.3 Condition (iv) in Proposition 4.2 first appeared in [20] and does not seem to have gotten much notice in the literature. The cocoercivity condition (vi)[a] was first used in [17] to prove the weak convergence of the classical forward-backward method in Hilbert spaces. Finally, in reflexive Banach space minimization problems, (iii) appears in [18]; see also [3] for the Euclidean case.

Remark 4.4 Condition (iii) is satisfied in particular when $\mathcal{X}$ is a Hilbert space, $f = \|\cdot\|^2/2$, $\operatorname{dom}\psi = \mathcal{X}$, and $\nabla\psi$ is Lipschitzian [7, Theorem 18.15], in which case it is known as the "descent lemma." Condition (ii) can be viewed as an extension of this standard descent lemma involving triples (x, y, z) and an arbitrary Bregman distance D_f in reflexive Banach spaces. Let us underline that (ii) is more general than (iii). Indeed, consider the setting of Problem 4.1 with the following additional assumptions: $\mathcal{X}$ is a Hilbert space, $0 \in \operatorname{int}\operatorname{dom} f$, A is the normal cone operator of some self-dual cone K, and there exists a Gâteaux differentiable convex function $\psi\colon \mathcal{X} \to \mathbb{R}$ such that

$$B = \nabla\psi, \quad \operatorname{Argmin}\psi = \{0\}, \quad \text{and} \quad \nabla\psi(K) \subset K. \tag{4.23}$$

Then $C = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom} A \subset K$ and $\mathscr{S} = \{0\}$. Further, for every $x \in C$ and every $y \in C$, (4.23) yields $D_\psi(x, y) - D_\psi(x, 0) - D_\psi(0, y) = \langle -x \mid \nabla\psi(y) - \nabla\psi(0)\rangle = \langle -x \mid \nabla\psi(y)\rangle \leqslant 0 \leqslant D_f(x, y)$. Therefore, (4.12) is satisfied. On the other hand, (iii) does not hold in general. For instance, take $\mathcal{X} = \mathbb{R}$, $K = [0, +\infty[$, $f = |\cdot|^2/2$, and $\psi = |\cdot|^{3/2}$.

4.2.2.3 Forward-backward splitting for monotone inclusions

The formal setting of the proposed Bregman forward-backward splitting method is as follows.

Algorithm 4.5 Consider the setting of Problem 4.1. Let $\alpha \in]0, +\infty[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be in $]0, +\infty[$, and let $(f_n)_{n\in\mathbb{N}}$ be in $\mathcal{C}_\alpha(f)$. Suppose that the following hold:

[a] $\inf_{n\in\mathbb{N}} \gamma_n > 0$, $\sup_{n\in\mathbb{N}}(\kappa\gamma_n) \leqslant \alpha$, and $\sup_{n\in\mathbb{N}}(\delta_1\gamma_{n+1}/\gamma_n) < 1$.

[b] There exists a summable sequence $(\eta_n)_{n\in\mathbb{N}}$ in $[0, +\infty[$ such that $(\forall n \in \mathbb{N})$ $D_{f_{n+1}} \leqslant (1 + \eta_n)D_{f_n}$.

[c] For every $n \in \mathbb{N}$, ∇f_n is strictly monotone on C and $(\nabla f_n - \gamma_n B)(C) \subset \operatorname{ran}(\nabla f_n + \gamma_n A)$.

Take $x_0 \in C$ and set $(\forall n \in \mathbb{N})$ $x_{n+1} = (\nabla f_n + \gamma_n A)^{-1}(\nabla f_n(x_n) - \gamma_n B x_n)$.

Let us establish basic asymptotic properties of Algorithm 4.5, starting with the fact that its viability domain is C.

Proposition 4.6 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 4.5 and let $z \in \mathscr{S}$. Then $(x_n)_{n\in\mathbb{N}}$ is a well-defined sequence in C and the following hold:*

(i) *$(D_{f_n}(z, x_n))_{n\in\mathbb{N}}$ converges.*

(ii) *$\sum_{n\in\mathbb{N}}(1 - \kappa\gamma_n/\alpha)D_{f_n}(x_{n+1}, x_n) < +\infty$ and $\sum_{n\in\mathbb{N}}(1 - \kappa\gamma_n/\alpha)D_f(x_{n+1}, x_n) < +\infty$.*

(iii) $\sum_{n\in\mathbb{N}}\langle x_{n+1}-z, \gamma_n^{-1}(\nabla f_n(x_n)-\nabla f_n(x_{n+1}))-Bx_n+Bz\rangle < +\infty$.

(iv) $\sum_{n\in\mathbb{N}}(1-\delta_2)\langle x_n - z, Bx_n - Bz\rangle < +\infty$.

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

[a] *C is bounded.*

[b] *f is supercoercive.*

[c] *f is uniformly convex.*

[d] *f is essentially strictly convex with* $\operatorname{dom} f^*$ *open and ∇f^* weakly sequentially continuous.*

[e] *$\mathcal{X}$ is finite-dimensional and* $\operatorname{dom} f^*$ *is open.*

[f] *f is essentially strictly convex and* $\rho = \inf\limits_{\substack{x\in\operatorname{int}\operatorname{dom} f\\ y\in\operatorname{int}\operatorname{dom} f\\ x\neq y}} \dfrac{D_f(x,y)}{D_f(y,x)} \in \left]0,+\infty\right[$.

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

Proof. Take $n\in\mathbb{N}$, and suppose that (y^*,y_1) and (y^*,y_2) belong to $\operatorname{gra}(\nabla f_n+\gamma_n A)^{-1}$. Then $y^*\in(\nabla f_n+\gamma_n A)y_1$ and $y^*\in(\nabla f_n+\gamma_n A)y_2$. However, by virtue of condition [c] in Algorithm 4.5, $\nabla f_n+\gamma_n A$ is strictly monotone. Therefore, since $\langle y_1-y_2, y^*-y^*\rangle = 0$, we infer that $y_1=y_2$. Hence

$$(\nabla f_n+\gamma_n A)^{-1} \text{ is single-valued on } \operatorname{dom}(\nabla f_n+\gamma_n A)^{-1} = \operatorname{ran}(\nabla f_n+\gamma_n A). \tag{4.24}$$

Moreover, it follows from [9, Proposition 4.2.2] and (4.11) that

$$\operatorname{ran}(\nabla f_n+\gamma_n A)^{-1} = \operatorname{dom}\nabla f_n \cap \operatorname{dom} A = (\operatorname{int}\operatorname{dom} f_n)\cap\operatorname{dom} A = C. \tag{4.25}$$

Next, we observe that, since $x_0\in C\subset\operatorname{int}\operatorname{dom} B$, $\nabla f_0(x_0)-\gamma_0 Bx_0$ is a singleton. Furthermore, in view of condition [c] in Algorithm 4.5, $\nabla f_0(x_0)-\gamma_0 Bx_0\in\operatorname{ran}(\nabla f_0+\gamma_0 A)$. We thus deduce from (4.24) that $x_1=(\nabla f_0+\gamma_0 A)^{-1}(\nabla f_0(x_0)-\gamma_0 Bx_0)$ is uniquely defined. In addition, (4.25) yields $x_1\in\operatorname{ran}(\nabla f_0+\gamma_0 A)^{-1}=C$. The conclusion that $(x_n)_{n\in\mathbb{N}}$ is a well-defined sequence in C follows by invoking these facts inductively.

(i)–(iv): Condition [a] in Algorithm 4.5 entails that there exists $\varepsilon\in\left]0,1\right[$ such that

$$\delta_1\gamma_{n+1}\leqslant(1-\varepsilon)\gamma_n. \tag{4.26}$$

Now take $x_0^*\in Ax_0$ and set

$$\begin{cases} x_{n+1}^* = \gamma_n^{-1}\big(\nabla f_n(x_n)-\nabla f_n(x_{n+1})\big)-Bx_n\\ \Delta_n = D_{f_n}(z,x_n)+\delta_1\gamma_n\langle x_n-z, x_n^*+Bz\rangle\\ \theta_n = (1-\kappa\gamma_n/\alpha)D_{f_n}(x_{n+1},x_n)\\ \qquad +\varepsilon\gamma_n\langle x_{n+1}-z, x_{n+1}^*+Bz\rangle+(1-\delta_2)\gamma_n\langle x_n-z, Bx_n-Bz\rangle. \end{cases} \tag{4.27}$$

In view of (4.27),

$$(x_{n+1}, x^*_{n+1}) \in \operatorname{gra} A. \tag{4.28}$$

In turn, since $(z, -Bz) \in \operatorname{gra} A$ and A is monotone,

$$\langle x_{n+1} - z, x^*_{n+1} + Bz \rangle \geqslant 0. \tag{4.29}$$

Hence, invoking condition [a] in Algorithm 4.5 and the monotonicity of B, we obtain $\theta_n \geqslant 0$. Next, since $z \in \operatorname{int}\operatorname{dom} f = \operatorname{int}\operatorname{dom} f_n$ by (4.11), we derive from (4.27) and [6, Proposition 2.3(ii)] that

$$\begin{aligned}
0 &= \big\langle x_{n+1} - z, \nabla f_n(x_n) - \nabla f_n(x_{n+1}) - \gamma_n B x_n - \gamma_n x^*_{n+1} \big\rangle \\
&= \big\langle x_{n+1} - z, \nabla f_n(x_n) - \nabla f_n(x_{n+1}) \big\rangle + \gamma_n \langle z - x_{n+1}, Bx_n - Bz \rangle - \gamma_n \langle x_{n+1} - z, x^*_{n+1} + Bz \rangle \\
&= D_{f_n}(z, x_n) - D_{f_n}(z, x_{n+1}) - D_{f_n}(x_{n+1}, x_n) + \gamma_n \langle z - x_{n+1}, Bx_n - Bz \rangle \\
&\quad - \gamma_n \langle x_{n+1} - z, x^*_{n+1} + Bz \rangle.
\end{aligned} \tag{4.30}$$

Thus, since $(z, -Bz) \in \operatorname{gra} A$ and $f_n \in \mathcal{C}_\alpha(f)$, we infer from (4.26), (4.29), (4.28), and (4.1) that

$$\begin{aligned}
&D_{f_n}(z, x_{n+1}) + \delta_1 \gamma_{n+1} \langle x_{n+1} - z, x^*_{n+1} + Bz \rangle \\
&\quad \leqslant D_{f_n}(z, x_{n+1}) + \gamma_n \langle x_{n+1} - z, x^*_{n+1} + Bz \rangle - \varepsilon \gamma_n \langle x_{n+1} - z, x^*_{n+1} + Bz \rangle \\
&\quad = D_{f_n}(z, x_n) - D_{f_n}(x_{n+1}, x_n) + \gamma_n \langle z - x_{n+1}, Bx_n - Bz \rangle - \varepsilon \gamma_n \langle x_{n+1} - z, x^*_{n+1} + Bz \rangle \\
&\quad = D_{f_n}(z, x_n) - D_{f_n}(x_{n+1}, x_n) + \gamma_n \langle x_n - x_{n+1}, Bx_n - Bz \rangle - \gamma_n \langle x_n - z, Bx_n - Bz \rangle \\
&\quad\quad - \varepsilon \gamma_n \langle x_{n+1} - z, x^*_{n+1} + Bz \rangle \\
&\quad \leqslant D_{f_n}(z, x_n) - D_{f_n}(x_{n+1}, x_n) + \kappa \gamma_n D_f(x_{n+1}, x_n) + \delta_1 \gamma_n \langle x_n - z, x^*_n + Bz \rangle \\
&\quad\quad + \delta_2 \gamma_n \langle x_n - z, Bx_n - Bz \rangle - \gamma_n \langle x_n - z, Bx_n - Bz \rangle - \varepsilon \gamma_n \langle x_{n+1} - z, x^*_{n+1} + Bz \rangle \\
&\quad \leqslant D_{f_n}(z, x_n) + \delta_1 \gamma_n \langle x_n - z, x^*_n + Bz \rangle - (1 - \kappa \gamma_n / \alpha) D_{f_n}(x_{n+1}, x_n) \\
&\quad\quad - \varepsilon \gamma_n \langle x_{n+1} - z, x^*_{n+1} + Bz \rangle - (1 - \delta_2) \gamma_n \langle x_n - z, Bx_n - Bz \rangle \\
&\quad = \Delta_n - \theta_n.
\end{aligned} \tag{4.31}$$

Consequently, by condition [b] in Algorithm 4.5 and (4.29),

$$\begin{aligned}
\Delta_{n+1} &= D_{f_{n+1}}(z, x_{n+1}) + \delta_1 \gamma_{n+1} \langle x_{n+1} - z, x^*_{n+1} + Bz \rangle \\
&\leqslant (1 + \eta_n) \big(D_{f_n}(z, x_{n+1}) + \delta_1 \gamma_{n+1} \langle x_{n+1} - z, x^*_{n+1} + Bz \rangle \big) \\
&\leqslant (1 + \eta_n)(\Delta_n - \theta_n) \\
&\leqslant (1 + \eta_n)\Delta_n - \theta_n.
\end{aligned} \tag{4.32}$$

Hence, [7, Lemma 5.31] asserts that

$$(\Delta_n)_{n\in\mathbb{N}} \text{ converges and } \sum_{n\in\mathbb{N}} \theta_n < +\infty. \tag{4.33}$$

In turn, we infer from (4.27) and condition [a] in Algorithm 4.5 that

$$\begin{cases} \sum_{n\in\mathbb{N}} (1-\kappa\gamma_n/\alpha) D_{f_n}(x_{n+1}, x_n) < +\infty \\ \sum_{n\in\mathbb{N}} \langle x_{n+1} - z, x_{n+1}^* + Bz\rangle < +\infty \\ \sum_{n\in\mathbb{N}} (1-\delta_2)\langle x_n - z, Bx_n - Bz\rangle < +\infty. \end{cases} \tag{4.34}$$

Thus, since $(f_n)_{n\in\mathbb{N}}$ lies in $\mathcal{C}_\alpha(f)$, we obtain $\sum_{n\in\mathbb{N}}(1-\kappa\gamma_n/\alpha)D_f(x_{n+1},x_n) < +\infty$. It results from (4.33) and (4.27) that $(D_{f_n}(z,x_n))_{n\in\mathbb{N}}$ converges.

(v): Recall that $(x_n)_{n\in\mathbb{N}}$ lies in C.

[a]: Clear.

[b]: We derive from (i) that $(D_f(z,x_n))_{n\in\mathbb{N}}$ is bounded. In turn, [5, Lemma 7.3(viii)] asserts that $(x_n)_{n\in\mathbb{N}}$ is bounded.

[c]: It results from [24, Theorem 3.5.10] that there exists a function $\phi\colon\ [0,+\infty[\ \to [0,+\infty]$ that vanishes only at 0 such that $\lim_{t\to+\infty} \phi(t)/t \to +\infty$ and

$$(\forall x \in \operatorname{int}\operatorname{dom} f)(\forall y \in \operatorname{dom} f) \quad \langle y - x, \nabla f(x)\rangle + f(x) + \phi\big(\|x-y\|\big) \leqslant f(y). \tag{4.35}$$

Hence, in the light of (i), $\sup_{n\in\mathbb{N}} \phi(\|x_n - z\|) \leqslant \sup_{n\in\mathbb{N}} D_f(z,x_n) \leqslant (1/\alpha)\sup_{n\in\mathbb{N}} D_{f_n}(z,x_n) < +\infty$ and $(x_n)_{n\in\mathbb{N}}$ is therefore bounded.

[d]: Suppose that there exists a subsequence $(x_{k_n})_{n\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$ such that $\|x_{k_n}\| \to +\infty$. We deduce from [5, Lemma 7.3(vii)] and (i) that

$$\sup_{n\in\mathbb{N}} D_{f^*}\big(\nabla f(x_n), \nabla f(z)\big) = \sup_{n\in\mathbb{N}} D_f(z,x_n) \leqslant \frac{1}{\alpha}\sup_{n\in\mathbb{N}} D_{f_n}(z,x_n) < +\infty. \tag{4.36}$$

However, f^* is a Legendre function by virtue of [5, Corollary 5.5] and $\nabla f(z) \in \operatorname{int}\operatorname{dom} f^*$ by virtue of [5, Theorem 5.10]. Thus, [5, Lemma 7.3(v)] guarantees that $D_{f^*}(\cdot, \nabla f(z))$ is coercive. It therefore follows from (4.36) that $(\nabla f(x_{k_n}))_{n\in\mathbb{N}}$ is bounded, and then from the reflexivity of $\mathcal{X}^*$ that $\mathfrak{W}(\nabla f(x_{k_n}))_{n\in\mathbb{N}} \neq \varnothing$. In turn, there exist a subsequence $(x_{l_{k_n}})_{n\in\mathbb{N}}$ of $(x_{k_n})_{n\in\mathbb{N}}$ and $x^* \in \mathcal{X}^*$ such that $\nabla f(x_{l_{k_n}}) \rightharpoonup x^*$. The weak lower semicontinuity of f^* and (4.36) yield $D_{f^*}(x^*, \nabla f(z)) \leqslant \underline{\lim}\, D_{f^*}(\nabla f(x_{l_{k_n}}), \nabla f(z)) < +\infty$. Therefore

$$\nabla f(x_{l_{k_n}}) \rightharpoonup x^* \in \operatorname{dom} f^* = \operatorname{int}\operatorname{dom} f^*. \tag{4.37}$$

Moreover, [5, Theorem 5.10] asserts that $\nabla f^*(x^*) \in \operatorname{int}\operatorname{dom} f$ and $(\forall n \in \mathbb{N})\ \nabla f^*\big(\nabla f(x_n)\big) =$

x_n. Hence, (4.37) and the weak sequential continuity of ∇f^* imply that $x_{l_{k_n}} = \nabla f^*(\nabla f(x_{l_{k_n}})) \rightharpoonup \nabla f^*(x^*)$. This yields $\sup_{n\in\mathbb{N}} \|x_{l_{k_n}}\| < +\infty$ and we reach a contradiction.

[e]: A consequence of [5, Lemma 7.3(ix)] and (i).

[f]: It results from [5, Lemma 7.3(v)] that $D_f(\cdot, z)$ is coercive. In turn, since $\sup_{n\in\mathbb{N}} D_f(x_n, z) \leqslant (1/\rho) \sup_{n\in\mathbb{N}} D_f(z, x_n) < +\infty$ by (i), $(x_n)_{n\in\mathbb{N}}$ is bounded. ☐

As seen in Proposition 4.6, by construction, an orbit of Algorithm 4.5 lies in C and therefore in $\operatorname{int\,dom} f$. Next, we proceed to identify sufficient conditions that guarantee that their weak sequential cluster points are also in $\operatorname{int\,dom} f$.

Proposition 4.7 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 4.5 and suppose that one of the following holds:*

[a] $\overline{\operatorname{dom}}\, f \cap \overline{\operatorname{dom}}\, A \subset \operatorname{int\,dom} f$.

[b] *f is essentially strictly convex with* $\operatorname{dom} f^*$ *open and ∇f^* weakly sequentially continuous.*

[c] *f is strictly convex on* $\operatorname{int\,dom} f$ *and* $\rho = \inf\limits_{\substack{x\in \operatorname{int\,dom} f\\ y\in\operatorname{int\,dom} f\\ x\neq y}} \dfrac{D_f(x,y)}{D_f(y,x)} \in \left]0, +\infty\right[$.

[d] *$\mathcal{X}$ is finite-dimensional.*

Then $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \operatorname{int\,dom} f$.

Proof. Suppose that $x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$, and fix $z \in \mathscr{S}$.

[a]: Since $\overline{\operatorname{dom}}\, f$ is closed and convex, it is weakly closed [10, Corollary II.6.3.3(i)]. Hence, since Proposition 4.6 asserts that $(x_n)_{n\in\mathbb{N}}$ lies in $C \subset \operatorname{dom} f$, we infer that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \overline{\operatorname{dom}}\, f$. Likewise, since $\overline{\operatorname{dom}}\, A$ is a closed convex set [24, Theorem 3.11.12] and $(x_n)_{n\in\mathbb{N}}$ lies in $C \subset \operatorname{dom} A$, we obtain $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \overline{\operatorname{dom}}\, A$. Altogether, $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \overline{\operatorname{dom}}\, f \cap \overline{\operatorname{dom}}\, A \subset \operatorname{int\,dom} f$.

[b]: Using an argument similar to that of the proof of Proposition 4.6(v)[d], we infer that there exist a strictly increasing sequence $(l_{k_n})_{n\in\mathbb{N}}$ in $\mathbb{N}$ and $x^* \in \operatorname{int\,dom} f^*$ such that $x_{l_{k_n}} \rightharpoonup \nabla f^*(x^*)$. Thus, appealing to [5, Theorem 5.10], we conclude that $x = \nabla f^*(x^*) \in \operatorname{int\,dom} f$.

[c]: Proposition 4.6(i) and the weak lower semicontinuity of $D_f(\cdot, z)$ yield

$$D_f(x,z) \leqslant \underline{\lim}\, D_f(x_{k_n}, z) \leqslant (1/\rho)\, \underline{\lim}\, D_f(z, x_{k_n}) \leqslant (\alpha\rho)^{-1} \lim D_{f_{k_n}}(z, x_{k_n}) < +\infty. \tag{4.38}$$

Thus $x \in \operatorname{dom} f$. We show that $\operatorname{dom} f$ is open. Suppose that there exists $y \in \operatorname{dom} f \smallsetminus \operatorname{int\,dom} f$, let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,1\right[$ such that $\alpha_n \to 1$, and set $(\forall n \in \mathbb{N})\ y_n = \alpha_n y + (1-\alpha_n) z$. Then $\{y_n\}_{n\in\mathbb{N}} \subset \left]y, z\right[\subset (\operatorname{int\,dom} f) \smallsetminus \{z\}$ [10, Proposition II.2.6.16]. Moreover, $y_n \to y$ and, by convexity of f, $(\forall n \in \mathbb{N})\ D_f(y_n, z) \leqslant \alpha_n(f(y) - f(z) - \langle y - z, \nabla f(z)\rangle)$. Hence

$$\overline{\lim}\, D_f(y_n, z) \leqslant f(y) - f(z) - \langle y - z, \nabla f(z)\rangle = D_f(y,z). \tag{4.39}$$

However, it results from the lower semicontinuity of f that $\underline{\lim}\, D_f(y_n, z) = \underline{\lim}(f(y_n) - f(z)) -$

$\underline{\lim}\langle y_n - z, \nabla f(z)\rangle \geqslant f(y) - f(z) - \langle y - z, \nabla f(z)\rangle = D_f(y,z)$. Hence, (4.39) forces

$$\lim D_f(y_n, z) = D_f(y,z). \tag{4.40}$$

In addition, by convexity of f, $(\forall n \in \mathbb{N})\ D_f(z, y_n) \geqslant \alpha_n\big(f(z) - f(y) - \langle z - y, \nabla f(y_n)\rangle\big)$. However, [5, Theorem 5.6] and the essential smoothness of f entail that

$$\langle z - y, \nabla f(y_n)\rangle = \langle z - y, \nabla f(y + (1-\alpha_n)(z-y))\rangle \to -\infty. \tag{4.41}$$

Thus,

$$+\infty = \lim \Big(\alpha_n\big(f(z) - f(y) - \langle z - y, \nabla f(y_n)\rangle\big)\Big) \leqslant \underline{\lim}\, D_f(z, y_n). \tag{4.42}$$

It results from (4.40) and (4.42) that $0 < \rho \leqslant \lim D_f(y_n, z)/D_f(z, y_n) = 0$, so that we reach a contradiction. Consequently, $\operatorname{dom} f$ is open and hence $x \in \operatorname{dom} f = \operatorname{int}\operatorname{dom} f$.

[d]: Proposition 4.6(i) ensures that $(x_{k_n})_{n\in\mathbb{N}}$ is a sequence in $\operatorname{int}\operatorname{dom} f$ such that $(D_f(z, x_{k_n}))_{n\in\mathbb{N}}$ is bounded. Therefore, [4, Theorem 3.8(ii)] and the essential smoothness of f yield $x \in \operatorname{int}\operatorname{dom} f$. $\square$

Definition 4.8 Algorithm 4.5 is focusing if, for every $z \in \mathscr{S}$,

$$\begin{cases} \big(D_{f_n}(z, x_n)\big)_{n\in\mathbb{N}} \text{ converges} \\ \sum_{n\in\mathbb{N}} \big\langle x_{n+1} - z, \gamma_n^{-1}\big(\nabla f_n(x_n) - \nabla f_n(x_{n+1})\big) - Bx_n + Bz\big\rangle < +\infty \\ \sum_{n\in\mathbb{N}} (1-\delta_2)\big\langle x_n - z, Bx_n - Bz\big\rangle < +\infty \\ \sum_{n\in\mathbb{N}} (1 - \kappa\gamma_n/\alpha) D_{f_n}(x_{n+1}, x_n) < +\infty \end{cases}$$
$$\Rightarrow \quad \mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \operatorname{zer}(A+B). \tag{4.43}$$

Our main result establishes the weak convergence of the orbits of Algorithm 4.5.

Theorem 4.9 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 4.5 and suppose that the following hold:*

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

[b] $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \operatorname{int}\operatorname{dom} f$.

[c] *Algorithm 4.5 is focusing.*

[d] *One of the following is satisfied:*

1/ $\mathscr{S}$ *is a singleton.*

2/ *There exists a function g in $\Gamma_0(\mathcal{X})$ which is Gâteaux differentiable on $\operatorname{int}\operatorname{dom} g \supset C$, with ∇g strictly monotone on C, and such that, for every sequence $(y_n)_{n\in\mathbb{N}}$ in C and every $y \in \mathfrak{W}(y_n)_{n\in\mathbb{N}} \cap C$, $y_{k_n} \rightharpoonup y \Rightarrow \nabla f_{k_n}(y_{k_n}) \rightharpoonup \nabla g(y)$.*

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

Proof. It results from [a] and the reflexivity of $\mathcal{X}$ that

$$(x_n)_{n\in\mathbb{N}} \text{ lies in a weakly sequentially compact set.} \tag{4.44}$$

On the other hand, [c] and items (i)–(iv) in Proposition 4.6 yield $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \operatorname{zer}(A+B)$. In turn, it results from [b] that

$$\varnothing \neq \mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathscr{S} \subset C. \tag{4.45}$$

In view of [7, Lemma 1.35] applied in $\mathcal{X}^{\text{weak}}$, it remains to show that $\mathfrak{W}(x_n)_{n\in\mathbb{N}}$ is a singleton. If [d]1/ holds, this follows from (4.45). Now suppose that [d]2/ holds, and take y_1 and y_2 in $\mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup y_1$ and $x_{l_n} \rightharpoonup y_2$. Then $y_1 \in \mathscr{S}$ and $y_2 \in \mathscr{S}$ by virtue of (4.45), and we therefore deduce from Proposition 4.6(i) that $(D_{f_n}(y_1, x_n))_{n\in\mathbb{N}}$ and $(D_{f_n}(y_2, x_n))_{n\in\mathbb{N}}$ converge. However, condition [b] in Algorithm 4.5 and [7, Lemma 5.31] assert that $(D_{f_n}(y_1, y_2))_{n\in\mathbb{N}}$ converges. Hence, appealing to [6, Proposition 2.3(ii)], it follows that $(\langle y_1 - y_2, \nabla f_n(x_n) - \nabla f_n(y_2)\rangle)_{n\in\mathbb{N}} = (D_{f_n}(y_2, x_n) + D_{f_n}(y_1, y_2) - D_{f_n}(y_1, x_n))_{n\in\mathbb{N}}$ converges. Set $\ell = \lim\langle y_1 - y_2, \nabla f_n(x_n) - \nabla f_n(y_2)\rangle$. Since $(x_n)_{n\in\mathbb{N}}$ is a sequence in C, we infer from (4.45) and [d]2/ that $\ell \leftarrow \langle y_1 - y_2, \nabla f_{l_n}(x_{l_n}) - \nabla f_{l_n}(y_2)\rangle \to \langle y_1 - y_2, \nabla g(y_2) - \nabla g(y_2)\rangle = 0$, which yields $\ell = 0$. However, invoking [d]2/, we obtain $\ell \leftarrow \langle y_1 - y_2, \nabla f_{k_n}(x_{k_n}) - \nabla f_{k_n}(y_2)\rangle \to \langle y_1 - y_2, \nabla g(y_1) - \nabla g(y_2)\rangle$. It therefore follows that $\langle y_1 - y_2, \nabla g(y_1) - \nabla g(y_2)\rangle = 0$ and hence from the strict monotonicity of ∇g on C that $y_1 = y_2$. ☐

Example 4.10 We provide an example with operating conditions that are not captured by any of the methods described in (4.4)–(4.7). Let $p \in\,]1, +\infty[$, let $(\chi_n)_{n\in\mathbb{N}}$ be a sequence in $[1, +\infty[$ such that $\chi_n \to 1$, and let $(\eta_n)_{n\in\mathbb{N}}$ be a summable sequence in $[0, +\infty[$ such that $(\forall n \in \mathbb{N})$ $\chi_{n+1} \leqslant (1 + \eta_n)\chi_n$. We denote by $z = (\zeta_k)_{k\in\mathbb{N}}$ a sequence in $\ell^p(\mathbb{N})$. Set $\mathcal{X} = \ell^p(\mathbb{N}) \times \mathbb{R}$, hence $\mathcal{X}^* = \ell^{p/(p-1)}(\mathbb{N}) \times \mathbb{R}$, and define the Legendre functions

$$(\forall n \in \mathbb{N}) \quad f_n\colon \mathcal{X} \to\,]{-\infty}, +\infty] : (z, \xi) \mapsto \begin{cases} \dfrac{\chi_n}{p}\|z\|^p + 1 - \xi + \xi\ln\xi, & \text{if } \xi > 0;\\[2mm] \dfrac{\chi_n}{p}\|z\|^p + 1, & \text{if } \xi = 0;\\[2mm] +\infty, & \text{if } \xi \leqslant 0 \end{cases} \tag{4.46}$$

and

$$f = g\colon \mathcal{X} \to\,]{-\infty}, +\infty] : (z, \xi) \mapsto \begin{cases} \dfrac{1}{p}\|z\|^p - \xi + \xi\ln\xi, & \text{if } \xi > 0;\\[2mm] \dfrac{1}{p}\|z\|^p, & \text{if } \xi = 0;\\[2mm] +\infty, & \text{if } \xi \leqslant 0. \end{cases} \tag{4.47}$$

Now let $\psi\colon \mathcal{X} \to [0, +\infty[\, : (z, \xi) \mapsto \|z\|^p/p$, set $B = \nabla\psi$, and let $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be any maximally monotone operator such that

$$\operatorname{dom} A \subset \ell^p(\mathbb{N}) \times \left]0,+\infty\right[\quad \text{and} \quad \operatorname{zer}(A+B) \neq \varnothing. \tag{4.48}$$

Let us check that this setting conforms to that of Theorem 4.9. First, Proposition 4.2(iii) implies that (4.1) is satisfied with $\delta_1 = 0$ and $\delta_2 = \kappa = 1$. Next, we note that $\operatorname{int}\operatorname{dom} f = \ell^p(\mathbb{N}) \times \left]0,+\infty\right[$, that $(f_n)_{n\in\mathbb{N}}$ lies in $\mathcal{C}_1(f)$, and that condition [b] in Algorithm 4.5 holds. Furthermore, we derive from (4.46) that

$$(\forall n \in \mathbb{N}) \quad \nabla f_n \colon \ell^p(\mathbb{N}) \times \left]0,+\infty\right[\to \mathcal{X}^* \colon (z,\xi) \mapsto \Big(\chi_n \big(\operatorname{sign}(\zeta_k)|\zeta_k|^{p-1}\big)_{k\in\mathbb{N}}, \ln \xi\Big) \tag{4.49}$$

and we observe that

$$(\forall n \in \mathbb{N}) \quad \operatorname{ran} \nabla f_n = \mathcal{X}^* \quad \text{and} \quad \operatorname{dom}(\gamma_n A) \subset \operatorname{dom} \nabla f_n. \tag{4.50}$$

It therefore follows from the Brézis–Haraux theorem [11, Théorème 4] that

$$(\forall n \in \mathbb{N}) \quad \operatorname{ran}(\nabla f_n + \gamma_n A) = \mathcal{X}^*, \tag{4.51}$$

and hence that condition [c] in Algorithm 4.5 holds. It remains to verify condition [d]2/ in Theorem 4.9. Set $\varphi \colon \ell^p(\mathbb{N}) \to \left[0,+\infty\right[\colon z \mapsto \|z\|^p/p$ and $(\forall n \in \mathbb{N})$ $\varphi_n \colon \ell^p(\mathbb{N}) \to \left[0,+\infty\right[\colon z \mapsto \chi_n\|z\|^p/p$. Take a sequence $(z_n,\xi_n)_{n\in\mathbb{N}}$ in $\operatorname{dom} A$ and a point $(z,\xi) \in \operatorname{dom} A$ such that $(z_n,\xi_n) \rightharpoonup (z,\xi)$. We have $\xi_n \to \xi$ and $(\forall k \in \mathbb{N})$ $\zeta_{n,k} \to \zeta_k$. Now let $(e_k)_{k\in\mathbb{N}}$ be the canonical Schauder basis of $\ell^p(\mathbb{N})$. Then

$$(\forall k \in \mathbb{N}) \quad \big\langle e_k, \nabla\varphi_n(z_n)\big\rangle = \chi_n \operatorname{sign}(\zeta_{n,k})|\zeta_{n,k}|^{p-1} \to \operatorname{sign}(\zeta_k)|\zeta_k|^{p-1} = \big\langle e_k, \nabla\varphi(z)\big\rangle \tag{4.52}$$

and $(\nabla\varphi_n(z_n))_{n\in\mathbb{N}}$ is bounded. It therefore follows from [2, Théorème VIII-2] that $\nabla\varphi_n(z_n) \rightharpoonup \nabla\varphi(z)$ and, in turn, that $\nabla f_n(z_n,\xi_n) \rightharpoonup \nabla g(z,\xi)$ by (4.47) and (4.49). Note that the above setting is not covered by the assumptions underlying (4.4)–(4.7): the fact that $B \neq 0$ excludes [6], the fact that $\mathcal{X}$ is not a Hilbert space excludes [15] and [20], and [18] is excluded because A is not a subdifferential.

4.2.3 Special cases and applications

We illustrate the general scope of Theorem 4.9 by recovering apparently unrelated results and also by deriving new ones. Sufficient conditions for [a] and [b] in Theorem 4.9 to hold can be found in Propositions 4.6(v) and 4.7, respectively. As to checking the focusing condition [c], the following fact will be useful.

Lemma 4.11 [13, Proposition 2.1(iii)] *Let $M_1 \colon \mathcal{X} \to 2^{\mathcal{X}^*}$ and $M_2 \colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be maximally monotone, let $(a_n,a_n^*)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} M_1$, let $(b_n,b_n^*)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} M_2$, let*

$x \in \mathcal{X}$, *and let* $y^* \in \mathcal{X}^*$. *Suppose that* $a_n \rightharpoonup x$, $b_n^* \rightharpoonup y^*$, $a_n^* + b_n^* \to 0$, *and* $a_n - b_n \to 0$. *Then* $x \in \operatorname{zer}(M_1 + M_2)$.

4.2.3.1 Recovering existing frameworks for monotone inclusions

In this section, we show that the existing results of [6, 15, 20] discussed in the Introduction can be recovered from Theorem 4.9. As will be clear from the proofs, more general versions of these results can also be derived at once from Theorem 4.9. First, we derive from Theorem 4.9 the convergence of the Bregman-based proximal point algorithm (4.4) studied in [6, Section 5.5].

Corollary 4.12 *Let* $A\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ *be maximally monotone, let* $f \in \Gamma_0(\mathcal{X})$ *be a supercoercive Legendre function such that* $\varnothing \neq \operatorname{zer} A \subset \operatorname{dom} A \subset \operatorname{int}\operatorname{dom} f$ *and* ∇f *is weakly sequentially continuous, and let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0, +\infty[$ *such that* $\inf_{n\in\mathbb{N}} \gamma_n > 0$. *Suppose that, for every bounded sequence* $(y_n)_{n\in\mathbb{N}}$ *in* $\operatorname{int}\operatorname{dom} f$,

$$D_f(y_{n+1}, y_n) \to 0 \quad \Rightarrow \quad \nabla f(y_{n+1}) - \nabla f(y_n) \to 0. \tag{4.53}$$

Take $x_0 \in C$ *and set* $(\forall n \in \mathbb{N})$ $x_{n+1} = \big(\nabla f + \gamma_n A\big)^{-1}(\nabla f(x_n))$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\operatorname{zer} A$.

Proof. We apply Theorem 4.9 with $B = 0$, $\alpha = 1$, $\kappa = \delta_1 = \delta_2 = 0$, and $(\forall n \in \mathbb{N})$ $f_n = f$. First, (4.1) together with conditions [a] and [b] in Algorithm 4.5 are trivially fulfilled. On the other hand, since f is a Legendre function and $\operatorname{dom} A \subset \operatorname{int}\operatorname{dom} f$, condition [c] in Algorithm 4.5 follows from [6, Theorem 3.13(iv)(d)]. Next, condition [a] in Theorem 4.9 follows from Proposition 4.6(v)[b]. Furthermore, in view of the weak sequential continuity of ∇f, condition [d]2/ in Theorem 4.9 is satisfied with $g = f$. Next, to show that the algorithm is focusing, suppose that $\sum_{n\in\mathbb{N}} D_f(x_{n+1}, x_n) < +\infty$ and take $x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$. Since $(x_n)_{n\in\mathbb{N}}$ is a bounded sequence in $\operatorname{int}\operatorname{dom} f$, we derive from (4.53) that $\nabla f(x_{n+1}) - \nabla f(x_n) \to 0$. In turn, since $\inf_{n\in\mathbb{N}} \gamma_n > 0$, it follows that $\gamma_n^{-1}(\nabla f(x_{n+1}) - \nabla f(x_n)) \to 0$. However, by construction, $(\forall n \in \mathbb{N})$ $\gamma_{k_n-1}^{-1}(\nabla f(x_{k_n-1}) - \nabla f(x_{k_n})) \in A x_{k_n}$. Therefore, upon invoking Lemma 4.11 (with $M_1 = A$ and $M_2 = 0$), we obtain $x \in \operatorname{zer} A$ and the algorithm is therefore focusing. This also shows that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \operatorname{zer} A \subset \operatorname{int}\operatorname{dom} f$. Condition [b] in Theorem 4.9 is thus satisfied. ☐

The next application of Theorem 4.9 is a variable metric version of the Hilbertian forward-backward method (4.5) established in [15, Theorem 4.1].

Corollary 4.13 *Let* $\mathcal{X}$ *be a real Hilbert space, let* $A\colon \mathcal{X} \to 2^{\mathcal{X}}$ *be maximally monotone, let* α *and* β *be in* $]0, +\infty[$, *and let* $B\colon \mathcal{X} \to \mathcal{X}$ *satisfy*

$$(\forall x \in \mathcal{X})(\forall y \in \mathcal{X}) \quad \langle x - y \mid Bx - By\rangle \geqslant \beta\|Bx - By\|^2. \tag{4.54}$$

Further, for every $n \in \mathbb{N}$, *let* $U_n\colon \mathcal{X} \to \mathcal{X}$ *be a bounded linear operator which is* α*-strongly monotone and self-adjoint. Suppose that* $\operatorname{zer}(A+B) \neq \varnothing$ *and that there exists a summable sequence*

$(\eta_n)_{n\in\mathbb{N}}$ *in* $[0,+\infty[$ *such that*

$$(\forall n\in\mathbb{N})(\forall x\in\mathcal{X})\quad \langle x\mid U_{n+1}x\rangle\leqslant(1+\eta_n)\langle x\mid U_nx\rangle. \tag{4.55}$$

Let $\varepsilon\in\,]0,2\beta[$ *and let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,+\infty[$ *such that* $0<\inf_{n\in\mathbb{N}}\gamma_n\leqslant\sup_{n\in\mathbb{N}}\gamma_n\leqslant(2\beta-\varepsilon)\alpha$. *Define a sequence* $(x_n)_{n\in\mathbb{N}}$ *via the recursion*

$$x_0\in\operatorname{dom}A\quad\textit{and}\quad(\forall n\in\mathbb{N})\quad x_{n+1}=(U_n+\gamma_nA)^{-1}(U_nx_n-\gamma_nBx_n). \tag{4.56}$$

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

Proof. Set $f=\|\cdot\|^2/2$, $C=\operatorname{dom}A$, and $\mathscr{S}=\operatorname{zer}(A+B)$. In addition, for every $n\in\mathbb{N}$, define $f_n\colon\mathcal{X}\to\mathbb{R}\colon x\mapsto\langle x\mid U_nx\rangle/2$. Let us apply Theorem 4.9 with $\kappa=1/(2\beta-\varepsilon)$, $\delta_1=0$, and $\delta_2=(2\beta-\varepsilon)/(2\beta)\in\,]0,1[$. First, $f\in\Gamma_0(\mathcal{X})$ is a supercoercive Legendre function with $\operatorname{dom}f=\mathcal{X}$ and, for every $n\in\mathbb{N}$, since $\nabla f_n=U_n$ is α-strongly monotone, $f_n\in\mathcal{C}_\alpha(f)$. Furthermore, it follows from Proposition 4.2(vi)[a] that (4.1) is fulfilled. We also observe that condition [a] in Algorithm 4.5 is satisfied. Next, by (4.55) and the assumption that the operators $(U_n)_{n\in\mathbb{N}}$ are self-adjoint,

$$\begin{aligned}(\forall n\in\mathbb{N})(\forall x\in\mathcal{X})(\forall y\in\mathcal{X})\quad D_{f_{n+1}}(x,y)&=\frac{1}{2}\langle x-y\mid U_{n+1}(x-y)\rangle\\&\leqslant\frac{1+\eta_n}{2}\langle x-y\mid U_n(x-y)\rangle\\&=D_{f_n}(x,y)\end{aligned} \tag{4.57}$$

and condition [b] in Algorithm 4.5 therefore holds. Now take $n\in\mathbb{N}$. Since $\nabla f_n=U_n$ is maximally monotone with $\operatorname{dom}\nabla f_n=\mathcal{X}$ and A is maximally monotone, [7, Corollary 25.5(i)] entails that $\nabla f_n+\gamma_nA$ is maximally monotone. Thus, since $\nabla f_n+\gamma_nA$ is α-strongly monotone, [7, Proposition 22.11(ii)] implies that $\operatorname{ran}(\nabla f_n+\gamma_nA)=\mathcal{X}$ and it follows that condition [c] in Algorithm 4.5 is satisfied. Next, in view of Proposition 4.6(v)[b], $(x_n)_{n\in\mathbb{N}}$ is bounded, while $\mathfrak{W}(x_n)_{n\in\mathbb{N}}\subset\mathcal{X}=\operatorname{int}\operatorname{dom}f$. Now set $\mu=\sup_{n\in\mathbb{N}}\|U_n\|$. For every $n\in\mathbb{N}$, since it results from (4.55) and [7, Fact 2.25(iii)] that

$$(\forall x\in\mathcal{X})\quad\|x\|\leqslant1\quad\Rightarrow\quad\langle x\mid U_nx\rangle\leqslant\Bigg(\prod_{k\in\mathbb{N}}(1+\eta_k)\Bigg)\langle x\mid U_0x\rangle\leqslant\Bigg(\prod_{k\in\mathbb{N}}(1+\eta_k)\Bigg)\|U_0\|, \tag{4.58}$$

we derive from [7, Fact 2.25(iii)] that $\|U_n\|\leqslant\|U_0\|\prod_{k\in\mathbb{N}}(1+\eta_k)$. Hence $\mu<+\infty$ and therefore, appealing to [14, Lemma 2.3(i)], there exists an α-strongly monotone self-adjoint bounded linear operator $U\colon\mathcal{X}\to\mathcal{X}$ such that $(\forall w\in\mathcal{X})\;U_nw\to Uw$. Define $g\colon\mathcal{X}\to\mathbb{R}\colon x\mapsto\langle x\mid Ux\rangle/2$. Then $\nabla g=U$ is strongly monotone (and thus strictly monotone). Furthermore, given $(y_n)_{n\in\mathbb{N}}$ in C and $y\in\mathfrak{W}(y_n)_{n\in\mathbb{N}}\cap C$, say $y_{k_n}\rightharpoonup y$, we have

$$(\forall w\in\mathcal{X})\quad\langle w\mid\nabla f_{k_n}(y_{k_n})\rangle=\langle U_{k_n}w\mid y_{k_n}\rangle\to\langle Uw\mid y\rangle=\langle w\mid Uy\rangle=\langle w\mid\nabla g(y)\rangle \tag{4.59}$$

and thus $\nabla f_{k_n}(y_{k_n}) \rightharpoonup \nabla g(y)$. Therefore, condition [d]2/ in Theorem 4.9 is satisfied. Let us now verify that (4.56) is focusing. Towards this goal, take $z \in \mathscr{S}$ and suppose that $\sum_{n\in\mathbb{N}}(1-\delta_2)\langle x_n - z \mid Bx_n - Bz\rangle < +\infty$ and $\sum_{n\in\mathbb{N}}(1-\kappa\gamma_n/\alpha)D_{f_n}(x_{n+1}, x_n) < +\infty$. Since $\delta_2 < 1$ and $\sup_{n\in\mathbb{N}}(\kappa\gamma_n) < \alpha$, we infer from (4.54) that

$$\sum_{n\in\mathbb{N}} \|Bx_n - Bz\|^2 \leqslant \frac{1}{\beta}\sum_{n\in\mathbb{N}} \langle x_n - z \mid Bx_n - Bz\rangle < +\infty \tag{4.60}$$

and $\sum_{n\in\mathbb{N}} \|x_{n+1}-x_n\|^2 = 2\sum_{n\in\mathbb{N}} D_f(x_{n+1}, x_n) \leqslant (2/\alpha)\sum_{n\in\mathbb{N}} D_{f_n}(x_{n+1}, x_n) < +\infty$. It follows that

$$\|U_n(x_{n+1} - x_n)\| \leqslant \mu\|x_{n+1} - x_n\| \to 0. \tag{4.61}$$

Now take $x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$, and set $(\forall n \in \mathbb{N})$ $x^*_{n+1} = \gamma_n^{-1}U_n(x_n - x_{n+1}) - Bx_n$. It results from (4.56) that $(x_{k_n+1}, x^*_{k_n+1})_{n\in\mathbb{N}}$ lies in $\operatorname{gra} A$ and from (4.61) that $x_{k_n+1} \rightharpoonup x$. Moreover, (4.61) yields $x^*_{k_n+1} + Bx_{k_n} \to 0$. Altogether, Lemma 4.11 (applied to the sequences $(x_{k_n+1}, x^*_{k_n+1})_{n\in\mathbb{N}}$ in $\operatorname{gra} A$ and $(x_{k_n}, Bx_{k_n})_{n\in\mathbb{N}}$ in $\operatorname{gra} B$) guarantees that $x \in \operatorname{zer}(A+B)$. Consequently, Theorem 4.9 asserts that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\mathscr{S}$. $\square$

Example 4.14 The classical forward-backward method is obtained by setting $U_n \equiv \mathrm{Id}$ in Corollary 4.13, which yields

$$x_0 \in \operatorname{dom} A \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = (\mathrm{Id} + \gamma_n A)^{-1}(x_n - \gamma_n Bx_n). \tag{4.62}$$

The case when the proximal parameters $(\gamma_n)_{n\in\mathbb{N}}$ are constant was first addressed in [17].

We now turn to the Renaud–Cohen algorithm (4.7) and recover [20, Theorem 3.4].

Corollary 4.15 *Let $\mathcal{X}$ be a real Hilbert space, let $A\colon \mathcal{X} \to 2^{\mathcal{X}}$ and $B\colon \mathcal{X} \to \mathcal{X}$ be maximally monotone, and let $f\colon \mathcal{X} \to \mathbb{R}$ be convex and Fréchet differentiable. Suppose that $\operatorname{zer}(A+B) \neq \varnothing$, that ∇f is 1-strongly monotone on $\operatorname{dom} A$ and Lipschitzian on bounded sets, and that there exists $\beta \in \left]0,+\infty\right[$ such that*

$$\big(\forall (x,x^*) \in \operatorname{gra}(A+B)\big)\big(\forall (y,y^*) \in \operatorname{gra}(A+B)\big) \quad \langle x - y \mid x^* - y^*\rangle \geqslant \beta\|Bx - By\|^2. \tag{4.63}$$

Let $\gamma \in \left]0, 2\beta\right[$, take $x_0 \in \operatorname{dom} A$, and set $(\forall n \in \mathbb{N})$ $x_{n+1} = (\nabla f + \gamma A)^{-1}(\nabla f(x_n) - \gamma Bx_n)$. Suppose, in addition, that ∇f is weakly sequentially continuous. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}(A+B)$.

Proof. Let $\varepsilon \in \left]0, 2\beta\right[$ be such that $\gamma < 2\beta - \varepsilon$. We apply Theorem 4.9 with $C = \operatorname{dom} A$, $\alpha = 1$, $\kappa = 1/(2\beta-\varepsilon)$, $\delta_1 = \delta_2 = (2\beta-\varepsilon)/(2\beta) \in \left]0,1\right[$, and $(\forall n \in \mathbb{N})$ $f_n = f$ and $\eta_n = 0$. Proposition 4.2(iv) asserts that (4.1) is satisfied. Furthermore, as shown in the proof of Proposition 4.2(iv),

$$(\forall x \in \overline{\operatorname{dom}}\, A)(\forall y \in \overline{\operatorname{dom}}\, A) \quad D_f(x,y) \geqslant \frac{1}{2}\|x-y\|^2. \tag{4.64}$$

Next, note that conditions [a] and [b] in Algorithm 4.5 are trivially satisfied. Since $\nabla f + \gamma A$ is strongly monotone and since, by [7, Corollary 25.5(i)], $\nabla f + \gamma A$ is maximally monotone, it follows from [7, Proposition 22.11(ii)] that $\operatorname{ran}(\nabla f + \gamma A) = \mathcal{X}$ and therefore that condition [c] in Algorithm 4.5 holds. We observe that condition [b] in Theorem 4.9 is trivially satisfied and that condition [a] in Theorem 4.9 follows from (4.64) and Proposition 4.6(i). Furthermore, since ∇f is weakly sequentially continuous and 1-strongly monotone on C, condition [d]2/ in Theorem 4.9 is satisfied with $g = f$. Now take $z \in \operatorname{zer}(A + B)$ and suppose that $\sum_{n\in\mathbb{N}}(1-\kappa\gamma)D_f(x_{n+1},x_n) < +\infty$, $\sum_{n\in\mathbb{N}}(1-\delta_2)\langle x_n - z \mid Bx_n - Bz\rangle < +\infty$, and $\sum_{n\in\mathbb{N}}\langle x_{n+1} - z \mid \gamma^{-1}(\nabla f(x_n) - \nabla f(x_{n+1})) - Bx_n + Bz\rangle < +\infty$. Then, since $\kappa\gamma < 1$ and $\delta_2 < 1$, it follows that

$$\sum_{n\in\mathbb{N}} D_f(x_{n+1},x_n) < +\infty \quad \text{and} \quad \sum_{n\in\mathbb{N}} \langle x_n - z \mid Bx_n - Bz\rangle < +\infty, \tag{4.65}$$

and therefore that

$$\sum_{n\in\mathbb{N}} \big\langle x_{n+1} - z \mid \gamma^{-1}(\nabla f(x_n) - \nabla f(x_{n+1})) - Bx_n + Bx_{n+1}\big\rangle < +\infty. \tag{4.66}$$

Since $(z,0) \in \operatorname{gra}(A+B)$ and since the sequence $(x_{n+1}, \gamma^{-1}(\nabla f(x_n) - \nabla f(x_{n+1})) - Bx_n + Bx_{n+1})_{n\in\mathbb{N}}$ lies in $\operatorname{gra}(A+B)$ by construction, it follows from (4.63) and (4.66) that $\sum_{n\in\mathbb{N}}\|Bx_n - Bz\|^2 < +\infty$. On the other hand, since $(x_n)_{n\in\mathbb{N}}$ lies in $\operatorname{dom} A$ by Proposition 4.6, we deduce from (4.64) and (4.65) that $x_{n+1} - x_n \to 0$. In turn, it results from the Lipschitz continuity of ∇f on the bounded set $\{x_n\}_{n\in\mathbb{N}}$ that $\nabla f(x_n) - \nabla f(x_{n+1}) \to 0$. Now take $x \in \mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \rightharpoonup x$, and set $(\forall n \in \mathbb{N})$ $x^*_{n+1} = \gamma^{-1}(\nabla f(x_n) - \nabla f(x_{n+1})) - Bx_n$. Then $(x_{k_n+1}, x^*_{k_n+1})_{n\in\mathbb{N}}$ lies in $\operatorname{gra} A$. Furthermore, $x^*_{k_n+1} + Bx_{k_n} = \gamma^{-1}(\nabla f(x_{k_n}) - \nabla f(x_{k_n+1})) \to 0$ and, since $x_n - x_{n+1} \to 0$, $x_{k_n+1} \rightharpoonup x$. Thus, applying Lemma 4.11 with the sequences $(x_{k_n+1}, x^*_{k_n+1})_{n\in\mathbb{N}}$ and $(x_{k_n}, Bx_{k_n})_{n\in\mathbb{N}}$ yields $x \in \operatorname{zer}(A+B)$, and we conclude that condition [c] in Theorem 4.9 is satisfied as well. ☐

4.2.3.2 The finite-dimensional case

We discuss the finite-dimensional case, a setting in which the assumptions can be greatly simplified and the results presented below are new.

Corollary 4.16 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 4.5. In addition, suppose that the following hold:*

[a] *$\mathcal{X}$ is finite-dimensional.*

[b] *f is essentially strictly convex and* dom f^* *is open.*

[c] $(\operatorname{int}\operatorname{dom} f) \cap \overline{\operatorname{dom} A} \subset \operatorname{int}\operatorname{dom} B$.

[d] $\sup_{n\in\mathbb{N}}(\kappa\gamma_n) < \alpha$.

[e] *There exists a function g in $\Gamma_0(\mathcal{X})$ which is differentiable on* $\operatorname{int}\operatorname{dom} g \supset \operatorname{int}\operatorname{dom} f$, *with ∇g strictly monotone on C, and such that, for every sequence $(y_n)_{n\in\mathbb{N}}$ in C and every sequential cluster point $y \in \operatorname{int}\operatorname{dom} f$ of $(y_n)_{n\in\mathbb{N}}$, $y_{k_n} \to y \Rightarrow \nabla f_{k_n}(y_{k_n}) \to \nabla g(y)$.*

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

Proof. It follows from Proposition 4.6(v)[e] that $(x_n)_{n\in\mathbb{N}}$ is bounded and from Proposition 4.7[d] that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \operatorname{int}\operatorname{dom} f$. In view of Theorem 4.9, it remains to show that Algorithm 4.5 is focusing. Towards this goal, let $z \in \mathscr{S}$, and suppose that $(D_{f_n}(z, x_n))_{n\in\mathbb{N}}$ converges and $\sum_{n\in\mathbb{N}}(1-\kappa\gamma_n/\alpha)D_{f_n}(x_{n+1}, x_n) < +\infty$, and let x be a sequential cluster point of $(x_n)_{n\in\mathbb{N}}$, say $x_{k_n} \to x$. Using [d] and the fact that $(f_n)_{n\in\mathbb{N}}$ lies in $\mathcal{C}_\alpha(f)$, we obtain

$$\big(D_f(z, x_n)\big)_{n\in\mathbb{N}} \text{ is bounded} \quad \text{and} \quad \sum_{n\in\mathbb{N}} D_{f_n}(x_{n+1}, x_n) < +\infty. \tag{4.67}$$

Since $(x_{k_n})_{n\in\mathbb{N}}$ lies in $\operatorname{int}\operatorname{dom} f$, [4, Theorem 3.8(ii)] and (4.67) imply that

$$x \in \operatorname{int}\operatorname{dom} f \tag{4.68}$$

and [5, Theorem 5.10] thus yields

$$\nabla f(x_{k_n}) \to \nabla f(x) \in \operatorname{int}\operatorname{dom} f^*. \tag{4.69}$$

Next, it results from [b], [5, Lemma 7.3(vii)], and (4.67) that

$$\big(D_{f^*}(\nabla f(x_n), \nabla f(z))\big)_{n\in\mathbb{N}} = \big(D_f(z, x_n)\big)_{n\in\mathbb{N}} \text{ is bounded.} \tag{4.70}$$

Therefore, since $\nabla f(z) \in \operatorname{int}\operatorname{dom} f^*$ [5, Theorem 5.10] and since f^* is a Legendre function [5, Corollary 5.5], it results from [5, Lemma 7.3(v)] that $(\nabla f(x_{k_n+1}))_{n\in\mathbb{N}}$ is bounded. In turn, there exists a strictly increasing sequence $(l_{k_n})_{n\in\mathbb{N}}$ in $\mathbb{N}$ and a point $x^* \in \mathcal{X}^*$ such that

$$\nabla f(x_{l_{k_n}+1}) \to x^*. \tag{4.71}$$

By lower semicontinuity of $D_{f^*}(\cdot, \nabla f(z))$ and (4.70), $x^* \in \operatorname{dom} f^*$. On the other hand, appealing to [5, Lemma 7.3(vii)] and (4.67), we obtain

$$0 \leqslant D_{f^*}\big(\nabla f(x_{l_{k_n}}), \nabla f(x_{l_{k_n}+1})\big) = D_f\big(x_{l_{k_n}+1}, x_{l_{k_n}}\big) \leqslant \frac{1}{\alpha} D_{f_{l_{k_n}}}\big(x_{l_{k_n}+1}, x_{l_{k_n}}\big) \to 0. \tag{4.72}$$

Thus, since $(\nabla f(x_n))_{n\in\mathbb{N}}$ lies in $\operatorname{int}\operatorname{dom} f^*$ by virtue of Proposition 4.6 and [5, Theorem 5.10], we derive from [4, Theorem 3.9(iii)], (4.69), and (4.71) that $x^* = \nabla f(x)$ and, hence, from (4.71) that $\nabla f(x_{l_{k_n}+1}) \to \nabla f(x)$. It thus follows from [5, Theorem 5.10] that $x_{l_{k_n}+1} \to x$. In turn, by using respectively [e] with the sequences $(x_n)_{n\in\mathbb{N}}$ and $(x_{n+1})_{n\in\mathbb{N}}$, we get $\nabla f_{l_{k_n}}(x_{l_{k_n}}) \to \nabla g(x)$ and $\nabla f_{l_{k_n}}(x_{l_{k_n}+1}) \to \nabla g(x)$. Now set $(\forall n \in \mathbb{N})$ $x^*_{n+1} = \gamma_n^{-1}(\nabla f_n(x_n) - \nabla f_n(x_{n+1})) - Bx_n$. Then, by construction of $(x_n)_{n\in\mathbb{N}}$, $(\forall n \in \mathbb{N})$ $(x_{n+1}, x^*_{n+1}) \in \operatorname{gra} A$. In addition, since $\inf_{n\in\mathbb{N}} \gamma_n >$

0 and $\nabla f_{l_{k_n}}(x_{l_{k_n}}) - \nabla f_{l_{k_n}}(x_{l_{k_n}+1}) \to \nabla g(x) - \nabla g(x) = 0$, we deduce that $x^*_{l_{k_n}+1} + Bx_{l_{k_n}} \to 0$. On the other hand, since $(x_n)_{n\in\mathbb{N}}$ lies in $\operatorname{dom} A$ and $x_{k_n} \to x$, it follows that $x \in \overline{\operatorname{dom}}\, A$ and therefore, by (4.68) and [c], that $x \in \operatorname{int}\operatorname{dom} B$. Hence, using [21, Corollary 1.1], we obtain $Bx_{l_{k_n}} \to Bx$. Altogether, Lemma 4.11 (applied to the sequence $(x_{l_{k_n}+1}, x^*_{l_{k_n}+1})_{n\in\mathbb{N}}$ in $\operatorname{gra} A$ and the sequence $(x_{l_{k_n}}, Bx_{l_{k_n}})_{n\in\mathbb{N}}$ in $\operatorname{gra} B$) asserts that $x \in \operatorname{zer}(A + B)$. In view of Theorem 4.9, we conclude that $(x_n)_{n\in\mathbb{N}}$ converges to a point in $\mathscr{S}$. ☐

4.2.3.3 Forward-backward splitting for convex minimization

In this section, we study the convergence of (4.6). Our results improve on and complement those of [18].

Problem 4.17 Let $\varphi \in \Gamma_0(\mathcal{X})$, let $\psi \in \Gamma_0(\mathcal{X})$, and let $f \in \Gamma_0(\mathcal{X})$ be essentially smooth. Set $C = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom}\partial\varphi$ and $\mathscr{S} = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{Argmin}(\varphi + \psi)$. Suppose that $\varphi + \psi$ is coercive, $\varnothing \neq C \subset \operatorname{int}\operatorname{dom}\psi$, $\mathscr{S} \neq \varnothing$, ψ is Gâteaux differentiable on $\operatorname{int}\operatorname{dom}\psi$, and there exists $\kappa \in \left]0, +\infty\right[$ such that

$$(\forall x \in C)(\forall y \in C) \quad D_\psi(x, y) \leqslant \kappa D_f(x, y). \tag{4.73}$$

The objective is to find a point in $\mathscr{S}$.

In the context of Problem 4.17, given $\gamma \in \left]0, +\infty\right[$ and $g \in \mathcal{C}_\alpha(f)$, we define $\operatorname{prox}^g_{\gamma\varphi} = (\nabla g + \gamma\partial\varphi)^{-1}$.

Algorithm 4.18 Consider the setting of Problem 4.17. Let $\alpha \in \left]0, +\infty\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be in $\left]0, +\infty\right[$, and let $(f_n)_{n\in\mathbb{N}}$ be in $\mathcal{C}_\alpha(f)$. Suppose that the following hold:

[a] There exists $\varepsilon \in \left]0, 1\right[$ such that $0 < \inf_{n\in\mathbb{N}} \gamma_n \leqslant \sup_{n\in\mathbb{N}} \gamma_n \leqslant \alpha(1 - \varepsilon)/\kappa$.

[b] There exists a summable sequence $(\eta_n)_{n\in\mathbb{N}}$ in $\left[0, +\infty\right[$ such that $(\forall n \in \mathbb{N})\ D_{f_{n+1}} \leqslant (1 + \eta_n)D_{f_n}$.

[c] For every $n \in \mathbb{N}$, $\operatorname{int}\operatorname{dom} f_n = \operatorname{dom}\partial f_n$ and ∇f_n is strictly monotone on C.

Take $x_0 \in C$ and set $(\forall n \in \mathbb{N})\ x_{n+1} = \operatorname{prox}^{f_n}_{\gamma_n\varphi}(\nabla f_n(x_n) - \gamma_n\nabla\psi(x_n))$.

Theorem 4.19 *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 4.18 and suppose that the following hold:*

[a] $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \operatorname{int}\operatorname{dom} f$.

[b] *One of the following is satisfied:*

1/ *$\mathscr{S}$ is a singleton.*

2/ *There exists a function g in $\Gamma_0(\mathcal{X})$ which is Gâteaux differentiable on $\operatorname{int}\operatorname{dom} g \supset C$, with ∇g strictly monotone on C, and such that, for every sequence $(y_n)_{n\in\mathbb{N}}$ in C and every $y \in \mathfrak{W}(y_n)_{n\in\mathbb{N}} \cap C$, $y_{k_n} \rightharpoonup y \Rightarrow \nabla f_{k_n}(y_{k_n}) \rightharpoonup \nabla g(y)$.*

Then the following hold:

(i) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\mathscr{S}$.

(ii) $(x_n)_{n\in\mathbb{N}}$ *is a monotone minimizing sequence:* $\varphi(x_n)+\psi(x_n)\downarrow\min(\varphi+\psi)(\mathcal{X})$.

(iii) $\sum_{n\in\mathbb{N}}((\varphi+\psi)(x_n)-\min(\varphi+\psi)(\mathcal{X}))<+\infty$ *and* $(\varphi+\psi)(x_n)-\min(\varphi+\psi)(\mathcal{X})=o(1/n)$.

(iv) $\sum_{n\in\mathbb{N}} n(D_{f_n}(x_{n+1},x_n)+D_{f_n}(x_n,x_{n+1}))<+\infty$.

Proof. (i): We shall derive this result from Theorem 4.9 with $A=\partial\varphi$, $B=\partial\psi$, $\delta_1=0$, and $\delta_2=1$. First, appealing to [24, Theorem 2.4.4(i)], B is single-valued on $\operatorname{int}\operatorname{dom}B=\operatorname{int}\operatorname{dom}\psi$ and $B=\nabla\psi$ on $\operatorname{int}\operatorname{dom}B$. Next, set $\theta=\varphi+\psi$. Since $\varnothing\neq(\operatorname{int}\operatorname{dom}f)\cap\operatorname{dom}\partial\varphi\subset\operatorname{int}\operatorname{dom}\psi$, we have $\operatorname{dom}\varphi\cap\operatorname{int}\operatorname{dom}\psi\neq\varnothing$. Hence, [9, Theorem 4.1.19] yields $A+B=\partial\theta$. Therefore, $\operatorname{Argmin}\theta=\operatorname{zer}\partial\theta=\operatorname{zer}(A+B)$ and $\mathscr{S}=(\operatorname{int}\operatorname{dom}f)\cap\operatorname{zer}(A+B)$. Next, in view of Proposition 4.2(iii), (4.1) is fulfilled. On the other hand, conditions [a] and [b] in Algorithm 4.5 are trivially satisfied. To verify condition [c] in Algorithm 4.5, it suffices to show that, for every $n\in\mathbb{N}$, $(\nabla f_n-\gamma_nB)(C)\subset\operatorname{ran}(\nabla f_n+\gamma_nA)$, i.e., since $C\subset\operatorname{int}\operatorname{dom}B$ and $B=\nabla\psi$ on $\operatorname{int}\operatorname{dom}B$, that $(\nabla f_n-\gamma_n\nabla\psi)(C)\subset\operatorname{ran}(\nabla f_n+\gamma_nA)$. To do so, fix temporarily $n\in\mathbb{N}$, let $x\in C$, and set

$$A_n=\nabla f_n+\gamma_nA-\nabla f_n(x)+\gamma_n\nabla\psi(x). \tag{4.74}$$

Then, since $\operatorname{dom}\partial f_n\cap\operatorname{dom}A=(\operatorname{int}\operatorname{dom}f_n)\cap\operatorname{dom}A=(\operatorname{int}\operatorname{dom}f)\cap\operatorname{dom}A\neq\varnothing$ by condition [c] in Algorithm 4.18, it results from [6, Proposition 3.12] that A_n is maximally monotone. Next, we deduce from condition [a] in Algorithm 4.18 and (4.73) that

$$(\forall u\in C)(\forall v\in C)\quad \gamma_nD_\psi(u,v)\leqslant\alpha(1-\varepsilon)D_\psi(u,v)/\kappa\leqslant\alpha(1-\varepsilon)D_f(u,v)\leqslant(1-\varepsilon)D_{f_n}(u,v). \tag{4.75}$$

In turn,

$$\begin{aligned}(\forall u\in C)(\forall v\in C)\quad \gamma_n\langle u-v,\nabla\psi(u)-\nabla\psi(v)\rangle&=\gamma_n\big(D_\psi(u,v)+D_\psi(v,u)\big)\\ &\leqslant(1-\varepsilon)\big(D_f(u,v)+D_f(v,u)\big)\\ &=(1-\varepsilon)\langle u-v,\nabla f_n(u)-\nabla f_n(v)\rangle.\end{aligned} \tag{4.76}$$

However, by coercivity of θ, there exists $\rho\in\left]0,+\infty\right[$ such that

$$(\forall y\in\mathcal{X})\quad \|y\|\geqslant\rho\quad\Rightarrow\quad \inf\langle y,(A+B)(y+x)\rangle=\inf\langle y,\partial\theta(y+x)\rangle\geqslant\theta(y+x)-\theta(x)\geqslant0. \tag{4.77}$$

Now suppose that $(y,y^*)\in\operatorname{gra}A_n(\cdot+x)$ satisfies $\|y\|\geqslant\rho$. Then $y+x\in\operatorname{dom}\nabla f_n\cap\operatorname{dom}A=(\operatorname{int}\operatorname{dom}f_n)\cap\operatorname{dom}A=C$ and $y^*-\nabla f_n(y+x)+\gamma_n\nabla\psi(y+x)+\nabla f_n(x)-\gamma_n\nabla\psi(x)\in\gamma_n(A+B)(y+x)$. Thus, it follows from (4.77) and (4.76) that

$$0\leqslant\langle y,y^*\rangle-\big\langle(y+x)-x,(\nabla f_n-\gamma_n\nabla\psi)(y+x)-(\nabla f_n-\gamma_n\nabla\psi)(x)\big\rangle\leqslant\langle y,y^*\rangle. \tag{4.78}$$

Therefore, in view of [22, Proposition 2] and the maximal monotonicity of $A_n(\cdot+x)$, there exists $\overline{y} \in \mathcal{X}$ such that $0 \in A_n(\overline{y}+x)$. Hence $(\nabla f_n - \gamma_n \nabla\psi)(x) \in \nabla f_n(\overline{y}+x) + \gamma_n A(\overline{y}+x) \subset \operatorname{ran}(\nabla f_n + \gamma_n A)$, as desired. Since $(x_{n+1}, \gamma_n^{-1}(\nabla f_n(x_n) - \nabla f_n(x_{n+1})) - \nabla\psi(x_n))$ lies in $\operatorname{gra}\partial\varphi$ by construction, we derive from [6, Proposition 2.3(ii)] that

$$\begin{aligned}(\forall x \in C) \quad \varphi(x) &\geqslant \varphi(x_{n+1}) - \langle x - x_{n+1}, \nabla\psi(x_n)\rangle + \gamma_n^{-1}\langle x - x_{n+1}, \nabla f_n(x_n) - \nabla f_n(x_{n+1})\rangle \\ &\geqslant \varphi(x_{n+1}) - \langle x - x_{n+1}, \nabla\psi(x_n)\rangle \\ &\quad + \gamma_n^{-1}\big(D_{f_n}(x, x_{n+1}) + D_{f_n}(x_{n+1}, x_n) - D_{f_n}(x, x_n)\big). \end{aligned} \tag{4.79}$$

On the other hand, (4.75) and the convexity of ψ entail that

$$\begin{aligned}(\forall x \in C) \quad \psi(x_{n+1}) &\leqslant \psi(x_n) + \langle x_{n+1} - x_n, \nabla\psi(x_n)\rangle + (1-\varepsilon)\gamma_n^{-1} D_{f_n}(x_{n+1}, x_n) \\ &= \psi(x_n) + \langle x - x_n, \nabla\psi(x_n)\rangle + \langle x_{n+1} - x, \nabla\psi(x_n)\rangle \\ &\quad + (1-\varepsilon)\gamma_n^{-1} D_{f_n}(x_{n+1}, x_n) \\ &\leqslant \psi(x) + \langle x_{n+1} - x, \nabla\psi(x_n)\rangle + (1-\varepsilon)\gamma_n^{-1} D_{f_n}(x_{n+1}, x_n). \end{aligned} \tag{4.80}$$

Altogether, upon adding (4.79) and (4.80), we obtain

$$(\forall x \in C) \quad \theta(x_{n+1}) + \gamma_n^{-1} D_{f_n}(x, x_{n+1}) + \varepsilon\gamma_n^{-1} D_{f_n}(x_{n+1}, x_n) \leqslant \theta(x) + \gamma_n^{-1} D_{f_n}(x, x_n). \tag{4.81}$$

In particular, since $x_n \in C$,

$$\theta(x_{n+1}) + \gamma_n^{-1}\big(D_{f_n}(x_n, x_{n+1}) + \varepsilon D_{f_n}(x_{n+1}, x_n)\big) \leqslant \theta(x_n). \tag{4.82}$$

This shows that

$$\big(\theta(x_n)\big)_{n\in\mathbb{N}} \text{ decreases.} \tag{4.83}$$

In turn, using the coercivity of θ, we infer that $(x_n)_{n\in\mathbb{N}}$ is bounded, which secures [a] in Theorem 4.9. It remains to verify that Algorithm 4.18 is focusing. Towards this end, let $z \in \mathscr{S}$ and suppose that

$$\big(D_{f_n}(z, x_n)\big)_{n\in\mathbb{N}} \text{ converges} \tag{4.84}$$

and

$$\varepsilon \sum_{n\in\mathbb{N}} D_{f_n}(x_{n+1}, x_n) \leqslant \sum_{n\in\mathbb{N}} (1 - \kappa\gamma_n/\alpha) D_{f_n}(x_{n+1}, x_n) < +\infty. \tag{4.85}$$

Set $\gamma = \inf_{n\in\mathbb{N}} \gamma_n$ and $\ell = \lim D_{f_n}(z, x_n)$. It follows from (4.81) applied to $z \in C$ that

$$(\forall n \in \mathbb{N}) \quad \gamma\big(\theta(x_{n+1}) - \min\theta(\mathcal{X})\big) + D_{f_n}(z, x_{n+1}) + \varepsilon D_{f_n}(x_{n+1}, x_n) \leqslant D_{f_n}(z, x_n) \tag{4.86}$$

and therefore from condition [b] in Algorithm 4.18 that

$$(\forall n \in \mathbb{N}) \quad \gamma\big(\theta(x_{n+1}) - \min\theta(\mathcal{X})\big) + D_{f_{n+1}}(z, x_{n+1}) + \varepsilon D_{f_n}(x_{n+1}, x_n)$$

$$\leqslant (1+\eta_n)\Big(\gamma\big(\theta(x_{n+1})-\min\theta(\mathcal{X})\big)+D_{f_n}(z,x_{n+1})+\varepsilon D_{f_n}(x_{n+1},x_n)\Big)$$
$$\leqslant (1+\eta_n)D_{f_n}(z,x_n). \tag{4.87}$$

Hence, $\varlimsup\gamma(\theta(x_{n+1})-\min\theta(\mathcal{X}))+\ell\leqslant\ell$ and therefore $\varlimsup(\theta(x_{n+1})-\min\theta(\mathcal{X}))=0$. Thus

$$\theta(x_n)\to\min\theta(\mathcal{X}). \tag{4.88}$$

Now take $x\in\mathfrak{W}(x_n)_{n\in\mathbb{N}}$, say $x_{k_n}\rightharpoonup x$. By weak lower semicontinuity of θ, $\min\theta(\mathcal{X})\leqslant\theta(x)\leqslant\varliminf\theta(x_{k_n})=\min\theta(\mathcal{X})$ and it follows that $x\in\operatorname{Argmin}\theta=\operatorname{zer}(A+B)$. Consequently, Theorem 4.9 asserts that $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\mathscr{S}$.

(ii): Combine (4.83) and (4.88).

(iii)&(iv): Fix $z\in\mathscr{S}$ and set $\gamma=\inf_{n\in\mathbb{N}}\gamma_n$. Arguing along the same lines as above, we obtain

$$(\forall n\in\mathbb{N})\quad \gamma\big(\theta(x_{n+1})-\min\theta(\mathcal{X})\big)+D_{f_{n+1}}(z,x_{n+1})+\varepsilon D_{f_n}(x_{n+1},x_n)\leqslant(1+\eta_n)D_{f_n}(z,x_n) \tag{4.89}$$

and therefore [7, Lemma 5.31] guarantees that $\sum_{n\in\mathbb{N}}(\theta(x_n)-\min\theta(\mathcal{X}))<+\infty$. In addition, $(\theta(x_n)-\min\theta(\mathcal{X}))_{n\in\mathbb{N}}$ is decreasing by virtue of (4.83). However, recall that if $(\alpha_n)_{n\in\mathbb{N}}$ is a decreasing sequence in $[0,+\infty[$ such that $\sum_{n\in\mathbb{N}}\alpha_n<+\infty$, then

$$\alpha_n=o\left(\frac{1}{n}\right)\quad\text{and}\quad\sum_{n\in\mathbb{N}}n(\alpha_n-\alpha_{n+1})<+\infty. \tag{4.90}$$

Hence, $\theta(x_n)-\min\theta(\mathcal{X})=o(1/n)$ and $\sum_{n\in\mathbb{N}}n(\theta(x_n)-\theta(x_{n+1}))<+\infty$. Consequently, since (4.81) yields

$$(\forall n\in\mathbb{N})\quad \gamma_n^{-1}D_{f_n}(x_n,x_{n+1})+\varepsilon\gamma_n^{-1}D_{f_n}(x_{n+1},x_n)\leqslant\theta(x_n)-\theta(x_{n+1}), \tag{4.91}$$

we infer that $\sum_{n\in\mathbb{N}}n(D_{f_n}(x_{n+1},x_n)+D_{f_n}(x_n,x_{n+1}))<+\infty$. ☐

Remark 4.20 Let us relate Theorem 4.19 to the literature.

(i) The conclusions of items (i) and (ii) are obtained in [18, Theorem 1(2)] under more restrictive conditions on the sequences $(\gamma_n)_{n\in\mathbb{N}}$ and $(f_n)_{n\in\mathbb{N}}$. Thus, we do not require in Theorem 4.19 the additional condition $(\forall n\in\mathbb{N})$ $(1+\eta_n)\gamma_n-\gamma_{n+1}\leqslant\alpha\eta_n/\kappa$. Furthermore, we do not suppose either that $-\operatorname{ran}\nabla\psi\subset\operatorname{dom}\varphi^*$ or that the functions $(f_n)_{n\in\mathbb{N}}$ are cofinite.

(ii) Items (iii) and (iv) are new even in Euclidean spaces. In the finite-dimensional setting, partial results can be found in [3], where:

(a) A single convex function is used: $(\forall n\in\mathbb{N})$ $f_n=f$.

(b) The viability of the sequence $(x_n)_{n\in\mathbb{N}}$ is a blanket assumption, while it is guaranteed in Theorem 4.19.

(c) Only the rates $\sum_{n\in\mathbb{N}} D_f(x_{n+1},x_n) < +\infty$ and $(\varphi+\psi)(x_n)-\min(\varphi+\psi)(\mathcal{X}) = O(1/n)$ are obtained.

4.2.3.4 Further applications

Theorems 4.9 and 4.19 operate under broad assumptions which go beyond those of the existing forward-backward methods of [6, 15, 18, 20] described in (4.4)–(4.7). Here are two examples which do not fit the existing scenarios and exploit this generality.

Example 4.21 Consider the setting of Problem 4.1. Suppose, in addition, that the following hold:

[a] A is uniformly monotone on bounded sets.

[b] There exist $\psi \in \Gamma_0(\mathcal{X})$ and $\kappa \in \left]0,+\infty\right[$ such that $B = \partial\psi$ and $(\forall x \in C)(\forall y \in C)$ $D_\psi(x,y) \leqslant \kappa D_f(x,y)$.

[c] f is supercoercive.

[d] $\operatorname{zer}(A+B) \subset \operatorname{int}\operatorname{dom} f$.

Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,+\infty\right[$ such that $0 < \inf_{n\in\mathbb{N}}\gamma_n \leqslant \sup_{n\in\mathbb{N}}\gamma_n < 1/\kappa$, take $x_0 \in C$, and set $(\forall n \in \mathbb{N})$ $x_{n+1} = (\nabla f + \gamma_n A)^{-1}(\nabla f(x_n) - \gamma_n \nabla\psi(x_n))$. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to the unique zero of $A + \nabla\psi$.

The next example concerns variational inequalities.

Example 4.22 Let $\varphi \in \Gamma_0(\mathcal{X})$, let $B\colon \mathcal{X} \to 2^{\mathcal{X}^*}$ be maximally monotone, let $f \in \Gamma_0(\mathcal{X})$ be essentially smooth, and set $C = (\operatorname{int}\operatorname{dom} f) \cap \operatorname{dom}\partial\varphi$. Suppose that $C \subset \operatorname{int}\operatorname{dom} B$ and B is single-valued on $\operatorname{int}\operatorname{dom} B$. Consider the problem of finding a point in

$$\mathscr{S} = \big\{x \in C \mid (\forall y \in \mathcal{X})\ \langle x-y, Bx\rangle + \varphi(x) \leqslant \varphi(y)\big\}, \tag{4.92}$$

which is assumed to be nonempty. This is a special case of Problem 4.1 with $A = \partial\varphi$ and, given $x_0 \in C$, Algorithm 4.5 produces the iterations $(\forall n \in \mathbb{N})$ $x_{n+1} = \operatorname{prox}^{f_n}_{\gamma_n\varphi}(\nabla f_n(x_n) - \gamma_n B x_n)$. The weak convergence of $(x_n)_{n\in\mathbb{N}}$ to a point in $\mathscr{S}$ is discussed in Theorem 4.9. Even in Euclidean spaces, this scheme is new and of interest since, as shown in [3, 13, 18], the Bregman proximity operator $\operatorname{prox}^{f_n}_{\gamma_n\varphi}$ may be easier to compute for a particular f_n than for the standard kernel $\|\cdot\|^2/2$. Altogether, our framework makes it possible to solve variational inequalities by forward-backward splitting with non-cocoercive operators and/or outside of Hilbert spaces.

References

[1] 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.

[2] S. Banach, *Théorie des Opérations Linéaires.* Seminar. Matem. Univ. Warszawa, 1932.

[3] 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.

[4] H. H. Bauschke and J. M. Borwein, Legendre functions and the method of random Bregman projections, *J. Convex Anal.*, vol. 4, pp. 27–67, 1997.

[5] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Essential smoothness, essential strict convexity, and Legendre functions in Banach spaces, *Commun. Contemp. Math.*, vol. 3, pp. 615–647, 2001.

[6] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, Bregman monotone optimization algorithms, *SIAM J. Control Optim.*, vol. 42, pp. 596–636, 2003.

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

[8] H. H. Bauschke, M. N. Dao, and S. B. Lindstrom, Regularizing with Bregman–Moreau envelopes, *SIAM J. Optim.*, vol. 28, pp. 3208–3228, 2018.

[9] J. M. Borwein and J. D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples.* Cambridge University Press, 2010.

[10] N. Bourbaki, *Espaces Vectoriels Topologiques, Chapitres 1 à 5.* Masson, Paris, 1981. English translation: *Topological Vector Spaces, Chapters 1–5.* Springer-Verlag, New York, 1987.

[11] H. Brézis and A. Haraux, Image d'une somme d'opérateurs monotones et applications, *Israel J. Math.*, vol. 23, pp. 165–186, 1976.

[12] Y. Censor and S. A. Zenios, *Parallel Optimization – Theory, Algorithms and Applications.* Oxford University Press, New York, 1997.

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

[14] P. L. Combettes and B. C. Vũ, Variable metric quasi-Fejér monotonicity, *Nonlinear Anal.*, vol. 78, pp. 17–31, 2013.

[15] 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.

[16] J. Frecon, S. Salzo, and M. Pontil, Bilevel learning of the group lasso structure, *Adv. Neural Inform. Process. Syst.*, vol. 31, pp. 8301–8311, 2018.

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

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

[19] G. Ortiz-Jiménez, M. El Gheche, E. Simou, H. Petric Maretić, and P. Frossard, Forward-backward splitting for optimal transport based problems, *Proc. Intl. Conf. Acoust., Speech, Signal Process.*, pp. 5405–5409, 2020.

[20] 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.

[21] R. T. Rockafellar, Local boundedness of nonlinear, monotone operators, *Michigan Math. J.*, vol. 16, pp. 397–407, 1969.

[22] R. T. Rockafellar, On the maximality of sums of nonlinear monotone operators, *Trans. Amer. Math. Soc.*, vol. 149, pp. 75–88, 1970.

[23] S. Salzo, The variable metric forward-backward splitting algorithm under mild differentiability assumptions, *SIAM J. Optim.*, vol. 27, pp. 2153–2181, 2017.

[24] C. Zălinescu, *Convex Analysis in General Vector Spaces.* World Scientific Publishing, River Edge, NJ, 2002.

Chapter 5

PROJECTIVE SPLITTING AS A WARPED PROXIMAL ALGORITHM

5.1 Introduction and context

We complement Chapter 3 by showing that the asynchronous block-iterative algorithm [6, Algorithm 12] can be viewed as a special case of the warped proximal algorithm of Theorem 3.16. This answers question **(Q4)** of Chapter 1.

This chapter presents the following article:

> M. N. Bùi, Projective splitting as a warped proximal algorithm, submitted.

5.2 Article: Projective splitting as a warped proximal algorithm

Abstract. We show that the asynchronous block-iterative primal-dual projective splitting framework introduced by P. L. Combettes and J. Eckstein in their 2018 *Math. Program.* paper can be viewed as an instantiation of the recently proposed warped proximal algorithm.

In [4], the warped proximal algorithm was proposed and its pertinence was illustrated through the ability to unify existing methods such as those of [1, 2, 8, 9], and to design novel flexible ones for solving challenging monotone inclusions. Let us state a version of [4, Theorem 4.2].

Fact 5.1 *Let* H *be a real Hilbert space, let* $\mathsf{M}\colon \mathsf{H} \to 2^{\mathsf{H}}$ *be a maximally monotone operator such that* $\operatorname{zer}\mathsf{M} \neq \varnothing$*, let* $\mathsf{x}_0 \in \mathsf{H}$*, let* $\varepsilon \in \left]0,1\right[$*, let* $\alpha \in \left]0,+\infty\right[$*, and let* $\beta \in \left[\alpha,+\infty\right[$*. For every* $n \in \mathbb{N}$*,* *let* $\mathbf{K}_n\colon \mathbf{H}\to\mathbf{H}$ *be* α*-strongly monotone and* β*-Lipschitzian, and let* $\lambda_n\in[\varepsilon,2-\varepsilon]$. *Iterate*

$$
\begin{array}{l}
\textit{for } n=0,1,\ldots\\
\left\lfloor
\begin{array}{l}
\textit{take } \widetilde{\mathbf{x}}_n\in\mathbf{H}\\
\mathbf{y}_n=(\mathbf{K}_n+\mathbf{M})^{-1}(\mathbf{K}_n\widetilde{\mathbf{x}}_n)\\
\mathbf{y}_n^*=\mathbf{K}_n\widetilde{\mathbf{x}}_n-\mathbf{K}_n\mathbf{y}_n\\
\textit{if } \langle\mathbf{x}_n-\mathbf{y}_n\mid\mathbf{y}_n^*\rangle>0\\
\left\lfloor
\begin{array}{l}
\mathbf{x}_{n+1}=\mathbf{x}_n-\dfrac{\lambda_n\langle\mathbf{x}_n-\mathbf{y}_n\mid\mathbf{y}_n^*\rangle}{\|\mathbf{y}_n^*\|^2}\,\mathbf{y}_n^*
\end{array}
\right.\\
\textit{else}\\
\left\lfloor
\begin{array}{l}
\mathbf{x}_{n+1}=\mathbf{x}_n.
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{5.1}
$$

Then the following hold:

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

(ii) $\sum_{n\in\mathbb{N}}\|\mathbf{x}_{n+1}-\mathbf{x}_n\|^2<+\infty$.

(iii) $(\forall n\in\mathbb{N})\ \langle\mathbf{x}_n-\mathbf{y}_n\mid\mathbf{y}_n^*\rangle\leqslant\varepsilon^{-1}\|\mathbf{y}_n^*\|\,\|\mathbf{x}_{n+1}-\mathbf{x}_n\|$.

(iv) *Suppose that* $\widetilde{\mathbf{x}}_n-\mathbf{x}_n\to\mathbf{0}$. *Then* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\operatorname{zer}\mathbf{M}$.

A problem of interest in modern nonlinear analysis is the following (see, e.g, [1, 5, 6] and the references therein for discussions on this problem).

Problem 5.2 Let $(\mathcal{H}_i)_{i\in I}$ and $(\mathcal{G}_k)_{k\in K}$ be finite families of real Hilbert spaces. For every $i\in I$ and every $k\in K$, let $A_i\colon\mathcal{H}_i\to 2^{\mathcal{H}_i}$ and $B_k\colon\mathcal{G}_k\to 2^{\mathcal{G}_k}$ be maximally monotone, let $z_i^*\in\mathcal{H}_i$, let $r_k\in\mathcal{G}_k$, and let $L_{k,i}\colon\mathcal{H}_i\to\mathcal{G}_k$ be linear and bounded. The problem is to

$$
\text{find }\ (\overline{x}_i)_{i\in I}\in\underset{i\in I}{\times}\mathcal{H}_i\ \text{ and }\ (\overline{v}_k^*)_{k\in K}\in\underset{k\in K}{\times}\mathcal{G}_k\ \text{ such that }\
\begin{cases}
(\forall i\in I)\ \ z_i^*-\displaystyle\sum_{k\in K}L_{k,i}^*\overline{v}_k^*\in A_i\overline{x}_i\\
(\forall k\in K)\ \ \displaystyle\sum_{i\in I}L_{k,i}\overline{x}_i-r_k\in B_k^{-1}\overline{v}_k^*.
\end{cases}
\tag{5.2}
$$

The set of solutions to (5.2) is denoted by $\mathbf{Z}$.

The first asynchronous block-iterative algorithm to solve Problem 5.2 was proposed in [6, Algorithm 12] as an extension of projective splitting techniques found in [1, 7]. The present paper shows that [6, Algorithm 12] can be viewed as a special case of (5.1). Towards this goal, we first derive an abstract weak convergence principle from Fact 5.1. We refer the reader to [3] for background in monotone operator theory and nonlinear analysis.

Theorem 5.3 *Let* $\mathbf{H}$ *be a real Hilbert space, let* $\mathbf{A}\colon\mathbf{H}\to 2^{\mathbf{H}}$ *be a maximally monotone operator, and let* $\mathbf{S}\colon\mathbf{H}\to\mathbf{H}$ *be a bounded linear operator such that* $\mathbf{S}^*=-\mathbf{S}$. *In addition, let* $\mathbf{x}_0\in\mathbf{H}$, *let* $\varepsilon\in\,]0,1[$, *let* $\alpha\in\,]0,+\infty[$, *let* $\rho\in[\alpha,+\infty[$, *and for every* $n\in\mathbb{N}$, *let* $\mathbf{F}_n\colon\mathbf{H}\to\mathbf{H}$ *be* α*-strongly* *monotone and ρ-Lipschitzian, and let $\lambda_n \in [\varepsilon, 2-\varepsilon]$. Iterate*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\textit{take } \mathbf{u}_n \in \mathbf{H},\ \mathbf{e}_n^* \in \mathbf{H},\ \textit{and } \mathbf{f}_n^* \in \mathbf{H} \\
\mathbf{u}_n^* = \mathbf{F}_n \mathbf{u}_n - \mathbf{S}\mathbf{u}_n + \mathbf{e}_n^* + \mathbf{f}_n^* \\
\mathbf{y}_n = (\mathbf{F}_n + \mathbf{A})^{-1} \mathbf{u}_n^* \\
\mathbf{a}_n^* = \mathbf{u}_n^* - \mathbf{F}_n \mathbf{y}_n \\
\mathbf{y}_n^* = \mathbf{a}_n^* + \mathbf{S}\mathbf{y}_n \\
\pi_n = \langle \mathbf{x}_n \mid \mathbf{y}_n^* \rangle - \langle \mathbf{y}_n \mid \mathbf{a}_n^* \rangle \\
\textit{if } \pi_n > 0 \\
\left\lfloor
\begin{array}{l}
\tau_n = \|\mathbf{y}_n^*\|^2 \\
\theta_n = \lambda_n \pi_n / \tau_n \\
\mathbf{x}_{n+1} = \mathbf{x}_n - \theta_n \mathbf{y}_n^*
\end{array}
\right. \\
\textit{else} \\
\left\lfloor
\begin{array}{l}
\mathbf{x}_{n+1} = \mathbf{x}_n.
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{5.3}
$$

Suppose that $\operatorname{zer}(\mathbf{A} + \mathbf{S}) \neq \varnothing$. *Then the following hold:*

(i) $\sum_{n\in\mathbb{N}} \|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 < +\infty$.

(ii) *Suppose that $\mathbf{u}_n - \mathbf{x}_n \to \mathbf{0}$, that $\mathbf{e}_n^* \to \mathbf{0}$, that $(\mathbf{f}_n^*)_{n\in\mathbb{N}}$ is bounded, and that there exists $\delta \in \,]0,1[$ such that*

$$
(\forall n \in \mathbb{N}) \qquad
\begin{cases}
\langle \mathbf{u}_n - \mathbf{y}_n \mid \mathbf{f}_n^* \rangle \geqslant -\delta \langle \mathbf{u}_n - \mathbf{y}_n \mid \mathbf{F}_n \mathbf{u}_n - \mathbf{F}_n \mathbf{y}_n \rangle \\
\langle \mathbf{a}_n^* + \mathbf{S}\mathbf{u}_n - \mathbf{e}_n^* \mid \mathbf{f}_n^* \rangle \leqslant \delta \|\mathbf{a}_n^* + \mathbf{S}\mathbf{u}_n - \mathbf{e}_n^*\|^2.
\end{cases}
\tag{5.4}
$$

Then $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}(\mathbf{A} + \mathbf{S})$.

Proof. Set $\mathbf{M} = \mathbf{A} + \mathbf{S}$ and $(\forall n \in \mathbb{N})\ \mathbf{K}_n = \mathbf{F}_n - \mathbf{S}$. Then, it follows from [3, Example 20.35 and Corollary 25.5(i)] that $\mathbf{M}$ is maximally monotone with $\operatorname{zer} \mathbf{M} \neq \varnothing$. Now take $n \in \mathbb{N}$. We have

$$
\mathbf{K}_n + \mathbf{M} = \mathbf{F}_n + \mathbf{A}. \tag{5.5}
$$

Since $\mathbf{S}^* = -\mathbf{S}$, we deduce that

$$
\mathbf{K}_n \text{ is } \alpha\text{-strongly monotone and } \beta\text{-Lipschitzian,} \tag{5.6}
$$

where $\beta = \rho + \|\mathbf{S}\|$. Thus, [3, Corollary 20.28 and Proposition 22.11(ii)] guarantee that there exists $\widetilde{\mathbf{x}}_n \in \mathbf{H}$ such that

$$
\mathbf{u}_n^* = \mathbf{K}_n \widetilde{\mathbf{x}}_n. \tag{5.7}
$$

Hence, by (5.3) and (5.5),

$$
\mathbf{y}_n = (\mathbf{K}_n + \mathbf{M})^{-1}(\mathbf{K}_n \widetilde{\mathbf{x}}_n) \quad \text{and} \quad \mathbf{y}_n^* = \mathbf{u}_n^* - \mathbf{F}_n \mathbf{y}_n + \mathbf{S}\mathbf{y}_n = \mathbf{K}_n \widetilde{\mathbf{x}}_n - \mathbf{K}_n \mathbf{y}_n. \tag{5.8}
$$

At the same time, we have $\langle \mathbf{y}_n \mid \mathbf{S}\mathbf{y}_n \rangle = 0$ and it thus results from (5.3) that $\pi_n = \langle \mathbf{x}_n \mid \mathbf{y}_n^* \rangle - \langle \mathbf{y}_n \mid \mathbf{a}_n^* + \mathbf{S}\mathbf{y}_n \rangle = \langle \mathbf{x}_n - \mathbf{y}_n \mid \mathbf{y}_n^* \rangle$. Altogether, (5.3) is a special case of (5.1).

(i): Fact 5.1(ii).

(ii): In the light of Fact 5.1(iv), it suffices to verify that $\widetilde{\mathbf{x}}_n - \mathbf{x}_n \to \mathbf{0}$. For every $n \in \mathbb{N}$, since $\mathbf{K}_n + \mathbf{M}$ is maximally monotone [3, Corollary 25.5(i)] and α-strongly monotone, [3, Example 22.7 and Proposition 22.11(ii)] implies that $(\mathbf{K}_n + \mathbf{M})^{-1} \colon \mathbf{H} \to \mathbf{H}$ is $(1/\alpha)$-Lipschitzian. Therefore, we derive from (5.3), (5.5), [4, Proposition 3.10(i)], and (5.6) that $(\forall \mathbf{z} \in \operatorname{zer} \mathbf{M})(\forall n \in \mathbb{N})\ \alpha\|\mathbf{y}_n - \mathbf{z}\| = \alpha\|(\mathbf{K}_n + \mathbf{M})^{-1}\mathbf{u}_n^* - (\mathbf{K}_n + \mathbf{M})^{-1}(\mathbf{K}_n\mathbf{z})\| \leqslant \|\mathbf{u}_n^* - \mathbf{K}_n\mathbf{z}\| = \|\mathbf{K}_n\mathbf{u}_n - \mathbf{K}_n\mathbf{z} + \mathbf{e}_n^* + \mathbf{f}_n^*\| \leqslant \|\mathbf{K}_n\mathbf{u}_n - \mathbf{K}_n\mathbf{z}\| + \|\mathbf{e}_n^*\| + \|\mathbf{f}_n^*\| \leqslant \beta\|\mathbf{u}_n - \mathbf{z}\| + \|\mathbf{e}_n^*\| + \|\mathbf{f}_n^*\|$. Thus, since Fact 5.1(i) and our assumption imply that $(\mathbf{u}_n)_{n\in\mathbb{N}}$ is bounded, it follows that $(\mathbf{y}_n)_{n\in\mathbb{N}}$ is bounded. At the same time, for every $n \in \mathbb{N}$, we get from (5.3) that

$$\mathbf{y}_n^* = \mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n + \mathbf{e}_n^* + \mathbf{f}_n^* - (\mathbf{S}\mathbf{u}_n - \mathbf{S}\mathbf{y}_n) = \mathbf{K}_n\mathbf{u}_n - \mathbf{K}_n\mathbf{y}_n + \mathbf{e}_n^* + \mathbf{f}_n^* \tag{5.9}$$

and, thus, from (5.6) that $\|\mathbf{y}_n^*\| \leqslant \|\mathbf{K}_n\mathbf{u}_n - \mathbf{K}_n\mathbf{y}_n\| + \|\mathbf{e}_n^*\| + \|\mathbf{f}_n^*\| \leqslant \beta\|\mathbf{u}_n - \mathbf{y}_n\| + \|\mathbf{e}_n^*\| + \|\mathbf{f}_n^*\|$. Thus, $(\mathbf{y}_n^*)_{n\in\mathbb{N}}$ is bounded, from which, (i), and Fact 5.1(iii) we obtain $\varlimsup \langle \mathbf{x}_n - \mathbf{y}_n \mid \mathbf{y}_n^* \rangle \leqslant 0$. In turn, since $\mathbf{x}_n - \mathbf{u}_n \to \mathbf{0}$ and $\mathbf{e}_n^* \to \mathbf{0}$, it results from (5.9) and (5.4) that

$$\begin{aligned}
0 &\geqslant \varlimsup \langle \mathbf{x}_n - \mathbf{y}_n \mid \mathbf{y}_n^* \rangle \\
&= \varlimsup \big(\langle \mathbf{u}_n - \mathbf{y}_n \mid \mathbf{y}_n^* \rangle + \langle \mathbf{x}_n - \mathbf{u}_n \mid \mathbf{y}_n^* \rangle \big) \\
&= \varlimsup \langle \mathbf{u}_n - \mathbf{y}_n \mid \mathbf{y}_n^* \rangle \\
&= \varlimsup \big(\langle \mathbf{u}_n - \mathbf{y}_n \mid \mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n + \mathbf{e}_n^* + \mathbf{f}_n^* \rangle - \langle \mathbf{u}_n - \mathbf{y}_n \mid \mathbf{S}\mathbf{u}_n - \mathbf{S}\mathbf{y}_n \rangle \big) \\
&= \varlimsup \big(\langle \mathbf{u}_n - \mathbf{y}_n \mid \mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n + \mathbf{f}_n^* \rangle + \langle \mathbf{u}_n - \mathbf{y}_n \mid \mathbf{e}_n^* \rangle \big) \\
&\geqslant \varlimsup \big((1-\delta)\langle \mathbf{u}_n - \mathbf{y}_n \mid \mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n \rangle + \langle \mathbf{u}_n - \mathbf{y}_n \mid \mathbf{e}_n^* \rangle \big) \\
&\geqslant \varlimsup \alpha(1-\delta)\|\mathbf{u}_n - \mathbf{y}_n\|^2 \\
&\geqslant \varlimsup \alpha(1-\delta)\rho^{-2}\|\mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n\|^2.
\end{aligned} \tag{5.10}$$

Hence, $\mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n \to \mathbf{0}$. On the other hand, since $(\mathbf{f}_n^*)_{n\in\mathbb{N}}$ is bounded and since (5.3) yields $(\mathbf{a}_n^* + \mathbf{S}\mathbf{u}_n - \mathbf{e}_n^*)_{n\in\mathbb{N}} = (\mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n + \mathbf{f}_n^*)_{n\in\mathbb{N}}$, we derive from (5.4) that

$$\begin{aligned}
\varlimsup (1-\delta)\|\mathbf{f}_n^*\|^2 &= \varlimsup \big(\langle \mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n \mid \mathbf{f}_n^* \rangle + (1-\delta)\|\mathbf{f}_n^*\|^2 \big) \\
&= \varlimsup \big(\langle \mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n + \mathbf{f}_n^* \mid \mathbf{f}_n^* \rangle - \delta\|\mathbf{f}_n^*\|^2 \big) \\
&\leqslant \varlimsup \big(\delta\|\mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n + \mathbf{f}_n^*\|^2 - \delta\|\mathbf{f}_n^*\|^2 \big) \\
&= \varlimsup \big(\delta\|\mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n\|^2 + 2\delta\langle \mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n \mid \mathbf{f}_n^* \rangle \big) \\
&= 0.
\end{aligned} \tag{5.11}$$

Therefore, $\mathbf{f}_n^* \to \mathbf{0}$. Consequently, by (5.6), (5.7), and (5.3), $\alpha\|\widetilde{\mathbf{x}}_n - \mathbf{x}_n\| \leqslant \|\mathbf{K}_n\widetilde{\mathbf{x}}_n - \mathbf{K}_n\mathbf{x}_n\| = \|\mathbf{K}_n\mathbf{u}_n - \mathbf{K}_n\mathbf{x}_n + \mathbf{e}_n^* + \mathbf{f}_n^*\| \leqslant \beta\|\mathbf{u}_n - \mathbf{x}_n\| + \|\mathbf{e}_n^*\| + \|\mathbf{f}_n^*\| \to 0$. ☐

We are now ready to recover [6, Theorem 13]. Recall that, given a real Hilbert space $\mathcal{H}$ with identity operator Id, the resolvent of an operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is $J_A = (\mathrm{Id} + A)^{-1}$.

Corollary 5.4 ([6]) *Consider the setting of Problem 5.2 and suppose that* $\mathbf{Z} \neq \varnothing$*. Let* $(I_n)_{n\in\mathbb{N}}$ *be nonempty subsets of* I *and* $(K_n)_{n\in\mathbb{N}}$ *be nonempty subsets of* K *such that*

$$I_0 = I, \quad K_0 = K, \quad \textit{and} \quad (\exists\, T \in \mathbb{N})(\forall n \in \mathbb{N}) \ \bigcup_{j=n}^{n+T} I_j = I \ \textit{and} \ \bigcup_{j=n}^{n+T} K_j = K. \tag{5.12}$$

In addition, let $D \in \mathbb{N}$*, let* $\varepsilon \in \left]0,1\right[$*, let* $(\lambda_n)_{n\in\mathbb{N}}$ *be in* $[\varepsilon, 2-\varepsilon]$*, and for every* $i \in I$ *and every* $k \in K$*, let* $(c_i(n))_{n\in\mathbb{N}}$ *and* $(d_k(n))_{n\in\mathbb{N}}$ *be in* $\mathbb{N}$ *such that*

$$(\forall n \in \mathbb{N}) \quad n - D \leqslant c_i(n) \leqslant n \quad \textit{and} \quad n - D \leqslant d_k(n) \leqslant n, \tag{5.13}$$

let $(\gamma_{i,n})_{n\in\mathbb{N}}$ *and* $(\mu_{k,n})_{n\in\mathbb{N}}$ *be in* $[\varepsilon, 1/\varepsilon]$, *let* $x_{i,0} \in \mathcal{H}_i$, *and let* $v^*_{k,0} \in \mathcal{G}_k$. *Iterate*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\textit{for every } i \in I_n \\
\left\lfloor
\begin{array}{l}
\textit{take } e_{i,n} \in \mathcal{H}_i \\
l^*_{i,n} = \sum_{k\in K} L^*_{k,i} v^*_{k,c_i(n)} \\
a_{i,n} = J_{\gamma_{i,c_i(n)} A_i}\big(x_{i,c_i(n)} + \gamma_{i,c_i(n)}(z^*_i - l^*_{i,n}) + e_{i,n}\big) \\
a^*_{i,n} = \gamma^{-1}_{i,c_i(n)}(x_{i,c_i(n)} - a_{i,n} + e_{i,n}) - l^*_{i,n}
\end{array}
\right. \\
\textit{for every } i \in I \smallsetminus I_n \\
\left\lfloor
\begin{array}{l}
a_{i,n} = a_{i,n-1} \\
a^*_{i,n} = a^*_{i,n-1}
\end{array}
\right. \\
\textit{for every } k \in K_n \\
\left\lfloor
\begin{array}{l}
\textit{take } f_{k,n} \in \mathcal{G}_k \\
l_{k,n} = \sum_{i\in I} L_{k,i} x_{i,d_k(n)} \\
b_{k,n} = r_k + J_{\mu_{k,d_k(n)} B_k}\big(l_{k,n} + \mu_{k,d_k(n)} v^*_{k,d_k(n)} + f_{k,n} - r_k\big) \\
b^*_{k,n} = v^*_{k,d_k(n)} + \mu^{-1}_{k,d_k(n)}(l_{k,n} - b_{k,n} + f_{k,n}) \\
t_{k,n} = b_{k,n} - \sum_{i\in I} L_{k,i} a_{i,n}
\end{array}
\right. \\
\textit{for every } k \in K \smallsetminus K_n \\
\left\lfloor
\begin{array}{l}
b_{k,n} = b_{k,n-1} \\
b^*_{k,n} = b^*_{k,n-1} \\
t_{k,n} = b_{k,n} - \sum_{i\in I} L_{k,i} a_{i,n}
\end{array}
\right. \\
\textit{for every } i \in I \\
\left\lfloor
\begin{array}{l}
t^*_{i,n} = a^*_{i,n} + \sum_{k\in K} L^*_{k,i} b^*_{k,n}
\end{array}
\right. \\
\pi_n = \sum_{i\in I} \big(\langle x_{i,n} \mid t^*_{i,n}\rangle - \langle a_{i,n} \mid a^*_{i,n}\rangle\big) + \sum_{k\in K} \big(\langle t_{k,n} \mid v^*_{k,n}\rangle - \langle b_{k,n} \mid b^*_{k,n}\rangle\big) \\
\textit{if } \pi_n > 0 \\
\left\lfloor
\begin{array}{l}
\tau_n = \sum_{i\in I} \|t^*_{i,n}\|^2 + \sum_{k\in K} \|t_{k,n}\|^2 \\
\theta_n = \lambda_n \pi_n / \tau_n
\end{array}
\right. \\
\textit{else} \\
\left\lfloor
\begin{array}{l}
\theta_n = 0
\end{array}
\right. \\
\textit{for every } i \in I \\
\left\lfloor
\begin{array}{l}
x_{i,n+1} = x_{i,n} - \theta_n t^*_{i,n}
\end{array}
\right. \\
\textit{for every } k \in K \\
\left\lfloor
\begin{array}{l}
v^*_{k,n+1} = v^*_{k,n} - \theta_n t_{k,n}.
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{5.14}
$$

In addition, suppose that there exist $\eta \in \left]0,+\infty\right[$, $\chi \in \left]0,+\infty\right[$, $\sigma \in \left]0,1\right[$, and $\zeta \in \left]0,1\right[$ such that

$$(\forall n \in \mathbb{N})(\forall i \in I_n) \quad \begin{cases} \|e_{i,n}\| \leqslant \eta \\ \langle x_{i,c_i(n)} - a_{i,n} \mid e_{i,n}\rangle \geqslant -\sigma \|x_{i,c_i(n)} - a_{i,n}\|^2 \\ \langle e_{i,n} \mid a^*_{i,n} + l^*_{i,n}\rangle \leqslant \sigma \gamma_{i,c_i(n)} \|a^*_{i,n} + l^*_{i,n}\|^2 \end{cases} \tag{5.15}$$

and that

$$(\forall n \in \mathbb{N})(\forall k \in K_n) \quad \begin{cases} \|f_{k,n}\| \leqslant \chi \\ \langle l_{k,n} - b_{k,n} \mid f_{k,n}\rangle \geqslant -\zeta \|l_{k,n} - b_{k,n}\|^2 \\ \langle f_{k,n} \mid b^*_{k,n} - v^*_{k,d_k(n)}\rangle \leqslant \zeta \mu_{k,d_k(n)} \|b^*_{k,n} - v^*_{k,d_k(n)}\|^2. \end{cases} \tag{5.16}$$

Then $((x_{i,n})_{i\in I}, (v^*_{k,n})_{k\in K})_{n\in\mathbb{N}}$ converges weakly to a point in $\mathbf{Z}$.

Proof. Denote by $\boldsymbol{\mathcal{H}}$ and $\boldsymbol{\mathcal{G}}$ the Hilbert direct sums of $(\mathcal{H}_i)_{i\in I}$ and $(\mathcal{G}_k)_{k\in K}$, set $\mathbf{H} = \boldsymbol{\mathcal{H}} \oplus \boldsymbol{\mathcal{G}}$, and define the operators

$$\mathbf{A} \colon \mathbf{H} \to 2^{\mathbf{H}} \colon \big((x_i)_{i\in I}, (v^*_k)_{k\in K}\big) \mapsto \left(\bigtimes_{i\in I} \big(-z^*_i + A_i x_i\big)\right) \times \left(\bigtimes_{k\in K} \big(r_k + B_k^{-1} v^*_k\big)\right) \tag{5.17}$$

and

$$\mathbf{S} \colon \mathbf{H} \to \mathbf{H} \colon \big((x_i)_{i\in I}, (v^*_k)_{k\in K}\big) \mapsto \left(\left(\sum_{k\in K} L^*_{k,i} v^*_k\right)_{i\in I}, \left(-\sum_{i\in I} L_{k,i} x_i\right)_{k\in K}\right). \tag{5.18}$$

Using the maximal monotonicity of the operators $(A_i)_{i\in I}$ and $(B_k)_{k\in K}$, we deduce from [3, Proposition 20.23] that $\mathbf{A}$ is maximally monotone. In addition, we observe that $\mathbf{S}$ is a bounded linear operator with $\mathbf{S}^* = -\mathbf{S}$. At the same time, it results from (5.17), (5.18), and (5.2) that

$$\operatorname{zer}(\mathbf{A} + \mathbf{S}) = \mathbf{Z} \neq \varnothing. \tag{5.19}$$

Furthermore, (5.14) yields

$$\big[\, (\forall i \in I)(\forall n \in \mathbb{N}) \;\; a^*_{i,n} \in -z^*_i + A_i a_{i,n} \,\big] \quad \text{and} \quad \big[\, (\forall k \in K)(\forall n \in \mathbb{N}) \;\; b_{k,n} \in r_k + B_k^{-1} b^*_{k,n} \,\big]. \tag{5.20}$$

Next, define

$$(\forall i \in I)(\forall n \in \mathbb{N}) \quad \begin{cases} \overline{\ell}_i(n) = \max\big\{ j \in \mathbb{N} \mid j \leqslant n \;\text{ and }\; i \in I_j \big\}, \;\; \ell_i(n) = c_i\big(\overline{\ell}_i(n)\big) \\ u^*_{i,n} = \gamma^{-1}_{i,\ell_i(n)} x_{i,\ell_i(n)} - l^*_{i,\overline{\ell}_i(n)} + \gamma^{-1}_{i,\ell_i(n)} e_{i,\overline{\ell}_i(n)} \\ w^*_{i,n} = \sum_{k\in K} L^*_{k,i} v^*_{k,\vartheta_k(n)} - l^*_{i,\overline{\ell}_i(n)}. \end{cases} \tag{5.21}$$

Then, for every $i \in I$ and every $n \in \mathbb{N}$, it follows from (5.14) that

$$a_{i,n} = a_{i,\bar{\ell}_i(n)} = J_{\gamma_{i,\ell_i(n)}A_i}\big(\gamma_{i,\ell_i(n)}(u^*_{i,n} + z^*_i)\big) = \big(\gamma^{-1}_{i,\ell_i(n)}\mathrm{Id} - z^*_i + A_i\big)^{-1}u^*_{i,n} \tag{5.22}$$

and, therefore, that

$$a^*_{i,n} = a^*_{i,\bar{\ell}_i(n)} = u^*_{i,n} - \gamma^{-1}_{i,\ell_i(n)}a_{i,\bar{\ell}_i(n)} = u^*_{i,n} - \gamma^{-1}_{i,\ell_i(n)}a_{i,n}. \tag{5.23}$$

Likewise, for every $k \in K$ and every $n \in \mathbb{N}$, upon setting

$$\begin{cases} \overline{\vartheta}_k(n) = \max\big\{j \in \mathbb{N} \mid j \leqslant n \ \text{ and } \ k \in K_j\big\}, \ \ \vartheta_k(n) = d_k\big(\overline{\vartheta}_k(n)\big) \\ v_{k,n} = \mu_{k,\vartheta_k(n)}v^*_{k,\vartheta_k(n)} + l_{k,\overline{\vartheta}_k(n)} + f_{k,\overline{\vartheta}_k(n)} \\ w_{k,n} = l_{k,\overline{\vartheta}_k(n)} - \sum_{i\in I} L_{k,i}x_{i,\ell_i(n)}, \end{cases} \tag{5.24}$$

we get from (5.14) and [3, Proposition 23.17(iii)] that

$$b_{k,n} = b_{k,\overline{\vartheta}_k(n)} = J_{\mu_{k,\vartheta_k(n)}B_k(\cdot - r_k)}v_{k,n} \tag{5.25}$$

and, in turn, from (5.14) and [3, Proposition 23.20] that

$$\begin{aligned} b^*_{k,n} &= b^*_{k,\overline{\vartheta}_k(n)} && (5.26)\\ &= \mu^{-1}_{k,\vartheta_k(n)}\big(v_{k,n} - b_{k,\overline{\vartheta}_k(n)}\big) \\ &= \mu^{-1}_{k,\vartheta_k(n)}(v_{k,n} - b_{k,n}) && (5.27)\\ &= J_{\mu^{-1}_{k,\vartheta_k(n)}(r_k + B_k^{-1})}\big(\mu^{-1}_{k,\vartheta_k(n)}v_{k,n}\big) \\ &= \big(\mu_{k,\vartheta_k(n)}\mathrm{Id} + r_k + B_k^{-1}\big)^{-1}v_{k,n}. && (5.28) \end{aligned}$$

Let us set

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \mathbf{x}_n = \big((x_{i,n})_{i\in I}, (v^*_{k,n})_{k\in K}\big), \ \ \mathbf{u}_n = \big((x_{i,\ell_i(n)})_{i\in I}, (v^*_{k,\vartheta_k(n)})_{k\in K}\big) \\ \mathbf{e}^*_n = \big((w^*_{i,n})_{i\in I}, (w_{k,n})_{k\in K}\big), \ \ \mathbf{f}^*_n = \big((\gamma^{-1}_{i,\ell_i(n)}e_{i,\bar{\ell}_i(n)})_{i\in I}, (f_{k,\overline{\vartheta}_k(n)})_{k\in K}\big) \\ \mathbf{u}^*_n = \big((u^*_{i,n})_{i\in I}, (v_{k,n})_{k\in K}\big), \ \ \mathbf{y}_n = \big((a_{i,n})_{i\in I}, (b^*_{k,n})_{k\in K}\big) \\ \mathbf{a}^*_n = \big((a^*_{i,n})_{i\in I}, (b_{k,n})_{k\in K}\big), \ \ \mathbf{y}^*_n = \big((t^*_{i,n})_{i\in I}, (t_{k,n})_{k\in K}\big) \\ \mathbf{F}_n \colon \mathbf{H} \to \mathbf{H} \colon \big((x_i)_{i\in I}, (v^*_k)_{k\in K}\big) \mapsto \big((\gamma^{-1}_{i,\ell_i(n)}x_i)_{i\in I}, (\mu_{k,\vartheta_k(n)}v^*_k)_{k\in K}\big). \end{cases} \tag{5.29}$$

Then, the operators $(\mathbf{F}_n)_{n\in\mathbb{N}}$ are ε-strongly monotone and $(1/\varepsilon)$-Lipschitzian. For every $n \in \mathbb{N}$, by virtue of (5.21) and (5.24), we deduce from (5.18) that

$$\mathbf{S}\mathbf{u}_n - \mathbf{e}^*_n = \Big(\big(l^*_{i,\bar{\ell}_i(n)}\big)_{i\in I}, \big(-l_{k,\overline{\vartheta}_k(n)}\big)_{k\in K}\Big), \tag{5.30}$$

which yields

$$\mathbf{u}_n^* = \mathbf{F}_n\mathbf{u}_n - \mathbf{S}\mathbf{u}_n + \mathbf{e}_n^* + \mathbf{f}_n^*. \tag{5.31}$$

Furthermore, we infer from (5.22), (5.28), and (5.17) that

$$(\forall n \in \mathbb{N}) \quad \mathbf{y}_n = (\mathbf{F}_n + \mathbf{A})^{-1}\mathbf{u}_n^*. \tag{5.32}$$

At the same time, (5.23) and (5.27) imply that

$$(\forall n \in \mathbb{N}) \quad \mathbf{a}_n^* = \mathbf{u}_n^* - \mathbf{F}_n\mathbf{y}_n, \tag{5.33}$$

while (5.29), (5.14), and (5.18) guarantee that

$$(\forall n \in \mathbb{N}) \quad \mathbf{y}_n^* = \mathbf{a}_n^* + \mathbf{S}\mathbf{y}_n \quad \text{and} \quad \pi_n = \langle \mathbf{x}_n \mid \mathbf{y}_n^* \rangle - \langle \mathbf{y}_n \mid \mathbf{a}_n^* \rangle. \tag{5.34}$$

Altogether, it follows from (5.31)–(5.34) that (5.14) is an instantiation of (5.3). Hence, Theorem 5.3(i) yields $\sum_{n\in\mathbb{N}} \|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 < +\infty$. In turn, using (5.12), (5.13), (5.21), and (5.24), we deduce from [5, Lemma A.3] that, for every $i \in I$ and every $k \in K$, we have $\mathbf{x}_{\ell_i(n)} - \mathbf{x}_n \to \mathbf{0}$ and $\mathbf{x}_{\vartheta_k(n)} - \mathbf{x}_n \to \mathbf{0}$. This and (5.29) imply that

$$\mathbf{u}_n - \mathbf{x}_n \to \mathbf{0}. \tag{5.35}$$

Moreover, we deduce from (5.21) that

$$(\forall i \in I) \quad \|w_{i,n}^*\| \leqslant \sum_{k\in K} \|L_{k,i}^*\| \left\|v_{k,\vartheta_k(n)}^* - v_{k,\ell_i(n)}^*\right\| \leqslant \sum_{k\in K} \|L_{k,i}^*\| \, \|\mathbf{x}_{\vartheta_k(n)} - \mathbf{x}_{\ell_i(n)}\| \to 0 \tag{5.36}$$

and from (5.24) that

$$(\forall k \in K) \quad \|w_{k,n}\| \leqslant \sum_{i\in I} \|L_{k,i}\| \, \|x_{i,\vartheta_k(n)} - x_{i,\ell_i(n)}\| \leqslant \sum_{i\in I} \|L_{k,i}\| \, \|\mathbf{x}_{\vartheta_k(n)} - \mathbf{x}_{\ell_i(n)}\| \to 0. \tag{5.37}$$

Therefore, $\mathbf{e}_n^* \to \mathbf{0}$. By (5.15) and (5.16), $(\mathbf{f}_n^*)_{n\in\mathbb{N}}$ is bounded. In view of (5.29), (5.15), and (5.16), we get from (5.22) and (5.26) that

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \langle \mathbf{u}_n - \mathbf{y}_n \mid \mathbf{f}_n^* \rangle &= \sum_{i\in I} \left\langle x_{i,\ell_i(n)} - a_{i,n} \mid \gamma_{i,\ell_i(n)}^{-1} e_{i,\bar{\ell}_i(n)} \right\rangle + \sum_{k\in K} \left\langle v_{k,\vartheta_k(n)}^* - b_{k,n}^* \mid f_{k,\bar{\vartheta}_k(n)} \right\rangle \\ &\geqslant -\sigma \sum_{i\in I} \gamma_{i,\ell_i(n)}^{-1} \|x_{i,\ell_i(n)} - a_{i,n}\|^2 - \zeta \sum_{k\in K} \mu_{k,\vartheta_k(n)} \|v_{k,\vartheta_k(n)}^* - b_{k,n}^*\|^2 \\ &\geqslant -\max\{\sigma, \zeta\} \langle \mathbf{u}_n - \mathbf{y}_n \mid \mathbf{F}_n\mathbf{u}_n - \mathbf{F}_n\mathbf{y}_n \rangle, \end{aligned} \tag{5.38}$$

and from (5.30), (5.23), and (5.25) that

$$
\begin{aligned}
\langle \mathbf{a}_n^* + \mathbf{S}\mathbf{u}_n - \mathbf{e}_n^* \mid \mathbf{f}_n^* \rangle &= \sum_{i\in I} \big\langle a^*_{i,\overline{\ell}_i(n)} + l^*_{i,\overline{\ell}_i(n)} \mid \gamma^{-1}_{i,\ell_i(n)} e_{i,\overline{\ell}_i(n)} \big\rangle + \sum_{k\in K} \big\langle b_{k,\overline{\vartheta}_k(n)} - l_{k,\overline{\vartheta}_k(n)} \mid f_{k,\overline{\vartheta}_k(n)} \big\rangle \\
&\leqslant \sigma \sum_{i\in I} \big\| a^*_{i,\overline{\ell}_i(n)} + l^*_{i,\overline{\ell}_i(n)} \big\|^2 + \zeta \sum_{k\in K} \big\| b_{k,\overline{\vartheta}_k(n)} - l_{k,\overline{\vartheta}_k(n)} \big\|^2 \\
&\leqslant \max\{\sigma, \zeta\} \|\mathbf{a}_n^* + \mathbf{S}\mathbf{u}_n - \mathbf{e}_n^*\|^2.
\end{aligned}
\tag{5.39}
$$

Altogether, the conclusion follows from Theorem 5.3(ii). ☐

Remark 5.5 Using similar arguments, one can show that the asynchronous strongly convergent block-iterative method [6, Algorithm 14] can be viewed as an instance of [4, Theorem 4.8].

References

[1] 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.

[2] 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.

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

[4] M. N. Bùi and P. L. Combettes, Warped proximal iterations for monotone inclusions, *J. Math. Anal. Appl.*, vol. 491, art. 124315, 21 pp., 2020.

[5] M. N. Bùi and P. L. Combettes, Multivariate monotone inclusions in saddle form, *Math. Oper. Res.*, to appear.

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

[7] 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.

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

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

Chapter 6

A WARPED RESOLVENT ALGORITHM TO CONSTRUCT NASH EQUILIBRIA

6.1 Introduction and context

We address question **(Q5)** of Chapter 1 by providing an application of the warped resolvent framework of Theorem 3.16 to solving the Nash equilibrium model (6.3) in Problem 6.1.

This chapter presents the following article:

M. N. Bùi and P. L. Combettes, A warped resolvent algorithm to construct Nash equilibria, submitted.

6.2 Article: A warped resolvent algorithm to construct Nash equilibria

Abstract. We propose an asynchronous block-iterative decomposition algorithm to solve Nash equilibrium problems involving a mix of nonsmooth and smooth functions acting on linear mixtures of strategies. The methodology relies heavily on monotone operator theory and in particular on warped resolvents.

6.2.1 Introduction

We consider a noncooperative game with m players indexed by $I = \{1, \ldots, m\}$, in which the strategy x_i of player $i \in I$ lies in a real Hilbert space $\mathcal{H}_i$. A strategy profile is a point $\boldsymbol{x} = (x_i)_{i\in I}$ in the Hilbert direct sum $\boldsymbol{\mathcal{H}} = \bigoplus_{i\in I} \mathcal{H}_i$, and the associated profile of the players other than $i \in I$ is the vector $\boldsymbol{x}_{\smallsetminus i} = (x_j)_{j\in I\smallsetminus\{i\}}$ in $\boldsymbol{\mathcal{H}}_{\smallsetminus i} = \bigoplus_{j\in I\smallsetminus\{i\}} \mathcal{H}_j$. Given an index $i \in I$ and a vector $(x_i, \boldsymbol{y}) \in \mathcal{H}_i \times \boldsymbol{\mathcal{H}}$, we set $(x_i; \boldsymbol{y}_{\smallsetminus i}) = (y_1, \ldots, y_{i-1}, x_i, y_{i+1}, \ldots, y_m)$.

A fundamental equilibrium notion was introduced by Nash in [20, 21] to describe a state in which the loss of each player cannot be reduced by unilateral deviation. In our context, a formulation of the Nash equilibrium problem is

$$\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \text{ such that } (\forall i \in I)\ \overline{x}_i \in \underset{x_i \in \mathcal{H}_i}{\operatorname{Argmin}}\ \theta_i(x_i) + \boldsymbol{\ell}_i(x_i; \overline{\boldsymbol{x}}_{\smallsetminus i}), \tag{6.1}$$

where the global loss function of player $i \in I$ is the sum of an individual loss $\theta_i \colon \mathcal{H}_i \to \left]-\infty, +\infty\right]$ and a joint loss $\boldsymbol{\ell}_i \colon \boldsymbol{\mathcal{H}} \to \left]-\infty, +\infty\right]$ that models the interactions with the other players. Under convexity assumptions, numerical methods to solve (6.1) have been investigated since the early 1970s [4] and they have since involved increasingly sophisticated tools from nonlinear analysis; see [1, 5, 8, 11, 14–19, 25]. In the present paper, we consider the following highly modular Nash equilibrium problem wherein the functions $(\theta_i)_{i \in I}$ and $(\boldsymbol{\ell}_i)_{i \in I}$ of (6.1) are decomposed into elementary components that are easier to process numerically.

Problem 6.1 Let $(\mathcal{H}_i)_{i \in I}$, $(\mathcal{K}_i)_{i \in I}$, and $(\mathcal{G}_k)_{k \in K}$ be finite families of real Hilbert spaces, and set $\boldsymbol{\mathcal{H}} = \bigoplus_{i \in I} \mathcal{H}_i$, $\boldsymbol{\mathcal{K}} = \bigoplus_{i \in I} \mathcal{K}_i$, and $\boldsymbol{\mathcal{G}} = \bigoplus_{k \in K} \mathcal{G}_k$. Suppose that the following are satisfied:

[a] For every $i \in I$, $\varphi_i \colon \mathcal{H}_i \to \left]-\infty, +\infty\right]$ is proper, lower semicontinuous, and convex, $\alpha_i \in \left[0, +\infty\right[$, and $\psi_i \colon \mathcal{H}_i \to \mathbb{R}$ is convex and differentiable with an α_i-Lipschitzian gradient.

[b] For every $i \in I$, $\boldsymbol{f}_i \colon \boldsymbol{\mathcal{K}} \to \mathbb{R}$ is such that, for every $\boldsymbol{y} \in \boldsymbol{\mathcal{K}}$, $\boldsymbol{f}_i(\cdot\,; \boldsymbol{y}_{\smallsetminus i}) \colon \mathcal{K}_i \to \mathbb{R}$ is convex and Gâteaux differentiable, and we denote its gradient at $y_i \in \mathcal{K}_i$ by $\nabla_i \boldsymbol{f}_i(\boldsymbol{y})$. Further, the operator $\boldsymbol{Q} \colon \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}} \colon \boldsymbol{y} \mapsto (\nabla_i \boldsymbol{f}_i(\boldsymbol{y}))_{i \in I}$ is monotone and Lipschitzian. Finally, $(\chi_i)_{i \in I}$ are positive numbers such that

$$(\forall \boldsymbol{y} \in \boldsymbol{\mathcal{K}})(\forall \boldsymbol{y}' \in \boldsymbol{\mathcal{K}}) \quad \langle \boldsymbol{y} - \boldsymbol{y}' \mid \boldsymbol{Q}\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{y}' \rangle \leqslant \sum_{i \in I} \chi_i \|y_i - y_i'\|^2. \tag{6.2}$$

[c] For every $k \in K$, $g_k \colon \mathcal{G}_k \to \left]-\infty, +\infty\right]$ is proper, lower semicontinuous, and convex, $\beta_k \in \left[0, +\infty\right[$, and $h_k \colon \mathcal{G}_k \to \mathbb{R}$ is convex and differentiable with a β_k-Lipschitzian gradient.

[d] For every $i \in I$ and every $k \in K$, $M_i \colon \mathcal{H}_i \to \mathcal{K}_i$ and $L_{k,i} \colon \mathcal{H}_i \to \mathcal{G}_k$ are linear and bounded, and, for every $\boldsymbol{x} \in \boldsymbol{\mathcal{H}}$, we write $\boldsymbol{L}_{k,\smallsetminus i}\boldsymbol{x}_{\smallsetminus i} = \sum_{j \in I \smallsetminus \{i\}} L_{k,j} x_j$ and $\boldsymbol{M}\boldsymbol{x} = (M_j x_j)_{j \in I}$.

The goal is to

$$\begin{aligned} &\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \text{ such that } (\forall i \in I) \\ &\overline{x}_i \in \underset{x_i \in \mathcal{H}_i}{\operatorname{Argmin}}\ \varphi_i(x_i) + \psi_i(x_i) + \boldsymbol{f}_i\big(M_i x_i; (\boldsymbol{M}\overline{\boldsymbol{x}})_{\smallsetminus i}\big) + \sum_{k \in K} (g_k + h_k)(L_{k,i} x_i + \boldsymbol{L}_{k,\smallsetminus i}\overline{\boldsymbol{x}}_{\smallsetminus i}). \end{aligned} \tag{6.3}$$

In Problem 6.1, the individual loss of player $i \in I$ consists of a nonsmooth component φ_i and a smooth component ψ_i, while his joint loss is decomposed into a smooth function $\boldsymbol{f}_i$ and a sum of nonsmooth functions $(g_k)_{k \in K}$ and smooth functions $(h_k)_{k \in K}$ acting on linear mixtures of the strategies. We aim at solving (6.3) with a numerical procedure that can be implemented in a flexible fashion and that is able to cope with possibly very large scale problems. This leads us to adopt the following design principles:

- **Decomposition:** Each function and each linear operator in Problem 6.1 is activated separately.
- **Block-iterative implementation:** Only a subgroup of functions needs to be activated at any iteration. This makes it possible to best modulate and adapt the computational load of each iteration in large-scale problems.
- **Asynchronous implementation:** The computations are asynchronous in the sense that the result of calculations initiated at earlier iterations can be incorporated at the current one.

Our methodology is to first transform (6.3) into a system of monotone set-valued inclusions and then approach it via monotone operator splitting techniques. Since no splitting technique tailored to (6.3) and compliant with the above principles appears to be available, we adopt a fresh perspective hinging on the theory of warped resolvents [9]. In Section 6.2.2 we provide the necessary notation and background on monotone operator theory. Section 6.2.3 is devoted to the derivation of the proposed asynchronous block-iterative algorithm to solve Problem 6.1. Application examples are provided in Section 6.2.4.

6.2.2 Notation and background

General background on monotone operators and related notions can be found in [3].

Let $\mathcal{H}$ be a real Hilbert space. We denote by $2^{\mathcal{H}}$ the power set of $\mathcal{H}$ and by Id the identity operator on $\mathcal{H}$. The weak convergence and the strong convergence of a sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{H}$ to a point x in $\mathcal{H}$ are denoted by $x_n \rightharpoonup x$ and $x_n \to x$, respectively. Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$. The domain of A is $\operatorname{dom} A = \big\{x \in \mathcal{H} \mid Ax \neq \varnothing\big\}$, the range of A is $\operatorname{ran} A = \bigcup_{x\in\operatorname{dom} A} Ax$, the graph of A is $\operatorname{gra} A = \big\{(x, x^*) \in \mathcal{H} \times \mathcal{H} \mid x^* \in Ax\big\}$, the set of zeros of A is $\operatorname{zer} A = \big\{x \in \mathcal{H} \mid 0 \in Ax\big\}$, and the inverse of A is $A^{-1}\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x^* \mapsto \big\{x \in \mathcal{H} \mid x^* \in Ax\big\}$. Now suppose that A is monotone, that is,

$$\big(\forall (x, x^*) \in \operatorname{gra} A\big)\big(\forall (y, y^*) \in \operatorname{gra} A\big) \quad \langle x - y \mid x^* - y^*\rangle \geqslant 0. \tag{6.4}$$

Then A is maximally monotone if, for every monotone operator $\widetilde{A}\colon \mathcal{H} \to 2^{\mathcal{H}}$, $\operatorname{gra} A \subset \operatorname{gra} \widetilde{A} \Rightarrow A = \widetilde{A}$; A is strongly monotone with constant $\varsigma \in \,]0, +\infty[$ if $A - \varsigma\mathrm{Id}$ is monotone; and A is 3^* monotone if

$$(\forall x \in \operatorname{dom} A)(\forall x^* \in \operatorname{ran} A) \quad \sup_{(y,y^*)\in\operatorname{gra} A} \langle x - y \mid y^* - x^*\rangle < +\infty. \tag{6.5}$$

$\Gamma_0(\mathcal{H})$ is the set of lower semicontinuous convex functions $\varphi\colon \mathcal{H} \to \,]-\infty, +\infty]$ which are proper in the sense that $\operatorname{dom}\varphi = \big\{x \in \mathcal{H} \mid \varphi(x) < +\infty\big\} \neq \varnothing$. Let $\varphi \in \Gamma_0(\mathcal{H})$. Then φ is supercoercive if $\lim_{\|x\|\to+\infty} \varphi(x)/\|x\| = +\infty$ and uniformly convex if there exists an increasing function

$\phi\colon\ [0,+\infty[\ \to [0,+\infty]$ that vanishes only at 0 such that

$$
\begin{aligned}
&(\forall x \in \operatorname{dom}\varphi)(\forall y \in \operatorname{dom}\varphi)(\forall \alpha \in\ ]0,1[)\\
&\qquad\qquad \varphi\big(\alpha x + (1-\alpha)y\big) + \alpha(1-\alpha)\phi\big(\|x-y\|\big) \leqslant \alpha\varphi(x) + (1-\alpha)\varphi(y).
\end{aligned} \tag{6.6}
$$

For every $x \in \mathcal{H}$, $\operatorname{prox}_\varphi x$ denotes the unique minimizer of $\varphi + (1/2)\|\cdot - x\|^2$. The subdifferential of φ is the maximally monotone operator $\partial\varphi\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \big\{x^* \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ \langle y - x \mid x^*\rangle + \varphi(x) \leqslant \varphi(y)\big\}$. Let C be a convex subset of $\mathcal{H}$. The indicator function of C is

$$
\iota_C\colon \mathcal{H} \to [0,+\infty] : x \mapsto \begin{cases} 0, & \text{if } x \in C;\\ +\infty, & \text{otherwise,}\end{cases} \tag{6.7}
$$

and the strong relative interior of C is

$$
\operatorname{sri} C = \left\{ x \in C \;\middle|\; \bigcup_{\lambda \in]0,+\infty[} \lambda(C - x) \text{ is a closed vector subspace of } \mathcal{H} \right\}. \tag{6.8}
$$

The following notion of a warped resolvent will be instrumental to our approach.

Definition 6.2 ([9]) Suppose that $\boldsymbol{\mathcal{X}}$ is a real Hilbert space. Let $\mathbf{D}$ be a nonempty subset of $\boldsymbol{\mathcal{X}}$, let $\mathbf{K}\colon \mathbf{D} \to \boldsymbol{\mathcal{X}}$, and let $\mathbf{A}\colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}}$ be such that $\operatorname{ran}\mathbf{K} \subset \operatorname{ran}(\mathbf{K}+\mathbf{A})$ and $\mathbf{K}+\mathbf{A}$ is injective. The *warped resolvent* of $\mathbf{A}$ with kernel $\mathbf{K}$ is $J_{\mathbf{A}}^{\mathbf{K}} = (\mathbf{K}+\mathbf{A})^{-1} \circ \mathbf{K}$.

We now provide a warped resolvent algorithm to find a zero of a maximally monotone operator $\mathbf{A}\colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}}$, where $\boldsymbol{\mathcal{X}}$ is a real Hilbert space. This algorithm has a simple geometric interpretation: at iteration n, we use the evaluation of the warped resolvent $J_{\mathbf{A}}^{\mathbf{K}_n}$ at a perturbation $\widetilde{\mathbf{x}}_n$ of the current iterate $\mathbf{x}_n$ to construct a point $(\mathbf{y}_n, \mathbf{y}_n^*) \in \operatorname{gra}\mathbf{A}$. By monotonicity of $\mathbf{A}$,

$$
\operatorname{zer}\mathbf{A} \subset \mathbf{H}_n = \big\{\mathbf{z} \in \boldsymbol{\mathcal{X}} \mid \langle \mathbf{z} - \mathbf{y}_n \mid \mathbf{y}_n^*\rangle \leqslant 0\big\}, \tag{6.9}
$$

and the update $\mathbf{x}_{n+1}$ is a relaxed projection of $\mathbf{x}_n$ onto the half-space $\mathbf{H}_n$.

Proposition 6.3 *Let $\boldsymbol{\mathcal{X}}$ be a real Hilbert space and let $\mathbf{A}\colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}}$ be a maximally monotone operator such that $\operatorname{zer}\mathbf{A} \neq \varnothing$. Let $\mathbf{x}_0 \in \boldsymbol{\mathcal{X}}$, let $\varepsilon \in\ ]0,1[$, let $\varsigma \in\ ]0,+\infty[$, and let $\varpi \in\ ]\varsigma,+\infty[$. Further, for every $n \in \mathbb{N}$, let $\lambda_n \in [\varepsilon, 2-\varepsilon]$, let $\widetilde{\mathbf{x}}_n \in \boldsymbol{\mathcal{X}}$, and let $\mathbf{K}_n\colon \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{X}}$ be a ς-strongly* *monotone and ϖ-Lipschitzian operator. Iterate*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\mathsf{y}_n = J_{\mathsf{A}}^{\mathsf{K}_n} \widetilde{\mathsf{x}}_n; \\
\mathsf{y}_n^* = \mathsf{K}_n \widetilde{\mathsf{x}}_n - \mathsf{K}_n \mathsf{y}_n; \\
\textit{if } \langle \mathsf{y}_n - \mathsf{x}_n \mid \mathsf{y}_n^* \rangle < 0 \\
\left\lfloor
\begin{array}{l}
\mathsf{x}_{n+1} = \mathsf{x}_n + \dfrac{\lambda_n \langle \mathsf{y}_n - \mathsf{x}_n \mid \mathsf{y}_n^* \rangle}{\|\mathsf{y}_n^*\|^2} \mathsf{y}_n^*;
\end{array}
\right. \\
\textit{else} \\
\left\lfloor
\begin{array}{l}
\mathsf{x}_{n+1} = \mathsf{x}_n.
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{6.10}
$$

Then the following hold:

(i) $\sum_{n \in \mathbb{N}} \|\mathsf{x}_{n+1} - \mathsf{x}_n\|^2 < +\infty$.

(ii) *Suppose that* $\widetilde{\mathsf{x}}_n - \mathsf{x}_n \to \mathbf{0}$*. Then* $\mathsf{x}_n - \mathsf{y}_n \to \mathbf{0}$ *and* $(\mathsf{x}_n)_{n \in \mathbb{N}}$ *converges weakly to a point in* $\operatorname{zer} \mathsf{A}$.

Proof. It follows from [9, Proposition 3.9(i)[d]&(ii)[b]] that the warped resolvents $(J_{\mathsf{A}}^{\mathsf{K}_n})_{n \in \mathbb{N}}$ in (6.10) are well defined. In turn, we derive (i) and the weak convergence claim from [9, Theorem 4.2 and Remark 4.3]. It thus remains to prove that $\mathsf{x}_n - \mathsf{y}_n \to \mathbf{0}$. It is shown in the proof of [9, Theorem 4.2(ii)] that $\mathsf{K}_n \widetilde{\mathsf{x}}_n - \mathsf{K}_n \mathsf{y}_n \to \mathbf{0}$. At the same time, for every $n \in \mathbb{N}$, every $\mathsf{x} \in \boldsymbol{\mathcal{X}}$, and every $\mathsf{y} \in \boldsymbol{\mathcal{X}}$, we deduce from the Cauchy–Schwarz inequality that $\varsigma\|\mathsf{x} - \mathsf{y}\|^2 \leqslant \langle \mathsf{x} - \mathsf{y} \mid \mathsf{K}_n \mathsf{x} - \mathsf{K}_n \mathsf{y} \rangle \leqslant \|\mathsf{x} - \mathsf{y}\| \, \|\mathsf{K}_n \mathsf{x} - \mathsf{K}_n \mathsf{y}\|$, from which it follows that

$$
\varsigma\|\mathsf{x} - \mathsf{y}\| \leqslant \|\mathsf{K}_n \mathsf{x} - \mathsf{K}_n \mathsf{y}\|. \tag{6.11}
$$

Therefore, $\|\mathsf{x}_n - \mathsf{y}_n\| \leqslant \|\mathsf{x}_n - \widetilde{\mathsf{x}}_n\| + \|\widetilde{\mathsf{x}}_n - \mathsf{y}_n\| \leqslant \|\mathsf{x}_n - \widetilde{\mathsf{x}}_n\| + (1/\varsigma)\|\mathsf{K}_n \widetilde{\mathsf{x}}_n - \mathsf{K}_n \mathsf{y}_n\| \to 0$, as desired. ☐

6.2.3 Algorithm

As mentioned in Section 6.2.1, there exists no method tailored to the format of Problem 6.1 that can solve it in an asynchronous block-iterative fashion. Our methodology to design such an algorithm can be broken down in the following steps:

1. We rephrase (6.3) as a monotone inclusion problem in $\boldsymbol{\mathcal{H}}$, namely,

$$
\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \text{ such that } \mathbf{0} \in \boldsymbol{A}\overline{\boldsymbol{x}} + \boldsymbol{M}^*\big(\boldsymbol{Q}(\boldsymbol{M}\overline{\boldsymbol{x}})\big) + \boldsymbol{L}^*\big(\boldsymbol{B}(\boldsymbol{L}\overline{\boldsymbol{x}})\big), \tag{6.12}
$$

where $\boldsymbol{Q}$ and $\boldsymbol{M}$ are defined in Problem 6.1[b] and Problem 6.1[d], respectively, and

$$\begin{cases} \boldsymbol{A}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}\colon \boldsymbol{x} \mapsto \bigtimes_{i\in I} \big(\partial\varphi_i(x_i) + \nabla\psi_i(x_i)\big) \\ \boldsymbol{B}\colon \boldsymbol{\mathcal{G}} \to 2^{\boldsymbol{\mathcal{G}}}\colon \boldsymbol{z} \mapsto \bigtimes_{k\in K} \big(\partial g_k(z_k) + \nabla h_k(z_k)\big) \\ \boldsymbol{L}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{G}}\colon \boldsymbol{x} \mapsto \big(\sum_{i\in I} L_{k,i}x_i\big)_{k\in K}. \end{cases} \tag{6.13}$$

2. The inclusion in (6.12) involves more than two operators, namely $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{Q}$, $\boldsymbol{L}$, and $\boldsymbol{M}$. Hence, in the spirit of the decomposition methodologies of [9,12,13], a space bigger than $\boldsymbol{\mathcal{H}}$ is required to devise a splitting method to solve it. We set $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{H}} \oplus \boldsymbol{\mathcal{K}} \oplus \boldsymbol{\mathcal{G}} \oplus \boldsymbol{\mathcal{K}} \oplus \boldsymbol{\mathcal{G}}$ and consider the inclusion problem

$$\text{find } \overline{\mathbf{x}} \in \boldsymbol{\mathcal{X}} \text{ such that } \mathbf{0} \in \mathbf{A}\overline{\mathbf{x}}, \tag{6.14}$$

where

$$\mathbf{A}\colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}}\colon (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}^*, \boldsymbol{v}^*) \mapsto$$
$$(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{M}^*\boldsymbol{u}^* + \boldsymbol{L}^*\boldsymbol{v}^*) \times \{\boldsymbol{Q}\boldsymbol{y} - \boldsymbol{u}^*\} \times (\boldsymbol{B}\boldsymbol{z} - \boldsymbol{v}^*) \times \{\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\} \times \{\boldsymbol{z} - \boldsymbol{L}\boldsymbol{x}\}. \tag{6.15}$$

3. We show that, if $\mathbf{x} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}^*, \boldsymbol{v}^*)$ solves (6.14), then $\boldsymbol{x}$ solves (6.12) and, therefore, (6.3).

4. To solve (6.14), we implement the warped resolvent algorithm of Proposition 6.3 with a specific choice of the auxiliary points $(\widetilde{\mathbf{x}}_n)_{n\in\mathbb{N}}$ and the kernels $(\mathbf{K}_n)_{n\in\mathbb{N}}$ that will lead to an asynchronous block-iterative splitting algorithm.

The methodology just described is put in motion in our main theorem, which we now state and prove.

Theorem 6.4 *Consider the setting of Problem 6.1. Let $\eta \in \left]0, +\infty\right[$ and $\varepsilon \in \left]0, 1\right[$ be such that $1/\varepsilon > \max\{\alpha_i + \eta, \beta_k + \eta, \chi_i + \eta\}_{i\in I, k\in K}$, let $(\lambda_n)_{n\in\mathbb{N}}$ be in $[\varepsilon, 2-\varepsilon]$, and let $D \in \mathbb{N}$. Suppose that the following are satisfied:*

[a] *For every $i \in I$ and every $n \in \mathbb{N}$, $\tau_i(n) \in \mathbb{N}$ satisfies $n - D \leqslant \tau_i(n) \leqslant n$, $\gamma_{i,n} \in [\varepsilon, 1/(\alpha_i + \eta)]$, $\mu_{i,n} \in [\varepsilon, 1/(\chi_i + \eta)]$, $\sigma_{i,n} \in [\varepsilon, 1/\varepsilon]$, $x_{i,0} \in \mathcal{H}_i$, $y_{i,0} \in \mathcal{K}_i$, and $u^*_{i,0} \in \mathcal{K}_i$.*

[b] *For every $k \in K$ and every $n \in \mathbb{N}$, $\delta_k(n) \in \mathbb{N}$ satisfies $n - D \leqslant \delta_k(n) \leqslant n$, $\nu_{k,n} \in [\varepsilon, 1/(\beta_k + \eta)]$, $\varrho_{k,n} \in [\varepsilon, 1/\varepsilon]$, $z_{k,0} \in \mathcal{G}_k$, and $v^*_{k,0} \in \mathcal{G}_k$.*

[c] *$(I_n)_{n\in\mathbb{N}}$ are nonempty subsets of I and $(K_n)_{n\in\mathbb{N}}$ are nonempty subsets of K such that, for some $P \in \mathbb{N}$,*

$$I_0 = I, \quad K_0 = K, \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \bigcup_{j=n}^{n+P} I_j = I \ \text{ and } \bigcup_{j=n}^{n+P} K_j = K. \tag{6.16}$$

Iterate

for $n = 0, 1, \ldots$

for every $i \in I_n$

$q_{i,n} = y_{i,\tau_i(n)} + \mu_{i,\tau_i(n)}\big(u^*_{i,\tau_i(n)} - \nabla_i \boldsymbol{f}_i\big(\boldsymbol{y}_{\tau_i(n)}\big)\big);$

$c^*_{i,n} = u^*_{i,\tau_i(n)} + \sigma_{i,\tau_i(n)}\big(M_i x_{i,\tau_i(n)} - y_{i,\tau_i(n)}\big);$

$x^*_{i,n} = x_{i,\tau_i(n)} - \gamma_{i,\tau_i(n)}\big(\nabla\psi_i\big(x_{i,\tau_i(n)}\big) + M_i^* u^*_{i,\tau_i(n)} + \sum_{k\in K} L^*_{k,i} v^*_{k,\tau_i(n)}\big);$

$a_{i,n} = \operatorname{prox}_{\gamma_{i,\tau_i(n)}\varphi_i} x^*_{i,n};$

$s^*_{i,n} = \gamma^{-1}_{i,\tau_i(n)}(x^*_{i,n} - a_{i,n}) + \nabla\psi_i(a_{i,n}) + M_i^* c^*_{i,n};$

$c_{i,n} = q_{i,n} - M_i a_{i,n};$

for every $i \in I \smallsetminus I_n$

$q_{i,n} = q_{i,n-1};\ c^*_{i,n} = c^*_{i,n-1};\ a_{i,n} = a_{i,n-1};\ s^*_{i,n} = s^*_{i,n-1};\ c_{i,n} = c_{i,n-1};$

for every $k \in K_n$

$d^*_{k,n} = z_{k,\delta_k(n)} + \nu_{k,\delta_k(n)}\big(v^*_{k,\delta_k(n)} - \nabla h_k\big(z_{k,\delta_k(n)}\big)\big);$

$b_{k,n} = \operatorname{prox}_{\nu_{k,\delta_k(n)} g_k} d^*_{k,n};$

$e^*_{k,n} = v^*_{k,\delta_k(n)} + \varrho_{k,\delta_k(n)}\big(\sum_{i\in I} L_{k,i} x_{i,\delta_k(n)} - z_{k,\delta_k(n)}\big);$

$b^*_{k,n} = \nu^{-1}_{k,\delta_k(n)}(d^*_{k,n} - b_{k,n}) + \nabla h_k(b_{k,n}) - e^*_{k,n};$

$e_{k,n} = b_{k,n} - \sum_{i\in I} L_{k,i} a_{i,n};$

for every $k \in K \smallsetminus K_n$

$b_{k,n} = b_{k,n-1};\ e^*_{k,n} = e^*_{k,n-1};\ b^*_{k,n} = b^*_{k,n-1};\ e_{k,n} = b_{k,n} - \sum_{i\in I} L_{k,i} a_{i,n};$

for every $i \in I$

$a^*_{i,n} = s^*_{i,n} + \sum_{k\in K} L^*_{k,i} e^*_{k,n};$

$q^*_{i,n} = \nabla_i \boldsymbol{f}_i(\boldsymbol{q}_n) - c^*_{i,n};$

$$\pi_n = \sum_{i\in I} \big(\langle a_{i,n} - x_{i,n} \mid a^*_{i,n}\rangle + \langle q_{i,n} - y_{i,n} \mid q^*_{i,n}\rangle + \langle c_{i,n} \mid c^*_{i,n} - u^*_{i,n}\rangle\big) + \sum_{k\in K} \big(\langle b_{k,n} - z_{k,n} \mid b^*_{k,n}\rangle + \langle e_{k,n} \mid e^*_{k,n} - v^*_{k,n}\rangle\big);$$

if $\pi_n < 0$

$\theta_n = \lambda_n \pi_n / \big(\sum_{i\in I} \big(\|a^*_{i,n}\|^2 + \|q^*_{i,n}\|^2 + \|c_{i,n}\|^2\big) + \sum_{k\in K} \big(\|b^*_{k,n}\|^2 + \|e_{k,n}\|^2\big)\big);$

for every $i \in I$

$x_{i,n+1} = x_{i,n} + \theta_n a^*_{i,n};\ y_{i,n+1} = y_{i,n} + \theta_n q^*_{i,n};\ u^*_{i,n+1} = u^*_{i,n} + \theta_n c_{i,n};$

for every $k \in K$

$z_{k,n+1} = z_{k,n} + \theta_n b^*_{k,n};\ v^*_{k,n+1} = v^*_{k,n} + \theta_n e_{k,n};$

else

for every $i \in I$

$x_{i,n+1} = x_{i,n};\ y_{i,n+1} = y_{i,n};\ u^*_{i,n+1} = u^*_{i,n};$

for every $k \in K$

$z_{k,n+1} = z_{k,n};\ v^*_{k,n+1} = v^*_{k,n}.$

(6.17)

Furthermore, suppose that there exist $\widehat{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}}$*,* $\widehat{\boldsymbol{u}}^* \in \boldsymbol{\mathcal{K}}$*, and* $\widehat{\boldsymbol{v}}^* \in \boldsymbol{\mathcal{G}}$ *such that*

$$\begin{cases} (\forall i \in I) \;\; \widehat{u}_i^* = \nabla_i \boldsymbol{f}_i(\boldsymbol{M}\widehat{\boldsymbol{x}}) \\ (\forall k \in K) \;\; \widehat{v}_k^* \in (\partial g_k + \nabla h_k)\big(\sum_{j \in I} L_{k,j}\widehat{x}_j\big) \\ (\forall i \in I) \;\; -M_i^*\widehat{u}_i^* - \sum_{k \in K} L_{k,i}^*\widehat{v}_k^* \in \partial\varphi_i(\widehat{x}_i) + \nabla\psi_i(\widehat{x}_i). \end{cases} \tag{6.18}$$

Then $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ *converges weakly to a solution to Problem* 6.1.

Proof. Set $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{H}}\oplus\boldsymbol{\mathcal{K}}\oplus\boldsymbol{\mathcal{G}}\oplus\boldsymbol{\mathcal{K}}\oplus\boldsymbol{\mathcal{G}}$ and consider the operators defined in (6.13) and (6.15). Let us first examine some properties of the operator $\mathbf{A}$ in (6.15). For every $i \in I$, it results from Problem 6.1[a] and [3, Theorem 20.25 and Proposition 17.31(i)] that $\partial\varphi_i$ and $\nabla\psi_i$ are maximally monotone and, therefore, from [3, Corollary 25.5(i)] that $\partial\varphi_i + \nabla\psi_i$ is maximally monotone. Thus, in view of (6.13) and [3, Proposition 20.23], $\boldsymbol{A}$ is maximally monotone. Likewise, $\boldsymbol{B}$ is maximally monotone. Hence, since $\boldsymbol{Q}$ is maximally monotone by virtue of Problem 6.1[b] and [3, Corollary 20.28], [3, Proposition 20.23] implies that the operator

$$\mathbf{R}\colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}}\colon (\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{u}^*,\boldsymbol{v}^*) \mapsto \boldsymbol{A}\boldsymbol{x} \times \{\boldsymbol{Q}\boldsymbol{y}\} \times \boldsymbol{B}\boldsymbol{z} \times \{\boldsymbol{0}\} \times \{\boldsymbol{0}\} \tag{6.19}$$

is maximally monotone. On the other hand, since the operator

$$\mathbf{S}\colon \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{X}}\colon (\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{u}^*,\boldsymbol{v}^*) \mapsto (\boldsymbol{M}^*\boldsymbol{u}^*+\boldsymbol{L}^*\boldsymbol{v}^*)\times\{-\boldsymbol{u}^*\}\times\{-\boldsymbol{v}^*\}\times\{\boldsymbol{y}-\boldsymbol{M}\boldsymbol{x}\}\times\{\boldsymbol{z}-\boldsymbol{L}\boldsymbol{x}\} \tag{6.20}$$

is linear and bounded with

$$\mathbf{S}^* = -\mathbf{S}, \tag{6.21}$$

we deduce from [3, Example 20.35] that $\mathbf{S}$ is maximally monotone. In turn, it follows from (6.15), (6.19), and [3, Corollary 25.5(i)] that

$$\mathbf{A} = \mathbf{R} + \mathbf{S} \text{ is maximally monotone.} \tag{6.22}$$

Upon setting $\widehat{\boldsymbol{y}} = \boldsymbol{M}\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{z}} = \boldsymbol{L}\widehat{\boldsymbol{x}}$, we derive from (6.18) and (6.13) that $\widehat{\boldsymbol{u}}^* = \boldsymbol{Q}\widehat{\boldsymbol{y}}$ and $\widehat{\boldsymbol{v}}^* \in \boldsymbol{B}\widehat{\boldsymbol{z}}$. Further, since

$$\boldsymbol{M}^*\colon \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{H}}\colon \boldsymbol{u}^* \mapsto (M_i^*u_i^*)_{i\in I} \quad \text{and} \quad \boldsymbol{L}^*\colon \boldsymbol{\mathcal{G}} \to \boldsymbol{\mathcal{H}}\colon \boldsymbol{v}^* \mapsto \left(\sum_{k\in K} L_{k,i}^*v_k^*\right)_{i\in I}, \tag{6.23}$$

it results from (6.18) and (6.13) that $-\boldsymbol{M}^*\widehat{\boldsymbol{u}}^*-\boldsymbol{L}^*\widehat{\boldsymbol{v}}^* \in \boldsymbol{A}\widehat{\boldsymbol{x}}$. Therefore, we infer from (6.15) that $(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{y}},\widehat{\boldsymbol{z}},\widehat{\boldsymbol{u}}^*,\widehat{\boldsymbol{v}}^*) \in \operatorname{zer}\mathbf{A}$ and, hence, that

$$\operatorname{zer}\mathbf{A} \neq \varnothing. \tag{6.24}$$

Define

$$(\forall i \in I)(\forall n \in \mathbb{N}) \quad \overline{\ell}_i(n) = \max\big\{j \in \mathbb{N} \mid j \leqslant n \;\text{ and }\; i \in I_j\big\} \quad \text{and} \quad \ell_i(n) = \tau_i\big(\overline{\ell}_i(n)\big) \tag{6.25}$$

and

$$(\forall k \in K)(\forall n \in \mathbb{N}) \quad \overline{\vartheta}_k(n) = \max\big\{j \in \mathbb{N} \mid j \leqslant n \;\text{ and }\; k \in K_j\big\} \quad \text{and} \quad \vartheta_k(n) = \delta_k\big(\overline{\vartheta}_k(n)\big). \tag{6.26}$$

In addition, let $\kappa \in \left]0, +\infty\right[$ be a Lipschitz constant of $\boldsymbol{Q}$ in Problem 6.1[b], set

$$\begin{cases} \alpha = \sqrt{2\big(\varepsilon^{-2} + \max_{i \in I} \alpha_i^2\big)}, \;\; \beta = \sqrt{2\big(\varepsilon^{-2} + \max_{k \in K} \beta_k^2\big)}, \;\; \chi = \sqrt{2(\varepsilon^{-2} + \kappa^2)} \\ \varsigma = \min\{\varepsilon, \eta\}, \;\; \varpi = \|\mathbf{S}\| + \max\{\alpha, \beta, \chi, 1/\varepsilon\}, \end{cases} \tag{6.27}$$

and define

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \boldsymbol{E}_n \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{x} \mapsto \big(\gamma_{i,\ell_i(n)}^{-1} x_i - \nabla\psi_i(x_i)\big)_{i \in I} \\ \boldsymbol{F}_n \colon \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}} \colon \boldsymbol{y} \mapsto \big(\mu_{i,\ell_i(n)}^{-1} y_i - \nabla_i \boldsymbol{f}_i(\boldsymbol{y})\big)_{i \in I} \\ \boldsymbol{G}_n \colon \boldsymbol{\mathcal{G}} \to \boldsymbol{\mathcal{G}} \colon \boldsymbol{z} \mapsto \big(\nu_{k,\vartheta_k(n)}^{-1} z_k - \nabla h_k(z_k)\big)_{k \in K} \\ \mathbf{T}_n \colon \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{X}} \colon (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{u}^*, \boldsymbol{v}^*) \mapsto \Big(\boldsymbol{E}_n\boldsymbol{x}, \boldsymbol{F}_n\boldsymbol{y}, \boldsymbol{G}_n\boldsymbol{z}, \big(\sigma_{i,\ell_i(n)}^{-1} u_i^*\big)_{i \in I}, \big(\varrho_{k,\vartheta_k(n)}^{-1} v_k^*\big)_{k \in K}\Big) \\ \mathbf{K}_n = \mathbf{T}_n - \mathbf{S}. \end{cases} \tag{6.28}$$

Fix temporarily $n \in \mathbb{N}$. Then, using [a], the Cauchy–Schwarz inequality, and Problem 6.1[a], we obtain

$$\begin{aligned} (\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}})(\forall \boldsymbol{x}' \in \boldsymbol{\mathcal{H}}) \quad & \langle \boldsymbol{x} - \boldsymbol{x}' \mid \boldsymbol{E}_n\boldsymbol{x} - \boldsymbol{E}_n\boldsymbol{x}' \rangle \\ &= \sum_{i \in I} \big(\gamma_{i,\ell_i(n)}^{-1} \|x_i - x_i'\|^2 - \langle x_i - x_i' \mid \nabla\psi_i(x_i) - \nabla\psi_i(x_i') \rangle\big) \\ &\geqslant \sum_{i \in I} \big((\alpha_i + \eta)\|x_i - x_i'\|^2 - \|x_i - x_i'\| \, \|\nabla\psi_i(x_i) - \nabla\psi_i(x_i')\|\big) \\ &\geqslant \sum_{i \in I} \big((\alpha_i + \eta)\|x_i - x_i'\|^2 - \alpha_i\|x_i - x_i'\|^2\big) \\ &= \eta\|\boldsymbol{x} - \boldsymbol{x}'\|^2 \end{aligned} \tag{6.29}$$

and

$$\begin{aligned} (\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}})(\forall \boldsymbol{x}' \in \boldsymbol{\mathcal{H}}) \quad & \|\boldsymbol{E}_n\boldsymbol{x} - \boldsymbol{E}_n\boldsymbol{x}'\|^2 \\ &= \sum_{i \in I} \big\|\gamma_{i,\ell_i(n)}^{-1}(x_i - x_i') - \big(\nabla\psi_i(x_i) - \nabla\psi_i(x_i')\big)\big\|^2 \\ &\leqslant 2\sum_{i \in I} \big(\gamma_{i,\ell_i(n)}^{-2}\|x_i - x_i'\|^2 + \|\nabla\psi_i(x_i) - \nabla\psi_i(x_i')\|^2\big) \end{aligned}$$

$$
\begin{aligned}
&\leqslant 2\sum_{i\in I}\big(\varepsilon^{-2}\|x_i-x_i'\|^2+\alpha_i^2\|x_i-x_i'\|^2\big)\\
&\leqslant \alpha^2\|\boldsymbol{x}-\boldsymbol{x}'\|^2.
\end{aligned}
\tag{6.30}
$$

Thus,

$$
\boldsymbol{E}_n \text{ is } \eta\text{-strongly monotone and } \alpha\text{-Lipschitzian.} \tag{6.31}
$$

Similarly,

$$
\begin{cases}
\boldsymbol{F}_n \text{ is } \eta\text{-strongly monotone and } \chi\text{-Lipschitzian}\\
\boldsymbol{G}_n \text{ is } \eta\text{-strongly monotone and } \beta\text{-Lipschitzian.}
\end{cases}
\tag{6.32}
$$

In turn, invoking (6.28), [a], [b], and (6.27), we deduce that $\mathbf{T}_n$ is strongly monotone with constant ς and Lipschitzian with constant $\max\{\alpha,\beta,\chi,1/\varepsilon\}$. It therefore follows from (6.28) and (6.27) that

$$
\mathbf{K}_n \text{ is } \varsigma\text{-strongly monotone and } \varpi\text{-Lipschitzian.} \tag{6.33}
$$

Let us define

$$
\begin{cases}
(\forall i\in I)\ \ E_{i,n}\colon \mathcal{H}_i\to\mathcal{H}_i\colon x_i\mapsto \gamma_{i,\ell_i(n)}^{-1}x_i-\nabla\psi_i(x_i)\\
(\forall k\in K)\ \ G_{k,n}\colon \mathcal{G}_k\to\mathcal{G}_k\colon z_k\mapsto \nu_{k,\vartheta_k(n)}^{-1}z_k-\nabla h_k(z_k)
\end{cases}
\tag{6.34}
$$

and let us introduce the variables

$$
\begin{cases}
\mathbf{x}_n=(\boldsymbol{x}_n,\boldsymbol{y}_n,\boldsymbol{z}_n,\boldsymbol{u}_n^*,\boldsymbol{v}_n^*),\ \ \mathbf{y}_n=(\boldsymbol{a}_n,\boldsymbol{q}_n,\boldsymbol{b}_n,\boldsymbol{c}_n^*,\boldsymbol{e}_n^*),\ \ \mathbf{y}_n^*=(\boldsymbol{a}_n^*,\boldsymbol{q}_n^*,\boldsymbol{b}_n^*,\boldsymbol{c}_n,\boldsymbol{e}_n)\\
(\forall i\in I)\ \ \begin{cases}
\widetilde{x}_{i,n}^*=E_{i,n}x_{i,\ell_i(n)}-E_{i,n}x_n+M_i^*\big(u_{i,n}^*-u_{i,\ell_i(n)}^*\big)+\sum_{k\in K}L_{k,i}^*\big(v_{k,n}^*-v_{k,\ell_i(n)}^*\big)\\
\widetilde{q}_{i,n}^*=\mu_{i,\ell_i(n)}^{-1}\big(y_{i,\ell_i(n)}-y_{i,n}\big)+\nabla_i\boldsymbol{f}_i(\boldsymbol{y}_n)-\nabla_i\boldsymbol{f}_i\big(\boldsymbol{y}_{\ell_i(n)}\big)+u_{i,\ell_i(n)}^*-u_{i,n}^*\\
\widetilde{c}_{i,n}^*=\sigma_{i,\ell_i(n)}^{-1}\big(u_{i,\ell_i(n)}^*-u_{i,n}^*\big)+M_i\big(x_{i,\ell_i(n)}-x_{i,n}\big)-y_{i,\ell_i(n)}+y_{i,n}
\end{cases}\\
(\forall k\in K)\ \ \begin{cases}
\widetilde{d}_{k,n}^*=G_{k,n}z_{k,\vartheta_k(n)}-G_{k,n}z_{k,n}+v_{k,\vartheta_k(n)}^*-v_{k,n}^*\\
\widetilde{e}_{k,n}^*=\varrho_{k,\vartheta_k(n)}^{-1}\big(v_{k,\vartheta_k(n)}^*-v_{k,n}^*\big)-z_{k,\vartheta_k(n)}+z_{k,n}+\sum_{i\in I}L_{k,i}\big(x_{i,\vartheta_k(n)}-x_{i,n}\big)
\end{cases}\\
\mathbf{e}_n^*=(\widetilde{\boldsymbol{x}}_n^*,\widetilde{\boldsymbol{q}}_n^*,\widetilde{\boldsymbol{d}}_n^*,\widetilde{\boldsymbol{c}}_n^*,\widetilde{\boldsymbol{e}}_n^*).
\end{cases}
\tag{6.35}
$$

Note that, by (6.17), (6.25), and (6.26), we have

$$
\begin{cases}
(\forall i\in I)\ \ q_{i,n}=q_{i,\overline{\ell}_i(n)},\ \ c_{i,n}^*=c_{i,\overline{\ell}_i(n)}^*,\ \ a_{i,n}=a_{i,\overline{\ell}_i(n)},\ \ s_{i,n}^*=s_{i,\overline{\ell}_i(n)}^*,\ \ c_{i,n}=c_{i,\overline{\ell}_i(n)}\\
(\forall k\in K)\ \ b_{k,n}=b_{k,\overline{\vartheta}_k(n)},\ \ e_{k,n}^*=e_{k,\overline{\vartheta}_k(n)}^*,\ \ b_{k,n}^*=b_{k,\overline{\vartheta}_k(n)}^*.
\end{cases}
\tag{6.36}
$$

Hence, for every $i\in I$, we deduce from (6.17), (6.25), and (6.34) that

$$
\gamma_{i,\ell_i(n)}^{-1}x_{i,\overline{\ell}_i(n)}^*=\gamma_{i,\ell_i(n)}^{-1}x_{i,\ell_i(n)}-\nabla\psi_i\big(x_{i,\ell_i(n)}\big)-M_i^*u_{i,\ell_i(n)}^*-\sum_{k\in K}L_{k,i}^*v_{k,\ell_i(n)}^*
$$

$$
\begin{aligned}
&= E_{i,n}x_{i,\ell_i(n)} - M_i^* u^*_{i,\ell_i(n)} - \sum_{k\in K} L^*_{k,i} v^*_{k,\ell_i(n)}\\
&= E_{i,n}x_{i,n} - M_i^* u^*_{i,n} - \sum_{k\in K} L^*_{k,i} v^*_{k,n} + \widetilde{x}^*_{i,n},
\end{aligned}
\tag{6.37}
$$

that

$$
\begin{aligned}
\mu^{-1}_{i,\ell_i(n)} q_{i,n} &= \mu^{-1}_{i,\ell_i(n)} q_{i,\bar{\ell}_i(n)}\\
&= \mu^{-1}_{i,\ell_i(n)} y_{i,\ell_i(n)} - \nabla_i \boldsymbol{f}_i\big(\boldsymbol{y}_{\ell_i(n)}\big) + u^*_{i,\ell_i(n)}\\
&= \mu^{-1}_{i,\ell_i(n)} y_{i,n} - \nabla_i \boldsymbol{f}_i(\boldsymbol{y}_n) + u^*_{i,n} + \widetilde{q}^*_{i,n},
\end{aligned}
\tag{6.38}
$$

and that

$$
\begin{aligned}
\sigma^{-1}_{i,\ell_i(n)} c^*_{i,n} &= \sigma^{-1}_{i,\ell_i(n)} c^*_{i,\bar{\ell}_i(n)}\\
&= \sigma^{-1}_{i,\ell_i(n)} u^*_{i,\ell_i(n)} - y_{i,\ell_i(n)} + M_i x_{i,\ell_i(n)}\\
&= \sigma^{-1}_{i,\ell_i(n)} u^*_{i,n} - y_{i,n} + M_i x_{i,n} + \widetilde{c}^*_{i,n}.
\end{aligned}
\tag{6.39}
$$

In a similar fashion,

$$
(\forall k\in K)\quad
\begin{cases}
\nu^{-1}_{k,\vartheta_k(n)} d^*_{k,\bar{\vartheta}_k(n)} = G_{k,n} z_{k,n} + v^*_{k,n} + \widetilde{d}^*_{k,n}\\
\varrho^{-1}_{k,\vartheta_k(n)} e^*_{k,n} = \varrho^{-1}_{k,\vartheta_k(n)} v^*_{k,n} - z_{k,n} + \sum_{i\in I} L_{k,i} x_{i,n} + \widetilde{e}^*_{k,n}.
\end{cases}
\tag{6.40}
$$

Therefore, it results from (6.28), (6.34), (6.35), (6.20), (6.23), (6.13), and Problem 6.1[d] that

$$
\begin{aligned}
&\mathbf{K}_n \mathbf{x}_n + \mathbf{e}^*_n\\
&\quad= \mathbf{T}_n \mathbf{x}_n - \mathbf{S}\mathbf{x}_n + \mathbf{e}^*_n\\
&\quad= \Big(\big(\gamma^{-1}_{i,\ell_i(n)} x^*_{i,\bar{\ell}_i(n)}\big)_{i\in I}, \big(\mu^{-1}_{i,\ell_i(n)} q_{i,n}\big)_{i\in I}, \big(\nu^{-1}_{k,\vartheta_k(n)} d^*_{k,\bar{\vartheta}_k(n)}\big)_{k\in K}, \big(\sigma^{-1}_{i,\ell_i(n)} c^*_{i,n}\big)_{i\in I}, \big(\varrho^{-1}_{k,\vartheta_k(n)} e^*_{k,n}\big)_{k\in K}\Big).
\end{aligned}
\tag{6.41}
$$

On the other hand, in the light of (6.28), (6.22), (6.19), (6.13), and [3, Proposition 16.44] we get

$$
\begin{aligned}
(\mathbf{K}_n + \mathbf{A})^{-1}\colon \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{X}}\colon (\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{z}^*, \boldsymbol{u}, \boldsymbol{v}) \mapsto \Big(&\big(\operatorname{prox}_{\gamma_{i,\ell_i(n)}\varphi_i}(\gamma_{i,\ell_i(n)} x^*_i)\big)_{i\in I}, \big(\mu_{i,\ell_i(n)} y^*_i\big)_{i\in I},\\
&\big(\operatorname{prox}_{\nu_{k,\vartheta_k(n)} g_k}(\nu_{k,\vartheta_k(n)} z^*_k)\big)_{k\in K}, \big(\sigma_{i,\ell_i(n)} u_i\big)_{i\in I}, \big(\varrho_{k,\vartheta_k(n)} v_k\big)_{k\in K}\Big).
\end{aligned}
\tag{6.42}
$$

Hence, since (6.36), (6.17), (6.25), and (6.26) entail that

$$
\begin{cases}
(\forall i\in I)\ \ a_{i,n} = a_{i,\bar{\ell}_i(n)} = \operatorname{prox}_{\gamma_{i,\ell_i(n)}\varphi_i} x^*_{i,\bar{\ell}_i(n)}\\
(\forall k\in K)\ \ b_{k,n} = b_{k,\bar{\vartheta}_k(n)} = \operatorname{prox}_{\nu_{k,\vartheta_k(n)} g_k} d^*_{k,\bar{\vartheta}_k(n)},
\end{cases}
\tag{6.43}
$$

we invoke (6.35) to get

$$\mathbf{y}_n = (\mathbf{K}_n + \mathbf{A})^{-1}(\mathbf{K}_n\mathbf{x}_n + \mathbf{e}_n^*). \tag{6.44}$$

At the same time, it follows from (6.33) and [3, Corollary 20.28 and Proposition 22.11(ii)] that $\mathbf{K}_n$ is surjective and, in turn, that there exists $\widetilde{\mathbf{x}}_n \in \boldsymbol{\mathcal{X}}$ such that

$$\mathbf{K}_n\widetilde{\mathbf{x}}_n = \mathbf{K}_n\mathbf{x}_n + \mathbf{e}_n^*. \tag{6.45}$$

Thus, (6.44) and Definition 6.2 yield

$$\mathbf{y}_n = J_{\mathbf{A}}^{\mathbf{K}_n}\widetilde{\mathbf{x}}_n. \tag{6.46}$$

In view of (6.17), (6.36), and (6.25), we derive from (6.37) that

$$\begin{aligned}
(\forall i \in I)\quad a_{i,n}^* &= s_{i,n}^* + \sum_{k\in K} L_{k,i}^* e_{k,n}^* \\
&= s_{i,\bar{\ell}_i(n)}^* + \sum_{k\in K} L_{k,i}^* e_{k,n}^* \\
&= \gamma_{i,\ell_i(n)}^{-1}\big(x_{i,\bar{\ell}_i(n)}^* - a_{i,\bar{\ell}_i(n)}\big) + \nabla\psi_i\big(a_{i,\bar{\ell}_i(n)}\big) + M_i^* c_{i,\bar{\ell}_i(n)}^* + \sum_{k\in K} L_{k,i}^* e_{k,n}^* \\
&= \gamma_{i,\ell_i(n)}^{-1} x_{i,\bar{\ell}_i(n)}^* - \left(\gamma_{i,\ell_i(n)}^{-1} a_{i,n} - \nabla\psi_i(a_{i,n}) - M_i^* c_{i,n}^* - \sum_{k\in K} L_{k,i}^* e_{k,n}^*\right) \\
&= \left(E_{i,n}x_{i,n} - M_i^* u_{i,n}^* - \sum_{k\in K} L_{k,i}^* v_{k,n}^*\right) \\
&\quad - \left(E_{i,n}a_{i,n} - M_i^* c_{i,n}^* - \sum_{k\in K} L_{k,i}^* e_{k,n}^*\right) + \widetilde{x}_{i,n}^*,
\end{aligned} \tag{6.47}$$

from (6.38) that

$$\begin{aligned}
(\forall i \in I)\quad q_{i,n}^* &= \nabla_i \boldsymbol{f}_i(\boldsymbol{q}_n) - c_{i,n}^* \\
&= \big(\mu_{i,\ell_i(n)}^{-1} y_{i,n} - \nabla_i \boldsymbol{f}_i(\boldsymbol{y}_n) + u_{i,n}^*\big) - \big(\mu_{i,\ell_i(n)}^{-1} q_{i,n} - \nabla_i \boldsymbol{f}_i(\boldsymbol{q}_n) + c_{i,n}^*\big) + \widetilde{q}_{i,n}^*,
\end{aligned} \tag{6.48}$$

and from (6.39) that

$$\begin{aligned}
(\forall i \in I)\quad c_{i,n} &= c_{i,\bar{\ell}_i(n)} \\
&= q_{i,\bar{\ell}_i(n)} - M_i a_{i,\bar{\ell}_i(n)} \\
&= q_{i,n} - M_i a_{i,n} \\
&= \big(\sigma_{i,\ell_i(n)}^{-1} u_{i,n}^* - y_{i,n} + M_i x_{i,n}\big) - \big(\sigma_{i,\ell_i(n)}^{-1} c_{i,n}^* - q_{i,n} + M_i a_{i,n}\big) + \widetilde{c}_{i,n}^*.
\end{aligned} \tag{6.49}$$

A similar analysis shows that

$$(\forall k \in K) \quad b_{k,n}^* = \big(G_{k,n} z_{k,n} + v_{k,n}^*\big) - \big(G_{k,n} b_{k,n} + e_{k,n}^*\big) + \widetilde{d}_{k,n}^* \tag{6.50}$$

and

$$(\forall k \in K) \quad e_{k,n} = \left(\varrho_{k,\vartheta_k(n)}^{-1} v_{k,n}^* - z_{k,n} + \sum_{i \in I} L_{k,i} x_{i,n}\right) - \left(\varrho_{k,\vartheta_k(n)}^{-1} e_{k,n}^* - b_{k,n} + \sum_{i \in I} L_{k,i} a_{i,n}\right) + \widetilde{e}_{k,n}^*. \tag{6.51}$$

Altogether, it follows from (6.35), (6.47)–(6.51), (6.34), (6.28), (6.20), (6.23), (6.13), and (6.45) that

$$\mathbf{y}_n^* = (\mathbf{T}_n \mathbf{x}_n - \mathbf{S}\mathbf{x}_n) - (\mathbf{T}_n \mathbf{y}_n - \mathbf{S}\mathbf{y}_n) + \mathbf{e}_n^* = \mathbf{K}_n \mathbf{x}_n - \mathbf{K}_n \mathbf{y}_n + \mathbf{e}_n^* = \mathbf{K}_n \widetilde{\mathbf{x}}_n - \mathbf{K}_n \mathbf{y}_n. \tag{6.52}$$

Further, in view of (6.17) and (6.35), we have

$$\pi_n = \langle \mathbf{y}_n - \mathbf{x}_n \mid \mathbf{y}_n^* \rangle \quad \text{and} \quad \mathbf{x}_{n+1} = \begin{cases} \mathbf{x}_n + \dfrac{\lambda_n \pi_n}{\|\mathbf{y}_n^*\|^2} \mathbf{y}_n^*, & \text{if } \pi_n < 0; \\ \mathbf{x}_n, & \text{otherwise.} \end{cases} \tag{6.53}$$

Combining (6.22), (6.24), (6.33), (6.46), (6.52), and (6.53), we conclude that (6.17) is an instantiation of (6.10). Hence, Proposition 6.3(i) yields

$$\sum_{n \in \mathbb{N}} \|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 < +\infty. \tag{6.54}$$

For every $i \in I$ and every integer $n \geqslant P$, (6.16) entails that $i \in \bigcup_{j=n-P}^{n} I_j$ and, in turn, (6.25) and [a] imply that $n - P - D \leqslant \overline{\ell}_i(n) - D \leqslant \tau_i(\overline{\ell}_i(n)) = \ell_i(n) \leqslant \overline{\ell}_i(n) \leqslant n$. Consequently,

$$(\forall i \in I)(\forall n \in \mathbb{N}) \quad n \geqslant P + D \quad \Rightarrow \quad \|\mathbf{x}_n - \mathbf{x}_{\ell_i(n)}\| \leqslant \sum_{j=0}^{P+D} \|\mathbf{x}_n - \mathbf{x}_{n-j}\|, \tag{6.55}$$

and we therefore infer from (6.54) that

$$(\forall i \in I) \quad \mathbf{x}_n - \mathbf{x}_{\ell_i(n)} \to \mathbf{0}. \tag{6.56}$$

Likewise,

$$(\forall k \in K) \quad \mathbf{x}_n - \mathbf{x}_{\vartheta_k(n)} \to \mathbf{0}. \tag{6.57}$$

Hence, we deduce from (6.35), (6.34), (6.28), and (6.31) that

$$\begin{aligned}(\forall i \in I) \quad \|\widetilde{x}_{i,n}^*\| \leqslant \|E_{i,n} x_{i,\ell_i(n)} - E_{i,n} x_n\| + \|M_i^*\|\,\|u_{i,n}^* - u_{i,\ell_i(n)}^*\| \\ + \sum_{k \in K} \|L_{k,i}^*\|\,\|v_{k,n}^* - v_{k,\ell_i(n)}^*\|\end{aligned}$$

$$
\begin{aligned}
&\leqslant \|\boldsymbol{E}_n\boldsymbol{x}_{\ell_i(n)} - \boldsymbol{E}_n\boldsymbol{x}_n\| + \|M_i^*\|\,\|\boldsymbol{u}_n^* - \boldsymbol{u}_{\ell_i(n)}^*\| + \sum_{k\in K}\|L_{k,i}^*\|\,\|\boldsymbol{v}_n^* - \boldsymbol{v}_{\ell_i(n)}^*\|\\
&\leqslant \alpha\|\boldsymbol{x}_{\ell_i(n)} - \boldsymbol{x}_n\| + \|M_i^*\|\,\|\boldsymbol{u}_n^* - \boldsymbol{u}_{\ell_i(n)}^*\| + \sum_{k\in K}\|L_{k,i}^*\|\,\|\boldsymbol{v}_n^* - \boldsymbol{v}_{\ell_i(n)}^*\|\\
&\to 0.
\end{aligned}
\tag{6.58}
$$

Moreover, using (6.35), [a], and (6.56), we get

$$
\begin{aligned}
(\forall i\in I)\quad \|\widetilde{q}_{i,n}^*\| &\leqslant \mu_{i,\ell_i(n)}^{-1}\|y_{i,\ell_i(n)} - y_{i,n}\| + \big\|\nabla_i\boldsymbol{f}_i(\boldsymbol{y}_n) - \nabla_i\boldsymbol{f}_i\big(\boldsymbol{y}_{\ell_i(n)}\big)\big\| + \|u_{i,\ell_i(n)}^* - u_{i,n}^*\|\\
&\leqslant \varepsilon^{-1}\|\boldsymbol{y}_{\ell_i(n)} - \boldsymbol{y}_n\| + \|\boldsymbol{Q}\boldsymbol{y}_n - \boldsymbol{Q}\boldsymbol{y}_{\ell_i(n)}\| + \|\boldsymbol{u}_{\ell_i(n)}^* - \boldsymbol{u}_n^*\|\\
&\leqslant (\varepsilon^{-1} + \kappa)\|\boldsymbol{y}_{\ell_i(n)} - \boldsymbol{y}_n\| + \|\boldsymbol{u}_{\ell_i(n)}^* - \boldsymbol{u}_n^*\|\\
&\to 0
\end{aligned}
\tag{6.59}
$$

and

$$
\begin{aligned}
(\forall i\in I)\quad \|\widetilde{c}_{i,n}^*\| &\leqslant \sigma_{i,\ell_i(n)}^{-1}\|u_{i,\ell_i(n)}^* - u_{i,n}^*\| + \|M_i\|\,\|x_{i,\ell_i(n)} - x_{i,n}\| + \|y_{i,\ell_i(n)} - y_{i,n}\|\\
&\leqslant \varepsilon^{-1}\|\boldsymbol{u}_{\ell_i(n)}^* - \boldsymbol{u}_n^*\| + \|M_i\|\,\|\boldsymbol{x}_{\ell_i(n)} - \boldsymbol{x}_n\| + \|\boldsymbol{y}_{\ell_i(n)} - \boldsymbol{y}_n\|\\
&\to 0.
\end{aligned}
\tag{6.60}
$$

A similar analysis shows that

$$
(\forall k\in K)\quad \|\widetilde{d}_{k,n}^*\|\to 0 \quad\text{and}\quad \|\widetilde{e}_{k,n}^*\|\to 0.
\tag{6.61}
$$

Altogether, we invoke (6.35) and (6.58)–(6.61) to get

$$
\mathsf{e}_n^*\to\mathbf{0}.
\tag{6.62}
$$

Hence, arguing as in (6.11), (6.33) and (6.45) give

$$
\|\widetilde{\mathsf{x}}_n - \mathsf{x}_n\| \leqslant \frac{\|\mathsf{K}_n\widetilde{\mathsf{x}}_n - \mathsf{K}_n\mathsf{x}_n\|}{\varsigma} = \frac{\|\mathsf{e}_n^*\|}{\varsigma}\to 0.
\tag{6.63}
$$

Hence, Proposition 6.3(ii) asserts that there exists $\overline{\mathsf{x}} = (\overline{\boldsymbol{x}},\overline{\boldsymbol{y}},\overline{\boldsymbol{z}},\overline{\boldsymbol{u}}^*,\overline{\boldsymbol{v}}^*)\in\operatorname{zer}\mathsf{A}$ such that $\mathsf{x}_n\rightharpoonup\overline{\mathsf{x}}$. This yields $\boldsymbol{x}_n\rightharpoonup\overline{\boldsymbol{x}}$. It remains to verify that $\overline{\boldsymbol{x}}$ solves (6.3). Towards this end, let $i\in I$ and set

$$
f_i = \boldsymbol{f}_i\big(\cdot\,;(\boldsymbol{M}\overline{\boldsymbol{x}})_{\smallsetminus i}\big)\quad\text{and}\quad(\forall k\in K)\ \ \widetilde{g}_k = (g_k+h_k)(\cdot + \boldsymbol{L}_{k,\smallsetminus i}\overline{\boldsymbol{x}}_{\smallsetminus i}).
\tag{6.64}
$$

Then, by Problem 6.1[b], $f_i\colon\mathcal{K}_i\to\mathbb{R}$ is convex and Gâteaux differentiable, with $\nabla f_i(M_i\overline{x}_i) = \nabla_i\boldsymbol{f}_i(\boldsymbol{M}\overline{\boldsymbol{x}})$. In addition, $(\forall k\in K)(\forall z_k\in\mathcal{G}_k)\ \partial\widetilde{g}_k(z_k) = (\partial g_k+\nabla h_k)(z_k+\boldsymbol{L}_{k,\smallsetminus i}\overline{\boldsymbol{x}}_{\smallsetminus i})$. At the same time, we deduce from (6.15) that $\overline{\boldsymbol{u}}^* = \boldsymbol{Q}\overline{\boldsymbol{y}} = \boldsymbol{Q}(\boldsymbol{M}\overline{\boldsymbol{x}})$, $\overline{\boldsymbol{z}} = \boldsymbol{L}\overline{\boldsymbol{x}}$, $\overline{\boldsymbol{v}}^*\in\boldsymbol{B}\overline{\boldsymbol{z}}$, and $\mathbf{0}\in$

$\boldsymbol{A}\overline{\boldsymbol{x}} + \boldsymbol{M}^*\overline{\boldsymbol{u}}^* + \boldsymbol{L}^*\overline{\boldsymbol{v}}^*$. Thus, it results from (6.13) and Problem 6.1[d] that

$$\begin{cases} \overline{u}_i^* = \nabla_i \boldsymbol{f}_i(\boldsymbol{M}\overline{\boldsymbol{x}}) = \nabla f_i(M_i \overline{x}_i) \\ (\forall k \in K) \;\; \overline{z}_k = \sum_{j \in I} L_{k,j}\overline{x}_j = L_{k,i}\overline{x}_i + \boldsymbol{L}_{k,\smallsetminus i}\overline{\boldsymbol{x}}_{\smallsetminus i} \\ (\forall k \in K) \;\; \overline{v}_k^* \in \partial g_k(\overline{z}_k) + \nabla h_k(\overline{z}_k) = (\partial g_k + \nabla h_k)(L_{k,i}\overline{x}_i + \boldsymbol{L}_{k,\smallsetminus i}\overline{\boldsymbol{x}}_{\smallsetminus i}) = \partial \widetilde{g}_k(L_{k,i}\overline{x}_i) \end{cases} \tag{6.65}$$

and, in turn, from (6.23) and [3, Proposition 16.6(ii)] that

$$\begin{aligned} 0 &\in \partial\varphi_i(\overline{x}_i) + \nabla\psi_i(\overline{x}_i) + M_i^*\overline{u}_i^* + \sum_{k \in K} L_{k,i}^*\overline{v}_k^* \\ &\subset \partial\varphi_i(\overline{x}_i) + \nabla\psi_i(\overline{x}_i) + M_i^*\big(\nabla f_i(M_i\overline{x}_i)\big) + \sum_{k \in K} L_{k,i}^*\big(\partial\widetilde{g}_k(L_{k,i}\overline{x}_i)\big) \\ &\subset \partial\Bigg(\varphi_i + \psi_i + f_i \circ M_i + \sum_{k \in K} \widetilde{g}_k \circ L_{k,i}\Bigg)(\overline{x}_i). \end{aligned} \tag{6.66}$$

Consequently, appealing to Fermat's rule [3, Theorem 16.3] and (6.64), we arrive at

$$\begin{aligned} \overline{x}_i &\in \underset{x_i \in \mathcal{H}_i}{\text{Argmin}} \;\; \varphi_i(x_i) + \psi_i(x_i) + f_i(M_i x_i) + \sum_{k \in K} \widetilde{g}_k(L_{k,i}x_i) \\ &= \underset{x_i \in \mathcal{H}_i}{\text{Argmin}} \;\; \varphi_i(x_i) + \psi_i(x_i) + \boldsymbol{f}_i\big(M_i x_i; (\boldsymbol{M}\overline{\boldsymbol{x}})_{\smallsetminus i}\big) + \sum_{k \in K} (g_k + h_k)(L_{k,i}x_i + \boldsymbol{L}_{k,\smallsetminus i}\overline{\boldsymbol{x}}_{\smallsetminus i}), \end{aligned} \tag{6.67}$$

which completes the proof. □

Remark 6.5 Let us confirm that algorithm (6.17) complies with the principles laid out in Section 6.2.1.

- **Decomposition:** In (6.17), the nonsmooth functions $(\varphi_i)_{i \in I}$ and $(g_k)_{k \in K}$ are activated separately via their proximity operators, while the smooth functions $(\psi_i)_{i \in I}$, $(\boldsymbol{f}_i)_{i \in I}$, and $(h_k)_{k \in K}$ are activated separately via their gradients.
- **Block-iterative implementation:** At any iteration n, the functions $(\boldsymbol{f}_i)_{i \in I}$ are activated and we require only that the subfamilies $(\varphi_i)_{i \in I_n}$, $(\psi_i)_{i \in I_n}$, $(g_k)_{k \in K_n}$, and $(h_k)_{k \in K_n}$ be used. To guarantee convergence, we ask in condition [c] of Theorem 6.4 that each of these functions be activated frequently enough.
- **Asynchronous implementation:** Given $i \in I$ and $k \in K$, the asynchronous character of the algorithm is materialized by the variables $\tau_i(n)$ and $\delta_k(n)$ which signal when the underlying computations incorporated at iteration n were initiated. Conditions [a] and [b] of Theorem 6.4 ask that the lag between the initiation and the incorporation of such computations do not exceed D iterations. The introduction of such techniques in monotone operator splitting were initiated in [13].

Remark 6.6 Consider the proof of Theorem 6.4. Since Proposition 6.3(ii) yields $\mathsf{x}_n - \mathsf{y}_n \to \mathbf{0}$, we obtain $\boldsymbol{x}_n - \boldsymbol{a}_n \to \mathbf{0}$ via (6.35) and thus $\boldsymbol{a}_n \rightharpoonup \overline{\boldsymbol{x}}$. At the same time, by (6.17), given $i \in I$, the sequence $(a_{i,n})_{n\in\mathbb{N}}$ lies in $\operatorname{dom}\partial\varphi_i \subset \operatorname{dom}\varphi_i$. In particular, if a constraint on $\overline{x}_i$ is enforced via $\varphi_i = \iota_{C_i}$, then $(a_{i,n})_{n\in\mathbb{N}}$ converges to the ith component of a solution $\overline{\boldsymbol{x}}$ while being feasible in the sense that $C_i \ni a_{i,n} \rightharpoonup \overline{x}_i$.

Remark 6.7 The proof of Theorem 6.4 implicitly establishes the convergence of an asynchronous block-iterative algorithm to solve the more general system of monotone inclusions

$$\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \text{ such that}$$
$$(\forall i \in I)\;\; 0 \in A_i\overline{x}_i + R_i\overline{x}_i + M_i^*\big(Q_i(\boldsymbol{M}\overline{\boldsymbol{x}})\big) + \sum_{k\in K} L_{k,i}^*\Bigg((B_k + D_k)\Bigg(\sum_{j\in I} L_{k,j}\overline{x}_j\Bigg)\Bigg) \tag{6.68}$$

under the following assumptions:

[a] For every $i \in I$, $A_i\colon \mathcal{H}_i \to 2^{\mathcal{H}_i}$ is maximally monotone, $\alpha_i \in [0,+\infty[$, and $R_i\colon \mathcal{H}_i \to \mathcal{H}_i$ is monotone and α_i-Lipschitzian.

[b] For every $i \in I$, $Q_i\colon \boldsymbol{\mathcal{K}} \to \mathcal{K}_i$. It is assumed that the operator $\boldsymbol{Q}\colon \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}}\colon \boldsymbol{y} \mapsto (Q_i\boldsymbol{y})_{i\in I}$ is monotone and Lipschitzian. Furthermore, $(\chi_i)_{i\in I}$ are positive numbers such that

$$(\forall \boldsymbol{y} \in \boldsymbol{\mathcal{K}})(\forall \boldsymbol{y}' \in \boldsymbol{\mathcal{K}}) \quad \langle \boldsymbol{y} - \boldsymbol{y}' \mid \boldsymbol{Q}\boldsymbol{y} - \boldsymbol{Q}\boldsymbol{y}'\rangle \leqslant \sum_{i\in I} \chi_i\|y_i - y_i'\|^2. \tag{6.69}$$

[c] For every $k \in K$, $B_k\colon \mathcal{G}_k \to 2^{\mathcal{G}_k}$ is maximally monotone, $\beta_k \in [0,+\infty[$, and $D_k\colon \mathcal{G}_k \to \mathcal{G}_k$ is monotone and β_k-Lipschitzian.

[d] For every $i \in I$ and every $k \in K$, $M_i\colon \mathcal{H}_i \to \mathcal{K}_i$ and $L_{k,i}\colon \mathcal{H}_i \to \mathcal{G}_k$ are linear and bounded. Moreover, we set $\boldsymbol{M}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{K}}\colon \boldsymbol{x} \mapsto (M_ix_i)_{i\in I}$.

Indeed, denote by Z the set of points $(\boldsymbol{x}, \boldsymbol{u}^*, \boldsymbol{v}^*) \in \boldsymbol{\mathcal{H}} \oplus \boldsymbol{\mathcal{K}} \oplus \boldsymbol{\mathcal{G}}$ such that

$$\begin{cases} (\forall i \in I)\;\; u_i^* = Q_i(\boldsymbol{M}\boldsymbol{x}) \\ (\forall k \in K)\;\; v_k^* \in (B_k + D_k)\big(\sum_{j\in I} L_{k,j}x_j\big) \\ (\forall i \in I)\;\; -M_i^*u_i^* - \sum_{k\in K} L_{k,i}^*v_k^* \in A_ix_i + R_ix_i. \end{cases} \tag{6.70}$$

Suppose that $\mathsf{Z} \neq \varnothing$ and execute (6.17) with the following modifications:

- For every $i \in I$ and every $n \in \mathbb{N}$, $\operatorname{prox}_{\gamma_{i,n}\varphi_i}$ is replaced by $J_{\gamma_{i,n}A_i}^{\mathrm{Id}}$, $\nabla\psi_i$ by R_i, and $\nabla_i\boldsymbol{f}_i$ by Q_i.
- For every $k \in K$ and every $n \in \mathbb{N}$, $\operatorname{prox}_{\nu_{k,n}g_k}$ is replaced by $J_{\nu_{k,n}B_k}^{\mathrm{Id}}$, and ∇h_k by D_k.

Then there exists $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}^*, \overline{\boldsymbol{v}}^*) \in \mathsf{Z}$ such that $(\boldsymbol{x}_n, \boldsymbol{u}_n^*, \boldsymbol{v}_n^*) \rightharpoonup (\overline{\boldsymbol{x}}, \overline{\boldsymbol{u}}^*, \overline{\boldsymbol{v}}^*)$ and $\overline{\boldsymbol{x}}$ solves (6.68).

Remark 6.8 By invoking [9, Theorem 4.8] and arguing as in the proof of Proposition 6.3, we obtain a strongly convergent counterpart of Proposition 6.3 which, in turn, yields a strongly convergent version of Theorem 6.4.

Theorem 6.4 requires that (6.18) be satisfied. With the assistance of monotone operator theory arguments applied to a set of primal-dual inclusions, we provide below sufficient conditions for that. Let us start with a technical fact.

Lemma 6.9 *Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $\boldsymbol{B}\colon \mathcal{G} \to 2^{\mathcal{G}}$ be 3^* monotone, and let $\boldsymbol{L}\colon \mathcal{H} \to \mathcal{G}$ be linear and bounded. Then $\boldsymbol{L}^* \circ \boldsymbol{B} \circ \boldsymbol{L}$ is 3^* monotone.*

Proof. Set $\boldsymbol{A} = \boldsymbol{L}^* \circ \boldsymbol{B} \circ \boldsymbol{L}$. First, we deduce from [3, Proposition 20.10] that $\boldsymbol{A}$ is monotone. Next, take $\boldsymbol{x} \in \operatorname{dom} \boldsymbol{A}$ and $\boldsymbol{x}^* \in \operatorname{ran} \boldsymbol{A}$. On the one hand, $\boldsymbol{L}\boldsymbol{x} \in \operatorname{dom} \boldsymbol{B}$ and there exists $\boldsymbol{z}^* \in \operatorname{ran} \boldsymbol{B}$ such that $\boldsymbol{x}^* = \boldsymbol{L}^*\boldsymbol{z}^*$. On the other hand, for every $(\boldsymbol{y}, \boldsymbol{y}^*) \in \operatorname{gra} \boldsymbol{A}$, there exists $\boldsymbol{v}^* \in \mathcal{G}$ such that $(\boldsymbol{L}\boldsymbol{y}, \boldsymbol{v}^*) \in \operatorname{gra} \boldsymbol{B}$ and $\boldsymbol{y}^* = \boldsymbol{L}^*\boldsymbol{v}^*$, from which we obtain

$$\begin{aligned}\langle \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{y}^* - \boldsymbol{x}^* \rangle &= \langle \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{L}^*\boldsymbol{v}^* - \boldsymbol{L}^*\boldsymbol{z}^* \rangle \\ &= \langle \boldsymbol{L}\boldsymbol{x} - \boldsymbol{L}\boldsymbol{y} \mid \boldsymbol{v}^* - \boldsymbol{z}^* \rangle \\ &\leqslant \sup_{(\boldsymbol{w},\boldsymbol{w}^*) \in \operatorname{gra} \boldsymbol{B}} \langle \boldsymbol{L}\boldsymbol{x} - \boldsymbol{w} \mid \boldsymbol{w}^* - \boldsymbol{z}^* \rangle. \end{aligned} \tag{6.71}$$

Therefore, by 3^* monotonicity of $\boldsymbol{B}$,

$$\sup_{(\boldsymbol{y},\boldsymbol{y}^*) \in \operatorname{gra} \boldsymbol{A}} \langle \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{y}^* - \boldsymbol{x}^* \rangle \leqslant \sup_{(\boldsymbol{w},\boldsymbol{w}^*) \in \operatorname{gra} \boldsymbol{B}} \langle \boldsymbol{L}\boldsymbol{x} - \boldsymbol{w} \mid \boldsymbol{w}^* - \boldsymbol{z}^* \rangle < +\infty. \tag{6.72}$$

Consequently, $\boldsymbol{A}$ is 3^* monotone. □

Proposition 6.10 *Consider the setting of Problem 6.1 and set*

$$\boldsymbol{C} = \left\{ \left(\sum_{i \in I} L_{k,i} x_i - z_k \right)_{k \in K} \,\middle|\, (\forall i \in I) \;\; x_i \in \operatorname{dom} \varphi_i \;\; \textit{and} \;\; (\forall k \in K) \;\; z_k \in \operatorname{dom} g_k \right\}. \tag{6.73}$$

Suppose that $\boldsymbol{0} \in \operatorname{sri} \boldsymbol{C}$ and that one of the following is satisfied:

[a] *For every $i \in I$, one of the following holds:*

1/ *$\partial(\varphi_i + \psi_i)$ is surjective.*

2/ *$\varphi_i + \psi_i$ is supercoercive.*

3/ *$\operatorname{dom} \varphi_i$ is bounded.*

4/ *$\varphi_i + \psi_i$ is uniformly convex.*

[b] *$\boldsymbol{Q}$ is 3^* monotone and one of the following holds:*

1/ *$\boldsymbol{M}^* \circ \boldsymbol{Q} \circ \boldsymbol{M}$ is surjective.*

2/ *$\boldsymbol{Q}$ is surjective and, for every $i \in I$, M_i is bijective.*

Then (6.18) holds.

Proof. Let $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{L}$ be as in (6.13) and define

$$\boldsymbol{T}\colon \mathcal{H} \to 2^{\mathcal{H}}\colon \boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x} + \boldsymbol{L}^*\big(\boldsymbol{B}(\boldsymbol{L}\boldsymbol{x})\big) + \boldsymbol{M}^*\big(\boldsymbol{Q}(\boldsymbol{M}\boldsymbol{x})\big). \tag{6.74}$$

Suppose that $\widehat{\boldsymbol{x}} \in \operatorname{zer} \boldsymbol{T}$ and set $\widehat{\boldsymbol{u}}^* = \boldsymbol{Q}(\boldsymbol{M}\widehat{\boldsymbol{x}})$. On the one hand, in view of Problem 6.1[b], $(\forall i \in I)\ \widehat{u}_i^* = \nabla_i \boldsymbol{f}_i(\boldsymbol{M}\widehat{\boldsymbol{x}})$. On the other hand, it results from (6.74) that there exists $\widehat{\boldsymbol{v}}^* \in \boldsymbol{B}(\boldsymbol{L}\widehat{\boldsymbol{x}})$ such that $-\boldsymbol{M}^*\widehat{\boldsymbol{u}}^* - \boldsymbol{L}^*\widehat{\boldsymbol{v}}^* \in \boldsymbol{A}\widehat{\boldsymbol{x}}$ or, equivalently, by (6.23) and (6.13), $(\forall i \in I)\ -M_i^*\widehat{u}_i^* - \sum_{k\in K} L_{k,i}^*\widehat{v}_k^* \in \partial\varphi_i(\widehat{x}_i) + \nabla\psi_i(\widehat{x}_i)$. Further, using (6.13), we obtain $(\forall k \in K)\ \widehat{v}_k^* \in (\partial g_k + \nabla h_k)(\sum_{j\in I} L_{k,j}\widehat{x}_j)$. Altogether, we have shown that $\operatorname{zer}\boldsymbol{T} \neq \varnothing \Rightarrow$ (6.18) holds. Therefore, it suffices to show that $\operatorname{zer}\boldsymbol{T} \neq \varnothing$. To do so, define

$$
\begin{cases}
\boldsymbol{\varphi}\colon \boldsymbol{\mathcal{H}} \to]-\infty,+\infty] : \boldsymbol{x} \mapsto \sum_{i\in I}\big(\varphi_i(x_i) + \psi_i(x_i)\big)\\
\boldsymbol{g}\colon \boldsymbol{\mathcal{G}} \to]-\infty,+\infty] : \boldsymbol{z} \mapsto \sum_{k\in K}\big(g_k(z_k) + h_k(z_k)\big)\\
\boldsymbol{P} = \boldsymbol{A} + \boldsymbol{L}^* \circ \boldsymbol{B} \circ \boldsymbol{L}.
\end{cases}
\tag{6.75}
$$

Then, by (6.13) and [3, Proposition 16.9], $\boldsymbol{A} = \partial\boldsymbol{\varphi}$ and $\boldsymbol{B} = \partial\boldsymbol{g}$. In turn, since (6.73) and (6.13) imply that $\boldsymbol{0} \in \operatorname{sri}\boldsymbol{C} = \operatorname{sri}(\boldsymbol{L}(\operatorname{dom}\boldsymbol{\varphi}) - \operatorname{dom}\boldsymbol{g})$, we derive from [3, Theorem 16.47(i)] that $\boldsymbol{P} = \boldsymbol{A} + \boldsymbol{L}^*\circ\boldsymbol{B}\circ\boldsymbol{L} = \partial(\boldsymbol{\varphi} + \boldsymbol{g}\circ\boldsymbol{L})$. Therefore, in view of [3, Theorem 20.25 and Example 25.13],

$$
\boldsymbol{A},\ \boldsymbol{B}, \text{ and } \boldsymbol{P} \text{ are maximally monotone and } 3^* \text{ monotone.}
\tag{6.76}
$$

[a]: Fix temporarily $i \in I$. By [3, Theorem 20.25], $\partial(\varphi_i + \psi_i)$ is maximally monotone. First, if [a]2/ holds, then [3, Corollary 16.30, and Propositions 14.15 and 16.27] entail that $\operatorname{ran}\partial(\varphi_i + \psi_i) = \operatorname{dom}\partial(\varphi_i + \psi_i)^* = \mathcal{H}_i$ and, hence, [a]1/ holds. Second, if [a]3/ holds, then $\operatorname{dom}\partial(\varphi_i + \psi_i) \subset \operatorname{dom}(\varphi_i + \psi_i) = \operatorname{dom}\varphi_i$ is bounded and, therefore, it follows from [3, Corollary 21.25] that [a]1/ holds. Finally, if [a]4/ holds, then [3, Proposition 17.26(ii)] implies that [a]2/ holds and, in turn, that [a]1/ holds. Altogether, it is enough to assume that the operators $(\partial(\varphi_i + \psi_i))_{i\in I}$ are surjective and to show that $\operatorname{zer}\boldsymbol{T} \neq \varnothing$. Assume that $(\partial(\varphi_i + \psi_i))_{i\in I}$ are surjective and set $\boldsymbol{R} = -\boldsymbol{M}\circ\boldsymbol{P}^{-1}\circ(-\boldsymbol{M}^*) + \boldsymbol{Q}^{-1}$. Then we derive from (6.13) that $\boldsymbol{A}$ is surjective. On the other hand, Lemma 6.9 asserts that $\boldsymbol{L}^*\circ\boldsymbol{B}\circ\boldsymbol{L}$ is 3^* monotone. Hence, (6.76) and [3, Corollary 25.27(i)] yields $\operatorname{dom}\boldsymbol{P}^{-1} = \operatorname{ran}\boldsymbol{P} = \boldsymbol{\mathcal{H}}$. In turn, since $\boldsymbol{P}^{-1}$ and $\boldsymbol{Q}^{-1}$ are maximally monotone, [3, Theorem 25.3] implies that $\boldsymbol{R}$ is likewise. Furthermore, we observe that $\operatorname{dom}\boldsymbol{Q}^{-1} \subset \boldsymbol{\mathcal{K}} = \operatorname{dom}(-\boldsymbol{M}\circ\boldsymbol{P}^{-1}\circ(-\boldsymbol{M}^*))$ and, by virtue of (6.76), [3, Proposition 25.19(i)], and Lemma 6.9, that $-\boldsymbol{M}\circ\boldsymbol{P}^{-1}\circ(-\boldsymbol{M}^*)$ is 3^* monotone. Therefore, since $\operatorname{ran}\boldsymbol{Q}^{-1} = \operatorname{dom}\boldsymbol{Q} = \boldsymbol{\mathcal{K}}$, [3, Corollary 25.27(ii)] entails that $\boldsymbol{R}$ is surjective and, in turn, that $\operatorname{zer}\boldsymbol{R} \neq \varnothing$. Consequently, [3, Proposition 26.33(iii)] asserts that $\operatorname{zer}\boldsymbol{T} \neq \varnothing$.

[b]1/: Lemma 6.9 asserts that $\boldsymbol{M}^*\circ\boldsymbol{Q}\circ\boldsymbol{M}$ is 3^* monotone. At the same time, since $\boldsymbol{Q}$ is maximally monotone and $\operatorname{dom}\boldsymbol{Q} = \boldsymbol{\mathcal{K}}$, it results from (6.76) and [3, Theorem 25.3] that $\boldsymbol{T} = \boldsymbol{P} + \boldsymbol{M}^*\circ\boldsymbol{Q}\circ\boldsymbol{M}$ is maximally monotone. Hence, since $\boldsymbol{M}^*\circ\boldsymbol{Q}\circ\boldsymbol{M}$ is surjective, we derive from (6.76) and [3, Corollary 25.27(i)] that $\boldsymbol{T}$ is surjective and, therefore, that $\operatorname{zer}\boldsymbol{T} \neq \varnothing$.

[b]2/⇒[b]1/: Since the assumption implies that $\boldsymbol{M}$ is bijective, so is $\boldsymbol{M}^*$. This makes $\boldsymbol{M}^*\circ\boldsymbol{Q}\circ\boldsymbol{M}$ surjective. ☐

Remark 6.11 Sufficient conditions for $\mathbf{0} \in \operatorname{sri} \boldsymbol{C}$ to hold in Proposition 6.10 can be found in [12, Proposition 5.3].

6.2.4 Application examples

We discuss problems which are shown to be realizations of Problem 6.1 and which can therefore be solved by the asynchronous block-iterative algorithm (6.17) of Theorem 6.4.

Example 6.12 (quadratic coupling) Let $\mathcal{K}$ be a real Hilbert space and let I be a nonempty finite set. For every $i \in I$, let $\mathcal{H}_i$ be a real Hilbert space, let $\varphi_i \in \Gamma_0(\mathcal{H}_i)$, let $\alpha_i \in [0,+\infty[$, let $\psi_i\colon \mathcal{H}_i \to \mathbb{R}$ be convex and differentiable with an α_i-Lipschitzian gradient, let $M_i\colon \mathcal{H}_i \to \mathcal{K}$ be linear and bounded, let Λ_i be a nonempty finite set, let $(\omega_{i,\ell,j})_{\ell\in\Lambda_i, j\in I\smallsetminus\{i\}}$ be in $[0,+\infty[$, and let $(\kappa_{i,\ell})_{\ell\in\Lambda_i}$ be in $]0,+\infty[$. Additionally, set $\boldsymbol{\mathcal{H}} = \bigoplus_{i\in I}\mathcal{H}_i$ and $\boldsymbol{\mathcal{K}} = \bigoplus_{i\in I}\mathcal{K}$. The problem is to

$$\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \text{ such that}$$

$$(\forall i \in I)\ \ \overline{x}_i \in \underset{x_i\in\mathcal{H}_i}{\operatorname{Argmin}}\ \varphi_i(x_i) + \psi_i(x_i) + \sum_{\ell\in\Lambda_i} \frac{\kappa_{i,\ell}}{2}\left\| M_i x_i - \sum_{j\in I\smallsetminus\{i\}} \omega_{i,\ell,j} M_j \overline{x}_j \right\|^2. \tag{6.77}$$

It is assumed that

$$(\forall \boldsymbol{y} \in \boldsymbol{\mathcal{K}})(\forall \boldsymbol{y}' \in \boldsymbol{\mathcal{K}})\quad \sum_{i\in I}\sum_{\ell\in\Lambda_i} \kappa_{i,\ell}\left\langle y_i - y_i' \,\middle|\, y_i - y_i' - \sum_{j\in I\smallsetminus\{i\}} \omega_{i,\ell,j}(y_j - y_j')\right\rangle \geqslant 0. \tag{6.78}$$

Define

$$(\forall i \in I)\quad \boldsymbol{f}_i\colon \boldsymbol{\mathcal{K}} \to \mathbb{R}\colon \boldsymbol{y} \mapsto \sum_{\ell\in\Lambda_i} \frac{\kappa_{i,\ell}}{2}\left\| y_i - \sum_{j\in I\smallsetminus\{i\}} \omega_{i,\ell,j} y_j \right\|^2. \tag{6.79}$$

Then, for every $i \in I$ and every $\boldsymbol{y} \in \boldsymbol{\mathcal{K}}$, $\boldsymbol{f}_i(\cdot\,;\boldsymbol{y}_{\smallsetminus i})$ is convex and differentiable with

$$\nabla_i \boldsymbol{f}_i(\boldsymbol{y}) = \sum_{\ell\in\Lambda_i} \kappa_{i,\ell}\left(y_i - \sum_{j\in I\smallsetminus\{i\}} \omega_{i,\ell,j} y_j \right). \tag{6.80}$$

Hence, in view of (6.78), the operator $\boldsymbol{Q}\colon \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}}\colon \boldsymbol{y} \mapsto (\nabla_i \boldsymbol{f}_i(\boldsymbol{y}))_{i\in I}$ is monotone and Lipschitzian. Thus, (6.77) is a special case of (6.3) with $(\forall i \in I)\ \mathcal{K}_i = \mathcal{K}$ and $(\forall k \in K)\ g_k = h_k = 0$. In particular, suppose that, for every $i \in I$, $\mathcal{H}_i = \mathcal{K}$, C_i is a nonempty closed convex subset of $\mathcal{H}_i$, $\varphi_i = \iota_{C_i}$, $\psi_i = 0$, $M_i = \mathrm{Id}$, $\Lambda_i \subset I \smallsetminus \{i\}$, and

$$(\forall \ell \in \Lambda_i)\quad \begin{cases} \kappa_{i,\ell} = 1 \\ (\forall j \in I \smallsetminus \{i\})\ \ \omega_{i,\ell,j} = \begin{cases} 1, & \text{if } j = \ell; \\ 0, & \text{if } j \neq \ell. \end{cases} \end{cases} \tag{6.81}$$

Then (6.77) becomes

$$\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \text{ such that } (\forall i \in I)\ \overline{x}_i \in \underset{x_i \in C_i}{\operatorname{Argmin}}\ \frac{1}{2} \sum_{\ell \in \Lambda_i} \|x_i - \overline{x}_\ell\|^2. \tag{6.82}$$

This unifies models found in [2].

Example 6.13 (minimax) Let I be a finite set and suppose that $\varnothing \neq J \subset I$. Let $(\mathcal{H}_i)_{i\in I}$ be real Hilbert spaces, and set $\boldsymbol{\mathcal{U}} = \bigoplus_{i \in I\smallsetminus J} \mathcal{H}_i$ and $\boldsymbol{\mathcal{V}} = \bigoplus_{j\in J} \mathcal{H}_j$. For every $i \in I$, let $\varphi_i \in \Gamma_0(\mathcal{H}_i)$, let $\alpha_i \in [0, +\infty[$, let $\psi_i \colon \mathcal{H}_i \to \mathbb{R}$ be convex and differentiable with an α_i-Lipschitzian gradient. Further, let $\boldsymbol{\mathcal{L}} \colon \boldsymbol{\mathcal{U}} \oplus \boldsymbol{\mathcal{V}} \to \mathbb{R}$ 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 $-\boldsymbol{\mathcal{L}}(\boldsymbol{u}, \cdot)$ and $\boldsymbol{\mathcal{L}}(\cdot, \boldsymbol{v})$ are convex. Finally, for every $i \in I \smallsetminus J$ and every $j \in J$, let $L_{j,i} \colon \mathcal{H}_i \to \mathcal{H}_j$ be linear and bounded. Consider the multivariate minimax problem

$$\underset{\boldsymbol{u}\in\boldsymbol{\mathcal{U}}}{\text{minimize}}\ \underset{\boldsymbol{v}\in\boldsymbol{\mathcal{V}}}{\text{maximize}}\ \sum_{i\in I\smallsetminus J} \big(\varphi_i(u_i)+\psi_i(u_i)\big) - \sum_{j\in J} \big(\varphi_j(v_j)+\psi_j(v_j)\big) + \boldsymbol{\mathcal{L}}(\boldsymbol{u},\boldsymbol{v}) + \sum_{i\in I\smallsetminus J}\sum_{j\in J} \langle L_{j,i}u_i \mid v_j\rangle. \tag{6.83}$$

Now set $\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{U}} \oplus \boldsymbol{\mathcal{V}}$ and define

$$(\forall i \in I) \quad \boldsymbol{f}_i \colon \boldsymbol{\mathcal{H}} \to \mathbb{R} \colon (\boldsymbol{u}, \boldsymbol{v}) \mapsto \begin{cases} \boldsymbol{\mathcal{L}}(\boldsymbol{u},\boldsymbol{v}) + \big\langle u_i \mid \sum_{j\in J} L_{j,i}^* v_j \big\rangle, & \text{if } i \in I \smallsetminus J;\\ -\boldsymbol{\mathcal{L}}(\boldsymbol{u},\boldsymbol{v}) - \big\langle \sum_{k\in I\smallsetminus J} L_{i,k}u_k \mid v_i \big\rangle, & \text{if } i \in J. \end{cases} \tag{6.84}$$

Then $\boldsymbol{\mathcal{H}} = \bigoplus_{i\in I} \mathcal{H}_i$ and (6.83) can be put in the form

$$\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \text{ such that } (\forall i \in I)\ \overline{x}_i \in \underset{x_i \in \mathcal{H}_i}{\operatorname{Argmin}}\ \varphi_i(x_i) + \psi_i(x_i) + \boldsymbol{f}_i(x_i; \overline{\boldsymbol{x}}_{\smallsetminus i}). \tag{6.85}$$

Let us verify Problem 6.1[b]. On the one hand, we have

$$(\forall i \in I)(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}}) \quad \nabla_i \boldsymbol{f}_i(\boldsymbol{x}) = \begin{cases} \nabla_i \boldsymbol{\mathcal{L}}(\boldsymbol{x}) + \sum_{j\in J} L_{j,i}^* x_j, & \text{if } i \in I \smallsetminus J;\\ -\nabla_i \boldsymbol{\mathcal{L}}(\boldsymbol{x}) - \sum_{k\in I\smallsetminus J} L_{i,k} x_k, & \text{if } i \in J. \end{cases} \tag{6.86}$$

On the other hand, the operator

$$\boldsymbol{R} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{x} \mapsto \Big(\big(\nabla_i \boldsymbol{\mathcal{L}}(\boldsymbol{x})\big)_{i\in I\smallsetminus J}, \big(-\nabla_j \boldsymbol{\mathcal{L}}(\boldsymbol{x})\big)_{j\in J} \Big) \tag{6.87}$$

is monotone [22, 23] and Lipschitzian, while the bounded linear operator

$$\boldsymbol{S} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{x} \mapsto \left(\left(\sum_{j\in J} L_{j,i}^* x_j \right)_{i\in I\smallsetminus J}, \left(-\sum_{k\in I\smallsetminus J} L_{i,k} x_k \right)_{i\in J} \right) \tag{6.88}$$

satisfies $\boldsymbol{S}^* = -\boldsymbol{S}$ and it is therefore monotone [3, Example 20.35]. Hence, since the operator $\boldsymbol{Q}$ in Problem 6.1[b] can be written as $\boldsymbol{Q} = \boldsymbol{R} + \boldsymbol{S}$, it is therefore monotone and Lipschitzian.

Altogether, (6.83) is an instantiation of (6.3). Special cases of (6.83) can be found in [14, 24].

Example 6.14 In Problem 6.1, consider the following scenario: $K = \{1\}$, $\mathcal{G}_1$ is the standard Euclidean space $\mathbb{R}^M$, $r \in \mathcal{G}_1$, $g_1 = \iota_E$, where $E = r + [0, +\infty[^M$, $h_1 = 0$, and, for every $i \in I$, $\mathcal{H}_i$ is the standard Euclidean space $\mathbb{R}^{N_i}$, $\psi_i = 0$, and $\varphi_i = \iota_{C_i}$, where C_i is a nonempty closed convex subset of $\mathcal{H}_i$. Then, upon setting $N = \sum_{i \in I} N_i$, we obtain the model

$$\text{find } \overline{\boldsymbol{x}} \in \mathbb{R}^N \text{ such that } (\forall i \in I)\ \overline{x}_i \in \underset{\substack{x_i \in C_i \\ L_{1,i} x_i + \boldsymbol{L}_{1,\smallsetminus i} \overline{\boldsymbol{x}}_{\smallsetminus i} \in E}}{\operatorname{Argmin}} \boldsymbol{f}_i(x_i; \overline{\boldsymbol{x}}_{\smallsetminus i}) \tag{6.89}$$

investigated in [25].

Example 6.15 (minimization) Consider the setting of Problem 6.1 where [b] is replaced by

[b'] For every $i \in I$, $\boldsymbol{f}_i = \boldsymbol{f}$, where $\boldsymbol{f} \colon \boldsymbol{\mathcal{K}} \to \mathbb{R}$ is a differentiable convex function such that $\boldsymbol{Q} = \nabla \boldsymbol{f}$ is Lipschitzian,

and, in addition, the following is satisfied:

[e] For every $k \in K$, $g_k \colon \mathcal{G}_k \to \mathbb{R}$ is Gâteaux differentiable.

Then (6.3) reduces to the multivariate minimization problem

$$\underset{\boldsymbol{x} \in \boldsymbol{\mathcal{H}}}{\text{minimize}} \sum_{i \in I} \big(\varphi_i(x_i) + \psi_i(x_i)\big) + \boldsymbol{f}(\boldsymbol{M}\boldsymbol{x}) + \sum_{k \in K} (g_k + h_k)\Bigg(\sum_{j \in I} L_{k,j} x_j\Bigg). \tag{6.90}$$

The only asynchronous block-iterative algorithm we know of to solve (6.90) is [10, Algorithm 4.5], which is based on different decomposition principles. Special cases of (6.90) are found in partial differential equations [1], machine learning [6], and signal recovery [7], where they were solved using synchronous and non block-iterative methods.

References

[1] 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.

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

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

[4] A. Bensoussan, J.-L. Lions, and R. Temam, Sur les méthodes de décomposition, de décentralisation et de coordination, et applications, *Cahier IRIA*, no. 11, pp. 5–189, 1972.

[5] M. Bravo, D. Leslie, and P. Mertikopoulos, Bandit learning in concave N-person games, *Proc. Adv. Neural Inform. Process. Syst. Conf.*, vol. 31, pp. 5661–5671, 2018.

[6] 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.

[7] 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.

[8] 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.

[9] M. N. Bùi and P. L. Combettes, Warped proximal iterations for monotone inclusions, *J. Math. Anal. Appl.*, vol. 491, art. 124315, 21 pp., 2020.

[10] M. N. Bùi and P. L. Combettes, Multivariate monotone inclusions in saddle form. `https://arxiv.org/abs/2002.06135`

[11] G. Cohen, Nash equilibria: Gradient and decomposition algorithms, *Large Scale Syst.*, vol. 12, pp. 173–184, 1987.

[12] P. L. Combettes, Systems of structured monotone inclusions: Duality, algorithms, and applications, *SIAM J. Optim.*, vol. 23, pp. 2420–2447, 2013.

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

[14] P. L. Combettes and J.-C. Pesquet, Fixed point strategies in data science. `https://arxiv.org/abs/2008.02260`

[15] R. Cominetti, F. Facchinei, and J. B. Lasserre, *Modern Optimization Modelling Techniques*. Birkhäuser, Basel, 2012.

[16] F. Facchinei, A. Fischer, and V. Piccialli, On generalized Nash games and variational inequalities, *Oper. Res. Lett.*, vol. 35, pp. 159–164, 2007.

[17] A. von Heusinger and C. Kanzow, Relaxation methods for generalized Nash equilibrium problems with inexact line search, *J. Optim. Theory Appl.*, vol. 143, pp. 159–183, 2009.

[18] S. Hoda, A. Gilpin, J. Peña, and T. Sandholm, Smoothing techniques for computing Nash equilibria of sequential games, *Math. Oper. Res.*, vol. 35, pp. 494–512, 2010.

[19] A. Kannan and U. V. Shanbhag, Distributed computation of equilibria in monotone Nash games via iterative regularization techniques, *SIAM J. Optim.*, vol. 22, pp. 1177–1205, 2012.

[20] J. F. Nash, Jr., Equilibrium points in n-person games, *Proc. Nat. Acad. Sci. USA*, vol. 36, pp. 48–49, 1950.

[21] J. Nash, Non-cooperative games, *Ann. Math.*, vol. 54, pp. 286–295, 1951.

[22] 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.

[23] 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.

[24] 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.

[25] P. Yi and L. Pavel, An operator splitting approach for distributed generalized Nash equilibria computation, *Automatica*, vol. 102, pp. 111–121, 2019.

Chapter 7

A DECOMPOSITION METHOD FOR SOLVING MULTICOMMODITY NETWORK EQUILIBRIUM

7.1 Introduction and context

This chapter addresses question **(Q6)** of Chapter 1. We devise a flexible algorithm for solving the multicommodity network equilibrium model proposed by Rockafellar in [13].

This chapter presents the following article:

> M. N. Bùi, A decomposition method for solving multicommodity network equilibrium, submitted.

7.2 Article: A decomposition method for solving multicommodity network equilibrium

Abstract. We consider the numerical aspect of the multicommodity network equilibrium problem proposed by Rockafellar in 1995. Our method relies on the flexible monotone operator splitting framework recently proposed by Combettes and Eckstein.

7.2.1 Problem formulation

Rockafellar proposed in [13] the important multicommodity network equilibrium model (see (7.6) in Problem 7.2) and studied some of its properties. In the present paper, we devise a flexible numerical method for solving this problem based on the asynchronous block-iterative decomposition framework of [6].

The following notion of a network from [12, Section 1A] plays a central role in our problem.

Definition 7.1 A network consists of nonempty finite sets $\mathscr{N}$ and $\mathscr{A}$ — whose elements are called nodes and arcs, respectively — and a mapping $\vartheta\colon \mathscr{A} \to \mathscr{N} \times \mathscr{N} : j \mapsto (\vartheta_1(j), \vartheta_2(j))$ such that, for every $j \in \mathscr{A}$, $\vartheta_1(j) \neq \vartheta_2(j)$. We call $\vartheta_1(j)$ and $\vartheta_2(j)$ the initial node and the terminal node of arc j, respectively. In addition, we set

$$(\forall i \in \mathscr{N}) \quad \begin{cases} \mathscr{A}^+(i) = \big\{ j \in \mathscr{A} \mid \text{node } i \text{ is the initial node of arc } j \big\} \\ \mathscr{A}^-(i) = \big\{ j \in \mathscr{A} \mid \text{node } i \text{ is the terminal node of arc } j \big\}. \end{cases} \tag{7.1}$$

Recall that, given a Euclidean space $\mathcal{G}$ with scalar product $\langle \cdot \mid \cdot \rangle$, an operator $A\colon \mathcal{G} \to 2^{\mathcal{G}}$ is maximally monotone if

$$\big(\forall (x, x^*) \in \mathcal{G} \times \mathcal{G}\big) \quad (x, x^*) \in \operatorname{gra} A \quad \Leftrightarrow \quad \big[\, \big(\forall (y, y^*) \in \operatorname{gra} A\big) \;\; \langle x - y \mid x^* - y^* \rangle \geqslant 0 \, \big], \tag{7.2}$$

where $\operatorname{gra} A = \big\{ (x, x^*) \in \mathcal{G} \times \mathcal{G} \mid x^* \in Ax \big\}$ is the graph of A. (The reader is referred to [2] for background and complements on monotone operator theory and convex analysis.) The problem of interest is the following.

Problem 7.2 Under consideration is a network $(\mathscr{N}, \mathscr{A}, \vartheta)$, together with a nonempty finite set $\mathscr{C}$ of commodities transiting on the network. Equip $\mathcal{H} = \mathbb{R}^{\mathscr{C}}$ with the scalar product $((\xi_k)_{k \in \mathscr{C}}, (\eta_k)_{k \in \mathscr{C}}) \mapsto \sum_{k \in \mathscr{C}} \xi_k \eta_k$ and let us introduce the spaces

$$\begin{cases} \boldsymbol{\mathcal{X}} = \big\{ \boldsymbol{x} = (x_j)_{j \in \mathscr{A}} \mid (\forall j \in \mathscr{A}) \;\; x_j = (\xi_{j,k})_{k \in \mathscr{C}} \in \mathcal{H} \big\} \\ \boldsymbol{\mathcal{V}} = \big\{ \boldsymbol{v}^* = (v_i^*)_{i \in \mathscr{N}} \mid (\forall i \in \mathscr{N}) \;\; v_i^* = (\nu_{i,k}^*)_{k \in \mathscr{C}} \in \mathcal{H} \big\}. \end{cases} \tag{7.3}$$

An element $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ is called a flow on the network, where $\xi_{j,k}$ is the flux of commodity k on arc j. The divergence of a flow $\boldsymbol{x} \in \boldsymbol{\mathcal{X}}$ at node i is

$$\operatorname{div}_i \boldsymbol{x} = \sum_{j \in \mathscr{A}^+(i)} x_j - \sum_{j \in \mathscr{A}^-(i)} x_j. \tag{7.4}$$

We refer to an element $\boldsymbol{v}^* \in \boldsymbol{\mathcal{V}}$ as a potential on the network, where $\nu_{i,k}^*$ is the potential of commodity k at node i. Given $\boldsymbol{v}^* \in \boldsymbol{\mathcal{V}}$ and $j \in \mathscr{A}$, the tension (or potential difference) across arc j relative to the potential $\boldsymbol{v}^*$ is

$$\Delta_j \boldsymbol{v}^* = v_{\vartheta_2(j)}^* - v_{\vartheta_1(j)}^*. \tag{7.5}$$

For every $j \in \mathscr{A}$, the flow-tension relation on arc j is modeled by the sum $Q_j + R_j$ of maximally monotone operators $Q_j\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $R_j\colon \mathcal{H} \to 2^{\mathcal{H}}$. Further, for every $i \in \mathscr{N}$, the divergence-potential relation at node i is modeled by a maximally monotone operator $S_i\colon \mathcal{H} \to 2^{\mathcal{H}}$. The task is to

$$\text{find a flow } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{X}} \text{ and a potential } \overline{\boldsymbol{v}}^* \in \boldsymbol{\mathcal{V}} \text{ such that } \begin{cases} (\forall j \in \mathscr{A}) & \Delta_j \overline{\boldsymbol{v}}^* \in Q_j \overline{x}_j + R_j \overline{x}_j \\ (\forall i \in \mathscr{N}) & \operatorname{div}_i \overline{\boldsymbol{x}} \in S_i^{-1} \overline{v}_i^*, \end{cases} \tag{7.6}$$

under the assumption that (7.6) has a solution.

Remark 7.3 The pertinence of Problem 7.2 is demonstrated in [12, Chapter 8] and [13], where it is shown to capture formulations arising in areas such as traffic assignment, hydraulic networks, and price equilibrium.

7.2.2 A block-iterative decomposition method

Notation. Throughout, $\mathcal{G}$ is a Euclidean space. Let $A\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone and let $x \in \mathcal{G}$. Then, in terms of the variable $p \in \mathcal{G}$, the inclusion $x \in p + Ap$ has a unique solution, which is denoted by $J_A x$. The operator $J_A\colon \mathcal{G} \to \mathcal{G}\colon x \mapsto J_A x$ is called the resolvent of A.

Our algorithm (see (7.8) in Proposition 7.4) is derived from [6, Algorithm 12] and it thus inherits the following attractive features from the framework of [6]:

- No additional assumption, such as Lipschitz continuity or cocoercivity, is imposed on the underlying operators.
- Algorithm (7.8) achieves full splitting in the sense that the operators $(Q_j)_{j\in\mathscr{A}}$, $(R_j)_{j\in\mathscr{A}}$, and $(S_i)_{i\in\mathscr{N}}$ are activated independently via their resolvents.
- Algorithm (7.8) is block-iterative, that is, at iteration n, only blocks $(Q_j)_{j\in\mathscr{A}_n}$, $(R_j)_{j\in\mathscr{A}_n}$, and $(S_i)_{i\in\mathscr{N}_n}$ of operators need to be activated. To guarantee convergence of the iterates, the mild sweeping condition (7.7) needs to be fulfilled.

We shall denote elements in $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{V}}$ by bold letters, e.g., $\boldsymbol{q}_n = (q_{j,n})_{j\in\mathscr{A}}$ and $\boldsymbol{s}_n^* = (s_{i,n}^*)_{i\in\mathscr{N}}$.

Proposition 7.4 *Consider the setting of Problem 7.2. Let $T \in \mathbb{N}$, let $(\mathscr{A}_n)_{n\in\mathbb{N}}$ be nonempty subsets of $\mathscr{A}$, and let $(\mathscr{N}_n)_{n\in\mathbb{N}}$ be nonempty subsets of $\mathscr{N}$ such that $\mathscr{A}_0 = \mathscr{A}$, $\mathscr{N}_0 = \mathscr{N}$, and*

$$(\forall n \in \mathbb{N}) \quad \bigcup_{k=n}^{n+T} \mathscr{A}_k = \mathscr{A} \quad \text{and} \quad \bigcup_{k=n}^{n+T} \mathscr{N}_k = \mathscr{N}. \tag{7.7}$$

Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$ such that $\inf_{n\in\mathbb{N}} \lambda_n > 0$ and $\sup_{n\in\mathbb{N}} \lambda_n < 2$. Moreover, for every

$j \in \mathcal{A}$ and every $i \in \mathcal{N}$, let $(x_{j,0}, x^*_{j,0}, v^*_{i,0}) \in \mathcal{H}^3$ and $(\gamma_j, \mu_j, \sigma_i) \in \left]0, +\infty\right[^3$. Iterate

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\textit{for every } j \in \mathcal{A}_n \\
\left\lfloor
\begin{array}{l}
l^*_{j,n} = x^*_{j,n} - \Delta_j \boldsymbol{v}^*_n \\
q_{j,n} = J_{\gamma_j Q_j}(x_{j,n} - \gamma_j l^*_{j,n}) \\
q^*_{j,n} = \gamma_j^{-1}(x_{j,n} - q_{j,n}) - l^*_{j,n} \\
r_{j,n} = J_{\mu_j R_j}(x_{j,n} + \mu_j x^*_{j,n}) \\
r^*_{j,n} = x^*_{j,n} + \mu_j^{-1}(x_{j,n} - r_{j,n})
\end{array}
\right. \\
\textit{for every } j \in \mathcal{A} \smallsetminus \mathcal{A}_n \\
\left\lfloor
\begin{array}{l}
q_{j,n} = q_{j,n-1};\ q^*_{j,n} = q^*_{j,n-1};\ r_{j,n} = r_{j,n-1};\ r^*_{j,n} = r^*_{j,n-1}
\end{array}
\right. \\
\textit{for every } i \in \mathcal{N}_n \\
\left\lfloor
\begin{array}{l}
l_{i,n} = \operatorname{div}_i \boldsymbol{x}_n \\
s_{i,n} = J_{\sigma_i S_i}(l_{i,n} + \sigma_i v^*_{i,n}) \\
s^*_{i,n} = v^*_{i,n} + \sigma_i^{-1}(l_{i,n} - s_{i,n}) \\
t_{i,n} = s_{i,n} - \operatorname{div}_i \boldsymbol{q}_n
\end{array}
\right. \\
\textit{for every } i \in \mathcal{N} \smallsetminus \mathcal{N}_n \\
\left\lfloor
\begin{array}{l}
s_{i,n} = s_{i,n-1};\ s^*_{i,n} = s^*_{i,n-1} \\
t_{i,n} = s_{i,n} - \operatorname{div}_i \boldsymbol{q}_n
\end{array}
\right. \\
\textit{for every } j \in \mathcal{A} \\
\left\lfloor
\begin{array}{l}
t^*_{j,n} = q^*_{j,n} + r^*_{j,n} - \Delta_j \boldsymbol{s}^*_n \\
u_{j,n} = r_{j,n} - q_{j,n}
\end{array}
\right. \\
\tau_n = \sum_{j\in\mathcal{A}} \left(\|t^*_{j,n}\|^2 + \|u_{j,n}\|^2\right) + \sum_{i\in\mathcal{N}} \|t_{i,n}\|^2 \\
\textit{if } \tau_n > 0 \\
\left\lfloor
\begin{array}{l}
\pi_n = \sum_{j\in\mathcal{A}} \left(\langle x_{j,n} \mid t^*_{j,n}\rangle - \langle q_{j,n} \mid q^*_{j,n}\rangle + \langle u_{j,n} \mid x^*_{j,n}\rangle - \langle r_{j,n} \mid r^*_{j,n}\rangle\right) \\
\qquad\quad + \sum_{i\in\mathcal{N}} \left(\langle t_{i,n} \mid v^*_{i,n}\rangle - \langle s_{i,n} \mid s^*_{i,n}\rangle\right) \\
\theta_n = \lambda_n \max\{\pi_n, 0\}/\tau_n
\end{array}
\right. \\
\textit{else} \\
\left\lfloor
\begin{array}{l}
\theta_n = 0
\end{array}
\right. \\
\textit{for every } j \in \mathcal{A} \\
\left\lfloor
\begin{array}{l}
x_{j,n+1} = x_{j,n} - \theta_n t^*_{j,n} \\
x^*_{j,n+1} = x^*_{j,n+1} - \theta_n u_{j,n}
\end{array}
\right. \\
\textit{for every } i \in \mathcal{N} \\
\left\lfloor
\begin{array}{l}
v^*_{i,n+1} = v^*_{i,n} - \theta_n t_{i,n}.
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{7.8}
$$

Then $((x_{j,n})_{j\in\mathcal{A}}, (v^*_{i,n})_{i\in\mathcal{N}})_{n\in\mathbb{N}}$ converges to a solution to (7.6).

Proof. Let us consider the multivariate monotone inclusion problem

$$\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{X}},\ \overline{\boldsymbol{x}}^* \in \boldsymbol{\mathcal{X}},\ \text{and } \overline{\boldsymbol{v}}^* \in \boldsymbol{\mathcal{V}} \text{ such that } \begin{cases} (\forall j \in \mathscr{A})\ \ \Delta_j \overline{\boldsymbol{v}}^* - \overline{x}_j^* \in Q_j \overline{x}_j \ \text{ and } \ \overline{x}_j \in R_j^{-1} \overline{x}_j^* \\ (\forall i \in \mathscr{N})\ \ \mathrm{div}_i\, \overline{\boldsymbol{x}} \in S_i^{-1} \overline{v}_i^*. \end{cases} \tag{7.9}$$

Then

$$\begin{aligned} (\forall \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{X}})(\forall \overline{\boldsymbol{v}}^* \in \boldsymbol{\mathcal{V}}) \quad (\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}}^*) \ \text{solves (7.6)} & \\ \Leftrightarrow (\exists\, \overline{\boldsymbol{x}}^* \in \boldsymbol{\mathcal{X}}) \ & \begin{cases} (\forall j \in \mathscr{A})\ \ \Delta_j \overline{\boldsymbol{v}}^* \in Q_j \overline{x}_j + \overline{x}_j^* \ \text{ and } \ \overline{x}_j^* \in R_j \overline{x}_j \\ (\forall i \in \mathscr{N})\ \ \mathrm{div}_i\, \overline{\boldsymbol{x}} \in S_i^{-1} \overline{v}_i^* \end{cases} \\ \Leftrightarrow (\exists\, \overline{\boldsymbol{x}}^* \in \boldsymbol{\mathcal{X}}) \ & (\overline{\boldsymbol{x}}, \overline{\boldsymbol{x}}^*, \overline{\boldsymbol{v}}^*) \ \text{solves (7.9)}. \end{aligned} \tag{7.10}$$

Therefore, since (7.6) has a solution, so does (7.9). Next, define

$$(\forall i \in \mathscr{N})(\forall j \in \mathscr{A}) \quad \varepsilon_{i,j} = \begin{cases} 1, & \text{if node } i \text{ is the initial node of arc } j; \\ -1, & \text{if node } i \text{ is the terminal node of arc } j; \\ 0, & \text{otherwise.} \end{cases} \tag{7.11}$$

It results from (7.4) and (7.1) that

$$(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{X}})(\forall i \in \mathscr{N}) \quad \mathrm{div}_i\, \boldsymbol{x} = \sum_{j \in \mathscr{A}} \varepsilon_{i,j} x_j, \tag{7.12}$$

and from (7.5) that

$$(\forall \boldsymbol{v}^* \in \boldsymbol{\mathcal{V}})(\forall j \in \mathscr{A}) \quad \Delta_j \boldsymbol{v}^* = -\sum_{i \in \mathscr{N}} \varepsilon_{i,j} v_i^*. \tag{7.13}$$

We now verify that (7.9) is a special case of [6, Problem 1] with the setting

$$I = \mathscr{A}, \quad K = \mathscr{A} \cup \mathscr{N}, \quad \text{and} \quad (\forall j \in I)(\forall k \in K) \ \begin{cases} \mathcal{H}_j = \mathcal{G}_k = \mathcal{H} \\ A_j = Q_j \\ B_k = \begin{cases} R_k, & \text{if } \ k \in \mathscr{A}; \\ S_k, & \text{if } \ k \in \mathscr{N} \end{cases} \\ z_j^* = r_k = 0 \\ L_{k,j} = \begin{cases} \mathrm{Id}, & \text{if } \ k = j; \\ 0, & \text{if } \ k \in \mathscr{A} \ \text{ and } \ k \neq j; \\ \varepsilon_{k,j}\mathrm{Id}, & \text{if } \ k \in \mathscr{N}. \end{cases} \end{cases} \tag{7.14}$$

We deduce from (7.12) that

$$(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{X}})(\forall k \in K) \quad \sum_{j\in I} L_{k,j} x_j = \begin{cases} x_k, & \text{if } k \in \mathscr{A}; \\ \sum_{j\in I} \varepsilon_{k,j} x_j, & \text{if } k \in \mathscr{N} \end{cases}$$
$$= \begin{cases} x_k, & \text{if } k \in \mathscr{A}; \\ \mathrm{div}_k\, \boldsymbol{x}, & \text{if } k \in \mathscr{N}, \end{cases} \tag{7.15}$$

and from (7.13) that

$$(\forall \boldsymbol{x}^* \in \boldsymbol{\mathcal{X}})(\forall \boldsymbol{v}^* \in \boldsymbol{\mathcal{V}})(\forall j \in I) \quad \sum_{k\in\mathscr{A}} L_{k,j}^* x_k^* + \sum_{k\in\mathscr{N}} L_{k,j}^* v_k^* = x_j^* + \sum_{k\in\mathscr{N}} \varepsilon_{k,j} v_k^* = x_j^* - \Delta_j \boldsymbol{v}^*. \tag{7.16}$$

Hence, in the setting of (7.14), (7.9) is an instantiation of [6, Problem 1] and (7.8) is a realization of [6, Algorithm 12], where $(\forall n \in \mathbb{N})$ $I_n = \mathscr{A}_n$ and $K_n = \mathscr{A}_n \cup \mathscr{N}_n$. Thus, upon letting

$$(\forall n \in \mathbb{N}) \quad \boldsymbol{x}_n = (x_{j,n})_{j\in\mathscr{A}}, \quad \boldsymbol{x}_n^* = (x_{j,n}^*)_{j\in\mathscr{A}}, \quad \text{and} \quad \boldsymbol{v}_n^* = (v_{i,n}^*)_{i\in\mathscr{N}}, \tag{7.17}$$

we infer from [6, Theorem 13] that $(\boldsymbol{x}_n, \boldsymbol{x}_n^*, \boldsymbol{v}_n^*)_{n\in\mathbb{N}}$ converges to a solution $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{x}}^*, \overline{\boldsymbol{v}}^*)$ to (7.9). Consequently, (7.10) asserts that $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}}^*)$ solves (7.6). $\square$

Remark 7.5 Some comments are in order.

(i) One might be tempted to consider (7.6) as a special case of [6, Problem 1] with the setting

$$I=\mathscr{A},\quad K=\mathscr{N},\quad\text{and}\quad(\forall j\in I)(\forall k\in K)\;\begin{cases}\mathcal{H}_j=\mathcal{G}_k=\mathcal{H}\\ A_j=Q_j+R_j\\ B_k=S_k\\ z_j^*=r_k=0\\ L_{k,j}=\varepsilon_{k,j}\mathrm{Id},\end{cases}\tag{7.18}$$

where $(\varepsilon_{i,j})_{i\in\mathscr{N},j\in\mathscr{A}}$ are defined in (7.11), and then specialize [6, Algorithm 12] to (7.18). However, this approach necessitates the computation of the resolvents of the operators $(Q_j+R_j)_{j\in\mathscr{A}}$, which cannot be expressed in terms of the resolvents of $(Q_j)_{j\in\mathscr{A}}$ and $(R_j)_{j\in\mathscr{A}}$ in general (see Examples 7.6 and 7.7).

(ii) Algorithm (7.8) of Proposition 7.4 requires to evaluate the resolvents of the operators $(Q_j)_{j\in\mathscr{A}}$, $(R_j)_{j\in\mathscr{A}}$, and $(S_i)_{i\in\mathscr{N}}$. Illustrations of such calculations in some special cases of Problem 7.2 encountered in the literature are provided in Examples 7.6, 7.7 and 7.9–7.12.

(iii) Alternate algorithms [5,7,11] can also be used to solve (7.9) and, in turn, (7.6). Nevertheless, there are certain restrictions on the resulting algorithms. For example, the method of [5] must activate all the operators $(Q_j)_{j\in\mathscr{A}}$, $(R_j)_{j\in\mathscr{A}}$, and $(S_i)_{i\in\mathscr{N}}$ at every iteration, while the frameworks of [7,11] do not allow for deterministic selections of the blocks $(Q_j)_{j\in\mathscr{A}_n}$, $(R_j)_{j\in\mathscr{A}_n}$, and $(S_i)_{i\in\mathscr{N}_n}$. Finally, the algorithm resulted from [7] involves the inversion of a linear operator acting on $\mathbb{R}^{MN}$, where $N=\operatorname{card}\mathscr{C}$ and $M=2\operatorname{card}\mathscr{A}+\operatorname{card}\mathscr{N}$, which may not be favorable in large-scale problems, e.g., [8].

Notation. Before proceeding further, let us recall some basic notion of convex analysis (see [2] for details). Let $\varphi\colon\mathcal{G}\to\left]-\infty,+\infty\right]$ be proper, lower semicontinuous, and convex. The subdifferential of φ is the maximally monotone operator $\partial\varphi\colon\mathcal{G}\to2^{\mathcal{G}}\colon x\mapsto\big\{x^*\in\mathcal{G}\mid(\forall y\in\mathcal{G})\;\langle y-x\mid x^*\rangle+\varphi(x)\leqslant\varphi(y)\big\}$. For every $x\in\mathcal{G}$, the unique minimizer of $\varphi+(1/2)\|\cdot-x\|^2$ is denoted by $\operatorname{prox}_\varphi x$. Let C be a nonempty closed convex subset of $\mathcal{G}$. The indicator function of C is the proper lower semicontinuous convex function

$$\iota_C\colon\mathcal{G}\to[0,+\infty]:x\mapsto\begin{cases}0,&\text{if }x\in C;\\+\infty,&\text{otherwise,}\end{cases}\tag{7.19}$$

the normal cone operator of C is $N_C=\partial\iota_C$, and the projector onto C is $\operatorname{proj}_C=\operatorname{prox}_{\iota_C}$.

Example 7.6 (Separable multicommodity flows) Consider the setting of Problem 7.2 and suppose, in addition, that the following are satisfied:

[a] For every $j\in\mathscr{A}$, $\mathfrak{c}_j\colon\mathbb{R}\to2^{\mathbb{R}}$ is maximally monotone, C_j is a nonempty closed convex

subset of $\mathcal{H}$, and

$$Q_j\colon \mathcal{H}\to 2^{\mathcal{H}}\colon x_j=(\xi_{j,k})_{k\in\mathscr{C}}\mapsto \left(\mathfrak{c}_j\left(\sum_{k\in\mathscr{C}}\xi_{j,k}\right)\right)_{k\in\mathscr{C}} \quad\text{and}\quad R_j=N_{C_j}. \tag{7.20}$$

[b] For every $i\in\mathscr{N}$, $s_i\in\mathcal{H}$ and

$$S_i^{-1}\colon \mathcal{H}\to 2^{\mathcal{H}}\colon v_i^*\mapsto\{s_i\}. \tag{7.21}$$

Then (7.6) reduces to the separable multicommodity flow problem; see, e.g., [3, Section 8.3] and the references listed in [3, Section 8.9]. Take $j\in\mathscr{A}$, $i\in\mathscr{N}$, and $\gamma\in\,]0,+\infty[$. We have $J_{\gamma R_j}=\operatorname{proj}_{C_j}$ and $J_{\gamma S_i}=s_i$. To compute $J_{\gamma Q_j}$, define $L\colon\mathcal{H}\to\mathbb{R}\colon(\xi_k)_{k\in\mathscr{C}}\mapsto\sum_{k\in\mathscr{C}}\xi_k$ and set $N=\operatorname{card}\mathscr{C}$. Then $L^*\colon\mathbb{R}\to\mathcal{H}\colon\xi\mapsto(\xi)_{k\in\mathscr{C}}$ and, therefore, $L\circ L^*=N\mathrm{Id}$. At the same time, by (7.20), $Q_j=L^*\circ\mathfrak{c}_j\circ L_j$. Thus, we derive from [2, Proposition 23.25(iii)] that

$$\begin{aligned}(\forall x_j=(\xi_{j,k})_{k\in\mathscr{C}}\in\mathcal{H})\quad J_{\gamma Q_j}x_j=x_j+\frac{1}{N}\big(J_{N\gamma\mathfrak{c}_j}(Lx_j)-Lx_j\big)_{k\in\mathscr{C}}=(\xi_{j,k}+\eta)_{k\in\mathscr{C}},&\\ \text{where }\ \eta=\left(J_{N\gamma\mathfrak{c}_j}\left(\sum_{k\in\mathscr{C}}\xi_{j,k}\right)-\sum_{k\in\mathscr{C}}\xi_{j,k}\right)\Big/N.&\end{aligned} \tag{7.22}$$

Example 7.7 The separable multicommodity flow problem with arc capacity constraints (see, e.g., [3, Section 8.3]) is an instantiation of Example 7.6 with, for every $j\in\mathscr{A}$, $\mathfrak{c}_j=\partial(\phi_j+\iota_{\Omega_j})$, where $\phi_j\colon\mathbb{R}\to\,]-\infty,+\infty]$ is a proper lower semicontinuous convex function and Ω_j is a nonempty closed interval in $\mathbb{R}$ such that $\Omega_j\cap\operatorname{dom}\phi_j\neq\varnothing$. In this setting, it follows from [2, Example 23.3 and Proposition 24.47] that

$$(\forall j\in\mathscr{A})(\forall\gamma\in\,]0,+\infty[)\quad J_{\gamma\mathfrak{c}_j}=\operatorname{prox}_{\gamma(\phi_j+\iota_{\Omega_j})}=\operatorname{proj}_{\Omega_j}\circ\operatorname{prox}_{\gamma\phi_j}. \tag{7.23}$$

Remark 7.8 Consider the standard traffic assignment problem, that is, the special case of Example 7.6 where $(\forall j\in\mathscr{A})\ C_j=[0,+\infty[^{\mathscr{C}}$.

(i) In [1, Example 4.4], this problem was solved by an application of the forward-backward method [1, Theorem 2.8], where it is further assumed that, for every $j\in\mathscr{A}$, $\operatorname{dom}\mathfrak{c}_j=\mathbb{R}$ and $\mathfrak{c}_j$ is Lipschitzian. However, some common operators found in the literature of traffic assignment [4] do not fulfill this requirement; their resolvents are provided in Examples 7.9–7.12.

(ii) The method of [9], which is an application of the Douglas–Rachford algorithm [10], requires to compute the projectors onto polyhedral sets of the form

$$\left\{(\xi_j)_{j\in\mathscr{A}}\in[0,+\infty[^{\mathscr{A}}\;\middle|\;(\forall i\in\mathscr{N})\ \sum_{j\in\mathscr{A}}\varepsilon_{i,j}\xi_j=\delta_i\right\}, \tag{7.24}$$

where $(\varepsilon_{i,j})_{i\in\mathcal{N},j\in\mathcal{A}}$ are defined in (7.11). This results in solving a subproblem at every iteration because there is no closed-form expression for such projectors.

Example 7.9 (Bureau of Public Roads capacity operator) Let $(\alpha,\varrho,\theta,p)\in\,]0,+\infty[^4$ and define

$$\mathfrak{c}\colon\mathbb{R}\to\mathbb{R}\colon\xi\mapsto\begin{cases}\theta\Bigg(1+\alpha\bigg(\dfrac{\xi}{\varrho}\bigg)^{p}\Bigg), & \text{if }\ \xi\geqslant 0;\\ \theta, & \text{if }\ \xi<0.\end{cases}\tag{7.25}$$

In addition, let $\gamma\in\,]0,+\infty[$ and $\xi\in\mathbb{R}$. Then the following hold:

(i) Suppose that $\xi\geqslant\gamma\theta$. Then, in terms of the variable $s\in\mathbb{R}$, the equation

$$\frac{\alpha\gamma\theta}{\varrho^p}s^p+s+\gamma\theta-\xi=0\tag{7.26}$$

has a unique solution $\overline{s}$ and $J_{\gamma\mathfrak{c}}\xi=\overline{s}$.

(ii) Suppose that $\xi<\gamma\theta$. Then $J_{\gamma\mathfrak{c}}\xi=\xi-\gamma\theta$.

Example 7.10 (Logarithmic capacity operator) Let $\omega\in\,]0,+\infty[$, let $\theta\in[0,+\infty[$, and define

$$\mathfrak{c}\colon\mathbb{R}\to 2^{\mathbb{R}}\colon\xi\mapsto\begin{cases}\left\{\theta+\ln\dfrac{\omega}{\omega-\xi}\right\}, & \text{if }\ \xi<\omega;\\ \varnothing, & \text{if }\ \xi\geqslant\omega.\end{cases}\tag{7.27}$$

Then

$$\big(\forall\gamma\in\,]0,+\infty[\big)(\forall\xi\in\mathbb{R})\quad J_{\gamma\mathfrak{c}}\xi=\omega-\gamma\mathcal{W}\big(\omega\gamma^{-1}\exp(\theta-\xi/\gamma+\omega/\gamma)\big),\tag{7.28}$$

where $\mathcal{W}$ is the Lambert W-function, that is, the inverse of $[-1,+\infty[\;\to[-1/e,+\infty[\,:\xi\mapsto\xi\exp(\xi)$.

Example 7.11 (Traffic Research Corporation capacity operator) Let $(\alpha,\beta,\delta,\omega)\in\,]0,+\infty[^4$ and define

$$\mathfrak{c}\colon\mathbb{R}\to\mathbb{R}\colon\xi\mapsto\delta+\alpha(\xi-\omega)+\sqrt{\alpha^2(\xi-\omega)^2+\beta}.\tag{7.29}$$

Then

$$\begin{aligned}\big(\forall\gamma\in\,]0,+\infty[\big)(\forall\xi\in\mathbb{R})\quad & J_{\gamma\mathfrak{c}}\xi\\ &=\frac{-\sqrt{\gamma^2\alpha^2(\xi-\gamma\delta-\omega)^2+(2\gamma\alpha+1)\gamma^2\beta}+\gamma\alpha(\xi-\gamma\delta+\omega)+\xi-\gamma\delta}{2\gamma\alpha+1}.\end{aligned}\tag{7.30}$$

Example 7.12 Let $\alpha\in\,]1,+\infty[$, let $\theta\in\,]0,+\infty[$, let $p\in\,]0,+\infty[$, and define

$$\mathfrak{c}\colon\mathbb{R}\to\mathbb{R}\colon\xi\mapsto\theta\alpha^{p\xi}.\tag{7.31}$$

Then

$$\big(\forall\gamma\in\,]0,+\infty[\,\big)(\forall\xi\in\mathbb{R})\quad J_{\gamma\mathfrak{c}}\xi=\xi-\frac{\mathcal{W}\big(\gamma\theta\alpha^{p\xi}p\ln\alpha\big)}{p\ln\alpha}. \tag{7.32}$$

References

[1] 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.

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

[3] D. P. Bertsekas, *Network Optimization: Continuous and Discrete Models*. Athena Scientific, Belmont, MA, 1998.

[4] D. Branston, Link capacity functions: A review, *Transpn. Res.*, vol. 10, pp. 223–236, 1976.

[5] P. L. Combettes, Systems of structured monotone inclusions: Duality, algorithms, and applications, *SIAM J. Optim.*, vol. 23, pp. 2420–2447, 2013.

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

[7] 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.

[8] M. Florian and S. Nguyen, An application and validation of equilibrium trip assignment methods, *Trans. Sci.*, vol. 10, pp. 374–390, 1976.

[9] 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.

[10] 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.

[11] 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.

[12] R. T. Rockafellar, *Network Flows and Monotropic Optimization*. Wiley, New York, 1984.

[13] 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.

Chapter 8

BLOCK-ACTIVATED ALGORITHMS FOR MULTICOMPONENT FULLY NONSMOOTH MINIMIZATION

8.1 Introduction and context

Our focus in this chapter is question **(Q7)** of Chapter 1. The numerical experiments concern machine learning and image recovery.

This chapter presents the following article:

> M. N. Bùi, P. L. Combettes, and Z. C. Woodstock, Block-activated algorithms for multicomponent fully nonsmooth minimization, submitted.

8.2 Article: Block-activated algorithms for multicomponent fully nonsmooth minimization

Abstract. Under consideration are multicomponent minimization problems involving a separable nonsmooth convex function penalizing the components individually, and nonsmooth convex coupling terms penalizing linear mixtures of the components. We investigate block-activated proximal algorithms for solving such problems, i.e., algorithms which, at each iteration, need to use only a block of the underlying functions, as opposed to all of them as in standard methods. For smooth coupling functions, several block-activated algorithms exist and they are well understood. By contrast, in the fully nonsmooth case, few block-activated methods are available and little effort has been devoted to assessing them. Our goal is to shed more light on the implementation, the features, and the behavior of these algorithms, compare their merits, and provide machine learning and image recovery experiments illustrating their performance.

8.2.1 Introduction

The goal of many signal processing and machine learning tasks is to exploit the observed data and the prior knowledge to produce a solution that represents information of interest. In this process of extracting information from data, structured convex optimization has established itself as an effective modeling and algorithmic framework; see, for instance, [3, 5, 8, 14, 19]. In state-of-the-art applications, the sought solution is often a tuple of vectors which reside in different spaces [1, 2, 4, 6, 12, 13, 16, 17, 20]. The following multicomponent minimization formulation captures such problems. It consists of a separable term penalizing the components individually, and of coupling terms penalizing linear mixtures of the components.

Problem 8.1 Let $(\mathcal{H}_i)_{1\leqslant i\leqslant m}$ and $(\mathcal{G}_k)_{1\leqslant k\leqslant p}$ be Euclidean spaces. For every $i \in \{1,\ldots,m\}$ and every $k \in \{1,\ldots,p\}$, let $f_i\colon \mathcal{H}_i \to \left]-\infty,+\infty\right]$ and $g_k\colon \mathcal{G}_k \to \left]-\infty,+\infty\right]$ be proper lower semicontinuous convex functions, and let $L_{k,i}\colon \mathcal{H}_i \to \mathcal{G}_k$ be a linear operator. The objective is to

$$\underset{x_1\in\mathcal{H}_1,\ldots,x_m\in\mathcal{H}_m}{\text{minimize}} \quad \underbrace{\sum_{i=1}^{m} f_i(x_i)}_{\text{separable term}} + \sum_{k=1}^{p} \underbrace{g_k\left(\sum_{i=1}^{m} L_{k,i}x_i\right)}_{k\text{th coupling term}}. \tag{8.1}$$

To solve Problem 8.1 reliably without adding restrictions (for instance, smoothness or strong convexity of some functions involved in the model), we focus on flexible proximal algorithms that have the following features:

① **Nondifferentiability:** None of the functions $f_1,\ldots, f_m, g_1,\ldots,g_p$ needs to be differentiable.

② **Splitting:** The functions $f_1,\ldots,f_m,g_1,\ldots,g_p$ and the linear operators are activated separately.

③ **Block activation:** Only a block of the functions $f_1,\ldots, f_m, g_1,\ldots,g_p$ is activated at each iteration. This is in contrast with most splitting methods which require full activation, i.e., that all the functions be used at every iteration.

④ **Operator norms:** Bounds on the norms of the linear operators involved in Problem 8.1 are not assumed since they can be hard to compute.

⑤ **Convergence guarantee:** The algorithm produces a sequence which converges (possibly almost surely) to a solution to Problem 8.1.

In view of features ① and ②, the algorithms of interest should activate the functions $f_1,\ldots,f_m,g_1,\ldots,g_p$ via their proximity operators (even if some functions happened to be smooth, proximal activation is often preferable [6, 10]). The motivation for ② is that proximity operators of composite functions are typically not known explicitly. Feature ③ is geared towards current large-scale problems. In such scenarios, memory and computing power limitations make the execution of standard proximal splitting algorithms, which require activating all the functions at each iteration, inefficient or simply impossible. We must therefore turn our attention to algorithms which employ only blocks of functions $(f_i)_{i\in I_n}$ and $(g_k)_{k\in K_n}$ at iteration n. If the functions $(g_k)_{1\leqslant k\leqslant p}$ were all smooth, one could use block-activated versions of the forward-backward algorithm proposed in [15, 25] and the references therein; in particular, when $m = 1$, methods such as those of [11, 18, 23, 26] would be pertinent. As noted in [15, Remark 5.10(iv)], another candidate of interest could be the randomly block-activated algorithm of [15, Section 5.2], which leads to block-activated versions of several primal-dual methods (see [24] for detailed developments and [7] for an inertial version when $m = 1$). However, this approach violates ④ as it imposes bounds on the proximal scaling parameters which depend on the norms of the linear operators. Finally, ⑤ rules out methods that guarantee merely minimizing sequences or ergodic convergence.

To the best of our knowledge, there are two primary methods that fulfill ①–⑤:

- Algorithm 8.2: The stochastic primal-dual Douglas–Rachford algorithm of [15].
- Algorithm 8.3: The deterministic primal-dual projective splitting algorithm of [9].

In the case of smooth coupling functions $(g_k)_{1\leqslant k\leqslant p}$, in (8.1), extensive numerical experience has been accumulated to understand the behavior of block-activated methods, especially in the case of stochastic gradient methods. By contrast, to date, very few numerical experiments with the recent, fully nonsmooth Algorithms 8.2 and 8.3 have been conducted and no comparison of their merits and performance has been undertaken. Thus far, Algorithm 8.2 has been employed only in the context of machine learning (see also the variant of 8.2 in [6] for partially smooth problems). On the other hand, Algorithm 8.3 has been used in image recovery in [10], but only in full activation mode, and in feature selection in [22], but with $m = 1$.

Contributions and novelty: This paper investigates for the first time the use of block-activated methods in fully nonsmooth multivariate minimization problems. It sheds more light on the implementation, the features, and the behavior of Algorithms 8.2 and 8.3, compares their merits, and provides experiments illustrating their performance.

Outline: Algorithms 8.2 and 8.3 are presented in Section 8.2.2. In Section 8.2.3, we analyze and compare their features, implementation, and asymptotic properties. This investigation is complemented in Section 8.2.4 by numerical experiments in the context of machine learning and image recovery.

8.2.2 Block-activated algorithms for Problem 8.1

The subdifferential, the conjugate, and the proximity operator of a proper lower semicontinuous convex function $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ are denoted by ∂f, f^*, and prox_f, respectively. Let us consider the setting of Problem 8.1 and let us set $\boldsymbol{\mathcal{H}} = \mathcal{H}_1 \times \cdots \times \mathcal{H}_m$ and $\boldsymbol{\mathcal{G}} = \mathcal{G}_1 \times \cdots \times \mathcal{G}_p$. A generic element in $\mathcal{H}$ is denoted by $\boldsymbol{x} = (x_i)_{1\leqslant i\leqslant m}$ and a generic element in $\mathcal{G}$ by $\boldsymbol{y} = (y_k)_{1\leqslant k\leqslant p}$.

As discussed in Section 8.2.1, two primary algorithms fulfill requirements ①–⑤. Both operate in the product space $\mathcal{H} \times \mathcal{G}$. The first one employs random activation of the blocks. To present it, let us introduce

$$\begin{cases} \boldsymbol{L}\colon \mathcal{H} \to \mathcal{G}\colon \boldsymbol{x} \mapsto \left(\displaystyle\sum_{i=1}^{m} L_{1,i}x_i, \ldots, \sum_{i=1}^{m} L_{p,i}x_i\right) \\ \boldsymbol{V} = \big\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{H} \times \mathcal{G} \mid \boldsymbol{y} = \boldsymbol{L}\boldsymbol{x}\big\} \\ \boldsymbol{F}\colon \mathcal{H} \times \mathcal{G} \to \left]-\infty, +\infty\right] \\ \qquad (\boldsymbol{x}, \boldsymbol{y}) \mapsto \sum_{i=1}^{m} f_i(x_i) + \sum_{k=1}^{p} g_k(y_k). \end{cases} \tag{8.2}$$

Then (8.1) is equivalent to

$$\underset{(\boldsymbol{x},\boldsymbol{y})\in\boldsymbol{V}}{\text{minimize}}\;\; \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{y}). \tag{8.3}$$

The idea is then to apply the Douglas–Rachford algorithm in block form to this problem [15]. To this end, we need $\text{prox}_{\boldsymbol{F}}$ and $\text{prox}_{\iota_{\boldsymbol{V}}} = \text{proj}_{\boldsymbol{V}}$. Note that $\text{prox}_{\boldsymbol{F}}\colon (\boldsymbol{x}, \boldsymbol{y}) \mapsto ((\text{prox}_{f_i} x_i)_{1\leqslant i\leqslant m}, (\text{prox}_{g_k} y_k)_{1\leqslant k\leqslant p})$. Now let $\boldsymbol{x} \in \mathcal{H}$ and $\boldsymbol{y} \in \mathcal{G}$, and set $\boldsymbol{t} = (\mathbf{Id} + \boldsymbol{L}^*\boldsymbol{L})^{-1}(\boldsymbol{x} + \boldsymbol{L}^*\boldsymbol{y})$ and $\boldsymbol{s} = (\mathbf{Id} + \boldsymbol{L}\boldsymbol{L}^*)^{-1}(\boldsymbol{L}\boldsymbol{x} - \boldsymbol{y})$. Then

$$\text{proj}_{\boldsymbol{V}}(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{t}, \boldsymbol{L}\boldsymbol{t}) = (\boldsymbol{x} - \boldsymbol{L}^*\boldsymbol{s}, \boldsymbol{y} + \boldsymbol{s}), \tag{8.4}$$

and we write it coordinate-wise as

$$\text{proj}_{\boldsymbol{V}}(\boldsymbol{x}, \boldsymbol{y}) = \big(Q_1(\boldsymbol{x}, \boldsymbol{y}), \ldots, Q_{m+p}(\boldsymbol{x}, \boldsymbol{y})\big). \tag{8.5}$$

Thus, given $\gamma \in \left]0, +\infty\right[$, $\boldsymbol{z}_0 \in \mathcal{H}$, and $\boldsymbol{y}_0 \in \mathcal{G}$, the standard Douglas–Rachford algorithm for (8.3) is

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \lambda_n \in \left]0, 2\right[\\ \text{for every } i \in \{1, \ldots, m\} \\ \left\lfloor \begin{array}{l} x_{i,n+1} = Q_i(\boldsymbol{z}_n, \boldsymbol{y}_n) \\ z_{i,n+1} = z_{i,n} + \lambda_n\big(\text{prox}_{\gamma f_i}(2x_{i,n+1} - z_{i,n}) - x_{i,n+1}\big) \end{array} \right. \\ \text{for every } k \in \{1, \ldots, p\} \\ \left\lfloor \begin{array}{l} w_{k,n+1} = Q_{m+k}(\boldsymbol{z}_n, \boldsymbol{y}_n) \\ y_{k,n+1} = y_{k,n} + \lambda_n\big(\text{prox}_{\gamma g_k}(2w_{k,n+1} - y_{k,n}) - w_{k,n+1}\big). \end{array} \right. \end{array} \right. \end{array} \tag{8.6}$$

The block-activated version of this algorithm is as follows.

Algorithm 8.2 ([15]) Let $\gamma \in \left]0, +\infty\right[$, let $\boldsymbol{x}_0$ and $\boldsymbol{z}_0$ be $\mathcal{H}$-valued random variables (r.v.), let

$\boldsymbol{y}_0$ and $\boldsymbol{w}_0$ be $\boldsymbol{\mathcal{G}}$-valued r.v. Iterate

$$
\begin{array}{l}
\text{for } j = 1, \dots, m+p \\
\lfloor \text{compute } Q_j \text{ as in (8.4)–(8.5)} \\
\text{for } n = 0, 1, \dots \\
\left\lfloor
\begin{array}{l}
\lambda_n \in \left]0, 2\right[\\
\text{select randomly } \varnothing \neq I_n \subset \{1, \dots, m\} \text{ and } \varnothing \neq K_n \subset \{1, \dots, p\} \\
\text{for every } i \in I_n \\
\left\lfloor
\begin{array}{l}
x_{i,n+1} = Q_i(\boldsymbol{z}_n, \boldsymbol{y}_n) \\
z_{i,n+1} = z_{i,n} + \lambda_n \big(\mathrm{prox}_{\gamma f_i}(2x_{i,n+1} - z_{i,n}) - x_{i,n+1}\big)
\end{array}
\right. \\
\text{for every } i \in \{1, \dots, m\} \smallsetminus I_n \\
\left\lfloor
(x_{i,n+1}, z_{i,n+1}) = (x_{i,n}, z_{i,n})
\right. \\
\text{for every } k \in K_n \\
\left\lfloor
\begin{array}{l}
w_{k,n+1} = Q_{m+k}(\boldsymbol{z}_n, \boldsymbol{y}_n) \\
y_{k,n+1} = y_{k,n} + \lambda_n \big(\mathrm{prox}_{\gamma g_k}(2w_{k,n+1} - y_{k,n}) - w_{k,n+1}\big)
\end{array}
\right. \\
\text{for every } k \in \{1, \dots, p\} \smallsetminus K_n \\
\left\lfloor
(w_{k,n+1}, y_{k,n+1}) = (w_{k,n}, y_{k,n}).
\right.
\end{array}
\right.
\end{array}
$$

The second algorithm operates by projecting onto hyperplanes which separate the current iterate from the set **Z** of Kuhn–Tucker points of Problem 8.1, i.e., the points $\widetilde{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}}$ and $\widetilde{\boldsymbol{v}}^* \in \boldsymbol{\mathcal{G}}$ such that

$$
\begin{cases}
(\forall i \in \{1, \dots, m\}) & -\sum_{k=1}^{p} L_{k,i}^* \widetilde{v}_k^* \in \partial f_i(\widetilde{x}_i) \\
(\forall k \in \{1, \dots, p\}) & \sum_{i=1}^{m} L_{k,i} \widetilde{x}_i \in \partial g_k^*(\widetilde{v}_k^*).
\end{cases}
\tag{8.7}
$$

This process is explained in Fig. 8.1.

Algorithm 8.3 ([9]) Set $I_0 = \{1, \dots, m\}$ and $K_0 = \{1, \dots, p\}$. For every $i \in I_0$ and every

$k \in K_0$, let $\{\gamma_i, \mu_k\} \subset \left]0,+\infty\right[$, $x_{i,0} \in \mathcal{H}_i$, and $v_{k,0}^* \in \mathcal{G}_k$. Iterate

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
\lambda_n \in \left]0,2\right[\\
\text{if } n > 0\\
\left\lfloor
\begin{array}{l}
\text{select } \varnothing \neq I_n \subset I_0 \text{ and } \varnothing \neq K_n \subset K_0
\end{array}
\right.\\
\text{for every } i \in I_n\\
\left\lfloor
\begin{array}{l}
x_{i,n}^* = x_{i,n} - \gamma_i \sum_{k=1}^{p} L_{k,i}^* v_{k,n}^*\\
a_{i,n} = \operatorname{prox}_{\gamma_i f_i} x_{i,n}^*\\
a_{i,n}^* = \gamma_i^{-1}(x_{i,n}^* - a_{i,n})
\end{array}
\right.\\
\text{for every } i \in I_0 \smallsetminus I_n\\
\left\lfloor
\begin{array}{l}
(a_{i,n}, a_{i,n}^*) = (a_{i,n-1}, a_{i,n-1}^*)
\end{array}
\right.\\
\text{for every } k \in K_n\\
\left\lfloor
\begin{array}{l}
y_{k,n}^* = \mu_k v_{k,n}^* + \sum_{i=1}^{m} L_{k,i} x_{i,n}\\
b_{k,n} = \operatorname{prox}_{\mu_k g_k} y_{k,n}^*\\
b_{k,n}^* = \mu_k^{-1}(y_{k,n}^* - b_{k,n})\\
t_{k,n} = b_{k,n} - \sum_{i=1}^{m} L_{k,i} a_{i,n}
\end{array}
\right.\\
\text{for every } k \in K_0 \smallsetminus K_n\\
\left\lfloor
\begin{array}{l}
(b_{k,n}, b_{k,n}^*) = (b_{k,n-1}, b_{k,n-1}^*)\\
t_{k,n} = b_{k,n} - \sum_{i=1}^{m} L_{k,i} a_{i,n}
\end{array}
\right.\\
\text{for every } i \in I_0\\
\left\lfloor
\begin{array}{l}
t_{i,n}^* = a_{i,n}^* + \sum_{k=1}^{p} L_{k,i}^* b_{k,n}^*
\end{array}
\right.\\
\tau_n = \sum_{i=1}^{m} \|t_{i,n}^*\|^2 + \sum_{k=1}^{p} \|t_{k,n}\|^2\\
\text{if } \tau_n > 0\\
\left\lfloor
\begin{array}{l}
\pi_n = \sum_{i=1}^{m} \big(\langle x_{i,n} \mid t_{i,n}^*\rangle - \langle a_{i,n} \mid a_{i,n}^*\rangle\big) + \sum_{k=1}^{p} \big(\big\langle t_{k,n} \mid v_{k,n}^*\big\rangle - \big\langle b_{k,n} \mid b_{k,n}^*\big\rangle\big)
\end{array}
\right.\\
\text{if } \tau_n > 0 \text{ and } \pi_n > 0\\
\left\lfloor
\begin{array}{l}
\theta_n = \lambda_n \pi_n / \tau_n\\
\text{for every } i \in I_0\\
\left\lfloor
\begin{array}{l}
x_{i,n+1} = x_{i,n} - \theta_n t_{i,n}^*
\end{array}
\right.\\
\text{for every } k \in K_0\\
\left\lfloor
\begin{array}{l}
v_{k,n+1}^* = v_{k,n}^* - \theta_n t_{k,n}
\end{array}
\right.
\end{array}
\right.\\
\text{else}\\
\left\lfloor
\begin{array}{l}
\text{for every } i \in I_0\\
\left\lfloor
\begin{array}{l}
x_{i,n+1} = x_{i,n}
\end{array}
\right.\\
\text{for every } k \in K_0\\
\left\lfloor
\begin{array}{l}
v_{k,n+1}^* = v_{k,n}^*.
\end{array}
\right.
\end{array}
\right.
\end{array}
\right.
\end{array}
$$

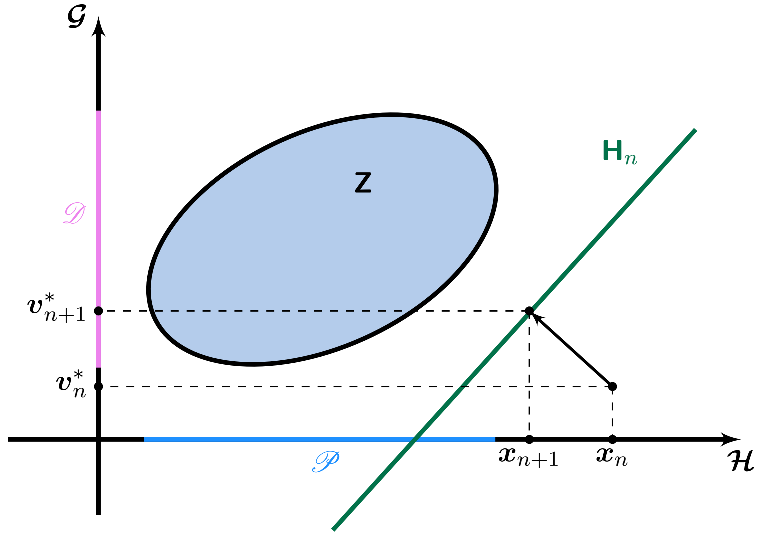

Figure 8.1 Let $\mathscr{P}$ be the set of solutions to Problem 8.1 and let $\mathscr{D}$ be the set of solutions to its dual. Then the Kuhn–Tucker set Z is a subset of $\mathscr{P} \times \mathscr{D}$. At iteration n, the proximity operators of blocks of functions $(f_i)_{i\in I_n}$ and $(g_k)_{k\in K_n}$ are used to construct a hyperplane H_n separating the current primal-dual iterate $(\boldsymbol{x}_n, \boldsymbol{v}_n^*)$ from Z, and the update $(\boldsymbol{x}_{n+1}, \boldsymbol{v}_{n+1}^*)$ is obtained as its projection onto H_n [9].

8.2.3 Asymptotic behavior and comparisons

Let us first state the convergence results available for Algorithms 8.2 and 8.3. We make the standing assumption that $\mathsf{Z} \neq \varnothing$ (see (8.7)), which implies that the solution set $\mathscr{P}$ of Problem 8.1 is nonempty.

Theorem 8.4 ([15]) *In the setting of Algorithm 8.2, define, for every $n \in \mathbb{N}$ and every $j \in \{1, \ldots, m+p\}$,*

$$\varepsilon_{j,n} = \begin{cases} 1, & \text{if } j \in I_n \text{ or } j - m \in K_n; \\ 0, & \text{otherwise}. \end{cases} \tag{8.8}$$

Suppose that the following hold:

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

(ii) *The r.v. $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ are identically distributed.*

(iii) *For every $n \in \mathbb{N}$, the r.v. $\boldsymbol{\varepsilon}_n$ and $(\boldsymbol{z}_j, \boldsymbol{y}_j)_{0\leqslant j\leqslant n}$ are mutually independent.*

(iv) $(\forall j \in \{1, \ldots, m+p\})\ \mathsf{Prob}[\varepsilon_{j,0} = 1] > 0$.

Then $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges almost surely to a $\mathscr{P}$-valued r.v.

Theorem 8.5 ([9]) *In the setting of Algorithm 8.3, suppose that the following hold:*

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

(ii) *There exists $T \in \mathbb{N}$ such that, for every $n \in \mathbb{N}$, $\bigcup_{j=n}^{n+T} I_j = \{1, \ldots, m\}$ and $\bigcup_{j=n}^{n+T} K_j = \{1, \ldots, p\}$.*

Then $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges to a point in $\mathscr{P}$.

Let us compare Algorithms 8.2 and 8.3.

a/ **Auxiliary tasks:** 8.2 requires the construction and storage of the operators $(Q_j)_{1\leqslant j\leqslant m+p}$ of (8.4)–(8.5), which can be quite demanding as they involve inversion of a linear operator acting on the product space $\boldsymbol{\mathcal{H}}$ or $\boldsymbol{\mathcal{G}}$. By contrast, 8.3 does not require such tasks.

b/ **Proximity operators:** Both algorithms are block-activated: only the blocks of functions $(f_i)_{i\in I_n}$ and $(g_k)_{k\in K_n}$ need to be activated at iteration n.

c/ **Linear operators:** In 8.2, the operators $(Q_i)_{i\in I_n}$ and $(Q_{m+k})_{k\in K_n}$ selected at iteration n are evaluated at $(z_{1,n},\ldots,z_{m,n},y_{1,n},\ldots,y_{p,n})\in\boldsymbol{\mathcal{H}}\times\boldsymbol{\mathcal{G}}$. On the other hand, 8.3 activates the local operators $L_{k,i}\colon\mathcal{H}_i\to\mathcal{G}_k$ and $L_{k,i}^*\colon\mathcal{G}_k\to\mathcal{H}_i$ once or twice, depending on whether they are selected. For instance, if we set $N=\dim\boldsymbol{\mathcal{H}}$ and $M=\dim\boldsymbol{\mathcal{G}}$ and if all the linear operators are implemented in matrix form, then the corresponding load per iteration in full activation mode of 8.2 is $\mathcal{O}((M+N)^2)$ versus $\mathcal{O}(MN)$ in 8.3.

d/ **Activation scheme:** As 8.2 selects the blocks randomly, the user does not have complete control of the computational load of an iteration, whereas the load of 8.3 is more predictable because of its deterministic activation scheme.

e/ **Parameters:** A single scale parameter γ is used in 8.2, while 8.3 allows the proximity operators to have their own scale parameters $(\gamma_1,\ldots,\gamma_m,\mu_1,\ldots,\mu_p)$. This gives 8.3 more flexibility, but more effort may be needed *a priori* to find efficient parameters. Further, in both algorithms, there is no restriction on the parameter values.

f/ **Convergence:** 8.3 guarantees sure convergence under the mild sweeping condition (ii) in Theorem 8.5, while 8.2 guarantees only almost sure convergence.

g/ **Other features:** Although this point is omitted for brevity, unlike 8.2, 8.3 can be executed asynchronously with iteration-dependent scale parameters [9].

8.2.4 Numerical experiments

We present two experiments which are reflective of our numerical investigations in solving various problems using Algorithms 8.2 and 8.3. The main objective is to illustrate the block processing ability of the algorithms (when implemented with full activation, i.e., $I_n=I_0$ and $K_n=K_0$, Algorithm 8.3 was already shown in [10] to be quite competitive compared to existing methods).

8.2.4.1 Experiment 1: Group-sparse binary classification

We revisit the classification problem of [12], which is based on the latent group lasso formulation in machine learning [21]. Let $\{G_1,\ldots,G_m\}$ be a covering of $\{1,\ldots,d\}$ and define

$$X=\big\{(x_1,\ldots,x_m)\mid x_i\in\mathbb{R}^d,\ \text{support}(x_i)\subset G_i\big\}. \tag{8.9}$$

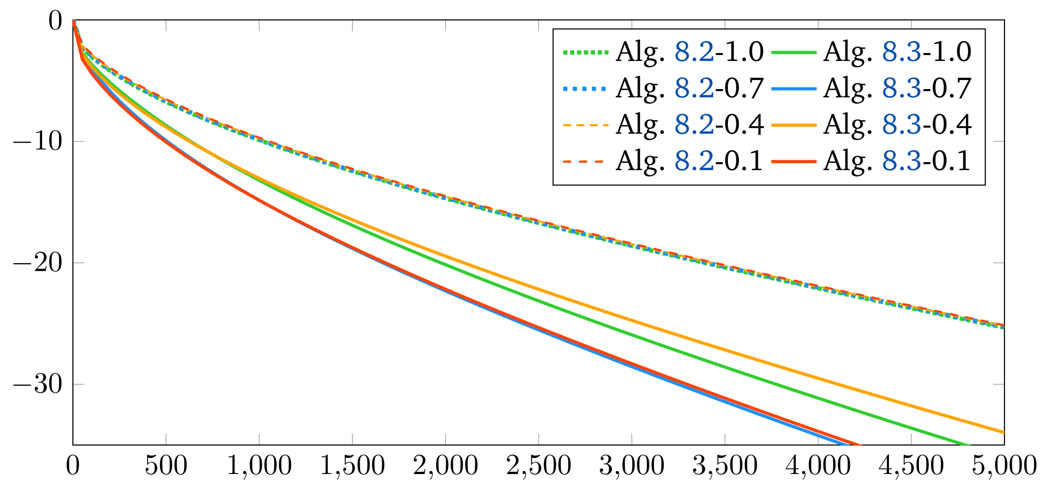

Figure 8.2 Normalized error $20\log_{10}(\|\boldsymbol{x}_n-\boldsymbol{x}_\infty\|/\|\boldsymbol{x}_0-\boldsymbol{x}_\infty\|)$ (dB), averaged over 20 runs, versus epoch count in Experiment 1. The variations around the averages were not significant. The computational load per epoch for both algorithms is comparable.

The sought vector is $\widetilde{y}=\sum_{i=1}^m \widetilde{x}_i$, where $(\widetilde{x}_1,\ldots,\widetilde{x}_m)$ solves

$$\underset{(x_1,\ldots,x_m)\in X}{\text{minimize}} \quad \sum_{i=1}^m \|x_i\|_2+\sum_{k=1}^p g_k\left(\sum_{i=1}^m \langle x_i \mid u_k\rangle\right), \tag{8.10}$$

with $u_k\in\mathbb{R}^d$ and $g_k\colon\mathbb{R}\to\mathbb{R}\colon\xi\mapsto 10\max\{0,1-\beta_k\xi\}$, where $\beta_k=\omega_k\,\mathrm{sign}(\langle\overline{y}\mid u_k\rangle)$ is the kth measurement of the true vector $\overline{y}\in\mathbb{R}^d$ $(d=10000)$ and $\omega_k\in\{-1,1\}$ induces 25% classification error. There are $p=1000$ measurements and the goal is to reconstruct the group-sparse vector $\overline{y}$. There are $m=1429$ groups. For every $i\in\{1,\ldots,m-1\}$, each G_i has 10 consecutive integers and an overlap with G_{i+1} of length 3. We obtain an instance of (8.1), where $\mathcal{H}_i=\mathbb{R}^{10}$, $f_i=\|\cdot\|_2$, and $L_{k,i}=\langle\cdot\mid u_k|_{G_i}\rangle$. The auxiliary tasks for Algorithm 8.2 (see a/) are negligible [12]. For each $\alpha\in\{0.1,0.4,0.7,1.0\}$, at iteration $n\in\mathbb{N}$, I_n has $\lceil\alpha m\rceil$ elements and the proximity operators of the scalar functions $(g_k)_{1\leqslant k\leqslant p}$ are all used, i.e., $K_n=\{1,\ldots,p\}$. We display in Fig. 8.2 the normalized error versus the epoch, that is, the cumulative number of activated blocks in $\{1,\ldots,m\}$ divided by m.

8.2.4.2 Experiment 2: Image recovery

We revisit the image interpolation problem of [10, Section 4.3]. The objective is to recover the image $\overline{x}\in C=[0,255]^N$ $(N=96^2)$ of Fig. 8.3(a), given a noisy masked observation $b=M\overline{x}+w_1\in\mathbb{R}^N$ and a noisy blurred observation $c=H\overline{x}+w_2\in\mathbb{R}^N$. Here, M masks all but $q=39$ rows $(x^{(r_k)})_{1\leqslant k\leqslant q}$ of an image x, and H is a nonstationary blurring operator, while w_1 and w_2 yield signal-to-noise ratios of 28.5 dB and 27.8 dB, respectively. Since H is sizable, we split it into $s=384$ subblocks: for every $k\in\{1,\ldots,s\}$, $H_k\in\mathbb{R}^{24\times N}$ and the corresponding block of c is denoted c_k. The goal is to

$$\underset{x\in C}{\text{minimize}}\;\; \|Dx\|_{1,2} + 10\sum_{k=1}^{q}\left\|x^{(r_k)} - b^{(r_k)}\right\|_2 + 5\sum_{k=1}^{s}\|H_k x - c_k\|_2^2, \tag{8.11}$$

where $D\colon \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N$ models finite differences and $\|\cdot\|_{1,2}\colon (y_1, y_2) \mapsto \sum_{j=1}^{N} \|(\eta_{1,j}, \eta_{2,j})\|_2$. Thus, (8.11) is an instance of Problem 8.1, where $m = 1$; $p = q+s+1$; for every $k \in \{1,\ldots,q\}$, $L_{k,1}\colon \mathbb{R}^N \to \mathbb{R}^{\sqrt{N}}\colon x \mapsto x^{(r_k)}$ and $g_k\colon y_k \mapsto 10\|y_k - b^{(r_k)}\|_2$; for every $k \in \{q+1,\ldots,q+s\}$, $L_{k,1} = H_{k-q}$, $g_k = 5\|\cdot - c_k\|_2^2$, and $g_p = \|\cdot\|_{1,2}$; $L_{p,1} = D$; $f_1\colon x \mapsto 0$ if $x \in C$; $+\infty$ if $x \notin C$. At iteration n, K_n has $\lceil \alpha p \rceil$ elements, where $\alpha \in \{0.1, 0.4, 0.7, 1.0\}$. The results are shown in Figs. 8.3–8.4, where the epoch is the cumulative number of activated blocks in $\{1,\ldots,p\}$ divided by p.

8.2.4.3 Discussion

Our first finding is that, for both Algorithms 8.2 and 8.3, even when full activation is computationally possible, it may not be the best strategy (see Figs. 8.2 and 8.4). Second, a/–g/ and our experiments suggest that 8.3 is preferable to 8.2. Let us add that, in general, 8.2 does not scale as well as 8.3. For instance, in Experiment 2, if the image size scales up, 8.3 can still operate since it involves only individual applications of the local $L_{k,i}$ operators, while 8.2 becomes unmanageable because of the size of the Q_j operators (see a/ and [6]).

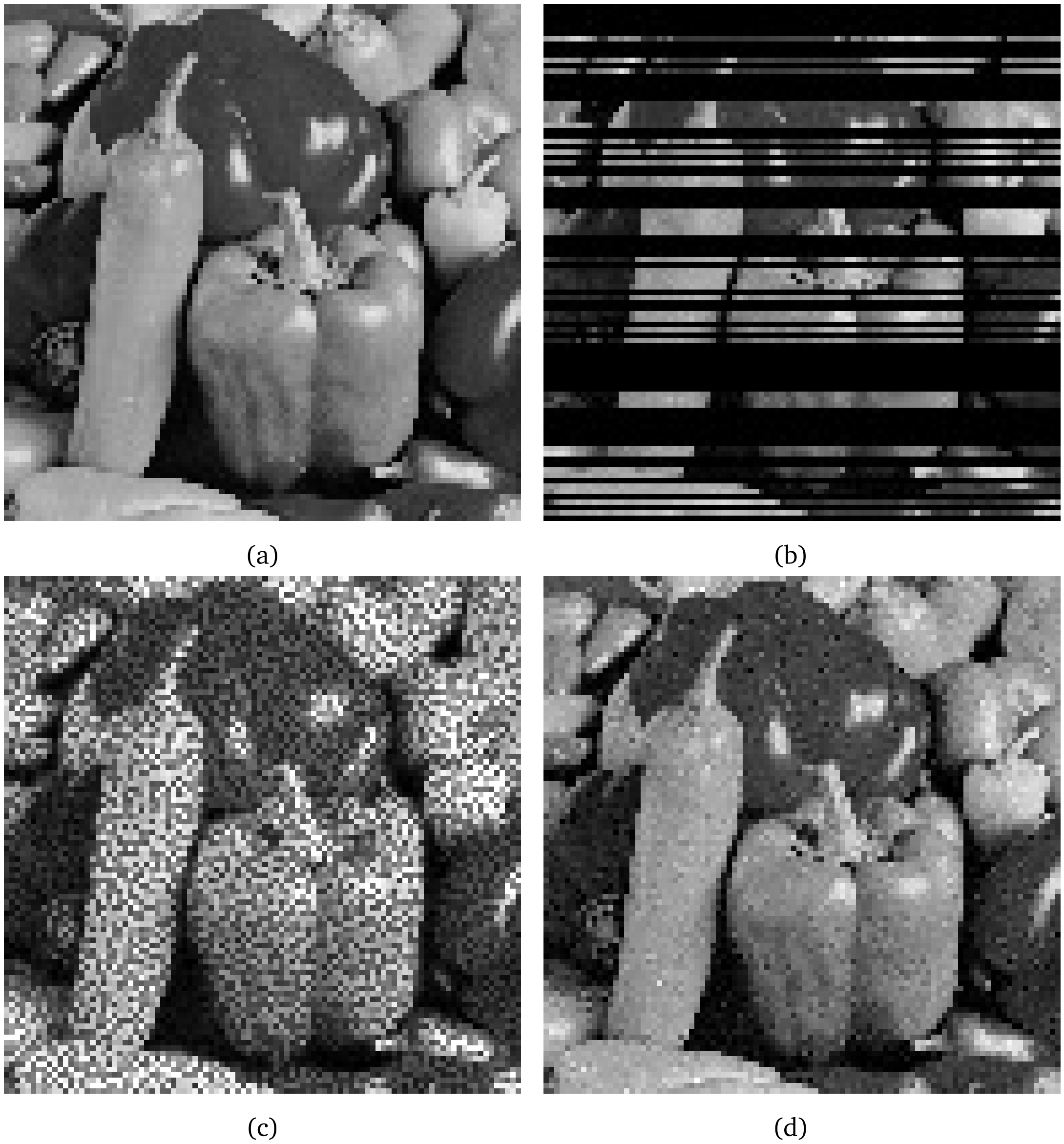

Figure 8.3 Experiment 2: (a) Original $\overline{x}$. (b) Observation b. (c) Observation c. (d) Recovery (all recoveries were visually indistinguishable).

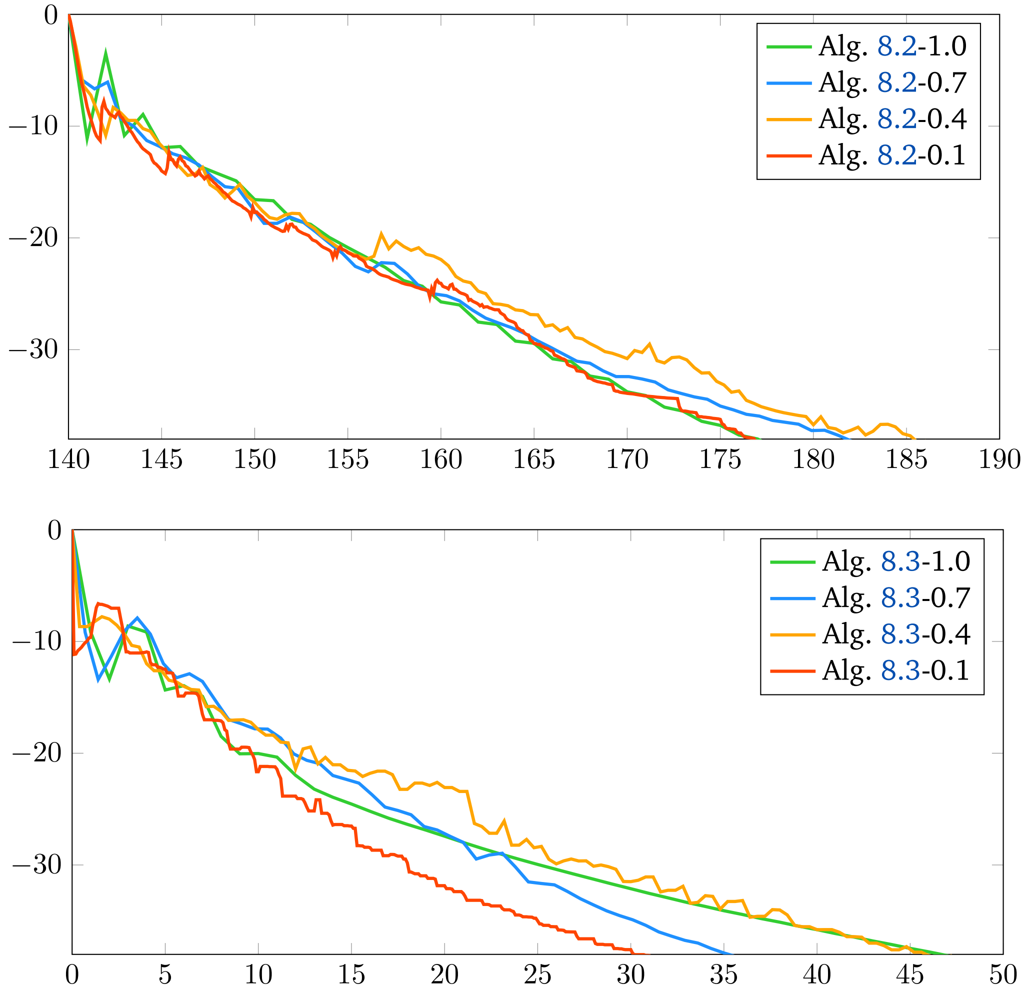

Figure 8.4 Normalized error $20\log_{10}(\|\boldsymbol{x}_n - \boldsymbol{x}_\infty\|/\|\boldsymbol{x}_0 - \boldsymbol{x}_\infty\|)$ (dB) versus epoch count in Experiment 2. Top: Algorithm 8.2. The horizontal axis starts at 140 epochs to account for the auxiliary tasks (see a/). Bottom: Algorithm 8.3. The computational load per epoch for Algorithm 8.3 was about twice that of Algorithm 8.2.

References

[1] 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.

[2] J.-F. Aujol and A. Chambolle, Dual norms and image decomposition models, *Int. J. Comput. Vision*, vol. 63, pp. 85–104, 2005.

[3] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, Optimization with sparsity-inducing penalties, *Found. Trends Machine Learn.*, vol. 4, pp. 1–106, 2012.

[4] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches, *IEEE J. Select. Topics Appl. Earth Observ. Remote Sensing*, vol. 5, pp. 354–379, 2012.

[5] 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.

[6] 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.

[7] A. Chambolle, M. J. Ehrhardt, P. Richtárik, and C.-B. Schönlieb, Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications, *SIAM J. Optim.*, vol. 28, pp. 2783–2808, 2018.

[8] A. Chambolle and T. Pock, An introduction to continuous optimization for imaging, *Acta Numer.*, vol. 25, pp. 161–319, 2016.

[9] P. L. Combettes and J. Eckstein, Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions, *Math. Program. Ser. B*, vol. 168, pp. 645–672, 2018.

[10] 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.

[11] P. L. Combettes and L. E. Glaudin, Solving composite fixed point problems with block updates, *Adv. Nonlinear Anal.*, vol. 10, 2021.

[12] P. L. Combettes, A. M. McDonald, C. A. Micchelli, and M. Pontil, Learning with optimal interpolation norms, *Numer. Algorithms*, vol. 81, pp. 695–717, 2019.

[13] P. L. Combettes and C. L. Müller, Perspective maximum likelihood-type estimation via proximal decomposition, *Electron. J. Stat.*, vol. 14, pp. 207–238, 2020.

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

[15] 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.

[16] P. L. Combettes and J.-C. Pesquet, Fixed point strategies in data science, *IEEE Trans. Signal Process.*, vol. 69, pp. 3878–3905, 2021.

[17] J. Darbon and T. Meng, On decomposition models in imaging sciences and multi-time Hamilton–Jacobi partial differential equations, *SIAM J. Imaging Sci.*, vol. 13, pp. 971–1014, 2020.

[18] A. J. Defazio, T. S. Caetano, and J. Domke, Finito: A faster, permutable incremental gradient method for big data problems, *Proc. Intl. Conf. Machine Learn.*, pp. 1125–1133, 2014.

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

[20] M. Hintermüller and G. Stadler, An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration, *SIAM J. Sci. Comput.*, vol. 28, pp. 1–23, 2006.

[21] L. Jacob, G. Obozinski, and J.-Ph. Vert, Group lasso with overlap and graph lasso, *Proc. Int. Conf. Machine Learn.*, pp. 433–440, 2009.

[22] P. R. Johnstone and J. Eckstein, Projective splitting with forward steps, *Math. Program. Ser. A*, published online 2020-09-30.

[23] K. Mishchenko, F. Iutzeler, and J. Malick, A distributed flexible delay-tolerant proximal gradient algorithm, *SIAM J. Optim.*, vol. 30, pp. 933–959, 2020.

[24] 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.

[25] S. Salzo and S. Villa, Parallel random block-coordinate forward-backward algorithm: A unified convergence analysis, *Math. Program. Ser. A*, published online 2021-04-11.

[26] M. Schmidt, N. Le Roux, and F. Bach, Minimizing finite sums with the stochastic average gradient, *Math. Program. Ser. A*, vol. 162, pp. 83–112, 2017.

Chapter 9

MULTIVARIATE MONOTONE INCLUSIONS IN SADDLE FORM

9.1 Introduction and context

To answer question **(Q8)** of Chapter 1, we introduce a saddle formalism for systems of monotone inclusions, study its properties, and provide abstract principles for finding a zero of the associated saddle operator. This, in turn, leads to a flexible algorithm for solving systems of monotone inclusions. Various applications are discussed in Section 9.2.4.

This chapter presents the following article:

> M. N. Bùi and P. L. Combettes, Multivariate monotone inclusions in saddle form, *Mathematics of Operations Research*, to appear.

9.2 Article: Multivariate monotone inclusions in saddle form

Abstract. We propose a novel approach to monotone operator splitting based on the notion of a saddle operator. Under investigation is a highly structured multivariate monotone inclusion problem involving a mix of set-valued, cocoercive, and Lipschitzian monotone operators, as well as various monotonicity-preserving operations among them. This model encompasses most formulations found in the literature. A limitation of existing primal-dual algorithms is that they operate in a product space that is too small to achieve full splitting of our problem in the sense that each operator is used individually. To circumvent this difficulty, we recast the problem as that of finding a zero of a saddle operator that acts on a bigger space. This leads to an algorithm of unprecedented flexibility, which achieves full splitting, exploits the specific attributes of each operator, is asynchronous, and requires to activate only blocks of operators at each iteration, as opposed to activating all of them. The latter feature is of critical importance in large-scale problems. Weak convergence of the main algorithm is established, as well as the

strong convergence of a variant. Various applications are discussed, and instantiations of the proposed framework in the context of variational inequalities and minimization problems are presented.

9.2.1 Introduction

In 1979, several methods appeared to solve the basic problem of finding a zero of the sum of two maximally monotone operators in a real Hilbert space [37, 38, 43]. Over the past forty years, increasingly complex inclusion problems and solution techniques have been considered [10, 14, 17, 19, 23, 25, 29, 34, 53] to address concrete problems in fields as diverse as game theory [2, 15, 56], evolution inclusions [3], traffic equilibrium [3, 31], domain decomposition [4], machine learning [6, 12], image recovery [7, 11, 16, 33], mean field games [18], convex programming [24, 36], statistics [26, 55], neural networks [27], signal processing [28], partial differential equations [32], tensor completion [39], and optimal transport [42]. In our view, two challenging issues in the field of monotone operator splitting algorithms are the following:

- A number of independent monotone inclusion models coexist with various assumptions on the operators and different types of operation among these operators. At the same time, as will be seen in Section 9.2.4, they are not sufficiently general to cover important applications.
- Most algorithms do not allow asynchrony and impose that all the operators be activated at each iteration. They can therefore not handle efficiently modern large-scale problems. The only methods that are asynchronous and block-iterative are limited to specific scenarios [25, 29, 34] and they do not cover inclusion models such as that of [23].

In an attempt to bring together and extend the application scope of the wide variety of unrelated models that coexist in the literature, we propose the following multivariate formulation which involves a mix of set-valued, cocoercive, and Lipschitzian monotone operators, as well as various monotonicity-preserving operations among them.

Problem 9.1 Let $(\mathcal{H}_i)_{i\in I}$ and $(\mathcal{G}_k)_{k\in K}$ be finite families of real Hilbert spaces with Hilbert direct sums $\boldsymbol{\mathcal{H}} = \bigoplus_{i\in I}\mathcal{H}_i$ and $\boldsymbol{\mathcal{G}} = \bigoplus_{k\in K}\mathcal{G}_k$. Denote by $\boldsymbol{x} = (x_i)_{i\in I}$ a generic element in $\boldsymbol{\mathcal{H}}$. For every $i \in I$ and every $k \in K$, let $s_i^* \in \mathcal{H}_i$, let $r_k \in \mathcal{G}_k$, and 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] $L_{ki}\colon \mathcal{H}_i \to \mathcal{G}_k$ is linear and bounded.

In addition, it is assumed that

[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

$$\begin{aligned}&\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \text{ such that } (\forall i \in I)\ \ s_i^* \in A_i\overline{x}_i + C_i\overline{x}_i + Q_i\overline{x}_i + R_i\overline{\boldsymbol{x}}\\ &\qquad + \sum_{k\in K} L_{ki}^*\Biggl(\Bigl(\bigl(B_k^m + B_k^c + B_k^\ell\bigr) \,\Box\, \bigl(D_k^m + D_k^c + D_k^\ell\bigr)\Bigr)\Biggl(\sum_{j\in I} L_{kj}\overline{x}_j - r_k\Biggr)\Biggr)\end{aligned} \tag{9.1}$$

and the associated dual problem

$$\begin{aligned}&\text{find } \overline{\boldsymbol{v}}^* \in \boldsymbol{\mathcal{G}} \text{ such that } (\exists\, \boldsymbol{x} \in \boldsymbol{\mathcal{H}})(\forall i \in I)(\forall k \in K)\\ &\qquad\begin{cases} s_i^* - \displaystyle\sum_{j\in K} L_{ji}^*\overline{v}_j^* \in A_ix_i + C_ix_i + Q_ix_i + R_i\boldsymbol{x}\\ \overline{v}_k^* \in \Bigl(\bigl(B_k^m + B_k^c + B_k^\ell\bigr) \,\Box\, \bigl(D_k^m + D_k^c + D_k^\ell\bigr)\Bigr)\Biggl(\displaystyle\sum_{j\in I} L_{kj}x_j - r_k\Biggr).\end{cases}\end{aligned} \tag{9.2}$$

Our highly structured model involves three basic monotonicity preserving operations, namely addition, composition with linear operators, and parallel sum. It extends the state-of-the-art model of [23], where the simpler form

$$(\forall i \in I) \quad s_i^* \in A_i\overline{x}_i + Q_i\overline{x}_i + \sum_{k\in K} L_{ki}^*\Biggl(\bigl(B_k^m \,\Box\, D_k^m\bigr)\Biggl(\sum_{j\in I} L_{kj}\overline{x}_j - r_k\Biggr)\Biggr) \tag{9.3}$$

of the system in (9.1) was investigated; see also [3, 25] for special cases. In an increasing number of applications, the sets I and K can be sizable. To handle such large-scale problems, it is critical to implement block-iterative solution algorithms, in which only subgroups of the operators involved in the problem need to be activated at each iteration. In addition, it is desirable that the algorithm be asynchronous in the sense that, at any iteration, it has the ability to incorporate the result of calculations initiated at earlier iterations. Such methods have been proposed for special cases of Problem 9.1: first in [25] for the system

$$\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \text{ such that } (\forall i \in I)\ \ s_i^* \in A_i\overline{x}_i + \sum_{k\in K} L_{ki}^*\Biggl(B_k^m\Biggl(\sum_{j\in I} L_{kj}\overline{x}_j - r_k\Biggr)\Biggr), \tag{9.4}$$

then in [29] for the inclusion (we omit the subscript '1')

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } 0 \in \sum_{k\in K} L_k^*\bigl(B_k^m(L_k\overline{x})\bigr), \tag{9.5}$$

and more recently in [34] for the inclusion

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } 0 \in A\overline{x} + Q\overline{x} + \sum_{k\in K} L_k^*\big((B_k^m + B_k^\ell)(L_k\overline{x})\big). \tag{9.6}$$

It is clear that the formulations (9.4) and (9.6) are not interdependent. Furthermore, as we shall see in Section 9.2.4, many applications of interest are not covered by either of them. From both a theoretical and a practical viewpoint, it is therefore important to unify and extend these approaches. To achieve this goal, we propose to design an algorithm for solving the general Problem 9.1 which possesses simultaneously the following features:

① It has the ability to process all the operators individually and exploit their specific attributes, e.g., set-valuedness, cocoercivity, Lipschitz continuity, and linearity.

② It is block-iterative in the sense that it does not need to activate all the operators at each iteration, but only a subgroup of them.

③ It is asynchronous.

④ Each set-valued monotone operator is scaled by its own, iteration-dependent, parameter.

⑤ It does not require any knowledge of the norms of the linear operators involved in the model.

Let us observe that the method of [25] has features ①–⑤, but it is restricted to (9.4). Likewise, the method of [34] has features ①–⑤, but it is restricted to (9.6).

Solving the intricate Problem 9.1 with the requirement ① does not seem possible with existing tools. The presence of requirements ②–⑤ further complicates this task. In particular, the Kuhn–Tucker approach initiated in [14] — and further developed in [1, 10, 23, 25, 34, 35] — relies on finding a zero of an operator acting on the primal-dual space $\boldsymbol{\mathcal{H}} \oplus \boldsymbol{\mathcal{G}}$. However, in the context of Problem 9.1, this primal-dual space is too small to achieve full splitting in the sense that each operator is used individually. To circumvent this difficulty, we propose a novel splitting strategy that consists of recasting the problem as that of finding a zero of a saddle operator acting on the bigger space $\boldsymbol{\mathcal{H}} \oplus \boldsymbol{\mathcal{G}} \oplus \boldsymbol{\mathcal{G}} \oplus \boldsymbol{\mathcal{G}}$. This is done in Section 9.2.2, where we define the saddle form of Problem 9.1, study its properties, and propose outer approximation principles to solve it. In Section 9.2.3, the main asynchronous block-iterative algorithm is presented and we establish its weak convergence under mild conditions on the frequency at which the operators are selected. We also present a strongly convergent variant. The specializations to variational inequalities and multivariate minimization are discussed in Section 9.2.4, along with several applications. Section 9.2.5 contains auxiliary results.

Notation. The notation used in this paper is standard and follows [9], to which one can refer for background and complements on monotone operators and convex analysis. Let $\mathcal{K}$ be a real Hilbert space. The symbols $\langle \cdot \mid \cdot \rangle$ and $\|\cdot\|$ denote the scalar product of $\mathcal{K}$ and the associated norm, respectively. The expressions $x_n \rightharpoonup x$ and $x_n \to x$ denote, respectively, the weak and the strong convergence of a sequence $(x_n)_{n\in\mathbb{N}}$ to x in $\mathcal{K}$, and $2^{\mathcal{K}}$ denotes the family of all subsets of $\mathcal{K}$. Let $A\colon \mathcal{K}\to 2^{\mathcal{K}}$. The graph of A is $\text{gra}\,A=\big\{(x,x^*)\in\mathcal{K}\times\mathcal{K} \mid x^*\in Ax\big\}$, the set of zeros of A is $\text{zer}\,A=\big\{x\in\mathcal{K}\mid 0\in Ax\big\}$, the inverse of A is $A^{-1}\colon\mathcal{K}\to 2^{\mathcal{K}}\colon x^*\mapsto\big\{x\in\mathcal{K}\mid x^*\in Ax\big\}$, and the resolvent of A is $J_A=(\text{Id}+A)^{-1}$, where Id is the identity operator on $\mathcal{K}$. Further, A is monotone if

$$\big(\forall (x,x^*)\in\text{gra}\,A\big)\big(\forall (y,y^*)\in\text{gra}\,A\big)\quad \langle x-y\mid x^*-y^*\rangle\geqslant 0, \tag{9.7}$$

and it is maximally monotone if, for every $(x,x^*)\in\mathcal{K}\times\mathcal{K}$,

$$(x,x^*)\in\text{gra}\,A\quad\Leftrightarrow\quad \big(\forall (y,y^*)\in\text{gra}\,A\big)\ \langle x-y\mid x^*-y^*\rangle\geqslant 0. \tag{9.8}$$

If A is maximally monotone, then J_A is a single-valued operator defined on $\mathcal{K}$. The parallel sum of $B\colon\mathcal{K}\to 2^{\mathcal{K}}$ and $D\colon\mathcal{K}\to 2^{\mathcal{K}}$ is $B\,\square\,D=(B^{-1}+D^{-1})^{-1}$. An operator $C\colon\mathcal{K}\to\mathcal{K}$ is cocoercive with constant $\alpha\in\left]0,+\infty\right[$ if $(\forall x\in\mathcal{K})(\forall y\in\mathcal{K})\ \langle x-y\mid Cx-Cy\rangle\geqslant\alpha\|Cx-Cy\|^2$. We denote by $\Gamma_0(\mathcal{K})$ the class of lower semicontinuous convex functions $f\colon\mathcal{K}\to\left]-\infty,+\infty\right]$ such that $\text{dom}\,f=\big\{x\in\mathcal{K}\mid f(x)<+\infty\big\}\neq\varnothing$. Let $f\in\Gamma_0(\mathcal{K})$. The conjugate of f is the function $\Gamma_0(\mathcal{K})\ni f^*\colon x^*\mapsto\sup_{x\in\mathcal{K}}(\langle x\mid x^*\rangle-f(x))$ and the subdifferential of f is the maximally monotone operator $\partial f\colon\mathcal{K}\to 2^{\mathcal{K}}\colon x\mapsto\big\{x^*\in\mathcal{K}\mid(\forall y\in\mathcal{K})\ \langle y-x\mid x^*\rangle+f(x)\leqslant f(y)\big\}$. In addition, $\text{epi}\,f$ is the epigraph of f. For every $x\in\mathcal{K}$, the unique minimizer of $f+(1/2)\|\cdot-x\|^2$ is denoted by $\text{prox}_f x$. We have $\text{prox}_f=J_{\partial f}$. Given $h\in\Gamma_0(\mathcal{K})$, the infimal convolution of f and h is $f\,\square\,h\colon\mathcal{K}\to[-\infty,+\infty]:x\mapsto\inf_{y\in\mathcal{K}}(f(y)+h(x-y))$; the infimal convolution $f\,\square\,h$ is exact if the infimum is achieved everywhere, in which case we write $f\,\boxdot\,h$. Now let $(\mathcal{K}_i)_{i\in I}$ be a finite family of real Hilbert spaces and, for every $i\in I$, let $f_i\colon\mathcal{K}_i\to\left]-\infty,+\infty\right]$. Then

$$\bigoplus_{i\in I}f_i\colon\boldsymbol{\mathcal{K}}=\bigoplus_{i\in I}\mathcal{K}_i\to\left]-\infty,+\infty\right]:\boldsymbol{x}\mapsto\sum_{i\in I}f_i(x_i). \tag{9.9}$$

The partial derivative of a differentiable function $\Theta\colon\boldsymbol{\mathcal{K}}\to\mathbb{R}$ relative to $\mathcal{K}_i$ is denoted by $\nabla_i\Theta$. Finally, let C be a nonempty convex subset of $\mathcal{K}$. A point $x\in C$ belongs to the strong relative interior of C, in symbols $x\in\text{sri}\,C$, if $\bigcup_{\lambda\in\left]0,+\infty\right[}\lambda(C-x)$ is a closed vector subspace of $\mathcal{K}$. If C is closed, the projection operator onto it is denoted by proj_C and the normal cone operator of C is the maximally monotone operator

$$N_C\colon\mathcal{K}\to 2^{\mathcal{K}}\colon x\mapsto\begin{cases}\big\{x^*\in\mathcal{K}\mid\sup\langle C-x\mid x^*\rangle\leqslant 0\big\}, & \text{if } x\in C;\\ \varnothing, & \text{otherwise.}\end{cases} \tag{9.10}$$

9.2.2 The saddle form of Problem 9.1

A classical Lagrangian setting for convex minimization is the following. Given real Hilbert spaces $\mathcal{H}$ and $\mathcal{G}$, $f\in\Gamma_0(\mathcal{H})$, $g\in\Gamma_0(\mathcal{G})$, and a bounded linear operator $L\colon\mathcal{H}\to\mathcal{G}$, consider the primal problem

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; f(x)+g(Lx) \tag{9.11}$$

together with its Fenchel–Rockafellar dual [47]

$$\underset{v^*\in\mathcal{G}}{\text{minimize}}\;\; f^*\big(-L^*v^*\big)+g^*(v^*). \tag{9.12}$$

The primal-dual pair (9.11)–(9.12) can be analyzed through the lens of Rockafellar's saddle formalism [49,50] as follows. Set $h\colon \mathcal{H}\oplus\mathcal{G}\to\left]-\infty,+\infty\right]\colon (x,y)\mapsto f(x)+g(y)$ and $U\colon \mathcal{H}\oplus\mathcal{G}\to\mathcal{G}\colon (x,y)\mapsto Lx-y$, and note that $U^*\colon\mathcal{G}\to\mathcal{H}\oplus\mathcal{G}\colon v^*\mapsto(L^*v^*,-v^*)$. Then, upon defining $\mathcal{K}=\mathcal{H}\oplus\mathcal{G}$ and introducing the variable $z=(x,y)\in\mathcal{K}$, (9.11) is equivalent to

$$\underset{z\in\mathcal{K},\,Uz=0}{\text{minimize}}\;\; h(z) \tag{9.13}$$

and (9.12) to

$$\underset{v^*\in\mathcal{G}}{\text{minimize}}\;\; h^*\big(-U^*v^*\big). \tag{9.14}$$

The Lagrangian associated with (9.13) is (see [51, Example 4'] or [9, Proposition 19.21])

$$\begin{aligned}\mathcal{L}\colon \mathcal{K}\oplus\mathcal{G}&\to\left]-\infty,+\infty\right]\\ (z,v^*)&\mapsto\begin{cases}h(z)+\langle Uz\mid v^*\rangle, & \text{if } z\in\operatorname{dom}h;\\ +\infty, & \text{otherwise,}\end{cases}\end{aligned} \tag{9.15}$$

and the associated saddle operator [49,50] is the maximally monotone operator

$$\mathcal{S}\colon \mathcal{K}\oplus\mathcal{G}\to 2^{\mathcal{K}\oplus\mathcal{G}}\colon (z,v^*)\mapsto\partial\mathcal{L}(\cdot,v^*)(z)\times\partial\big(-\mathcal{L}(z,\cdot)\big)(v^*)=\big(\partial h(z)+U^*v^*\big)\times\{-Uz\}. \tag{9.16}$$

As shown in [49], a zero $(\overline{z},\overline{v}^*)$ of $\mathcal{S}$ is a saddle point of $\mathcal{L}$, and it has the property that $\overline{z}$ solves (9.13) and $\overline{v}^*$ solves (9.14). Thus, going back to the original Fenchel–Rockafellar pair (9.11)–(9.12), we learn that, if $(\overline{x},\overline{y},\overline{v}^*)$ is a zero of the saddle operator

$$\mathcal{S}\colon \mathcal{H}\oplus\mathcal{G}\oplus\mathcal{G}\to 2^{\mathcal{H}\oplus\mathcal{G}\oplus\mathcal{G}}\colon (x,y,v^*)\mapsto\big(\partial f(x)+L^*v^*\big)\times\big(\partial g(y)-v^*\big)\times\{-Lx+y\}, \tag{9.17}$$

then $\overline{x}$ solves (9.11) and $\overline{v}^*$ solves (9.12). As shown in [24, Section 4.5], a suitable splitting of $\mathcal{S}$ leads to an implementable algorithm to solve (9.11)–(9.12).

A generalization of Fenchel–Rockafellar duality to monotone inclusions was proposed in [44,46] and further extended in [23]. Given maximally monotone operators $A\colon\mathcal{H}\to 2^{\mathcal{H}}$ and $B\colon\mathcal{G}\to 2^{\mathcal{G}}$, and a bounded linear operator $L\colon\mathcal{H}\to\mathcal{G}$, the primal problem

$$\text{find}\;\; \overline{x}\in\mathcal{H}\;\; \text{such that}\;\; 0\in A\overline{x}+L^*\big(B(L\overline{x})\big) \tag{9.18}$$

is paired with the dual problem

$$\text{find } \overline{v}^* \in \mathcal{G} \text{ such that } 0 \in -L\big(A^{-1}(-L^*\overline{v}^*)\big) + B^{-1}\overline{v}^*. \tag{9.19}$$

Following the same pattern as that described above, let us consider the *saddle operator*

$$\mathcal{S}\colon \mathcal{H} \oplus \mathcal{G} \oplus \mathcal{G} \to 2^{\mathcal{H}\oplus\mathcal{G}\oplus\mathcal{G}} \colon (x, y, v^*) \mapsto (Ax + L^*v^*) \times (By - v^*) \times \{-Lx + y\}. \tag{9.20}$$

It is readily shown that, if $(\overline{x}, \overline{y}, \overline{v}^*)$ is a zero of $\mathcal{S}$, then $\overline{x}$ solves (9.18) and $\overline{v}^*$ solves (9.19). We call the problem of finding a zero of $\mathcal{S}$ the *saddle form* of (9.18)–(9.19). We now introduce a saddle operator for the general Problem 9.1.

Definition 9.2 In the setting of Problem 9.1, let $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{H}} \oplus \boldsymbol{\mathcal{G}} \oplus \boldsymbol{\mathcal{G}} \oplus \boldsymbol{\mathcal{G}}$. The *saddle operator* associated with Problem 9.1 is

$$\begin{aligned}
&\mathcal{S}\colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}} \colon (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{v}^*) \mapsto \\
&\quad \Bigg(\bigtimes_{i\in I} \Big(-s_i^* + A_i x_i + C_i x_i + Q_i x_i + R_i \boldsymbol{x} + \sum_{k\in K} L_{ki}^* v_k^* \Big), \bigtimes_{k\in K} \big(B_k^m y_k + B_k^c y_k + B_k^\ell y_k - v_k^*\big), \\
&\qquad \bigtimes_{k\in K} \big(D_k^m z_k + D_k^c z_k + D_k^\ell z_k - v_k^*\big), \bigtimes_{k\in K} \Big\{ r_k + y_k + z_k - \sum_{i\in I} L_{ki} x_i \Big\} \Bigg),
\end{aligned} \tag{9.21}$$

and the *saddle form* of Problem 9.1 is to

$$\text{find } \overline{\mathbf{x}} \in \boldsymbol{\mathcal{X}} \text{ such that } \mathbf{0} \in \mathcal{S}\overline{\mathbf{x}}. \tag{9.22}$$

Next, we establish some properties of the saddle operator as well as connections with Problem 9.1.

Proposition 9.3 *Consider the setting of Problem 9.1 and Definition 9.2. Let $\mathscr{P}$ be the set of solutions to (9.1), let $\mathscr{D}$ be the set of solutions to (9.2), and let*

$$\begin{aligned}
\mathsf{Z} = \Bigg\{ (\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}}^*) \in \boldsymbol{\mathcal{H}} \oplus \boldsymbol{\mathcal{G}} \;\Bigg|\; &(\forall i \in I)(\forall k \in K) \quad s_i^* - \sum_{j\in K} L_{ji}^* \overline{v}_j^* \in A_i \overline{x}_i + C_i \overline{x}_i + Q_i \overline{x}_i + R_i \overline{\boldsymbol{x}} \text{ and} \\
&\sum_{j\in I} L_{kj} \overline{x}_j - r_k \in \big(B_k^m + B_k^c + B_k^\ell\big)^{-1} \overline{v}_k^* + \big(D_k^m + D_k^c + D_k^\ell\big)^{-1} \overline{v}_k^* \Bigg\}
\end{aligned} \tag{9.23}$$

be the associated Kuhn–Tucker set. Then the following hold:

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

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

(iii) *Suppose that $\overline{\mathbf{x}} = (\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}}, \overline{\boldsymbol{z}}, \overline{\boldsymbol{v}}^*) \in \operatorname{zer} \mathcal{S}$. Then $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}}^*) \in \mathsf{Z} \subset \mathscr{P} \times \mathscr{D}$.*

(iv) $\mathscr{D} \neq \varnothing \Leftrightarrow \operatorname{zer} \mathcal{S} \neq \varnothing \Leftrightarrow \mathsf{Z} \neq \varnothing \Rightarrow \mathscr{P} \neq \varnothing$.

(v) *Suppose that one of the following holds:*

[a] *I is a singleton.*

[b] *For every $k \in K$, $(B_k^m + B_k^c + B_k^\ell) \,\square\, (D_k^m + D_k^c + D_k^\ell)$ is at most single-valued.*

[c] *For every $k \in K$, $(D_k^m + D_k^c + D_k^\ell)^{-1}$ is strictly monotone.*

[d] *$I \subset K$, the operators $((B_k^m + B_k^c + B_k^\ell) \,\square\, (D_k^m + D_k^c + D_k^\ell))_{k \in K \smallsetminus I}$ are at most single-valued, and $(\forall i \in I)(\forall k \in I)\ k \neq i \Rightarrow L_{ki} = 0$.*

Then $\mathscr{P} \neq \varnothing \Rightarrow \mathsf{Z} \neq \varnothing$.

Proof. Define

$$\begin{cases} \boldsymbol{A} \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}} \colon \boldsymbol{x} \mapsto \boldsymbol{R}\boldsymbol{x} + \times_{i \in I} \big(A_i x_i + C_i x_i + Q_i x_i\big) \\ \boldsymbol{B} \colon \boldsymbol{\mathcal{G}} \to 2^{\boldsymbol{\mathcal{G}}} \colon \boldsymbol{y} \mapsto \times_{k \in K} \big(B_k^m y_k + B_k^c y_k + B_k^\ell y_k\big) \\ \boldsymbol{D} \colon \boldsymbol{\mathcal{G}} \to 2^{\boldsymbol{\mathcal{G}}} \colon \boldsymbol{z} \mapsto \times_{k \in K} \big(D_k^m z_k + D_k^c z_k + D_k^\ell z_k\big) \\ \boldsymbol{L} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{G}} \colon \boldsymbol{x} \mapsto \big(\sum_{i \in I} L_{ki} x_i\big)_{k \in K} \\ \boldsymbol{s}^* = (s_i^*)_{i \in I} \ \text{ and } \ \boldsymbol{r} = (r_k)_{k \in K}. \end{cases} \tag{9.24}$$

Then the adjoint of $\boldsymbol{L}$ is

$$\boldsymbol{L}^* \colon \boldsymbol{\mathcal{G}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{v}^* \mapsto \left(\sum_{k \in K} L_{ki}^* v_k^*\right)_{i \in I}. \tag{9.25}$$

Hence, in view of (9.21) and (9.24),

$$\boldsymbol{\mathcal{S}} \colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}} \colon (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{v}^*) \mapsto \big(-\boldsymbol{s}^* + \boldsymbol{A}\boldsymbol{x} + \boldsymbol{L}^*\boldsymbol{v}^*\big) \times \big(\boldsymbol{B}\boldsymbol{y} - \boldsymbol{v}^*\big) \times \big(\boldsymbol{D}\boldsymbol{z} - \boldsymbol{v}^*\big) \times \big\{\boldsymbol{r} - \boldsymbol{L}\boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}\big\}. \tag{9.26}$$

(i): Let us introduce the operators

$$\begin{cases} \mathbf{P} \colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}} \colon (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{v}^*) \mapsto (-\boldsymbol{s}^* + \boldsymbol{A}\boldsymbol{x}) \times \boldsymbol{B}\boldsymbol{y} \times \boldsymbol{D}\boldsymbol{z} \times \{\boldsymbol{r}\} \\ \mathbf{W} \colon \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{X}} \colon (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{v}^*) \mapsto (\boldsymbol{L}^*\boldsymbol{v}^*, -\boldsymbol{v}^*, -\boldsymbol{v}^*, -\boldsymbol{L}\boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}). \end{cases} \tag{9.27}$$

Using Problem 9.1[a]–[c], we derive from [9, Example 20.31, Corollaries 20.28 and 25.5(i)] that, for every $i \in I$ and every $k \in K$, the operators $A_i + C_i + Q_i$, $B_k^m + B_k^c + B_k^\ell$, and $D_k^m + D_k^c + D_k^\ell$ are maximally monotone. At the same time, Problem 9.1[e] and [9, Corollary 20.28] entail that $\boldsymbol{R}$ is maximally monotone. Therefore, it results from (9.24), [9, Proposition 20.23 and Corollary 25.5(i)], and (9.27) that $\mathbf{P}$ is maximally monotone. However, since Problem 9.1[d] and (9.27) imply that $\mathbf{W}$ is linear and bounded with $\mathbf{W}^* = -\mathbf{W}$, [9, Example 20.35] asserts that $\mathbf{W}$ is maximally monotone. Hence, in view of [9, Corollary 25.5(i)], we infer from (9.26)–(9.27) that $\boldsymbol{\mathcal{S}} = \mathbf{P} + \mathbf{W}$ is maximally monotone.

(ii): This follows from (i) and [9, Proposition 23.39].

(iii): Using (9.24) and (9.25), we deduce from (9.23) that

$$\mathsf{Z} = \big\{(\boldsymbol{x}, \boldsymbol{v}^*) \in \boldsymbol{\mathcal{H}} \oplus \boldsymbol{\mathcal{G}} \mid \boldsymbol{s}^* - \boldsymbol{L}^*\boldsymbol{v}^* \in \boldsymbol{A}\boldsymbol{x} \ \text{ and } \ \boldsymbol{L}\boldsymbol{x} - \boldsymbol{r} \in \boldsymbol{B}^{-1}\boldsymbol{v}^* + \boldsymbol{D}^{-1}\boldsymbol{v}^*\big\} \tag{9.28}$$

and from (9.2) that

$$\mathscr{D}=\big\{\boldsymbol{v}^*\in\boldsymbol{\mathcal{G}}\mid -\boldsymbol{r}\in-\boldsymbol{L}\big(\boldsymbol{A}^{-1}(\boldsymbol{s}^*-\boldsymbol{L}^*\boldsymbol{v}^*)\big)+\boldsymbol{B}^{-1}\boldsymbol{v}^*+\boldsymbol{D}^{-1}\boldsymbol{v}^*\big\}. \tag{9.29}$$

Suppose that $(\boldsymbol{x},\boldsymbol{v}^*)\in\mathsf{Z}$. Then it follows from (9.28) that $\boldsymbol{x}\in\boldsymbol{A}^{-1}(\boldsymbol{s}^*-\boldsymbol{L}^*\boldsymbol{v}^*)$ and, in turn, that $-\boldsymbol{r}\in-\boldsymbol{L}\boldsymbol{x}+\boldsymbol{B}^{-1}\boldsymbol{v}^*+\boldsymbol{D}^{-1}\boldsymbol{v}^*\subset-\boldsymbol{L}(\boldsymbol{A}^{-1}(\boldsymbol{s}^*-\boldsymbol{L}^*\boldsymbol{v}^*))+\boldsymbol{B}^{-1}\boldsymbol{v}^*+\boldsymbol{D}^{-1}\boldsymbol{v}^*$. Thus $\boldsymbol{v}^*\in\mathscr{D}$ by (9.29). In addition, (9.23) implies that

$$(\forall k\in K)\quad v_k^*\in\big((B_k^m+B_k^c+B_k^\ell)\,\square\,(D_k^m+D_k^c+D_k^\ell)\big)\Bigg(\sum_{j\in I}L_{kj}x_j-r_k\Bigg) \tag{9.30}$$

and, therefore, that

$$\begin{aligned}(\forall i\in I)\quad s_i^*&\in A_ix_i+C_ix_i+Q_ix_i+R_i\boldsymbol{x}+\sum_{k\in K}L_{ki}^*v_k^*\\ &\subset A_ix_i+C_ix_i+Q_ix_i+R_i\boldsymbol{x}\\ &\quad+\sum_{k\in K}L_{ki}^*\Bigg(\Big(\big(B_k^m+B_k^c+B_k^\ell\big)\,\square\,\big(D_k^m+D_k^c+D_k^\ell\big)\Big)\Bigg(\sum_{j\in I}L_{kj}x_j-r_k\Bigg)\Bigg).\end{aligned} \tag{9.31}$$

Hence, $\boldsymbol{x}\in\mathscr{P}$. To summarize, we have shown that $\mathsf{Z}\subset\mathscr{P}\times\mathscr{D}$. It remains to show that $(\overline{\boldsymbol{x}},\overline{\boldsymbol{v}}^*)\in\mathsf{Z}$. Since $\boldsymbol{0}\in\boldsymbol{\mathcal{S}}\overline{\mathbf{x}}$, we deduce from (9.26) that $\boldsymbol{s}^*-\boldsymbol{L}^*\overline{\boldsymbol{v}}^*\in\boldsymbol{A}\overline{\boldsymbol{x}}$, $\boldsymbol{L}\overline{\boldsymbol{x}}-\boldsymbol{r}=\overline{\boldsymbol{y}}+\overline{\boldsymbol{z}}$, $\boldsymbol{0}\in\boldsymbol{B}\overline{\boldsymbol{y}}-\overline{\boldsymbol{v}}^*$, and $\boldsymbol{0}\in\boldsymbol{D}\overline{\boldsymbol{z}}-\overline{\boldsymbol{v}}^*$. Therefore, $\boldsymbol{L}\overline{\boldsymbol{x}}-\boldsymbol{r}\in\boldsymbol{B}^{-1}\overline{\boldsymbol{v}}^*+\boldsymbol{D}^{-1}\overline{\boldsymbol{v}}^*$ and (9.28) thus yields $(\overline{\boldsymbol{x}},\overline{\boldsymbol{v}}^*)\in\mathsf{Z}$.

(iv): The implication $\operatorname{zer}\boldsymbol{\mathcal{S}}\neq\varnothing\Rightarrow\mathscr{P}\neq\varnothing$ follows from (iii). Next, we derive from (9.29) and (9.28) that

$$\begin{aligned}\mathscr{D}\neq\varnothing&\Leftrightarrow(\exists\,\overline{\boldsymbol{v}}^*\in\boldsymbol{\mathcal{G}})\ \ -\boldsymbol{r}\in-\boldsymbol{L}\big(\boldsymbol{A}^{-1}(\boldsymbol{s}^*-\boldsymbol{L}^*\overline{\boldsymbol{v}}^*)\big)+\boldsymbol{B}^{-1}\overline{\boldsymbol{v}}^*+\boldsymbol{D}^{-1}\overline{\boldsymbol{v}}^*\\ &\Leftrightarrow\big(\exists\,(\overline{\boldsymbol{v}}^*,\overline{\boldsymbol{x}})\in\boldsymbol{\mathcal{G}}\oplus\boldsymbol{\mathcal{H}}\big)\ \ -\boldsymbol{r}\in-\boldsymbol{L}\overline{\boldsymbol{x}}+\boldsymbol{B}^{-1}\overline{\boldsymbol{v}}^*+\boldsymbol{D}^{-1}\overline{\boldsymbol{v}}^*\ \text{ and }\ \overline{\boldsymbol{x}}\in\boldsymbol{A}^{-1}(\boldsymbol{s}^*-\boldsymbol{L}^*\overline{\boldsymbol{v}}^*)\\ &\Leftrightarrow\big(\exists\,(\overline{\boldsymbol{x}},\overline{\boldsymbol{v}}^*)\in\boldsymbol{\mathcal{H}}\oplus\boldsymbol{\mathcal{G}}\big)\ \ \boldsymbol{s}^*-\boldsymbol{L}^*\overline{\boldsymbol{v}}^*\in\boldsymbol{A}\overline{\boldsymbol{x}}\ \text{ and }\ \boldsymbol{L}\overline{\boldsymbol{x}}-\boldsymbol{r}\in\boldsymbol{B}^{-1}\overline{\boldsymbol{v}}^*+\boldsymbol{D}^{-1}\overline{\boldsymbol{v}}^*\\ &\Leftrightarrow\mathsf{Z}\neq\varnothing.\end{aligned} \tag{9.32}$$

However, (iii) asserts that $\operatorname{zer}\boldsymbol{\mathcal{S}}\neq\varnothing\Rightarrow\mathsf{Z}\neq\varnothing$. Therefore, it remains to show that $\mathsf{Z}\neq\varnothing\Rightarrow\operatorname{zer}\boldsymbol{\mathcal{S}}\neq\varnothing$. Towards this end, suppose that $(\overline{\boldsymbol{x}},\overline{\boldsymbol{v}}^*)\in\mathsf{Z}$. Then, by (9.28), $\boldsymbol{s}^*-\boldsymbol{L}^*\overline{\boldsymbol{v}}^*\in\boldsymbol{A}\overline{\boldsymbol{x}}$ and $\boldsymbol{L}\overline{\boldsymbol{x}}-\boldsymbol{r}\in\boldsymbol{B}^{-1}\overline{\boldsymbol{v}}^*+\boldsymbol{D}^{-1}\overline{\boldsymbol{v}}^*$. Hence, $\boldsymbol{0}\in-\boldsymbol{s}^*+\boldsymbol{A}\overline{\boldsymbol{x}}+\boldsymbol{L}^*\overline{\boldsymbol{v}}^*$ and there exists $(\overline{\boldsymbol{y}},\overline{\boldsymbol{z}})\in\boldsymbol{\mathcal{G}}\oplus\boldsymbol{\mathcal{G}}$ such that $\overline{\boldsymbol{y}}\in\boldsymbol{B}^{-1}\overline{\boldsymbol{v}}^*$, $\overline{\boldsymbol{z}}\in\boldsymbol{D}^{-1}\overline{\boldsymbol{v}}^*$, and $\boldsymbol{L}\overline{\boldsymbol{x}}-\boldsymbol{r}=\overline{\boldsymbol{y}}+\overline{\boldsymbol{z}}$. We thus deduce that $\boldsymbol{0}\in\boldsymbol{B}\overline{\boldsymbol{y}}-\overline{\boldsymbol{v}}^*$, $\boldsymbol{0}\in\boldsymbol{D}\overline{\boldsymbol{z}}-\overline{\boldsymbol{v}}^*$, and $\boldsymbol{r}-\boldsymbol{L}\overline{\boldsymbol{x}}+\overline{\boldsymbol{y}}+\overline{\boldsymbol{z}}=\boldsymbol{0}$. Consequently, (9.26) implies that $(\overline{\boldsymbol{x}},\overline{\boldsymbol{y}},\overline{\boldsymbol{z}},\overline{\boldsymbol{v}}^*)\in\operatorname{zer}\boldsymbol{\mathcal{S}}$.

(v): In view of (iv), it suffices to establish that $\mathscr{P}\neq\varnothing\Rightarrow\mathscr{D}\neq\varnothing$. Suppose that $\overline{\boldsymbol{x}}\in\mathscr{P}$.

[a]: Suppose that $I=\{1\}$. We then infer from (9.1) that there exists $\overline{\boldsymbol{v}}^*\in\boldsymbol{\mathcal{G}}$ such that

$$\begin{cases} s_1^*\in A_1\overline{x}_1+C_1\overline{x}_1+Q_1\overline{x}_1+R_1\overline{\boldsymbol{x}}+\sum\limits_{k\in K}L_{k1}^*\overline{v}_k^*\\ (\forall k\in K)\ \ \overline{v}_k^*\in\big((B_k^m+B_k^c+B_k^\ell)\,\square\,(D_k^m+D_k^c+D_k^\ell)\big)(L_{k1}\overline{x}_1-r_k). \end{cases} \tag{9.33}$$

Therefore, by (9.2), $\overline{\boldsymbol{v}}^*\in\mathscr{D}$.

[b]: Set $(\forall k\in K)$ $\overline{v}_k^*=((B_k^m+B_k^c+B_k^\ell)\,\square\,(D_k^m+D_k^c+D_k^\ell))(\sum_{j\in I}L_{kj}\overline{x}_j-r_k)$. Then $\overline{\boldsymbol{v}}^*$ solves (9.2).

[c]⇒[b]: See [23, Section 4].

[d]: Let $i\in I$. It results from our assumption that

$$\begin{aligned} s_i^*\in{}& A_i\overline{x}_i+C_i\overline{x}_i+Q_i\overline{x}_i+R_i\overline{\boldsymbol{x}}+L_{ii}^*\Big(\big((B_i^m+B_i^c+B_i^\ell)\,\square\,(D_i^m+D_i^c+D_i^\ell)\big)(L_{ii}\overline{x}_i-r_i)\Big)\\ &+\sum_{k\in K\smallsetminus I}L_{ki}^*\left(\Big((B_k^m+B_k^c+B_k^\ell)\,\square\,(D_k^m+D_k^c+D_k^\ell)\Big)\left(\sum_{j\in I}L_{kj}\overline{x}_j-r_k\right)\right). \end{aligned} \tag{9.34}$$

Thus, there exists $\overline{v}_i^*\in\mathcal{G}_i$ such that $\overline{v}_i^*\in((B_i^m+B_i^c+B_i^\ell)\,\square\,(D_i^m+D_i^c+D_i^\ell))(L_{ii}\overline{x}_i-r_i)$ and that

$$\begin{aligned} s_i^*\in{}& A_i\overline{x}_i+C_i\overline{x}_i+Q_i\overline{x}_i+R_i\overline{\boldsymbol{x}}+L_{ii}^*\overline{v}_i^*\\ &+\sum_{k\in K\smallsetminus I}L_{ki}^*\left(\Big((B_k^m+B_k^c+B_k^\ell)\,\square\,(D_k^m+D_k^c+D_k^\ell)\Big)\left(\sum_{j\in I}L_{kj}\overline{x}_j-r_k\right)\right). \end{aligned} \tag{9.35}$$

As a result, upon setting

$$(\forall k\in K\smallsetminus I)\quad \overline{v}_k^*=\big((B_k^m+B_k^c+B_k^\ell)\,\square\,(D_k^m+D_k^c+D_k^\ell)\big)\left(\sum_{j\in I}L_{kj}\overline{x}_j-r_k\right), \tag{9.36}$$

we conclude that $\overline{\boldsymbol{v}}^*\in\mathscr{D}$. ☐

Remark 9.4 Some noteworthy observations about Proposition 9.3 are the following.

(i) The Kuhn–Tucker set (9.23) extends to Problem 9.1 the corresponding notion introduced for some special cases in [1, 14, 25].

(ii) In connection with Proposition 9.3(v), we note that the implication $\mathscr{P}\neq\varnothing\Rightarrow\mathsf{Z}\neq\varnothing$ is implicitly used in [25, Theorems 13 and 15], where one requires $\mathsf{Z}\neq\varnothing$ but merely assumes $\mathscr{P}\neq\varnothing$. However, this implication is not true in general (a similar oversight is found in [1, 45, 52]). Indeed, consider as a special case of (9.1), the problem of solving the system

$$\begin{cases} 0\in B_1(x_1+x_2)+B_2(x_1-x_2)\\ 0\in B_1(x_1+x_2)-B_2(x_1-x_2) \end{cases} \tag{9.37}$$

in the Euclidean plane $\mathbb{R}^2$. Then, by choosing $B_1 = \{0\}^{-1}$ and $B_2 = 1$, we obtain $\mathscr{P} = \big\{(x_1, -x_1) \mid x_1 \in \mathbb{R}\big\}$, whereas $\mathsf{Z} = \varnothing$.

(iii) As stated in Proposition 9.3(iii), any Kuhn–Tucker point is a solution to (9.1)–(9.2). In the simpler setting considered in [25], a splitting algorithm was devised for finding such a point. However, in the more general context of Problem 9.1, there does not seem to exist a path from the Kuhn–Tucker formalism in $\boldsymbol{\mathcal{H}} \oplus \boldsymbol{\mathcal{G}}$ to an algorithm that is fully split in the sense of ①. This motivates our approach, which seeks a zero of the saddle operator $\boldsymbol{\mathcal{S}}$ defined on the bigger space $\boldsymbol{\mathcal{X}}$ and, thereby, offers more flexibility.

(iv) Special cases of Problem 9.1 can be found in [1, 25, 34, 35], where they were solved by algorithms that proceed by outer approximation of the Kuhn–Tucker set in $\boldsymbol{\mathcal{H}} \oplus \boldsymbol{\mathcal{G}}$. In those special cases, Algorithm 9.12 below does not reduce to those of [1, 25, 34, 35] since it operates by outer approximation of the set of zeros of the saddle operator $\boldsymbol{\mathcal{S}}$ in the bigger space $\boldsymbol{\mathcal{X}}$.

The following operators will induce a decomposition of the saddle operator that will lead to a splitting algorithm which complies with our requirements ①–⑤.

Definition 9.5 In the setting of Definition 9.2, set

$$\begin{aligned}\mathbf{M}\colon \boldsymbol{\mathcal{X}} \to 2^{\boldsymbol{\mathcal{X}}}\colon (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{v}^*) \mapsto &\\ \Bigg(\bigtimes_{i\in I} \bigg(-s_i^* + A_i x_i + Q_i x_i + R_i \boldsymbol{x} + \sum_{k\in K} L_{ki}^* v_k^* \bigg), &\bigtimes_{k\in K} \big(B_k^m y_k + B_k^\ell y_k - v_k^* \big),\\ \bigtimes_{k\in K} \big(D_k^m z_k + D_k^\ell z_k - v_k^* \big), &\bigtimes_{k\in K} \bigg\{ r_k + y_k + z_k - \sum_{i\in I} L_{ki} x_i \bigg\} \Bigg)\end{aligned} \tag{9.38}$$

and

$$\mathbf{C}\colon \boldsymbol{\mathcal{X}} \to \boldsymbol{\mathcal{X}}\colon (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{v}^*) \mapsto \Big(\big(C_i x_i\big)_{i\in I}, \big(B_k^c y_k\big)_{k\in K}, \big(D_k^c z_k\big)_{k\in K}, \mathbf{0} \Big). \tag{9.39}$$

Proposition 9.6 *In the setting of Problem 9.1 and of Definitions 9.2 and 9.5, the following hold:*

(i) $\boldsymbol{\mathcal{S}} = \mathbf{M} + \mathbf{C}$.

(ii) $\mathbf{M}$ *is maximally monotone.*

(iii) *Set* $\alpha = \min\{\alpha_i^c, \beta_k^c, \delta_k^c\}_{i\in I, k\in K}$. *Then the following hold:*

  (a) $\mathbf{C}$ *is* α*-cocoercive.*

  (b) *Let* $(\mathbf{p}, \mathbf{p}^*) \in \operatorname{gra} \mathbf{M}$ *and* $\mathbf{q} \in \boldsymbol{\mathcal{X}}$. *Then* $\operatorname{zer} \boldsymbol{\mathcal{S}} \subset \big\{ \mathbf{x} \in \boldsymbol{\mathcal{X}} \mid \langle \mathbf{x} - \mathbf{p} \mid \mathbf{p}^* + \mathbf{C}\mathbf{q} \rangle \leqslant (4\alpha)^{-1} \|\mathbf{p} - \mathbf{q}\|^2 \big\}$.

Proof. (i): Clear from (9.21), (9.38), and (9.39).

(ii): This is a special case of Proposition 9.3(i), where, for every $i \in I$ and every $k \in K$, $C_i = 0$ and $B_k^c = D_k^c = 0$.

(iii)(a): Take $\mathbf{x} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{v}^*)$ and $\mathbf{y} = (\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{w}^*)$ in $\boldsymbol{\mathcal{X}}$. By (9.39) and Problem 9.1[a]–[c],

$$\langle \mathbf{x} - \mathbf{y} \mid \mathbf{C}\mathbf{x} - \mathbf{C}\mathbf{y} \rangle$$

$$
\begin{aligned}
&= \sum_{i\in I} \langle x_i - a_i \mid C_i x_i - C_i a_i\rangle + \sum_{k\in K} \big(\langle y_k - b_k \mid B_k^c y_k - B_k^c b_k\rangle + \langle z_k - c_k \mid D_k^c z_k - D_k^c c_k\rangle\big)\\
&\geqslant \sum_{i\in I} \alpha_i^c \|C_i x_i - C_i a_i\|^2 + \sum_{k\in K} \big(\beta_k^c \|B_k^c y_k - B_k^c b_k\|^2 + \delta_k^c \|D_k^c z_k - D_k^c c_k\|^2\big)\\
&\geqslant \alpha \sum_{i\in I} \|C_i x_i - C_i a_i\|^2 + \alpha \sum_{k\in K} \big(\|B_k^c y_k - B_k^c b_k\|^2 + \|D_k^c z_k - D_k^c c_k\|^2\big)\\
&= \alpha \|\mathbf{Cx} - \mathbf{Cy}\|^2. \qquad (9.40)
\end{aligned}
$$

(iii)(b): Suppose that $\mathbf{z} \in \operatorname{zer} \mathcal{S}$. We deduce from (i) that $-\mathbf{Cz} \in \mathbf{Mz}$ and from our assumption that $\mathbf{p}^* \in \mathbf{Mp}$. Hence, (ii) implies that $\langle \mathbf{z} - \mathbf{p} \mid \mathbf{p}^* + \mathbf{Cz}\rangle \leqslant 0$. Thus, we infer from (iii)(a) and the Cauchy–Schwarz inequality that

$$
\begin{aligned}
\langle \mathbf{z} - \mathbf{p} \mid \mathbf{p}^* + \mathbf{Cq}\rangle &= \langle \mathbf{z} - \mathbf{p} \mid \mathbf{p}^* + \mathbf{Cz}\rangle - \langle \mathbf{z} - \mathbf{q} \mid \mathbf{Cz} - \mathbf{Cq}\rangle + \langle \mathbf{p} - \mathbf{q} \mid \mathbf{Cz} - \mathbf{Cq}\rangle\\
&\leqslant -\alpha \|\mathbf{Cz} - \mathbf{Cq}\|^2 + \|\mathbf{p} - \mathbf{q}\|\, \|\mathbf{Cz} - \mathbf{Cq}\|\\
&= (4\alpha)^{-1} \|\mathbf{p} - \mathbf{q}\|^2 - \Big|\big(2\sqrt{\alpha}\big)^{-1} \|\mathbf{p} - \mathbf{q}\| - \sqrt{\alpha}\|\mathbf{Cz} - \mathbf{Cq}\|\Big|^2\\
&\leqslant (4\alpha)^{-1} \|\mathbf{p} - \mathbf{q}\|^2, \qquad (9.41)
\end{aligned}
$$

which establishes the claim. ☐

Next, we solve the saddle form (9.22) of Problem 9.1 via successive projections onto the outer approximations constructed in Proposition 9.6(iii)(b).

Proposition 9.7 *Consider the setting of Problem 9.1 and of Definitions 9.2 and 9.5, and suppose that* $\operatorname{zer} \mathcal{S} \neq \varnothing$. *Set* $\alpha = \min\{\alpha_i^c, \beta_k^c, \delta_k^c\}_{i\in I, k\in K}$, *let* $\mathbf{x}_0 \in \boldsymbol{\mathcal{X}}$, *let* $\varepsilon \in \left]0,1\right[$, *and iterate*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots\\
\left\lfloor
\begin{array}{l}
(\mathbf{p}_n, \mathbf{p}_n^*) \in \operatorname{gra} \mathbf{M};\ \mathbf{q}_n \in \boldsymbol{\mathcal{X}};\\
\mathbf{t}_n^* = \mathbf{p}_n^* + \mathbf{C}\mathbf{q}_n;\\
\Delta_n = \langle \mathbf{x}_n - \mathbf{p}_n \mid \mathbf{t}_n^*\rangle - (4\alpha)^{-1} \|\mathbf{p}_n - \mathbf{q}_n\|^2;\\
\textit{if } \Delta_n > 0\\
\left\lfloor
\begin{array}{l}
\lambda_n \in [\varepsilon, 2-\varepsilon];\\
\mathbf{x}_{n+1} = \mathbf{x}_n - (\lambda_n \Delta_n / \|\mathbf{t}_n^*\|^2)\, \mathbf{t}_n^*;
\end{array}
\right.\\
\textit{else}\\
\left\lfloor
\begin{array}{l}
\mathbf{x}_{n+1} = \mathbf{x}_n.
\end{array}
\right.
\end{array}
\right.
\end{array}
\qquad (9.42)
$$

Then the following hold:

(i) $(\forall \mathbf{z} \in \operatorname{zer} \mathcal{S})(\forall n \in \mathbb{N})\ \|\mathbf{x}_{n+1} - \mathbf{z}\| \leqslant \|\mathbf{x}_n - \mathbf{z}\|$.

(ii) $\sum_{n\in\mathbb{N}} \|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 < +\infty$.

(iii) *Suppose that* $(\mathbf{t}_n^*)_{n\in\mathbb{N}}$ *is bounded. Then* $\varlimsup \Delta_n \leqslant 0$.

(iv) *Suppose that* $\mathbf{x}_n - \mathbf{p}_n \rightharpoonup \mathbf{0}$, $\mathbf{p}_n - \mathbf{q}_n \to \mathbf{0}$, *and* $\mathbf{t}_n^* \to \mathbf{0}$. *Then* $(\mathbf{x}_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* $\operatorname{zer} \mathcal{S}$.

Proof. (i)&(ii): Proposition 9.3(ii) and our assumption ensure that $\operatorname{zer}\mathcal{S}$ is a nonempty closed convex subset of $\boldsymbol{\mathcal{X}}$. Now, for every $n \in \mathbb{N}$, set $\eta_n = (4\alpha)^{-1}\|\mathbf{p}_n - \mathbf{q}_n\|^2 + \langle \mathbf{p}_n \mid \mathbf{t}_n^*\rangle$ and $\mathbf{H}_n = \{\mathbf{x} \in \boldsymbol{\mathcal{X}} \mid \langle \mathbf{x} \mid \mathbf{t}_n^*\rangle \leqslant \eta_n\}$. On the one hand, according to Proposition 9.6(iii)(b), $(\forall n \in \mathbb{N})$ $\operatorname{zer}\mathcal{S} \subset \mathbf{H}_n$. On the other hand, (9.42) gives $(\forall n \in \mathbb{N})$ $\Delta_n = \langle \mathbf{x}_n \mid \mathbf{t}_n^*\rangle - \eta_n$. Altogether, (9.42) is an instantiation of (9.142). The claims thus follow from Lemma 9.28(i)&(ii).

(iii): Set $\mu = \sup_{n\in\mathbb{N}}\|\mathbf{t}_n^*\|$. For every $n \in \mathbb{N}$, if $\Delta_n > 0$, then (9.42) yields $\Delta_n = \lambda_n^{-1}\|\mathbf{t}_n^*\|\,\|\mathbf{x}_{n+1} - \mathbf{x}_n\| \leqslant \varepsilon^{-1}\mu\|\mathbf{x}_{n+1} - \mathbf{x}_n\|$; otherwise, $\Delta_n \leqslant 0 = \varepsilon^{-1}\mu\|\mathbf{x}_{n+1} - \mathbf{x}_n\|$. We therefore invoke (ii) to get $\overline{\lim}\,\Delta_n \leqslant \lim \varepsilon^{-1}\mu\|\mathbf{x}_{n+1} - \mathbf{x}_n\| = 0$.

(iv): Let $\mathbf{x} \in \boldsymbol{\mathcal{X}}$, let $(k_n)_{n\in\mathbb{N}}$ be a strictly increasing sequence in $\mathbb{N}$, and suppose that $\mathbf{x}_{k_n} \rightharpoonup \mathbf{x}$. Then $\mathbf{p}_{k_n} = (\mathbf{p}_{k_n} - \mathbf{x}_{k_n}) + \mathbf{x}_{k_n} \rightharpoonup \mathbf{x}$. In addition, (9.42) and Proposition 9.6(i) imply that $(\mathbf{p}_{k_n}, \mathbf{p}_{k_n}^* + \mathbf{C}\mathbf{p}_{k_n})_{n\in\mathbb{N}}$ lies in $\operatorname{gra}(\mathbf{M}+\mathbf{C}) = \operatorname{gra}\mathcal{S}$. We also note that, since $\mathbf{C}$ is $(1/\alpha)$-Lipschitzian by Proposition 9.6(iii)(a), (9.42) yields $\|\mathbf{p}_n^* + \mathbf{C}\mathbf{p}_n\| = \|\mathbf{t}_n^* - \mathbf{C}\mathbf{q}_n + \mathbf{C}\mathbf{p}_n\| \leqslant \|\mathbf{t}_n^*\| + \|\mathbf{C}\mathbf{p}_n - \mathbf{C}\mathbf{q}_n\| \leqslant \|\mathbf{t}_n^*\| + \|\mathbf{p}_n - \mathbf{q}_n\|/\alpha \to 0$. Altogether, since $\mathcal{S}$ is maximally monotone by Proposition 9.3(i), [9, Proposition 20.38(ii)] yields $\mathbf{x} \in \operatorname{zer}\mathcal{S}$. In turn, Lemma 9.28(iii) guarantees that $(\mathbf{x}_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}\mathcal{S}$. ☐

The next outer approximation scheme is a variant of the previous one that guarantees strong convergence to a specific zero of the saddle operator.

Proposition 9.8 *Consider the setting of Problem 9.1 and of Definitions 9.2 and 9.5, and suppose that* $\operatorname{zer}\mathcal{S} \neq \varnothing$. *Define*

$$\Xi\colon\ ]0,+\infty[\,\times\,]0,+\infty[\,\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}^2$$
$$(\Delta,\tau,\varsigma,\chi)\mapsto\begin{cases}(1,\Delta/\tau), & \text{if } \rho = 0;\\ \big(0,(\Delta+\chi)/\tau\big), & \text{if } \rho\neq 0 \text{ and } \chi\Delta\geqslant\rho;\\ \big(1-\chi\Delta/\rho,\varsigma\Delta/\rho\big), & \text{if } \rho\neq 0 \text{ and } \chi\Delta<\rho,\end{cases}$$
$$\text{where } \rho = \tau\varsigma - \chi^2, \tag{9.43}$$

set $\alpha = \min\{\alpha_i^c, \beta_k^c, \delta_k^c\}_{i\in I,k\in K}$, *and let* $\mathbf{x}_0 \in \boldsymbol{\mathcal{X}}$. *Iterate*

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor\begin{array}{l}(\mathbf{p}_n,\mathbf{p}_n^*)\in\operatorname{gra}\mathbf{M};\ \mathbf{q}_n\in\boldsymbol{\mathcal{X}};\\ \mathbf{t}_n^* = \mathbf{p}_n^* + \mathbf{C}\mathbf{q}_n;\\ \Delta_n = \langle \mathbf{x}_n - \mathbf{p}_n \mid \mathbf{t}_n^*\rangle - (4\alpha)^{-1}\|\mathbf{p}_n - \mathbf{q}_n\|^2;\\ \text{if } \Delta_n > 0\\ \left\lfloor\begin{array}{l}\tau_n = \|\mathbf{t}_n^*\|^2;\ \varsigma_n = \|\mathbf{x}_0 - \mathbf{x}_n\|^2;\ \chi_n = \langle \mathbf{x}_0 - \mathbf{x}_n \mid \mathbf{t}_n^*\rangle;\\ (\kappa_n,\lambda_n) = \Xi(\Delta_n,\tau_n,\varsigma_n,\chi_n);\\ \mathbf{x}_{n+1} = (1-\kappa_n)\mathbf{x}_0 + \kappa_n\mathbf{x}_n - \lambda_n\mathbf{t}_n^*;\end{array}\right.\\ \text{else}\\ \left\lfloor\begin{array}{l}\mathbf{x}_{n+1} = \mathbf{x}_n.\end{array}\right.\end{array}\right.\end{array} \tag{9.44}$$

Then the following hold:

(i) $(\forall n \in \mathbb{N})\ \|\mathbf{x}_n - \mathbf{x}_0\| \leqslant \|\mathbf{x}_{n+1} - \mathbf{x}_0\| \leqslant \|\mathrm{proj}_{\mathrm{zer}\,\mathcal{S}}\mathbf{x}_0 - \mathbf{x}_0\|$.

(ii) $\sum_{n\in\mathbb{N}} \|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 < +\infty$.

(iii) *Suppose that* $(\mathbf{t}_n^*)_{n\in\mathbb{N}}$ *is bounded. Then* $\overline{\lim}\, \Delta_n \leqslant 0$.

(iv) *Suppose that* $\mathbf{x}_n - \mathbf{p}_n \rightharpoonup \mathbf{0}$*,* $\mathbf{p}_n - \mathbf{q}_n \to \mathbf{0}$*, and* $\mathbf{t}_n^* \to \mathbf{0}$. *Then* $\mathbf{x}_n \to \mathrm{proj}_{\mathrm{zer}\,\mathcal{S}}\mathbf{x}_0$.

Proof. Set $(\forall n \in \mathbb{N})\ \eta_n = (4\alpha)^{-1}\|\mathbf{p}_n - \mathbf{q}_n\|^2 + \langle \mathbf{p}_n \mid \mathbf{t}_n^*\rangle$ and $\mathbf{H}_n = \big\{\mathbf{x} \in \boldsymbol{\mathcal{X}} \mid \langle \mathbf{x} \mid \mathbf{t}_n^*\rangle \leqslant \eta_n\big\}$. As seen in the proof of Proposition 9.7, $\mathrm{zer}\,\mathcal{S}$ is a nonempty closed convex subset of $\boldsymbol{\mathcal{X}}$ and, for every $n \in \mathbb{N}$, $\mathrm{zer}\,\mathcal{S} \subset \mathbf{H}_n$ and $\Delta_n = \langle \mathbf{x}_n \mid \mathbf{t}_n^*\rangle - \eta_n$. This and (9.43) make (9.44) an instance of (9.143).

(i)&(ii): Apply Lemma 9.29(i)&(ii).

(iii): Set $\mu = \sup_{n\in\mathbb{N}} \|\mathbf{t}_n^*\|$. Take $n \in \mathbb{N}$. Suppose that $\Delta_n > 0$. Then, by construction of $\mathbf{H}_n$, $\mathrm{proj}_{\mathbf{H}_n}\mathbf{x}_n = \mathbf{x}_n - (\Delta_n/\|\mathbf{t}_n^*\|^2)\,\mathbf{t}_n^*$. This implies that $\Delta_n = \|\mathbf{t}_n^*\|\,\|\mathrm{proj}_{\mathbf{H}_n}\mathbf{x}_n - \mathbf{x}_n\| \leqslant \mu\|\mathrm{proj}_{\mathbf{H}_n}\mathbf{x}_n - \mathbf{x}_n\|$. Next, suppose that $\Delta_n \leqslant 0$. Then $\mathbf{x}_n \in \mathbf{H}_n$ and therefore $\Delta_n \leqslant 0 = \mu\|\mathrm{proj}_{\mathbf{H}_n}\mathbf{x}_n - \mathbf{x}_n\|$. Altogether, $(\forall n \in \mathbb{N})\ \Delta_n \leqslant \mu\|\mathrm{proj}_{\mathbf{H}_n}\mathbf{x}_n - \mathbf{x}_n\|$. Consequently, Lemma 9.29(ii) yields $\overline{\lim}\,\Delta_n \leqslant 0$.

(iv): Follow the same procedure as in the proof of Proposition 9.7(iv), invoking Lemma 9.29(iii) instead of Lemma 9.28(iii). ☐

9.2.3 Asynchronous block-iterative outer approximation methods

We exploit the saddle form of Problem 9.1 described in Definition 9.2 to obtain splitting algorithms with features ①–⑤. Let us comment on the impact of requirements ①–④.

① For every $i \in I$ and every $k \in K$, each single-valued operator C_i, Q_i, R_i, B_k^c, B_k^ℓ, D_k^c, D_k^ℓ, and L_{ki} must be activated individually via a forward step, whereas each of the set-valued operators A_i, B_k^m, and D_k^m must be activated individually via a backward resolvent step.

② At iteration n, only operators indexed by subgroups $I_n \subset I$ and $K_n \subset K$ of indices need to be involved in the sense that the results of their evaluations are incorporated. This considerably reduces the computational load compared to standard methods, which require the use of all the operators at every iteration. Assumption 9.10 below regulates the frequency at which the indices should be chosen over time.

③ When an operator is involved at iteration n, its evaluation can be made at a point based on data available at an earlier iteration. This makes it possible to initiate a computation at a given iteration and incorporate its result at a later time. Assumption 9.11 below controls the lag allowed in the process of using past data.

④ Assumption 9.9 below describes the range allowed for the various scaling parameters in terms of the cocoercivity and Lipschitz constants of the operators.

Assumption 9.9 *In the setting of Problem 9.1, set* $\alpha = \min\{\alpha_i^c, \beta_k^c, \delta_k^c\}_{i\in I,k\in K}$*, let* $\sigma \in \left]0,+\infty\right[$ *and* $\varepsilon \in \left]0,1\right[$ *be such that*

$$\sigma > 1/(4\alpha) \quad \text{and} \quad 1/\varepsilon > \max\left\{\alpha_i^\ell + \chi + \sigma, \beta_k^\ell + \sigma, \delta_k^\ell + \sigma\right\}_{i\in I,k\in K}, \tag{9.45}$$

and suppose that the following are satisfied:

[a] *For every* $i \in I$ *and every* $n \in \mathbb{N}$, $\gamma_{i,n} \in \left[\varepsilon, 1/(\alpha_i^\ell + \chi + \sigma)\right]$.

[b] *For every* $k \in K$ *and every* $n \in \mathbb{N}$, $\mu_{k,n} \in \left[\varepsilon, 1/(\beta_k^\ell + \sigma)\right]$, $\nu_{k,n} \in \left[\varepsilon, 1/(\delta_k^\ell + \sigma)\right]$, *and* $\sigma_{k,n} \in [\varepsilon, 1/\varepsilon]$.

[c] *For every* $i \in I$, $x_{i,0} \in \mathcal{H}_i$*; for every* $k \in K$, $\{y_{k,0}, z_{k,0}, v_{k,0}^*\} \subset \mathcal{G}_k$.

Assumption 9.10 *I and K are finite sets,* $P \in \mathbb{N}$, $(I_n)_{n\in\mathbb{N}}$ *are nonempty subsets of* I*, and* $(K_n)_{n\in\mathbb{N}}$ *are nonempty subsets of* K *such that*

$$I_0 = I, \quad K_0 = K, \quad \text{and} \quad (\forall n \in \mathbb{N}) \bigcup_{j=n}^{n+P} I_j = I \ \text{ and } \bigcup_{j=n}^{n+P} K_j = K. \tag{9.46}$$

Assumption 9.11 *I and K are finite sets,* $T \in \mathbb{N}$*, 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* $n - T \leqslant \omega_k(n) \leqslant n$.

Our first algorithm is patterned after the abstract geometric outer approximation principle described in Proposition 9.7. As before, bold letters denote product space elements, e.g., $\boldsymbol{x}_n = (x_{i,n})_{i\in I} \in \boldsymbol{\mathcal{H}}$.

Algorithm 9.12 Consider the setting of Problem 9.1 and suppose that Assumption 9.9–9.11 is in force. Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 2-\varepsilon]$ 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)}(s_i^* - l_{i,n}^* - C_i x_{i,\pi_i(n)})\big); \\
\quad\quad a_{i,n}^* = \gamma_{i,\pi_i(n)}^{-1}(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_k^\ell y_{k,\omega_k(n)}; \\
\quad\quad w_{k,n}^* = v_{k,\omega_k(n)}^* - D_k^\ell z_{k,\omega_k(n)}; \\
\quad\quad b_{k,n} = J_{\mu_{k,\omega_k(n)} B_k^m}\big(y_{k,\omega_k(n)} + \mu_{k,\omega_k(n)}(u_{k,n}^* - B_k^c y_{k,\omega_k(n)})\big); \\
\quad\quad d_{k,n} = J_{\nu_{k,\omega_k(n)} D_k^m}\big(z_{k,\omega_k(n)} + \nu_{k,\omega_k(n)}(w_{k,n}^* - D_k^c 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)} - r_k\big) + v_{k,\omega_k(n)}^*; \\
\quad\quad q_{k,n}^* = \mu_{k,\omega_k(n)}^{-1}(y_{k,\omega_k(n)} - b_{k,n}) + u_{k,n}^* + B_k^\ell b_{k,n} - e_{k,n}^*; \\
\quad\quad t_{k,n}^* = \nu_{k,\omega_k(n)}^{-1}(z_{k,\omega_k(n)} - d_{k,n}) + w_{k,n}^* + D_k^\ell 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} = r_k + 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}^*;\ q_{k,n}^* = q_{k,n-1}^*;\ t_{k,n}^* = t_{k,n-1}^*; \\
\quad\quad \eta_{k,n} = \eta_{k,n-1};\ e_{k,n} = r_k + 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\quad\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 + \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}^*;\ 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{9.47}
$$

The convergence properties of Algorithm 9.12 are laid out in the following theorem.

Theorem 9.13 *Consider the setting of Algorithm 9.12 and suppose that the dual problem (9.2) has a solution. Then the following hold:*

(i) *Let $i \in I$. Then $\sum_{n\in\mathbb{N}} \|x_{i,n+1} - x_{i,n}\|^2 < +\infty$.*

(ii) *Let $k \in K$. Then $\sum_{n\in\mathbb{N}} \|y_{k,n+1} - y_{k,n}\|^2 < +\infty$, $\sum_{n\in\mathbb{N}} \|z_{k,n+1} - z_{k,n}\|^2 < +\infty$, and $\sum_{n\in\mathbb{N}} \|v^*_{k,n+1} - v^*_{k,n}\|^2 < +\infty$.*

(iii) *Let $i \in I$ and $k \in K$. Then $x_{i,n} - a_{i,n} \to 0$, $y_{k,n} - b_{k,n} \to 0$, $z_{k,n} - d_{k,n} \to 0$, and $v^*_{k,n} - e^*_{k,n} \to 0$.*

(iv) *There exist a solution $\overline{\boldsymbol{x}}$ to (9.1) and a solution $\overline{\boldsymbol{v}}^*$ to (9.2) such that, for every $i \in I$ and every $k \in K$, $x_{i,n} \rightharpoonup \overline{x}_i$, $a_{i,n} \rightharpoonup \overline{x}_i$, and $v^*_{k,n} \rightharpoonup \overline{v}^*_k$. In addition, $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}}^*)$ is a Kuhn–Tucker point of Problem 9.1 in the sense of (9.23).*

Proof. We use the notation of Definitions 9.2 and 9.5. We first observe that $\operatorname{zer} \boldsymbol{\mathcal{S}} \neq \varnothing$ by virtue of Proposition 9.3(iv). Next, let us verify that (9.47) is a special case of (9.42). For every $i \in I$, denote by $\overline{\vartheta}_i(n)$ the most recent iteration preceding an iteration n at which the results of the evaluations of the operators A_i, C_i, Q_i, and R_i were incorporated, and by $\vartheta_i(n)$ the iteration at which the corresponding calculations were initiated, i.e.,

$$\overline{\vartheta}_i(n) = \max\big\{j \in \mathbb{N} \mid j \leqslant n \;\text{ and }\; i \in I_j\big\} \quad \text{and} \quad \vartheta_i(n) = \pi_i\big(\overline{\vartheta}_i(n)\big). \tag{9.48}$$

Similarly, we define

$$(\forall k \in K)(\forall n \in \mathbb{N}) \quad \overline{\varrho}_k(n) = \max\big\{j \in \mathbb{N} \mid j \leqslant n \;\text{ and }\; k \in K_j\big\} \quad \text{and} \quad \varrho_k(n) = \omega_k\big(\overline{\varrho}_k(n)\big). \tag{9.49}$$

By virtue of (9.47),

$$(\forall i \in I)(\forall n \in \mathbb{N}) \quad a_{i,n} = a_{i,\overline{\vartheta}_i(n)}, \quad a^*_{i,n} = a^*_{i,\overline{\vartheta}_i(n)}, \quad \xi_{i,n} = \xi_{i,\overline{\vartheta}_i(n)}, \tag{9.50}$$

and likewise

$$(\forall k \in K)(\forall n \in \mathbb{N}) \quad \begin{cases} b_{k,n} = b_{k,\overline{\varrho}_k(n)}, \;\; d_{k,n} = d_{k,\overline{\varrho}_k(n)}, \;\; \eta_{k,n} = \eta_{k,\overline{\varrho}_k(n)} \\ e^*_{k,n} = e^*_{k,\overline{\varrho}_k(n)}, \;\; q^*_{k,n} = q^*_{k,\overline{\varrho}_k(n)}, \;\; t^*_{k,n} = t^*_{k,\overline{\varrho}_k(n)}. \end{cases} \tag{9.51}$$

To proceed further, set

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \mathbf{x}_n = (\boldsymbol{x}_n, \boldsymbol{y}_n, \boldsymbol{z}_n, \boldsymbol{v}_n^*) \\ \mathbf{p}_n = (\boldsymbol{a}_n, \boldsymbol{b}_n, \boldsymbol{d}_n, \boldsymbol{e}_n^*) \\ \mathbf{p}_n^* = \big(\boldsymbol{p}_n^* - (C_i x_{i,\vartheta_i(n)})_{i\in I}, \boldsymbol{q}_n^* - (B_k^c y_{k,\varrho_k(n)})_{k\in K}, \boldsymbol{t}_n^* - (D_k^c z_{k,\varrho_k(n)})_{k\in K}, \boldsymbol{e}_n\big) \\ \mathbf{q}_n = \big((x_{i,\vartheta_i(n)})_{i\in I}, (y_{k,\varrho_k(n)})_{k\in K}, (z_{k,\varrho_k(n)})_{k\in K}, (e_{k,n}^*)_{k\in K}\big) \\ \mathbf{t}_n^* = (\boldsymbol{p}_n^*, \boldsymbol{q}_n^*, \boldsymbol{t}_n^*, \boldsymbol{e}_n). \end{cases} \tag{9.52}$$

For every $i \in I$ and every $n \in \mathbb{N}$, it follows from (9.50), (9.48), (9.47), and [9, Proposition 23.2(ii)] that

$$\begin{aligned} a_{i,n}^* - C_i x_{i,\vartheta_i(n)} &= a_{i,\overline{\vartheta}_i(n)}^* - C_i x_{i,\pi_i(\overline{\vartheta}_i(n))} \\ &= \gamma_{i,\pi_i(\overline{\vartheta}_i(n))}^{-1}\big(x_{i,\pi_i(\overline{\vartheta}_i(n))} - a_{i,\overline{\vartheta}_i(n)}\big) - l_{i,\overline{\vartheta}_i(n)}^* - C_i x_{i,\pi_i(\overline{\vartheta}_i(n))} + Q_i a_{i,\overline{\vartheta}_i(n)} \\ &\in -s_i^* + A_i a_{i,\overline{\vartheta}_i(n)} + Q_i a_{i,\overline{\vartheta}_i(n)} \\ &= -s_i^* + A_i a_{i,n} + Q_i a_{i,n} \end{aligned} \tag{9.53}$$

and, therefore, that

$$\begin{aligned} p_{i,n}^* - C_i x_{i,\vartheta_i(n)} &= a_{i,n}^* - C_i x_{i,\vartheta_i(n)} + R_i \boldsymbol{a}_n + \sum_{k\in K} L_{ki}^* e_{k,n}^* \\ &\in -s_i^* + A_i a_{i,n} + Q_i a_{i,n} + R_i \boldsymbol{a}_n + \sum_{k\in K} L_{ki}^* e_{k,n}^*. \end{aligned} \tag{9.54}$$

Analogously, we invoke (9.51), (9.49), and (9.47) to obtain

$$(\forall k \in K)(\forall n \in \mathbb{N}) \quad q_{k,n}^* - B_k^c y_{k,\varrho_k(n)} \in B_k^m b_{k,n} + B_k^\ell b_{k,n} - e_{k,n}^* \tag{9.55}$$

and

$$(\forall k \in K)(\forall n \in \mathbb{N}) \quad t_{k,n}^* - D_k^c z_{k,\varrho_k(n)} \in D_k^m d_{k,n} + D_k^\ell d_{k,n} - e_{k,n}^*. \tag{9.56}$$

In addition, (9.47) states that

$$(\forall k \in K)(\forall n \in \mathbb{N}) \quad e_{k,n} = r_k + b_{k,n} + d_{k,n} - \sum_{i\in I} L_{ki} a_{i,n}. \tag{9.57}$$

Hence, using (9.52) and (9.38), we deduce that $(\mathbf{p}_n, \mathbf{p}_n^*)_{n\in\mathbb{N}}$ lies in gra $\mathbf{M}$. Next, it results from (9.52) and (9.39) that $(\forall n \in \mathbb{N})$ $\mathbf{t}_n^* = \mathbf{p}_n^* + \mathbf{C}\mathbf{q}_n$. Moreover, for every $n \in \mathbb{N}$, (9.47)–(9.52) entail that

$$\sum_{i\in I} \xi_{i,n} + \sum_{k\in K} \eta_{k,n}$$

$$
\begin{aligned}
&= \sum_{i\in I} \xi_{i,\overline{\vartheta}_i(n)} + \sum_{k\in K} \eta_{k,\overline{\varrho}_k(n)} \\
&= \sum_{i\in I} \big\|a_{i,\overline{\vartheta}_i(n)} - x_{i,\pi_i(\overline{\vartheta}_i(n))}\big\|^2 + \sum_{k\in K} \Big(\big\|b_{k,\overline{\varrho}_k(n)} - y_{k,\omega_k(\overline{\varrho}_k(n))}\big\|^2 + \big\|d_{k,\overline{\varrho}_k(n)} - z_{k,\omega_k(\overline{\varrho}_k(n))}\big\|^2\Big) \\
&= \sum_{i\in I} \big\|a_{i,n} - x_{i,\vartheta_i(n)}\big\|^2 + \sum_{k\in K} \Big(\big\|b_{k,n} - y_{k,\varrho_k(n)}\big\|^2 + \big\|d_{k,n} - z_{k,\varrho_k(n)}\big\|^2\Big) \\
&= \|\mathbf{p}_n - \mathbf{q}_n\|^2
\end{aligned}
\tag{9.58}
$$

and, in turn, that

$$
\Delta_n = \langle \mathbf{x}_n - \mathbf{p}_n \mid \mathbf{t}_n^* \rangle - (4\alpha)^{-1}\|\mathbf{p}_n - \mathbf{q}_n\|^2. \tag{9.59}
$$

To sum up, (9.47) is an instantiation of (9.42). Therefore, Proposition 9.7(ii) asserts that

$$
\sum_{n\in\mathbb{N}} \|\mathbf{x}_{n+1} - \mathbf{x}_n\|^2 < +\infty. \tag{9.60}
$$

(i)&(ii): These follow from (9.60) and (9.52).

(iii)&(iv): Proposition 9.7(i) implies that $(\mathbf{x}_n)_{n\in\mathbb{N}}$ is bounded. It therefore results from (9.52) that

$$
(\boldsymbol{x}_n)_{n\in\mathbb{N}},\ \ (\boldsymbol{y}_n)_{n\in\mathbb{N}},\ \ (\boldsymbol{z}_n)_{n\in\mathbb{N}},\ \ \text{and}\ \ (\boldsymbol{v}_n^*)_{n\in\mathbb{N}}\ \ \text{are bounded.} \tag{9.61}
$$

Hence, (9.51), (9.47), (9.49), and Assumption 9.9[b] ensure that

$$
(\forall k \in K)\quad (e_{k,n}^*)_{n\in\mathbb{N}} = \left(\sigma_{k,\varrho_k(n)}\left(\sum_{i\in I} L_{ki}x_{i,\varrho_k(n)} - y_{k,\varrho_k(n)} - z_{k,\varrho_k(n)} - r_k\right) + v_{k,\varrho_k(n)}^*\right)_{n\in\mathbb{N}} \text{ is bounded.} \tag{9.62}
$$

Next, we deduce from (9.61) and Problem 9.1[e] that

$$
(\forall i \in I)\quad \big(R_i\boldsymbol{x}_{\vartheta_i(n)}\big)_{n\in\mathbb{N}}\ \text{is bounded.} \tag{9.63}
$$

In turn, it follows from (9.47), (9.61), the fact that $(Q_i)_{i\in I}$ and $(C_i)_{i\in I}$ are Lipschitzian, and Assumption 9.9[a] that

$$
(\forall i \in I)\quad \Big(x_{i,\vartheta_i(n)} + \gamma_{i,\vartheta_i(n)}\big(s_i^* - l_{i,\overline{\vartheta}_i(n)}^* - C_i x_{i,\vartheta_i(n)}\big)\Big)_{n\in\mathbb{N}}\ \text{is bounded.} \tag{9.64}
$$

An inspection of (9.50), (9.47), (9.48), and Lemma 9.25 reveals that

$$
(\forall i \in I)\quad (a_{i,n})_{n\in\mathbb{N}} = \Big(J_{\gamma_{i,\vartheta_i(n)}A_i}\big(x_{i,\vartheta_i(n)} + \gamma_{i,\vartheta_i(n)}\big(s_i^* - l_{i,\overline{\vartheta}_i(n)}^* - C_i x_{i,\vartheta_i(n)}\big)\big)\Big)_{n\in\mathbb{N}}\ \text{is bounded.} \tag{9.65}
$$

Hence, we infer from (9.50), (9.47), (9.61), and Assumption 9.9[a] that

$$
(\forall i \in I)\quad (a_{i,n}^*)_{n\in\mathbb{N}}\ \text{is bounded.} \tag{9.66}
$$

Accordingly, by (9.47), (9.61), and Assumption 9.9[b],

$$(\forall k\in K)\quad \Big(y_{k,\varrho_k(n)}+\mu_{k,\varrho_k(n)}\big(u^*_{k,\overline{\varrho}_k(n)}-B^c_k y_{k,\varrho_k(n)}\big)\Big)_{n\in\mathbb{N}}\ \text{ is bounded.}\tag{9.67}$$

Therefore, (9.51), (9.47), (9.49), and Lemma 9.25 imply that

$$(\forall k\in K)\quad (b_{k,n})_{n\in\mathbb{N}}=\Big(J_{\mu_{k,\varrho_k(n)}B^m_k}\big(y_{k,\varrho_k(n)}+\mu_{k,\varrho_k(n)}\big(u^*_{k,\overline{\varrho}_k(n)}-B^c_k y_{k,\varrho_k(n)}\big)\big)\Big)_{n\in\mathbb{N}}\ \text{ is bounded.}\tag{9.68}$$

Thus, (9.51), (9.47), (9.61), (9.62), and Assumption 9.9[b] yield

$$(\boldsymbol{q}^*_n)_{n\in\mathbb{N}}\ \text{ is bounded.}\tag{9.69}$$

Likewise,

$$(\boldsymbol{d}_n)_{n\in\mathbb{N}}\ \text{ and }\ (\boldsymbol{t}^*_n)_{n\in\mathbb{N}}\ \text{ are bounded.}\tag{9.70}$$

We deduce from (9.57), (9.68), (9.70), and (9.65) that

$$(\boldsymbol{e}_n)_{n\in\mathbb{N}}\ \text{ is bounded.}\tag{9.71}$$

On the other hand, (9.47), (9.66), (9.65), Problem 9.1[e], and (9.62) imply that

$$(\boldsymbol{p}^*_n)_{n\in\mathbb{N}}\ \text{ is bounded.}\tag{9.72}$$

Hence, we infer from (9.52) and (9.69)–(9.71) that $(\mathbf{t}^*_n)_{n\in\mathbb{N}}$ is bounded. Consequently, (9.59) and Proposition 9.7(iii) yield

$$\varlimsup\big(\langle\mathbf{x}_n-\mathbf{p}_n\mid\mathbf{t}^*_n\rangle-(4\alpha)^{-1}\|\mathbf{p}_n-\mathbf{q}_n\|^2\big)=\varlimsup\Delta_n\leqslant 0.\tag{9.73}$$

Let $\boldsymbol{L}$ and $\mathbf{W}$ be as in (9.24) and (9.27). For every $n\in\mathbb{N}$, set

$$\begin{cases}(\forall i\in I)\ \ E_{i,n}=\gamma^{-1}_{i,\vartheta_i(n)}\mathrm{Id}-Q_i\\ (\forall k\in K)\ \ F_{k,n}=\mu^{-1}_{k,\varrho_k(n)}\mathrm{Id}-B^\ell_k,\ \ G_{k,n}=\nu^{-1}_{k,\varrho_k(n)}\mathrm{Id}-D^\ell_k\\ \mathbf{E}_n\colon\boldsymbol{\mathcal{X}}\to\boldsymbol{\mathcal{X}}\colon(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{v}^*)\mapsto\big((E_{i,n}x_i)_{i\in I},(F_{k,n}y_k)_{k\in K},(G_{k,n}z_k)_{k\in K},(\sigma^{-1}_{k,\varrho_k(n)}v^*_k)_{k\in K}\big)\end{cases}\tag{9.74}$$

and

$$\begin{cases}\widetilde{\mathbf{x}}_n=\big((x_{i,\vartheta_i(n)})_{i\in I},(y_{k,\varrho_k(n)})_{k\in K},(z_{k,\varrho_k(n)})_{k\in K},(v^*_{k,\varrho_k(n)})_{k\in K}\big)\\ \mathbf{v}^*_n=\mathbf{E}_n\mathbf{x}_n-\mathbf{E}_n\mathbf{p}_n,\ \ \mathbf{w}^*_n=\mathbf{W}\mathbf{p}_n-\mathbf{W}\mathbf{x}_n\\ \mathbf{r}^*_n=\big((R_i\boldsymbol{a}_n-R_i\boldsymbol{x}_n)_{i\in I},\mathbf{0},\mathbf{0},\mathbf{0}\big),\ \ \widetilde{\mathbf{r}}^*_n=\big((R_i\boldsymbol{a}_n-R_i\boldsymbol{x}_{\vartheta_i(n)})_{i\in I},\mathbf{0},\mathbf{0},\mathbf{0}\big)\\ \mathbf{l}^*_n=\Big(\big(-\sum_{k\in K}L^*_{ki}v^*_{k,\vartheta_i(n)}\big)_{i\in I},\big(v^*_{k,\varrho_k(n)}\big)_{k\in K},\big(v^*_{k,\varrho_k(n)}\big)_{k\in K},\\ \qquad\qquad\big(\sum_{i\in I}L_{ki}x_{i,\varrho_k(n)}-y_{k,\varrho_k(n)}-z_{k,\varrho_k(n)}\big)_{k\in K}\Big).\end{cases}\tag{9.75}$$

In view of Problem 9.1[a]–[c] and Assumption 9.9[a]&[b], we deduce from Lemma 9.26 that

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \text{the operators } (E_{i,n})_{i \in I} \text{ are } (\chi+\sigma)\text{-strongly monotone} \\ \text{the operators } (F_{k,n})_{k \in K} \text{ and } (G_{k,n})_{k \in K} \text{ are } \sigma\text{-strongly monotone,} \end{cases} \tag{9.76}$$

and from (9.74) that there exists $\kappa \in \left]0,+\infty\right[$ such that

$$\text{the operators } (\boldsymbol{\mathsf{E}}_n)_{n\in\mathbb{N}} \text{ are } \kappa\text{-Lipschitzian.} \tag{9.77}$$

It results from (9.50), (9.47), (9.48), and (9.74) that

$$\begin{aligned}(\forall i \in I)(\forall n \in \mathbb{N}) \quad a^*_{i,n} &= a^*_{i,\overline{\vartheta}_i(n)} \\ &= \Big(\gamma^{-1}_{i,\pi_i(\overline{\vartheta}_i(n))} x_{i,\pi_i(\overline{\vartheta}_i(n))} - Q_i x_{i,\pi_i(\overline{\vartheta}_i(n))}\Big) - \Big(\gamma^{-1}_{i,\pi_i(\overline{\vartheta}_i(n))} a_{i,\overline{\vartheta}_i(n)} - Q_i a_{i,\overline{\vartheta}_i(n)}\Big) \\ &\quad - R_i \boldsymbol{x}_{\pi_i(\overline{\vartheta}_i(n))} - \sum_{k\in K} L^*_{ki} v^*_{k,\pi_i(\overline{\vartheta}_i(n))} \\ &= E_{i,n} x_{i,\vartheta_i(n)} - E_{i,n} a_{i,n} - R_i \boldsymbol{x}_{\vartheta_i(n)} - \sum_{k\in K} L^*_{ki} v^*_{k,\vartheta_i(n)} \end{aligned} \tag{9.78}$$

and, therefore, that

$$\begin{aligned}(\forall i \in I)(\forall n \in \mathbb{N}) \quad p^*_{i,n} &= a^*_{i,n} + R_i \boldsymbol{a}_n + \sum_{k\in K} L^*_{ki} e^*_{k,n} \\ &= E_{i,n} x_{i,\vartheta_i(n)} - E_{i,n} a_{i,n} + R_i \boldsymbol{a}_n - R_i \boldsymbol{x}_{\vartheta_i(n)} - \sum_{k\in K} L^*_{ki} v^*_{k,\vartheta_i(n)} + \sum_{k\in K} L^*_{ki} e^*_{k,n}. \end{aligned} \tag{9.79}$$

At the same time, (9.51), (9.47), (9.49), and (9.74) entail that

$$\begin{aligned}(\forall k \in K)(\forall n \in \mathbb{N}) \quad q^*_{k,n} &= q^*_{k,\overline{\varrho}_k(n)} \\ &= \Big(\mu^{-1}_{k,\omega_k(\overline{\varrho}_k(n))} y_{k,\omega_k(\overline{\varrho}_k(n))} - B^\ell_k y_{k,\omega_k(\overline{\varrho}_k(n))}\Big) \\ &\quad - \Big(\mu^{-1}_{k,\omega_k(\overline{\varrho}_k(n))} b_{k,\overline{\varrho}_k(n)} - B^\ell_k b_{k,\overline{\varrho}_k(n)}\Big) + v^*_{k,\omega_k(\overline{\varrho}_k(n))} - e^*_{k,\overline{\varrho}_k(n)} \\ &= F_{k,n} y_{k,\varrho_k(n)} - F_{k,n} b_{k,n} + v^*_{k,\varrho_k(n)} - e^*_{k,n} \end{aligned} \tag{9.80}$$

and that

$$(\forall k \in K)(\forall n \in \mathbb{N}) \quad t^*_{k,n} = G_{k,n} z_{k,\varrho_k(n)} - G_{k,n} d_{k,n} + v^*_{k,\varrho_k(n)} - e^*_{k,n}. \tag{9.81}$$

Further, we derive from (9.51), (9.47), and (9.49) that

$$(\forall k \in K)(\forall n \in \mathbb{N}) \quad r_k = \sigma^{-1}_{k,\varrho_k(n)} v^*_{k,\varrho_k(n)} - \sigma^{-1}_{k,\varrho_k(n)} e^*_{k,n} - y_{k,\varrho_k(n)} - z_{k,\varrho_k(n)} + \sum_{i\in I} L_{ki} x_{i,\varrho_k(n)} \tag{9.82}$$

and, in turn, from (9.57) that

$$(\forall k \in K)(\forall n \in \mathbb{N}) \quad e_{k,n} = \sigma_{k,\varrho_k(n)}^{-1} v_{k,\varrho_k(n)}^* - \sigma_{k,\varrho_k(n)}^{-1} e_{k,n}^* - y_{k,\varrho_k(n)} - z_{k,\varrho_k(n)} + \sum_{i \in I} L_{ki} x_{i,\varrho_k(n)} + b_{k,n} + d_{k,n} - \sum_{i \in I} L_{ki} a_{i,n}. \tag{9.83}$$

Altogether, it follows from (9.52), (9.79)–(9.81), (9.83), (9.74), (9.75), (9.27), and (9.25) that

$$(\forall n \in \mathbb{N}) \quad \mathbf{t}_n^* = \mathbf{E}_n \widetilde{\mathbf{x}}_n - \mathbf{E}_n \mathbf{p}_n + \widetilde{\mathbf{r}}_n^* + \mathbf{l}_n^* + \mathbf{W} \mathbf{p}_n. \tag{9.84}$$

Next, in view of (9.60), (9.48), (9.49), and Assumption 9.10–9.11, we learn from Lemma 9.27 that

$$(\forall i \in I)(\forall k \in K) \quad \begin{cases} \boldsymbol{x}_{\vartheta_i(n)} - \boldsymbol{x}_n \to \mathbf{0}, \;\; \boldsymbol{x}_{\varrho_k(n)} - \boldsymbol{x}_n \to \mathbf{0}, \;\; \text{and} \;\; \boldsymbol{v}_{\vartheta_i(n)}^* - \boldsymbol{v}_n^* \to \mathbf{0} \\ \boldsymbol{y}_{\varrho_k(n)} - \boldsymbol{y}_n \to \mathbf{0}, \;\; \boldsymbol{z}_{\varrho_k(n)} - \boldsymbol{z}_n \to \mathbf{0}, \;\; \text{and} \;\; \boldsymbol{v}_{\varrho_k(n)}^* - \boldsymbol{v}_n^* \to \mathbf{0}. \end{cases} \tag{9.85}$$

Thus, (9.75), (9.27), (9.25), and (9.24) yield

$$\mathbf{l}_n^* + \mathbf{W} \mathbf{x}_n \to \mathbf{0}, \tag{9.86}$$

while Problem 9.1[e] gives

$$(\forall i \in I) \quad \|R_i \boldsymbol{x}_{\vartheta_i(n)} - R_i \boldsymbol{x}_n\| \leqslant \chi \|\boldsymbol{x}_{\vartheta_i(n)} - \boldsymbol{x}_n\| \to 0. \tag{9.87}$$

On the other hand, we infer from (9.77), (9.75), and (9.85) that

$$\|\mathbf{E}_n \widetilde{\mathbf{x}}_n - \mathbf{E}_n \mathbf{x}_n\| \leqslant \kappa \|\widetilde{\mathbf{x}}_n - \mathbf{x}_n\| \to 0. \tag{9.88}$$

Combining (9.84), (9.75), and (9.86)–(9.88), we obtain

$$\mathbf{t}_n^* - \big(\mathbf{v}_n^* + \mathbf{r}_n^* + \mathbf{w}_n^*\big) = \mathbf{l}_n^* + \mathbf{W} \mathbf{x}_n + \mathbf{E}_n \widetilde{\mathbf{x}}_n - \mathbf{E}_n \mathbf{x}_n + \widetilde{\mathbf{r}}_n^* - \mathbf{r}_n^* \to \mathbf{0}. \tag{9.89}$$

Now set

$$(\forall n \in \mathbb{N}) \quad \widetilde{\mathbf{q}}_n = (\boldsymbol{x}_n, \boldsymbol{y}_n, \boldsymbol{z}_n, \boldsymbol{e}_n^*). \tag{9.90}$$

Then $(\widetilde{\mathbf{q}}_n)_{n \in \mathbb{N}}$ is bounded by virtue of (9.61) and (9.62). On the one hand, (9.52), (9.62), (9.65), (9.68), and (9.70) imply that $(\mathbf{p}_n)_{n \in \mathbb{N}}$ is bounded. On the other hand, (9.52) and (9.85) give

$$\widetilde{\mathbf{q}}_n - \mathbf{q}_n \to \mathbf{0}. \tag{9.91}$$

Therefore, appealing to the Cauchy–Schwarz inequality, we obtain

$$\big|\langle \mathbf{p}_n - \widetilde{\mathbf{q}}_n \mid \widetilde{\mathbf{q}}_n - \mathbf{q}_n \rangle\big| \leqslant \left(\sup_{m \in \mathbb{N}} \|\mathbf{p}_m\| + \sup_{m \in \mathbb{N}} \|\widetilde{\mathbf{q}}_m\| \right) \|\widetilde{\mathbf{q}}_n - \mathbf{q}_n\| \to 0 \tag{9.92}$$

and, by (9.89),

$$\big|\langle \mathbf{x}_n-\mathbf{p}_n \mid \mathbf{t}_n^*-(\mathbf{v}_n^*+\mathbf{r}_n^*+\mathbf{w}_n^*)\rangle\big| \leqslant \Big(\sup_{m\in\mathbb{N}}\|\mathbf{x}_m\|+\sup_{m\in\mathbb{N}}\|\mathbf{p}_m\|\Big)\|\mathbf{t}_n^*-(\mathbf{v}_n^*+\mathbf{r}_n^*+\mathbf{w}_n^*)\|\to 0. \tag{9.93}$$

However, since $\mathbf{W}^*=-\mathbf{W}$ by (9.27), it results from (9.75) that $(\forall n\in\mathbb{N})$ $\langle \mathbf{x}_n-\mathbf{p}_n \mid \mathbf{w}_n^*\rangle=0$. Thus, by (9.73) and (9.91)–(9.93),

$$\begin{aligned}
0 &\geqslant \varlimsup\big(\langle \mathbf{x}_n-\mathbf{p}_n \mid \mathbf{t}_n^*\rangle-(4\alpha)^{-1}\|\mathbf{p}_n-\mathbf{q}_n\|^2\big)\\
&=\varlimsup\big(\langle \mathbf{x}_n-\mathbf{p}_n \mid \mathbf{v}_n^*+\mathbf{r}_n^*+\mathbf{w}_n^*\rangle+\langle \mathbf{x}_n-\mathbf{p}_n \mid \mathbf{t}_n^*-(\mathbf{v}_n^*+\mathbf{r}_n^*+\mathbf{w}_n^*)\rangle-(4\alpha)^{-1}\|\mathbf{p}_n-\mathbf{q}_n\|^2\big)\\
&=\varlimsup\big(\langle \mathbf{x}_n-\mathbf{p}_n \mid \mathbf{v}_n^*+\mathbf{r}_n^*\rangle-(4\alpha)^{-1}\big(\|\mathbf{p}_n-\widetilde{\mathbf{q}}_n\|^2+2\langle \mathbf{p}_n-\widetilde{\mathbf{q}}_n \mid \widetilde{\mathbf{q}}_n-\mathbf{q}_n\rangle+\|\widetilde{\mathbf{q}}_n-\mathbf{q}_n\|^2\big)\big)\\
&=\varlimsup\big(\langle \mathbf{x}_n-\mathbf{p}_n \mid \mathbf{v}_n^*+\mathbf{r}_n^*\rangle-(4\alpha)^{-1}\|\mathbf{p}_n-\widetilde{\mathbf{q}}_n\|^2\big).
\end{aligned}\tag{9.94}$$

On the other hand, we deduce from (9.75), (9.52), (9.74), (9.76), Assumption 9.9[b], the Cauchy–Schwarz inequality, Problem 9.1[e], and (9.90) that, for every $n\in\mathbb{N}$,

$$\begin{aligned}
&\langle \mathbf{x}_n-\mathbf{p}_n \mid \mathbf{v}_n^*+\mathbf{r}_n^*\rangle-(4\alpha)^{-1}\|\mathbf{p}_n-\widetilde{\mathbf{q}}_n\|^2\\
&\quad=\langle \mathbf{x}_n-\mathbf{p}_n \mid \mathbf{E}_n\mathbf{x}_n-\mathbf{E}_n\mathbf{p}_n\rangle+\langle \mathbf{x}_n-\mathbf{p}_n \mid \mathbf{r}_n^*\rangle-(4\alpha)^{-1}\|\mathbf{p}_n-\widetilde{\mathbf{q}}_n\|^2\\
&\quad=\sum_{i\in I}\langle x_{i,n}-a_{i,n} \mid E_{i,n}x_{i,n}-E_{i,n}a_{i,n}\rangle+\sum_{k\in K}\langle y_{k,n}-b_{k,n} \mid F_{k,n}y_{k,n}-F_{k,n}b_{k,n}\rangle\\
&\quad\quad+\sum_{k\in K}\langle z_{k,n}-d_{k,n} \mid G_{k,n}z_{k,n}-G_{k,n}d_{k,n}\rangle+\sum_{k\in K}\sigma_{k,\varrho_k(n)}^{-1}\|v_{k,n}^*-e_{k,n}^*\|^2\\
&\quad\quad+\langle \boldsymbol{x}_n-\boldsymbol{a}_n \mid \boldsymbol{R}\boldsymbol{a}_n-\boldsymbol{R}\boldsymbol{x}_n\rangle-(4\alpha)^{-1}\|\mathbf{p}_n-\widetilde{\mathbf{q}}_n\|^2\\
&\quad\geqslant(\chi+\sigma)\|\boldsymbol{x}_n-\boldsymbol{a}_n\|^2+\sigma\|\boldsymbol{y}_n-\boldsymbol{b}_n\|^2+\sigma\|\boldsymbol{z}_n-\boldsymbol{d}_n\|^2\\
&\quad\quad+\varepsilon\|\boldsymbol{v}_n^*-\boldsymbol{e}_n^*\|^2-\|\boldsymbol{x}_n-\boldsymbol{a}_n\|\,\|\boldsymbol{R}\boldsymbol{a}_n-\boldsymbol{R}\boldsymbol{x}_n\|-(4\alpha)^{-1}\|\mathbf{p}_n-\widetilde{\mathbf{q}}_n\|^2\\
&\quad\geqslant(\chi+\sigma)\|\boldsymbol{x}_n-\boldsymbol{a}_n\|^2+\sigma\|\boldsymbol{y}_n-\boldsymbol{b}_n\|^2+\sigma\|\boldsymbol{z}_n-\boldsymbol{d}_n\|^2\\
&\quad\quad+\varepsilon\|\boldsymbol{v}_n^*-\boldsymbol{e}_n^*\|^2-\chi\|\boldsymbol{x}_n-\boldsymbol{a}_n\|^2-(4\alpha)^{-1}\|\mathbf{p}_n-\widetilde{\mathbf{q}}_n\|^2\\
&\quad=\big(\sigma-(4\alpha)^{-1}\big)\big(\|\boldsymbol{x}_n-\boldsymbol{a}_n\|^2+\|\boldsymbol{y}_n-\boldsymbol{b}_n\|^2+\|\boldsymbol{z}_n-\boldsymbol{d}_n\|^2\big)+\varepsilon\|\boldsymbol{v}_n^*-\boldsymbol{e}_n^*\|^2.
\end{aligned}\tag{9.95}$$

Hence, since $\sigma>1/(4\alpha)$ by (9.45), taking the limit superior in (9.95) and invoking (9.94) yield

$$\boldsymbol{x}_n-\boldsymbol{a}_n\to\boldsymbol{0},\ \ \boldsymbol{y}_n-\boldsymbol{b}_n\to\boldsymbol{0},\ \ \boldsymbol{z}_n-\boldsymbol{d}_n\to\boldsymbol{0},\ \ \text{and}\ \ \boldsymbol{v}_n^*-\boldsymbol{e}_n^*\to\boldsymbol{0}, \tag{9.96}$$

which establishes (iii). In turn, (9.52) and (9.77) force

$$\mathbf{x}_n-\mathbf{p}_n\to\mathbf{0}\quad\text{and}\quad\|\mathbf{E}_n\mathbf{x}_n-\mathbf{E}_n\mathbf{p}_n\|\leqslant\kappa\|\mathbf{x}_n-\mathbf{p}_n\|\to 0 \tag{9.97}$$

and (9.85) thus yields $\mathbf{p}_n-\mathbf{q}_n\to\mathbf{0}$. Further, we infer from (9.75), (9.96), and Problem 9.1[e] that

$$\|\mathbf{r}_n^*\|^2=\|\boldsymbol{R}\boldsymbol{a}_n-\boldsymbol{R}\boldsymbol{x}_n\|^2\leqslant\chi^2\|\boldsymbol{a}_n-\boldsymbol{x}_n\|^2\to 0. \tag{9.98}$$

Altogether, it follows from (9.75), (9.89), (9.97), and (9.98) that

$$\mathbf{t}_n^* = \big(\mathbf{t}_n^* - (\mathbf{v}_n^* + \mathbf{r}_n^* + \mathbf{w}_n^*)\big) + \big(\mathbf{E}_n\mathbf{x}_n - \mathbf{E}_n\mathbf{p}_n\big) + \mathbf{W}(\mathbf{p}_n - \mathbf{x}_n) + \mathbf{r}_n^* \to \mathbf{0}. \tag{9.99}$$

Hence, Proposition 9.7(iv) guarantees that there exists $\overline{\mathbf{x}} = (\overline{\boldsymbol{x}}, \overline{\boldsymbol{y}}, \overline{\boldsymbol{z}}, \overline{\boldsymbol{v}}^*) \in \operatorname{zer} \mathcal{S}$ such that $\mathbf{x}_n \rightharpoonup \overline{\mathbf{x}}$. This and (9.96) imply that, for every $i \in I$ and every $k \in K$, $x_{i,n} \rightharpoonup \overline{x}_i$, $a_{i,n} \rightharpoonup \overline{x}_i$, and $v_{k,n}^* \rightharpoonup \overline{v}_k^*$. Finally, Proposition 9.3(iii) asserts that $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}}^*)$ lies in the set of Kuhn–Tucker points (9.23), that $\overline{\boldsymbol{x}}$ solves (9.1), and that $\overline{\boldsymbol{v}}^*$ solves (9.2). $\square$

Some infinite-dimensional applications require strong convergence of the iterates; see, e.g., [3, 4]. This will be guaranteed by the following variant of Algorithm 9.12, which hinges on the principle outlined in Proposition 9.8.

Algorithm 9.14 Consider the setting of Problem 9.1, define Ξ as in (9.43), and suppose that Assumption 9.9–9.11 is in force. 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}^* = 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)}^*; \\
a_{i,n} = J_{\gamma_{i,\pi_i(n)} A_i}\big(x_{i,\pi_i(n)} + \gamma_{i,\pi_i(n)}(s_i^* - l_{i,n}^* - C_i x_{i,\pi_i(n)})\big); \\
a_{i,n}^* = \gamma_{i,\pi_i(n)}^{-1}(x_{i,\pi_i(n)} - a_{i,n}) - l_{i,n}^* + Q_i a_{i,n}; \\
\xi_{i,n} = \|a_{i,n} - x_{i,\pi_i(n)}\|^2;
\end{array}
\right. \\
\text{for every } i \in I \smallsetminus I_n \\
\left\lfloor
\begin{array}{l}
a_{i,n} = a_{i,n-1};\ a_{i,n}^* = a_{i,n-1}^*;\ \xi_{i,n} = \xi_{i,n-1};
\end{array}
\right. \\
\text{for every } k \in K_n \\
\left\lfloor
\begin{array}{l}
u_{k,n}^* = v_{k,\omega_k(n)}^* - B_k^\ell y_{k,\omega_k(n)}; \\
w_{k,n}^* = v_{k,\omega_k(n)}^* - D_k^\ell z_{k,\omega_k(n)}; \\
b_{k,n} = J_{\mu_{k,\omega_k(n)} B_k^m}\big(y_{k,\omega_k(n)} + \mu_{k,\omega_k(n)}(u_{k,n}^* - B_k^c y_{k,\omega_k(n)})\big); \\
d_{k,n} = J_{\nu_{k,\omega_k(n)} D_k^m}\big(z_{k,\omega_k(n)} + \nu_{k,\omega_k(n)}(w_{k,n}^* - D_k^c z_{k,\omega_k(n)})\big); \\
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)} - r_k\big) + v_{k,\omega_k(n)}^*; \\
q_{k,n}^* = \mu_{k,\omega_k(n)}^{-1}(y_{k,\omega_k(n)} - b_{k,n}) + u_{k,n}^* + B_k^\ell b_{k,n} - e_{k,n}^*; \\
t_{k,n}^* = \nu_{k,\omega_k(n)}^{-1}(z_{k,\omega_k(n)} - d_{k,n}) + w_{k,n}^* + D_k^\ell d_{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} = r_k + b_{k,n} + d_{k,n} - \sum_{i\in I} L_{ki} a_{i,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} = r_k + b_{k,n} + d_{k,n} - \sum_{i\in I} L_{ki} a_{i,n};
\end{array}
\right. \\
\text{for every } i \in I \\
\left\lfloor
\begin{array}{l}
p_{i,n}^* = a_{i,n}^* + R_i \boldsymbol{a}_n + \sum_{k\in K} L_{ki}^* e_{k,n}^*;
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{9.100}
$$

$$
\begin{array}{l}
\Delta_n = -(4\alpha)^{-1}\big(\sum_{i\in I}\xi_{i,n} + \sum_{k\in K}\eta_{k,n}\big) + \sum_{i\in I}\langle x_{i,n} - a_{i,n} \mid p^*_{i,n}\rangle \\
\qquad + \sum_{k\in K}\big(\langle y_{k,n} - b_{k,n} \mid q^*_{k,n}\rangle + \langle z_{k,n} - d_{k,n} \mid t^*_{k,n}\rangle + \langle e_{k,n} \mid v^*_{k,n} - e^*_{k,n}\rangle\big); \\
\text{if } \Delta_n > 0 \\
\quad \tau_n = \sum_{i\in I}\|p^*_{i,n}\|^2 + \sum_{k\in K}\big(\|q^*_{k,n}\|^2 + \|t^*_{k,n}\|^2 + \|e_{k,n}\|^2\big); \\
\quad \varsigma_n = \sum_{i\in I}\|x_{i,0} - x_{i,n}\|^2 \\
\qquad\quad + \sum_{k\in K}\big(\|y_{k,0} - y_{k,n}\|^2 + \|z_{k,0} - z_{k,n}\|^2 + \|v^*_{k,0} - v^*_{k,n}\|^2\big); \\
\quad \chi_n = \sum_{i\in I}\langle x_{i,0} - x_{i,n} \mid p^*_{i,n}\rangle \\
\qquad\quad + \sum_{k\in K}\big(\langle y_{k,0} - y_{k,n} \mid q^*_{k,n}\rangle + \langle z_{k,0} - z_{k,n} \mid t^*_{k,n}\rangle + \langle e_{k,n} \mid v^*_{k,0} - v^*_{k,n}\rangle\big); \\
\quad (\kappa_n, \lambda_n) = \Xi(\Delta_n, \tau_n, \varsigma_n, \chi_n); \\
\quad \text{for every } i \in I \\
\qquad x_{i,n+1} = (1-\kappa_n)x_{i,0} + \kappa_n x_{i,n} - \lambda_n p^*_{i,n}; \\
\quad \text{for every } k \in K \\
\qquad y_{k,n+1} = (1-\kappa_n)y_{k,0} + \kappa_n y_{k,n} - \lambda_n q^*_{k,n}; \\
\qquad z_{k,n+1} = (1-\kappa_n)z_{k,0} + \kappa_n z_{k,n} - \lambda_n t^*_{k,n}; \\
\qquad v^*_{k,n+1} = (1-\kappa_n)v^*_{k,0} + \kappa_n v^*_{k,n} - \lambda_n e_{k,n}; \\
\text{else} \\
\quad \text{for every } i \in I \\
\qquad x_{i,n+1} = x_{i,n}; \\
\quad \text{for every } k \in K \\
\qquad y_{k,n+1} = y_{k,n};\ z_{k,n+1} = z_{k,n};\ v^*_{k,n+1} = v^*_{k,n}.
\end{array}
$$

Theorem 9.15 *Consider the setting of Algorithm 9.14 and suppose that the dual problem (9.2) has a solution. Then the following hold:*

(i) *Let $i \in I$. Then $\sum_{n\in\mathbb{N}}\|x_{i,n+1} - x_{i,n}\|^2 < +\infty$.*

(ii) *Let $k \in K$. Then $\sum_{n\in\mathbb{N}}\|y_{k,n+1} - y_{k,n}\|^2 < +\infty$, $\sum_{n\in\mathbb{N}}\|z_{k,n+1} - z_{k,n}\|^2 < +\infty$, and $\sum_{n\in\mathbb{N}}\|v^*_{k,n+1} - v^*_{k,n}\|^2 < +\infty$.*

(iii) *Let $i \in I$ and $k \in K$. Then $x_{i,n} - a_{i,n} \to 0$, $y_{k,n} - b_{k,n} \to 0$, $z_{k,n} - d_{k,n} \to 0$, and $v^*_{k,n} - e^*_{k,n} \to 0$.*

(iv) *There exist a solution $\overline{\boldsymbol{x}}$ to (9.1) and a solution $\overline{\boldsymbol{v}}^*$ to (9.2) such that, for every $i \in I$ and every $k \in K$, $x_{i,n} \to \overline{x}_i$, $a_{i,n} \to \overline{x}_i$, and $v^*_{k,n} \to \overline{v}^*_k$. In addition, $(\overline{\boldsymbol{x}}, \overline{\boldsymbol{v}}^*)$ is a Kuhn–Tucker point of Problem 9.1 in the sense of (9.23).*

Proof. Proceed as in the proof of Theorem 9.13 and use Proposition 9.8 instead of Proposition 9.7. ☐

9.2.4 Applications

In nonlinear analysis and optimization, problems with multiple variables occur in areas such as game theory [2,15,56], evolution inclusions [3], traffic equilibrium [3,31], domain decomposition [4], machine learning [6,12], image recovery [13,16], infimal-convolution regularization

[23], statistics [26, 55], neural networks [27], and variational inequalities [31]. The numerical methods used in the above papers are limited to special cases of Problem 9.1 and they do not perform block iterations and they operate in synchronous mode. The methods presented in Theorems 9.13 and 9.15 provide a unified treatment of these problems as well as extensions, within a considerably more flexible algorithmic framework. In this section, we illustrate this in the context of variational inequalities and multivariate minimization. Below we present only the applications of Theorem 9.13 as similar applications of Theorem 9.15 follow using similar arguments.

9.2.4.1 Application to variational inequalities

The standard variational inequality problem associated with a closed convex subset D of a real Hilbert space $\mathcal{G}$ and a maximally monotone operator $B\colon \mathcal{G} \to \mathcal{G}$ is to

$$\text{find } \overline{y} \in D \text{ such that } (\forall y \in D)\ \langle \overline{y} - y \mid B\overline{y} \rangle \leqslant 0. \tag{9.101}$$

Classical methods require the ability to project onto D and specific assumptions on B such as cocoercivity, Lipschitz continuity, or the ability to compute the resolvent [9, 30, 53]. Let us consider a refined version of (9.101) in which B and D are decomposed into basic components, and for which these classical methods are not applicable.

Problem 9.16 Let I be a nonempty finite set and let $(\mathcal{H}_i)_{i\in I}$ and $\mathcal{G}$ be real Hilbert spaces. For every $i \in I$, let E_i and F_i be closed convex subsets of $\mathcal{H}_i$ such that $E_i \cap F_i \neq \varnothing$ and let $L_i\colon \mathcal{H}_i \to \mathcal{G}$ be linear and bounded. In addition, let $B^m\colon \mathcal{G} \to 2^{\mathcal{G}}$ be at most single-valued and maximally monotone, let $B^c\colon \mathcal{G} \to \mathcal{G}$ be cocoercive with constant $\beta^c \in \,]0, +\infty[$, and let $B^\ell\colon \mathcal{G} \to \mathcal{G}$ be Lipschitzian with constant $\beta^\ell \in [0, +\infty[$. The objective is to

$$\text{find } \overline{y} \in \sum_{i\in I} L_i(E_i \cap F_i) \text{ such that } \left(\forall y \in \sum_{i\in I} L_i(E_i \cap F_i)\right)\ \langle \overline{y} - y \mid B^m\overline{y} + B^c\overline{y} + B^\ell\overline{y} \rangle \leqslant 0. \tag{9.102}$$

To motivate our analysis, let us consider an illustration of (9.102).

Example 9.17 Let I be a nonempty finite set and let $(\mathcal{Z}_i)_{i\in I}$ and $\mathcal{K}$ be real Hilbert spaces. For every $i \in I$, let $S_i \subset \mathcal{Z}_i$ be closed and convex, and let $M_i\colon \mathcal{Z}_i \to \mathcal{K}$ be linear and bounded. In addition, let $f \in \Gamma_0(\mathcal{K})$ be Gâteaux differentiable on $\operatorname{dom}\partial f$, let $\varphi\colon \mathcal{K} \to \mathbb{R}$ be convex and differentiable with a Lipschitzian gradient, let $\mathcal{V}$ be a real Hilbert space, let $g \in \Gamma_0(\mathcal{V})$ be such that g^* is Gâteaux differentiable on $\operatorname{dom}\partial g^*$, let D be a closed convex subset of $\mathcal{V}$ such that

$$0 \in \operatorname{sri}(D - \operatorname{dom} g^*), \tag{9.103}$$

let $h \in \Gamma_0(\mathcal{V})$ be strongly convex, and let $L\colon \mathcal{K} \to \mathcal{V}$ be linear and bounded. Note that, by [9, Theorem 18.15], h^* is differentiable on $\mathcal{V}$ and ∇h^* is cocoercive. The objective is to solve the Kuhn–Tucker problem

$$
\begin{aligned}
&\text{find } (\overline{x},\overline{v}^*)\in\mathcal{K}\oplus\mathcal{V} \text{ such that}\\
&\begin{bmatrix}0\\0\end{bmatrix}\in\underbrace{\begin{bmatrix}\nabla f & 0\\ 0 & \nabla g^*\end{bmatrix}}_{\text{monotone}}\begin{bmatrix}\overline{x}\\ \overline{v}^*\end{bmatrix}+\underbrace{\begin{bmatrix}\nabla \varphi & 0\\ 0 & \nabla h^*\end{bmatrix}}_{\text{cocoercive}}\begin{bmatrix}\overline{x}\\ \overline{v}^*\end{bmatrix}+\underbrace{\begin{bmatrix}0 & L^*\\ -L & 0\end{bmatrix}}_{\text{Lipschitzian}}\begin{bmatrix}\overline{x}\\ \overline{v}^*\end{bmatrix}+\underbrace{\begin{bmatrix}N_C & 0\\ 0 & N_D\end{bmatrix}}_{\text{normal cone}}\begin{bmatrix}\overline{x}\\ \overline{v}^*\end{bmatrix},
\end{aligned}
\tag{9.104}
$$

where it is assumed that

$$
C=\sum_{i\in I}M_i(S_i)\ \text{ is closed and }\ 0\in\operatorname{sri}(C-\operatorname{dom}f). \tag{9.105}
$$

Since $\operatorname{dom}h^*=\mathcal{V}$, we deduce from (9.103) and [9, Proposition 15.7(i)] that $g\,\square\,h\,\square\,\sigma_D\in\Gamma_0(\mathcal{V})$. It follows from standard convex calculus [9] that a solution $(\overline{x},\overline{v}^*)$ to (9.104) provides a solution $\overline{x}$ to

$$
\underset{x\in C}{\text{minimize}}\ \ f(x)+\big(g\,\square\,h\,\square\,\sigma_D\big)(Lx)+\varphi(x), \tag{9.106}
$$

as well as a solution $\overline{v}^*$ to the associated Fenchel–Rockafellar dual

$$
\underset{v^*\in D}{\text{minimize}}\ \ \big((f+\varphi)^*\,\square\,\sigma_C\big)(-L^*v^*)+g^*(v^*)+h^*(v^*). \tag{9.107}
$$

To see that (9.104)–(9.105) is a special case of Problem 9.16, set $\mathcal{G}=\mathcal{K}\oplus\mathcal{V}$ and

$$
(\forall i\in I)\quad L_i\colon\mathcal{H}_i=\mathcal{Z}_i\oplus\mathcal{V}\to\mathcal{G}\colon(z_i,v^*)\mapsto(M_iz_i,v^*/\operatorname{card}I),\ \ E_i=S_i\times D,\ \text{ and }\ F_i=\mathcal{Z}_i\times\mathcal{V}. \tag{9.108}
$$

Note that

$$
C\times D=\sum_{i\in I}L_i(E_i\cap F_i). \tag{9.109}
$$

Further, in view of [9, Proposition 17.31(i)], let us define

$$
\begin{cases}
B^m\colon\mathcal{G}\to 2^{\mathcal{G}}\\
\quad(x,v^*)\mapsto\partial(f\oplus g^*)(x,v^*)=\begin{cases}\big(\nabla f(x),\nabla g^*(v^*)\big), & \text{if } (x,v^*)\in\operatorname{dom}\partial f\times\operatorname{dom}\partial g^*;\\ \varnothing, & \text{otherwise}\end{cases}\\
B^c\colon\mathcal{G}\to\mathcal{G}\colon(x,v^*)\mapsto\big(\nabla\varphi(x),\nabla h^*(v^*)\big)\\
B^\ell\colon\mathcal{G}\to\mathcal{G}\colon(x,v^*)\mapsto(L^*v^*,-Lx).
\end{cases}
\tag{9.110}
$$

Then B^m is maximally monotone [9, Theorem 20.25], B^c is cocoercive [9, Corollary 18.17], and B^ℓ is a skew bounded linear operator, hence monotone and Lipschitzian [9, Example 20.35]. In turn, combining (9.108) and (9.110), we conclude that (9.104) can be written as

$$\text{find } (\overline{x}, \overline{v}^*) \in \mathcal{K} \oplus \mathcal{V} \text{ such that } (0,0) \in B^m(\overline{x}, \overline{v}^*) + B^c(\overline{x}, \overline{v}^*) + B^\ell(\overline{x}, \overline{v}^*) + N_{C \times D}(\overline{x}, \overline{v}^*) \tag{9.111}$$

which, in the light of (9.109), fits the format of (9.102). Special cases of (9.106) involving minimization over Minkowski sum of sets are found in areas such as signal and image processing [5, 28, 41], location and network problems [40], as well as robotics and computational mechanics [54].

We are going to reformulate Problem 9.16 as a realization of Problem 9.1 and solve it via a block-iterative method derived from Algorithm 9.12. In addition, our approach employs the individual projection operators onto the sets $(E_i)_{i \in I}$ and $(F_i)_{i \in I}$, and the resolvents of the operator B^m. We are not aware of any method which features such flexibility. For instance, consider the special case discussed in [31, Section 4], where $\mathcal{G} = \mathbb{R}^N$, $B^c = B^\ell = 0$, $T \colon \mathbb{R}^N \to \mathbb{R}^M$ is a linear operator, and, for every $i \in I$, $\mathcal{H}_i = \mathbb{R}^N$, $L_i = \mathrm{Id}$, $E_i = T^{-1}(\{d_i\})$ for some $d_i \in \mathbb{R}^M$, and $F_i = [0, +\infty[^N$. There, the evaluations of all the projectors $(\mathrm{proj}_{E_i \cap F_i})_{i \in I}$ are required at every iteration. Note that there are no closed-form expressions for $(\mathrm{proj}_{E_i \cap F_i})_{i \in I}$ in general.

Corollary 9.18 *Consider the setting of Problem 9.16. Let* $\sigma \in \left]1/(4\beta^c), +\infty\right[$, $\varepsilon \in \left]0, \min\{1, 1/(\beta^\ell + \sigma)\}\right[$, *and* $K = I \cup \{\overline{k}\}$, *where* $\overline{k} \notin I$. *Suppose that Assumption 9.10 is in force, together with the following:*

[a] *For every* $i \in I$ *and every* $n \in \mathbb{N}$, $\{\gamma_{i,n}, \mu_{i,n}, \nu_{i,n}\} \subset [\varepsilon, 1/\sigma]$ *and* $\sigma_{i,n} \in [\varepsilon, 1/\varepsilon]$.

[b] *For every* $n \in \mathbb{N}$, $\lambda_n \in [\varepsilon, 2 - \varepsilon]$, $\mu_{\overline{k},n} \in \left[\varepsilon, 1/(\beta^\ell + \sigma)\right]$, $\nu_{\overline{k},n} \in [\varepsilon, 1/\sigma]$, *and* $\sigma_{\overline{k},n} \in [\varepsilon, 1/\varepsilon]$.

[c] *For every* $i \in I$, $\{x_{i,0}, y_{i,0}, z_{i,0}, v^*_{i,0}\} \subset \mathcal{H}_i$; $\{y_{\overline{k},0}, z_{\overline{k},0}, v^*_{\overline{k},0}\} \subset \mathcal{G}$.

Iterate

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\textit{for every } i \in I_n \\
\left\lfloor
\begin{array}{l}
l_{i,n}^* = v_{i,n}^* + L_i^* v_{\overline{k},n}^*; \\
a_{i,n} = \operatorname{proj}_{E_i}\big(x_{i,n} - \gamma_{i,n} l_{i,n}^*\big); \\
a_{i,n}^* = \gamma_{i,n}^{-1}(x_{i,n} - a_{i,n}) - l_{i,n}^*; \\
\xi_{i,n} = \|a_{i,n} - x_{i,n}\|^2;
\end{array}
\right. \\
\textit{for every } i \in I \smallsetminus I_n \\
\left\lfloor
\begin{array}{l}
a_{i,n} = a_{i,n-1};\ a_{i,n}^* = a_{i,n-1}^*;\ \xi_{i,n} = \xi_{i,n-1};
\end{array}
\right. \\
\textit{for every } k \in K_n \\
\left\lfloor
\begin{array}{l}
\textit{if } k \in I \\
\left\lfloor
\begin{array}{l}
b_{k,n} = \operatorname{proj}_{F_k}\big(y_{k,n} + \mu_{k,n} v_{k,n}^*\big); \\
e_{k,n}^* = \sigma_{k,n}(x_{k,n} - y_{k,n} - z_{k,n}) + v_{k,n}^*; \\
q_{k,n}^* = \mu_{k,n}^{-1}(y_{k,n} - b_{k,n}) + v_{k,n}^* - e_{k,n}^*; \\
e_{k,n} = b_{k,n} - a_{k,n};
\end{array}
\right. \\
\textit{if } k = \overline{k} \\
\left\lfloor
\begin{array}{l}
u_{k,n}^* = v_{k,n}^* - B^{\ell} y_{k,n}; \\
b_{k,n} = J_{\mu_{k,n} B^m}\big(y_{k,n} + \mu_{k,n}(u_{k,n}^* - B^{c} y_{k,n})\big); \\
e_{k,n}^* = \sigma_{k,n}\big(\sum_{i\in I} L_i x_{i,n} - y_{k,n} - z_{k,n}\big) + v_{k,n}^*; \\
q_{k,n}^* = \mu_{k,n}^{-1}(y_{k,n} - b_{k,n}) + u_{k,n}^* + B^{\ell} b_{k,n} - e_{k,n}^*; \\
e_{k,n} = b_{k,n} - \sum_{i\in I} L_i a_{i,n};
\end{array}
\right. \\
t_{k,n}^* = \nu_{k,n}^{-1} z_{k,n} + v_{k,n}^* - e_{k,n}^*; \\
\eta_{k,n} = \|b_{k,n} - y_{k,n}\|^2 + \|z_{k,n}\|^2;
\end{array}
\right. \\
\textit{for every } k \in K \smallsetminus K_n \\
\left\lfloor
\begin{array}{l}
b_{k,n} = b_{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}; \\
\textit{if } k \in I \\
\left\lfloor
\begin{array}{l}
e_{k,n} = b_{k,n} - a_{k,n};
\end{array}
\right. \\
\textit{if } k = \overline{k} \\
\left\lfloor
\begin{array}{l}
e_{k,n} = b_{k,n} - \sum_{i\in I} L_i a_{i,n};
\end{array}
\right.
\end{array}
\right. \\
\textit{for every } i \in I \\
\left\lfloor
\begin{array}{l}
p_{i,n}^* = a_{i,n}^* + e_{i,n}^* + L_i^* e_{\overline{k},n}^*;
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{9.112}
$$

$$
\begin{array}{l}
\Delta_n = -(4\beta^c)^{-1}\big(\sum_{i\in I}\xi_{i,n} + \sum_{k\in K}\eta_{k,n}\big) + \sum_{i\in I}\langle x_{i,n} - a_{i,n} \mid p^*_{i,n}\rangle \\
\qquad + \sum_{k\in K}\big(\langle y_{k,n} - b_{k,n} \mid q^*_{k,n}\rangle + \langle z_{k,n} \mid t^*_{k,n}\rangle + \langle e_{k,n} \mid v^*_{k,n} - e^*_{k,n}\rangle\big);\\
\textit{if } \Delta_n > 0\\
\quad \theta_n = \lambda_n\Delta_n/\big(\sum_{i\in I}\|p^*_{i,n}\|^2 + \sum_{k\in K}\big(\|q^*_{k,n}\|^2 + \|t^*_{k,n}\|^2 + \|e_{k,n}\|^2\big)\big);\\
\quad \textit{for every } i \in I\\
\quad\quad x_{i,n+1} = x_{i,n} - \theta_n p^*_{i,n};\\
\quad \textit{for every } k \in K\\
\quad\quad y_{k,n+1} = y_{k,n} - \theta_n q^*_{k,n};\; z_{k,n+1} = z_{k,n} - \theta_n t^*_{k,n};\; v^*_{k,n+1} = v^*_{k,n} - \theta_n e_{k,n};\\
\textit{else}\\
\quad \textit{for every } i \in I\\
\quad\quad x_{i,n+1} = x_{i,n};\\
\quad \textit{for every } k \in K\\
\quad\quad y_{k,n+1} = y_{k,n};\; z_{k,n+1} = z_{k,n};\; v^*_{k,n+1} = v^*_{k,n}.
\end{array}
$$

Furthermore, suppose that (9.102) has a solution and that

$$(\forall i \in I)\quad N_{E_i\cap F_i} = N_{E_i} + N_{F_i}. \tag{9.113}$$

Then there exists $(\overline{x}_i)_{i\in I} \in \bigoplus_{i\in I}\mathcal{H}_i$ such that $\sum_{i\in I} L_i\overline{x}_i$ solves (9.102) and, for every $i \in I$, $x_{i,n} \rightharpoonup \overline{x}_i$ and $a_{i,n} \rightharpoonup \overline{x}_i$.

Proof. Set $\boldsymbol{\mathcal{H}} = \bigoplus_{i\in I}\mathcal{H}_i$. Let us consider the problem

$$\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \text{ such that } (\forall i \in I)\;\; 0 \in N_{E_i}\overline{x}_i + N_{F_i}\overline{x}_i + L_i^*(B^m + B^c + B^\ell)\Bigg(\sum_{j\in I} L_j\overline{x}_j\Bigg) \tag{9.114}$$

together with the associated dual problem

$$\text{find } \big(\overline{\boldsymbol{x}}^*, \overline{v}^*\big) \in \boldsymbol{\mathcal{H}}\oplus\mathcal{G} \text{ such that } \big(\exists\, \boldsymbol{x} \in \boldsymbol{\mathcal{H}}\big)\;\; \begin{cases}(\forall i \in I)\;\; -\overline{x}^*_i - L_i^*\overline{v}^* \in N_{E_i}x_i \;\text{ and }\; \overline{x}^*_i \in N_{F_i}x_i\\ \overline{v}^* = (B^m + B^c + B^\ell)\big(\sum_{j\in I} L_j x_j\big).\end{cases} \tag{9.115}$$

Denote by $\mathscr{P}$ and $\mathscr{D}$ the sets of solutions to (9.114) and (9.115), respectively. We observe that the primal-dual problem (9.114)–(9.115) is a special case of Problem 9.1 with

$$(\forall i \in I)\quad A_i = N_{E_i},\quad C_i = Q_i = 0,\quad R_i = 0,\quad \text{and}\quad s^*_i = 0, \tag{9.116}$$

and

$$
(\forall k\in K)\quad
\begin{cases}
\mathcal{G}_k=\mathcal{H}_k,\;\; B_k^m=N_{F_k},\;\; B_k^c=B_k^\ell=0\;\text{ if }\; k\in I;\\
\mathcal{G}_{\overline{k}}=\mathcal{G},\;\; B_{\overline{k}}^m=B^m,\;\; B_{\overline{k}}^c=B^c,\;\; B_{\overline{k}}^\ell=B^\ell\\
D_k^m=\{0\}^{-1},\;\; D_k^c=D_k^\ell=0,\;\; r_k=0\\
(\forall j\in I)\;\; L_{kj}=
\begin{cases}
\mathrm{Id}, & \text{if }\; k=j;\\
0, & \text{if }\; k\in I\;\text{ and }\; k\neq j;\\
L_j, & \text{if }\; k=\overline{k}.
\end{cases}
\end{cases}
\tag{9.117}
$$

Further, we have

$$
\begin{cases}
(\forall i\in I)(\forall n\in\mathbb{N})\;\; J_{\gamma_{i,n}A_i}=\mathrm{proj}_{E_i}\\
(\forall k\in K)(\forall n\in\mathbb{N})\;\; J_{\nu_{k,n}D_k^m}=0\;\text{ and }\; J_{\mu_{k,n}B_k^m}=
\begin{cases}
\mathrm{proj}_{F_k}, & \text{if }\; k\in I;\\
J_{\mu_{k,n}B^m}, & \text{if }\; k=\overline{k}.
\end{cases}
\end{cases}
\tag{9.118}
$$

Therefore, (9.112) is a realization of Algorithm 9.12 in the context of (9.114)–(9.115). Now define $\boldsymbol{D}=\times_{i\in I}(E_i\cap F_i)$ and $\boldsymbol{L}\colon\boldsymbol{\mathcal{H}}\to\mathcal{G}\colon\boldsymbol{x}\mapsto\sum_{i\in I}L_ix_i$. Then $\boldsymbol{L}^*\colon\mathcal{G}\to\boldsymbol{\mathcal{H}}\colon y^*\mapsto(L_i^*y^*)_{i\in I}$. Hence, by (9.102), [9, Proposition 16.9], and (9.113),

$$
\begin{aligned}
(\forall\overline{y}\in\mathcal{G})\quad \overline{y}\;\text{ solves (9.102)}&\\
&\Leftrightarrow(\exists\,\overline{\boldsymbol{x}}\in\boldsymbol{D})\;
\begin{cases}
\overline{y}=\boldsymbol{L}\overline{\boldsymbol{x}}\\
(\forall\boldsymbol{x}\in\boldsymbol{D})\;\;\big\langle\boldsymbol{L}\overline{\boldsymbol{x}}-\boldsymbol{L}\boldsymbol{x}\mid(B^m+B^c+B^\ell)(\boldsymbol{L}\overline{\boldsymbol{x}})\big\rangle\leqslant 0
\end{cases}\\
&\Leftrightarrow(\exists\,\overline{\boldsymbol{x}}\in\boldsymbol{D})\;
\begin{cases}
\overline{y}=\boldsymbol{L}\overline{\boldsymbol{x}}\\
(\forall\boldsymbol{x}\in\boldsymbol{D})\;\;\big\langle\overline{\boldsymbol{x}}-\boldsymbol{x}\mid\boldsymbol{L}^*\big((B^m+B^c+B^\ell)(\boldsymbol{L}\overline{\boldsymbol{x}})\big)\big\rangle\leqslant 0
\end{cases}\\
&\Leftrightarrow(\exists\,\overline{\boldsymbol{x}}\in\boldsymbol{\mathcal{H}})\;
\begin{cases}
\overline{y}=\boldsymbol{L}\overline{\boldsymbol{x}}\\
\boldsymbol{0}\in N_{\boldsymbol{D}}\overline{\boldsymbol{x}}+\boldsymbol{L}^*\big((B^m+B^c+B^\ell)(\boldsymbol{L}\overline{\boldsymbol{x}})\big)
\end{cases}\\
&\Leftrightarrow(\exists\,\overline{\boldsymbol{x}}\in\boldsymbol{\mathcal{H}})\;
\begin{cases}
\overline{y}=\boldsymbol{L}\overline{\boldsymbol{x}}\\
(\forall i\in I)\;\;0\in N_{E_i\cap F_i}\overline{x}_i+L_i^*(B^m+B^c+B^\ell)\big(\sum_{j\in I}L_j\overline{x}_j\big)
\end{cases}\\
&\Leftrightarrow(\exists\,\overline{\boldsymbol{x}}\in\boldsymbol{\mathcal{H}})\;
\begin{cases}
\overline{y}=\boldsymbol{L}\overline{\boldsymbol{x}}\\
(\forall i\in I)\;\;0\in N_{E_i}\overline{x}_i+N_{F_i}\overline{x}_i+L_i^*(B^m+B^c+B^\ell)\big(\sum_{j\in I}L_j\overline{x}_j\big)
\end{cases}\\
&\Leftrightarrow(\exists\,\overline{\boldsymbol{x}}\in\mathscr{P})\;\;\overline{y}=\boldsymbol{L}\overline{\boldsymbol{x}}.
\end{aligned}
\tag{9.119}
$$

In turn, $\mathscr{P}\neq\varnothing$ since (9.102) has a solution. Therefore, in view of (9.117), Proposition 9.3(v)[d] yields $\mathscr{D}\neq\varnothing$. As a result, Theorem 9.13(iv) asserts that there exists $(\overline{x}_i)_{i\in I}\in\mathscr{P}$ such that, for every $i\in I$, $x_{i,n}\rightharpoonup\overline{x}_i$ and $a_{i,n}\rightharpoonup\overline{x}_i$. Finally, using (9.119), we conclude that $\sum_{i\in I}L_i\overline{x}_i$ solves (9.102). ☐

Remark 9.19 Theorem 9.13 allows us to tackle other types of variational inequalities. For instance, let $(\mathcal{H}_i)_{i\in I}$ be a finite family of real Hilbert spaces and set $\boldsymbol{\mathcal{H}} = \bigoplus_{i\in I} \mathcal{H}_i$. For every $i \in I$, let $\varphi_i \in \Gamma_0(\mathcal{H}_i)$ and let $R_i \colon \boldsymbol{\mathcal{H}} \to \mathcal{H}_i$ be such that Problem 9.1[e] holds. The objective is to

$$\text{find } \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \text{ such that } (\forall i \in I) \;\; 0 \in \partial\varphi_i(\overline{x}_i) + R_i\overline{\boldsymbol{x}}. \tag{9.120}$$

This simple instantiation of Problem 9.1 shows up in neural networks [27] and in game theory [2,15]. Thanks to Theorem 9.13, it can be solved using an asynchronous block-iterative strategy, which is not possible with current splitting techniques such as those of [25,34].

9.2.4.2 Application to multivariate minimization

We consider a composite multivariate minimization problem involving various types of convex functions and combinations between them.

Problem 9.20 Let $(\mathcal{H}_i)_{i\in I}$ and $(\mathcal{G}_k)_{k\in K}$ be finite families of real Hilbert spaces, and set $\boldsymbol{\mathcal{H}} = \bigoplus_{i\in I} \mathcal{H}_i$ and $\boldsymbol{\mathcal{G}} = \bigoplus_{k\in K} \mathcal{G}_k$. For every $i \in I$ and every $k \in K$, let $f_i \in \Gamma_0(\mathcal{H}_i)$, let $\alpha_i \in \left]0, +\infty\right[$, let $\varphi_i \colon \mathcal{H}_i \to \mathbb{R}$ be convex and differentiable with a $(1/\alpha_i)$-Lipschitzian gradient, let $g_k \in \Gamma_0(\mathcal{G}_k)$, let $h_k \in \Gamma_0(\mathcal{G}_k)$, let $\beta_k \in \left]0, +\infty\right[$, let $\psi_k \colon \mathcal{G}_k \to \mathbb{R}$ be convex and differentiable with a $(1/\beta_k)$-Lipschitzian gradient, and suppose that $L_{ki} \colon \mathcal{H}_i \to \mathcal{G}_k$ is linear and bounded. In addition, let $\chi \in \left[0, +\infty\right[$ and let $\Theta \colon \boldsymbol{\mathcal{H}} \to \mathbb{R}$ be convex and differentiable with a χ-Lipschitzian gradient. The objective is to

$$\underset{\boldsymbol{x}\in\boldsymbol{\mathcal{H}}}{\text{minimize}} \;\; \Theta(\boldsymbol{x}) + \sum_{i\in I} \big(f_i(x_i) + \varphi_i(x_i)\big) + \sum_{k\in K} \big((g_k + \psi_k) \,\Box\, h_k\big)\Bigg(\sum_{j\in I} L_{kj}x_j\Bigg). \tag{9.121}$$

Special cases of Problem 9.20 are found in various contexts, e.g., [13,16,23,25,33,34]. Formulation (9.121) brings together these disparate problems and the following algorithm makes it possible to solve them in an asynchronous block-iterative fashion in full generality.

Algorithm 9.21 Consider the setting of Problem 9.20 and suppose that Assumption 9.10–9.11 is in force. Set $\alpha = \min\{\alpha_i, \beta_k\}_{i\in I, k\in K}$, let $\sigma \in \left]1/(4\alpha), +\infty\right[$, and let $\varepsilon \in \left]0, \min\{1, 1/(\chi + \sigma)\}\right[$. For every $i \in I$, every $k \in K$, and every $n \in \mathbb{N}$, let $\gamma_{i,n} \in [\varepsilon, 1/(\chi + \sigma)]$, let $\{\mu_{k,n}, \nu_{k,n}\} \subset [\varepsilon, 1/\sigma]$, let $\sigma_{k,n} \in [\varepsilon, 1/\varepsilon]$, and let $\lambda_n \in [\varepsilon, 2-\varepsilon]$. In addition, let $\boldsymbol{x}_0 \in \boldsymbol{\mathcal{H}}$ and $\{\boldsymbol{y}_0, \boldsymbol{z}_0, \boldsymbol{v}_0^*\} \subset \boldsymbol{\mathcal{G}}$. Iterate

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\quad \text{for every } i \in I_n\\
\quad\quad l_{i,n}^* = \nabla_i \Theta(\boldsymbol{x}_{\pi_i(n)}) + \sum_{k\in K} L_{ki}^* v_{k,\pi_i(n)}^*;\\
\quad\quad a_{i,n} = \operatorname{prox}_{\gamma_{i,\pi_i(n)} f_i}\big(x_{i,\pi_i(n)} - \gamma_{i,\pi_i(n)}\big(l_{i,n}^* + \nabla\varphi_i(x_{i,\pi_i(n)})\big)\big);\\
\quad\quad a_{i,n}^* = \gamma_{i,\pi_i(n)}^{-1}(x_{i,\pi_i(n)} - a_{i,n}) - l_{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 b_{k,n} = \operatorname{prox}_{\mu_{k,\omega_k(n)} g_k}\big(y_{k,\omega_k(n)} + \mu_{k,\omega_k(n)}\big(v_{k,\omega_k(n)}^* - \nabla\psi_k(y_{k,\omega_k(n)})\big)\big);\\
\quad\quad d_{k,n} = \operatorname{prox}_{\nu_{k,\omega_k(n)} h_k}\big(z_{k,\omega_k(n)} + \nu_{k,\omega_k(n)} v_{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) + v_{k,\omega_k(n)}^*;\\
\quad\quad 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}^*;\\
\quad\quad t_{k,n}^* = \nu_{k,\omega_k(n)}^{-1}(z_{k,\omega_k(n)} - d_{k,n}) + v_{k,\omega_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}^*;\ q_{k,n}^* = q_{k,n-1}^*;\ t_{k,n}^* = t_{k,n-1}^*;\\
\quad\quad \eta_{k,n} = \eta_{k,n-1};\ 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}^* + \nabla_i \Theta(\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\quad\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 + \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}^*;\ 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{9.122}
$$

Corollary 9.22 *Consider the setting of Algorithm 9.21. Suppose that*

$$(\forall k \in K) \quad \operatorname{epi}(g_k + \psi_k) + \operatorname{epi} h_k \ \text{ is closed} \tag{9.123}$$

and that Problem 9.20 admits a Kuhn–Tucker point, that is, there exist $\widetilde{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}}$ and $\widetilde{\boldsymbol{v}}^* \in \boldsymbol{\mathcal{G}}$ such that

$$(\forall i \in I)(\forall k \in K) \quad \begin{cases} -\sum_{j\in K} L_{ji}^* \widetilde{v}_j^* \in \partial f_i(\widetilde{x}_i) + \nabla \varphi_i(\widetilde{x}_i) + \nabla_i \Theta(\widetilde{\boldsymbol{x}}) \\ \sum_{j\in I} L_{kj}\widetilde{x}_j \in \partial\big(g_k^* \,\Box\, \psi_k^*\big)(\widetilde{v}_k^*) + \partial h_k^*(\widetilde{v}_k^*). \end{cases} \tag{9.124}$$

Then there exists a solution $\overline{\boldsymbol{x}}$ to (9.121) such that, for every $i \in I$, $x_{i,n} \rightharpoonup \overline{x}_i$ and $a_{i,n} \rightharpoonup \overline{x}_i$.

Proof. Set

$$\begin{cases} (\forall i \in I) \ \ A_i = \partial f_i, \ \ C_i = \nabla \varphi_i, \ \text{ and } \ R_i = \nabla_i \Theta \\ (\forall k \in K) \ \ B_k^m = \partial g_k, \ \ B_k^c = \nabla \psi_k, \ \text{ and } \ D_k^m = \partial h_k. \end{cases} \tag{9.125}$$

First, [9, Theorem 20.25] asserts that the operators $(A_i)_{i\in I}$, $(B_k^m)_{k\in K}$, and $(D_k^m)_{k\in K}$ are maximally monotone. Second, it follows from [9, Corollary 18.17] that, for every $i \in I$, C_i is α_i-cocoercive and, for every $k \in K$, B_k^c is β_k-cocoercive. Third, in view of (9.125) and [9, Proposition 17.7], $\boldsymbol{R} = \nabla\Theta$ is monotone and χ-Lipschitzian. Now consider the problem

$$\text{find } \ \overline{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}} \ \text{ such that}$$
$$(\forall i \in I) \ \ 0 \in A_i\overline{x}_i + C_i\overline{x}_i + R_i\overline{\boldsymbol{x}} + \sum_{k\in K} L_{ki}^*\left(\Big(\big(B_k^m + B_k^c\big) \,\Box\, D_k^m\Big)\left(\sum_{j\in I} L_{kj}\overline{x}_j\right)\right) \tag{9.126}$$

together with its dual

$$\text{find } \ \overline{\boldsymbol{v}}^* \in \boldsymbol{\mathcal{G}} \ \text{ such that}$$
$$\big(\exists\, \boldsymbol{x} \in \boldsymbol{\mathcal{H}}\big)(\forall i \in I)(\forall k \in K) \ \ \begin{cases} -\displaystyle\sum_{j\in K} L_{ji}^*\overline{v}_j^* \in A_i x_i + C_i x_i + R_i \boldsymbol{x} \\ \overline{v}_k^* \in \Big(\big(B_k^m + B_k^c\big) \,\Box\, D_k^m\Big)\left(\displaystyle\sum_{j\in I} L_{kj}x_j\right). \end{cases} \tag{9.127}$$

Denote by $\mathscr{P}$ and $\mathscr{D}$ the sets of solutions to (9.126) and (9.127), respectively. We observe that, by (9.125) and [9, Example 23.3], Algorithm 9.21 is an application of Algorithm 9.12 to the primal-dual problem (9.126)–(9.127). Furthermore, it results from (9.124) and Proposition 9.3(iv) that $\mathscr{D} \neq \varnothing$. According to Theorem 9.13(iv), there exist $\overline{\boldsymbol{x}} \in \mathscr{P}$ and $\overline{\boldsymbol{v}}^* \in \mathscr{D}$ such that, for every $i \in I$ and every $k \in K$,

$$x_{i,n} \rightharpoonup \overline{x}_i, \quad a_{i,n} \rightharpoonup \overline{x}_i, \quad \text{and} \quad \begin{cases} -\displaystyle\sum_{j\in K} L_{ji}^*\overline{v}_j^* \in A_i\overline{x}_i + C_i\overline{x}_i + R_i\overline{\boldsymbol{x}} \\ \overline{v}_k^* \in \Big(\big(B_k^m + B_k^c\big) \,\Box\, D_k^m\Big)\left(\displaystyle\sum_{j\in I} L_{kj}\overline{x}_j\right). \end{cases} \tag{9.128}$$

It remains to show that $\overline{\boldsymbol{x}}$ solves (9.121). Define

$$\begin{cases}\boldsymbol{f} = \bigoplus_{i\in I} f_i, \;\; \boldsymbol{\varphi} = \bigoplus_{i\in I} \varphi_i, \;\; \boldsymbol{g} = \bigoplus_{k\in K} g_k, \;\; \boldsymbol{h} = \bigoplus_{k\in K} h_k, \;\; \text{and} \;\; \boldsymbol{\psi} = \bigoplus_{k\in K} \psi_k \\ \boldsymbol{L}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{G}}\colon \boldsymbol{x} \mapsto \big(\sum_{i\in I} L_{ki}x_i\big)_{k\in K}.\end{cases} \tag{9.129}$$

We deduce from [9, Theorem 15.3] that $(\forall k \in K)\; (g_k+\psi_k)^* = g_k^* \boxdot \psi_k^*$. In turn, (9.124) implies that

$$(\forall k \in K) \quad \varnothing \neq \operatorname{dom}\big(g_k^* \boxdot \psi_k^*\big) \cap \operatorname{dom} h_k^* = \operatorname{dom}(g_k+\psi_k)^* \cap \operatorname{dom} h_k^*. \tag{9.130}$$

On the other hand, since the sets $(\operatorname{epi}(g_k+\psi_k)+\operatorname{epi} h_k)_{k\in K}$ are convex, it follows from (9.123) and [9, Theorem 3.34] that they are weakly closed. Therefore, [20, Theorem 1] and the Fenchel–Moreau theorem [9, Theorem 13.37] imply that

$$(\forall k \in K) \quad \big((g_k+\psi_k)^* + h_k^*\big)^* = (g_k+\psi_k)^{**} \boxdot h_k^{**} = (g_k+\psi_k) \boxdot h_k. \tag{9.131}$$

Hence, we derive from (9.125), [9, Corollaries 16.48(iii) and 16.30], (9.131), and [9, Proposition 16.42] that

$$\begin{aligned}(\forall k \in K) \quad \big(B_k^m + B_k^c\big) \,\Box\, D_k^m &= \big(\partial g_k + \nabla \psi_k\big) \,\Box\, (\partial h_k) \\ &= \Big(\big(\partial(g_k+\psi_k)\big)^{-1} + (\partial h_k)^{-1}\Big)^{-1} \\ &= \big(\partial(g_k+\psi_k)^* + \partial h_k^*\big)^{-1} \\ &= \Big(\partial\big((g_k+\psi_k)^* + h_k^*\big)\Big)^{-1} \\ &= \partial\big((g_k+\psi_k)^* + h_k^*\big)^* \\ &= \partial\big((g_k+\psi_k) \boxdot h_k\big).\end{aligned} \tag{9.132}$$

Since it results from (9.129) and (9.131) that

$$(\boldsymbol{g}+\boldsymbol{\psi}) \,\Box\, \boldsymbol{h} = (\boldsymbol{g}+\boldsymbol{\psi}) \boxdot \boldsymbol{h} = \bigoplus_{k\in K} \big((g_k+\psi_k) \boxdot h_k\big), \tag{9.133}$$

we deduce from [9, Proposition 16.9] and (9.132) that

$$\partial\big((\boldsymbol{g}+\boldsymbol{\psi}) \boxdot \boldsymbol{h}\big) = \bigtimes_{k\in K} \partial\big((g_k+\psi_k) \boxdot h_k\big) = \bigtimes_{k\in K} \big((B_k^m + B_k^c) \,\Box\, D_k^m\big). \tag{9.134}$$

It thus follows from (9.128) and (9.129) that $\overline{\boldsymbol{v}}^* \in \partial((\boldsymbol{g}+\boldsymbol{\psi}) \boxdot \boldsymbol{h})(\boldsymbol{L}\overline{\boldsymbol{x}})$. On the other hand, since $\boldsymbol{L}^*\colon \boldsymbol{\mathcal{G}} \to \boldsymbol{\mathcal{H}}\colon \boldsymbol{v}^* \mapsto (\sum_{k\in K} L_{ki}^* v_k^*)_{i\in I}$, we infer from (9.128), (9.125), (9.129), and [9, Proposition 16.9] that $-\boldsymbol{L}^*\overline{\boldsymbol{v}}^* \in (C_i\overline{x}_i)_{i\in I} + \boldsymbol{R}\overline{\boldsymbol{x}} + \bigtimes_{i\in I} A_i\overline{x}_i = \nabla\boldsymbol{\varphi}(\overline{\boldsymbol{x}}) + \nabla\Theta(\overline{\boldsymbol{x}}) + \partial\boldsymbol{f}(\overline{\boldsymbol{x}})$. Hence, we invoke [9, Proposition 16.6(ii)] to obtain

$$\boldsymbol{0} \in \partial\boldsymbol{f}(\overline{\boldsymbol{x}}) + \nabla\boldsymbol{\varphi}(\overline{\boldsymbol{x}}) + \nabla\Theta(\overline{\boldsymbol{x}}) + \boldsymbol{L}^*\overline{\boldsymbol{v}}^*$$

$$
\begin{aligned}
&\subset \partial \boldsymbol{f}(\overline{\boldsymbol{x}}) + \nabla \boldsymbol{\varphi}(\overline{\boldsymbol{x}}) + \nabla \Theta(\overline{\boldsymbol{x}}) + \boldsymbol{L}^*\Big(\partial\big((\boldsymbol{g}+\boldsymbol{\psi})\,\square\,\boldsymbol{h}\big)(\boldsymbol{L}\overline{\boldsymbol{x}})\Big)\\
&\subset \partial\Big(\boldsymbol{f}+\boldsymbol{\varphi}+\Theta+\big((\boldsymbol{g}+\boldsymbol{\psi})\,\square\,\boldsymbol{h}\big)\circ \boldsymbol{L}\Big)(\overline{\boldsymbol{x}}).
\end{aligned}
\tag{9.135}
$$

However, thanks to (9.129) and (9.133), (9.121) is equivalent to

$$
\underset{\boldsymbol{x}\in\boldsymbol{\mathcal{H}}}{\text{minimize}}\;\; \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\varphi}(\boldsymbol{x}) + \Theta(\boldsymbol{x}) + \big((\boldsymbol{g}+\boldsymbol{\psi})\,\square\,\boldsymbol{h}\big)(\boldsymbol{L}\boldsymbol{x}). \tag{9.136}
$$

Consequently, in view of Fermat's rule [9, Theorem 16.3], (9.135) implies that $\overline{\boldsymbol{x}}$ solves (9.121). $\square$

Remark 9.23 In [16], multicomponent image recovery problems were approached by applying the forward-backward and the Douglas–Rachford algorithms in a product space. Using Corollary 9.22, we can now solve these problems with asynchronous block-iterative algorithms and more sophisticated formulations. For instance, the standard total variation loss used in [16] can be replaced by the pth order Huber total variation penalty of [33], which turns out to involve an infimal convolution.

To conclude, we provide some scenarios in which condition (9.123) is satisfied.

Proposition 9.24 *Consider the setting of Problem 9.20. Suppose that there exist $\widetilde{\boldsymbol{x}} \in \boldsymbol{\mathcal{H}}$ and $\widetilde{\boldsymbol{v}}^* \in \boldsymbol{\mathcal{G}}$ such that*

$$
(\forall i \in I)(\forall k \in K)\quad
\begin{cases}
-\sum_{j\in K} L_{ji}^* \widetilde{v}_j^* \in \partial f_i(\widetilde{x}_i) + \nabla \varphi_i(\widetilde{x}_i) + \nabla_i \Theta(\widetilde{\boldsymbol{x}})\\
\sum_{j\in I} L_{kj}\widetilde{x}_j \in \partial\big(g_k^*\,\square\,\psi_k^*\big)(\widetilde{v}_k^*) + \partial h_k^*(\widetilde{v}_k^*)
\end{cases}
\tag{9.137}
$$

and that, for every $k \in K$, one of the following is satisfied:

[a] $0 \in \operatorname{sri}(\operatorname{dom} g_k^* + \operatorname{dom}\psi_k^* - \operatorname{dom} h_k^*)$.

[b] *$\mathcal{G}_k$ is finite-dimensional, h_k is polyhedral, and* $\operatorname{dom} h_k^* \cap \operatorname{ri}\operatorname{dom}(g_k+\psi_k)^* \neq \varnothing$.

[c] *$\mathcal{G}_k$ is finite-dimensional, g_k and h_k are polyhedral, and $\psi_k = 0$.*

Then, for every $k \in K$, $\operatorname{epi}(g_k+\psi_k) + \operatorname{epi} h_k$ *is closed.*

Proof. Let $k \in K$. Since $\operatorname{dom}\psi_k = \mathcal{G}_k$, [9, Theorem 15.3] yields

$$
(g_k+\psi_k)^* = g_k^*\,\square\,\psi_k^*. \tag{9.138}
$$

Therefore, (9.137) implies that

$$
\varnothing \neq \operatorname{dom}\big(g_k^*\,\square\,\psi_k^*\big) \cap \operatorname{dom} h_k^* = \operatorname{dom}(g_k+\psi_k)^* \cap \operatorname{dom} h_k^*. \tag{9.139}
$$

In view of (9.139), [20, Theorem 1], and [9, Theorem 3.34], it suffices to show that $((g_k + \psi_k)^* + h_k^*)^* = (g_k+\psi_k)^{**}\,\square\,h_k^{**}$.

[a]: We deduce from [9, Proposition 12.6(ii)] and (9.138) that $0 \in \operatorname{sri}(\operatorname{dom}(g_k^* \,\square\, \psi_k^*) - \operatorname{dom} h_k^*) = \operatorname{sri}(\operatorname{dom}(g_k + \psi_k)^* - \operatorname{dom} h_k^*)$. In turn, [9, Theorem 15.3] gives $((g_k + \psi_k)^* + h_k^*)^* = (g_k + \psi_k)^{**} \,\square\, h_k^{**}$.

[b]: Since [48, Theorem 19.2] asserts that h_k^* is polyhedral, we infer from [48, Theorem 20.1] that $((g_k + \psi_k)^* + h_k^*)^* = (g_k + \psi_k)^{**} \,\square\, h_k^{**}$.

[c]: Since g_k^* and h_k^* are polyhedral by [48, Theorem 19.2], it follows from (9.139) and [48, Theorem 20.1] that $(g_k^* + h_k^*)^* = g_k^{**} \,\square\, h_k^{**}$. □

9.2.5 Appendix

In this section, $\mathcal{K}$ is a real Hilbert space.

Lemma 9.25 *Let $A\colon \mathcal{K} \to 2^{\mathcal{K}}$ be maximally monotone, let $(x_n)_{n\in\mathbb{N}}$ be a bounded sequence in $\mathcal{K}$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a bounded sequence in $]0, +\infty[$. Then $(J_{\gamma_n A} x_n)_{n\in\mathbb{N}}$ is bounded.*

Proof. Fix $x \in \mathcal{K}$. Using the triangle inequality, the nonexpansiveness of $(J_{\gamma_n A})_{n\in\mathbb{N}}$, and [9, Proposition 23.31(iii)], we obtain $(\forall n \in \mathbb{N})\ \|J_{\gamma_n A} x_n - J_A x\| \leqslant \|J_{\gamma_n A} x_n - J_{\gamma_n A} x\| + \|J_{\gamma_n A} x - J_A x\| \leqslant \|x_n - x\| + |1 - \gamma_n|\, \|J_A x - x\| \leqslant \|x\| + \sup_{m\in\mathbb{N}} \|x_m\| + (1 + \sup_{m\in\mathbb{N}} \gamma_m)\|J_A x - x\|$. □

Lemma 9.26 *Let $\alpha \in [0, +\infty[$, let $A\colon \mathcal{K} \to \mathcal{K}$ be α-Lipschitzian, let $\sigma \in]0, +\infty[$, and let $\gamma \in]0, 1/(\alpha + \sigma)]$. Then $\gamma^{-1}\mathrm{Id} - A$ is σ-strongly monotone.*

Proof. By Cauchy–Schwarz,

$$
\begin{aligned}
(\forall x \in \mathcal{K})(\forall y \in \mathcal{K}) \quad & \big\langle x - y \mid \big(\gamma^{-1}\mathrm{Id} - A\big)x - \big(\gamma^{-1}\mathrm{Id} - A\big)y \big\rangle \\
&= \gamma^{-1}\|x - y\|^2 - \langle x - y \mid Ax - Ay \rangle \\
&\geqslant (\alpha + \sigma)\|x - y\|^2 - \|x - y\|\,\|Ax - Ay\| \\
&\geqslant (\alpha + \sigma)\|x - y\|^2 - \alpha\|x - y\|^2 \\
&= \sigma\|x - y\|^2,
\end{aligned} \tag{9.140}
$$

which proves the assertion. □

Lemma 9.27 *Let I be a nonempty finite set, let $(I_n)_{n\in\mathbb{N}}$ be nonempty subsets of I, let $P \in \mathbb{N}$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{K}$. Suppose that $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$, $I_0 = I$, and $(\forall n \in \mathbb{N})$ $\bigcup_{j=n}^{n+P} I_j = I$. Furthermore, let $T \in \mathbb{N}$, let $i \in I$, and let $(\pi_i(n))_{n\in\mathbb{N}}$ be a sequence in $\mathbb{N}$ such that $(\forall n \in \mathbb{N})\ n - T \leqslant \pi_i(n) \leqslant n$. For every $n \in \mathbb{N}$, set $\overline{\vartheta}_i(n) = \max\big\{j \in \mathbb{N} \mid j \leqslant n \text{ and } i \in I_j\big\}$ and $\vartheta_i(n) = \pi_i(\overline{\vartheta}_i(n))$. Then $x_{\vartheta_i(n)} - x_n \to 0$.*

Proof. For every integer $n \geqslant P$, since $i \in \bigcup_{j=n-P}^{n} I_j$, we have $n \leqslant \overline{\vartheta}_i(n) + P \leqslant \pi_i(\overline{\vartheta}_i(n)) + P + T = \vartheta_i(n) + P + T$. Hence $\vartheta_i(n) \to +\infty$ and therefore $\sum_{j=\vartheta_i(n)}^{\vartheta_i(n)+P+T} \|x_{j+1} - x_j\|^2 \to 0$. However, it results from our assumption that $(\forall n \in \mathbb{N})\ \vartheta_i(n) = \pi_i(\overline{\vartheta}_i(n)) \leqslant \overline{\vartheta}_i(n) \leqslant n$. We thus deduce from the triangle and Cauchy–Schwarz inequalities that

$$\|x_n - x_{\vartheta_i(n)}\|^2 \leqslant \left|\sum_{j=\vartheta_i(n)}^{\vartheta_i(n)+P+T} \|x_{j+1} - x_j\|\right|^2 \leqslant (P+T+1) \sum_{j=\vartheta_i(n)}^{\vartheta_i(n)+P+T} \|x_{j+1} - x_j\|^2 \to 0. \quad (9.141)$$

Consequently, $x_{\vartheta_i(n)} - x_n \to 0$. ☐

Lemma 9.28 ([22]) *Let Z be a nonempty closed convex subset of $\mathcal{K}$, $x_0 \in \mathcal{K}$, and $\varepsilon \in \left]0,1\right[$. Suppose that*

$$\begin{array}{l}
\textit{for } n = 0,1,\ldots \\
\left\lfloor \begin{array}{l}
t_n^* \in \mathcal{K} \textit{ and } \eta_n \in \mathbb{R} \textit{ satisfy } Z \subset H_n = \big\{x \in \mathcal{K} \mid \langle x \mid t_n^*\rangle \leqslant \eta_n\big\}; \\
\Delta_n = \langle x_n \mid t_n^*\rangle - \eta_n; \\
\textit{if } \Delta_n > 0 \\
\left\lfloor \begin{array}{l}
\lambda_n \in [\varepsilon, 2-\varepsilon]\,; \\
x_{n+1} = x_n - (\lambda_n \Delta_n / \|t_n^*\|^2)\, t_n^*;
\end{array}\right. \\
\textit{else} \\
\left\lfloor \begin{array}{l}
x_{n+1} = x_n.
\end{array}\right.
\end{array}\right.
\end{array} \quad (9.142)$$

Then the following hold:

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

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

(iii) *Suppose that, for every $x \in \mathcal{K}$ and every strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$, $x_{k_n} \rightharpoonup x$ $\Rightarrow x \in Z$. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in Z.*

We now revisit ideas found in [8, 21] in a format that is be more suited for our purposes.

Lemma 9.29 *Let Z be a nonempty closed convex subset of $\mathcal{K}$ and let $x_0 \in \mathcal{K}$. Suppose that*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\quad t_n^* \in \mathcal{K} \textit{ and } \eta_n \in \mathbb{R} \textit{ satisfy } Z \subset H_n = \big\{ x \in \mathcal{K} \mid \langle x \mid t_n^* \rangle \leqslant \eta_n \big\}; \\
\quad \Delta_n = \langle x_n \mid t_n^* \rangle - \eta_n; \\
\quad \textit{if } \Delta_n > 0 \\
\quad\quad \tau_n = \|t_n^*\|^2;\ \varsigma_n = \|x_0 - x_n\|^2;\ \chi_n = \langle x_0 - x_n \mid t_n^* \rangle;\ \rho_n = \tau_n \varsigma_n - \chi_n^2; \\
\quad\quad \textit{if } \rho_n = 0 \\
\quad\quad\quad \kappa_n = 1;\ \lambda_n = \Delta_n / \tau_n; \\
\quad\quad \textit{else} \\
\quad\quad\quad \textit{if } \chi_n \Delta_n \geqslant \rho_n \\
\quad\quad\quad\quad \kappa_n = 0;\ \lambda_n = \big(\Delta_n + \chi_n \big) / \tau_n; \\
\quad\quad\quad \textit{else} \\
\quad\quad\quad\quad \kappa_n = 1 - \chi_n \Delta_n / \rho_n;\ \lambda_n = \varsigma_n \Delta_n / \rho_n; \\
\quad\quad x_{n+1} = (1 - \kappa_n) x_0 + \kappa_n x_n - \lambda_n t_n^*; \\
\quad \textit{else} \\
\quad\quad x_{n+1} = x_n.
\end{array}
\tag{9.143}
$$

Then the following hold:

(i) $(\forall n \in \mathbb{N})\ \|x_n - x_0\| \leqslant \|x_{n+1} - x_0\| \leqslant \|\mathrm{proj}_Z x_0 - x_0\|$.

(ii) $\sum_{n \in \mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$ *and* $\sum_{n \in \mathbb{N}} \|\mathrm{proj}_{H_n} x_n - x_n\|^2 < +\infty$.

(iii) *Suppose that, for every $x \in \mathcal{K}$ and every strictly increasing sequence $(k_n)_{n \in \mathbb{N}}$ in $\mathbb{N}$, $x_{k_n} \rightharpoonup x$ $\Rightarrow x \in Z$. Then $x_n \to \mathrm{proj}_Z x_0$.*

Proof. Define $(\forall n \in \mathbb{N})\ G_n = \big\{ x \in \mathcal{K} \mid \langle x - x_n \mid x_0 - x_n \rangle \leqslant 0 \big\}$. Then, by virtue of (9.143),

$$
(\forall n \in \mathbb{N}) \quad x_n = \mathrm{proj}_{G_n} x_0 \quad \text{and} \quad \big[\, \Delta_n > 0 \ \Rightarrow\ \mathrm{proj}_{H_n} x_n = x_n - \big(\Delta_n / \|t_n^*\|^2 \big)\, t_n^* \,\big]. \tag{9.144}
$$

Let us establish that

$$
(\forall n \in \mathbb{N}) \quad Z \subset H_n \cap G_n \quad \text{and} \quad x_{n+1} = \mathrm{proj}_{H_n \cap G_n} x_0. \tag{9.145}
$$

Since $G_0 = \mathcal{K}$, (9.143) yields $Z \subset H_0 = H_0 \cap G_0$. Hence, we derive from (9.144) and (9.143) that $\Delta_0 > 0 \Rightarrow [\,\mathrm{proj}_{H_0} x_0 = x_0 - (\Delta_0/\tau_0)\, t_0^*$ and $\rho_0 = 0\,] \Rightarrow [\,\mathrm{proj}_{H_0} x_0 = x_0 - (\Delta_0/\tau_0)\, t_0^*$, $\kappa_0 = 1$, and $\lambda_0 = \Delta_0/\tau_0\,] \Rightarrow x_1 = x_0 - (\Delta_0/\tau_0)\, t_0^* = \mathrm{proj}_{H_0} x_0 = \mathrm{proj}_{H_0 \cap G_0} x_0$. On the other hand, $\Delta_0 \leqslant 0 \Rightarrow x_1 = x_0 \in H_0 = H_0 \cap G_0 \Rightarrow x_1 = \mathrm{proj}_{H_0 \cap G_0} x_0$. Now assume that, for some integer $n \geqslant 1$, $Z \subset H_{n-1} \cap G_{n-1}$ and $x_n = \mathrm{proj}_{H_{n-1} \cap G_{n-1}} x_0$. Then, according to [9, Theorem 3.16], $Z \subset H_{n-1} \cap G_{n-1} \subset \big\{ x \in \mathcal{K} \mid \langle x - x_n \mid x_0 - x_n \rangle \leqslant 0 \big\} = G_n$. In turn, (9.143) entails that $Z \subset H_n \cap G_n$. Next, it follows from (9.143), (9.144), and [9, Proposition 29.5] that $\Delta_n \leqslant 0$ $\Rightarrow [\, x_{n+1} = x_n$ and $\mathrm{proj}_{G_n} x_0 = x_n \in H_n\,] \Rightarrow x_{n+1} = \mathrm{proj}_{G_n} x_0 = \mathrm{proj}_{H_n \cap G_n} x_0$. To complete the induction argument, it remains to verify that $\Delta_n > 0 \Rightarrow x_{n+1} = \text{proj}_{H_n \cap G_n} x_0$. Assume that $\Delta_n > 0$ and set

$$y_n = \text{proj}_{H_n} x_n, \quad \widetilde{\chi}_n = \langle x_0 - x_n \mid x_n - y_n \rangle, \quad \widetilde{\nu}_n = \|x_n - y_n\|^2, \quad \text{and} \quad \widetilde{\rho}_n = \varsigma_n \widetilde{\nu}_n - \widetilde{\chi}_n^2. \tag{9.146}$$

Since $\Delta_n > 0$, we have $H_n = \{x \in \mathcal{K} \mid \langle x - y_n \mid x_n - y_n \rangle \leqslant 0\}$ and $y_n = x_n - \theta_n t_n^*$, where $\theta_n = \Delta_n/\tau_n > 0$. In turn, we infer from (9.146) and (9.143) that

$$\widetilde{\chi}_n = \theta_n \chi_n, \quad \widetilde{\nu}_n = \theta_n^2 \tau_n = \theta_n \Delta_n, \quad \text{and} \quad \widetilde{\rho}_n = \theta_n^2 \rho_n. \tag{9.147}$$

Furthermore, (9.143) and the Cauchy–Schwarz inequality ensure that $\rho_n \geqslant 0$, which leads to two cases.

- $\rho_n = 0$: On the one hand, (9.143) asserts that $x_{n+1} = x_n - (\Delta_n/\tau_n)\, t_n^* = y_n$. On the other hand, (9.147) yields $\widetilde{\rho}_n = 0$ and, therefore, since $H_n \cap G_n \neq \varnothing$, [9, Corollary 29.25(ii)] yields $\text{proj}_{H_n \cap G_n} x_0 = y_n$. Altogether, $x_{n+1} = \text{proj}_{H_n \cap G_n} x_0$.
- $\rho_n > 0$: By (9.147), $\widetilde{\rho}_n > 0$. First, suppose that $\chi_n \Delta_n \geqslant \rho_n$. It follows from (9.143) that $x_{n+1} = x_0 - ((\Delta_n + \chi_n)/\tau_n)\, t_n^*$ and from (9.147) that $\widetilde{\chi}_n \widetilde{\nu}_n = \theta_n^2 \chi_n \Delta_n \geqslant \theta_n^2 \rho_n = \widetilde{\rho}_n$. Thus [9, Corollary 29.25(ii)] and (9.147) imply that

$$\begin{aligned}
\text{proj}_{H_n \cap G_n} x_0 &= x_0 + \left(1 + \frac{\widetilde{\chi}_n}{\widetilde{\nu}_n}\right)(y_n - x_n) \\
&= x_0 - \left(1 + \frac{\chi_n}{\theta_n \tau_n}\right)\theta_n t_n^* \\
&= x_0 - \frac{\theta_n \tau_n + \chi_n}{\tau_n}\, t_n^* \\
&= x_0 - \frac{\Delta_n + \chi_n}{\tau_n}\, t_n^* \\
&= x_{n+1}.
\end{aligned} \tag{9.148}$$

  Now suppose that $\chi_n \Delta_n < \rho_n$. Then $\widetilde{\chi}_n \widetilde{\nu}_n < \widetilde{\rho}_n$ and hence it results from [9, Corollary 29.25(ii)], (9.147), and (9.143) that

$$\begin{aligned}
\text{proj}_{H_n \cap G_n} x_0 &= x_n + \frac{\widetilde{\nu}_n}{\widetilde{\rho}_n}\Big(\widetilde{\chi}_n (x_0 - x_n) + \varsigma_n (y_n - x_n)\Big) \\
&= \frac{\widetilde{\chi}_n \widetilde{\nu}_n}{\widetilde{\rho}_n}\, x_0 + \left(1 - \frac{\widetilde{\chi}_n \widetilde{\nu}_n}{\widetilde{\rho}_n}\right) x_n + \frac{\widetilde{\nu}_n \varsigma_n}{\widetilde{\rho}_n}(y_n - x_n) \\
&= \frac{\chi_n \Delta_n}{\rho_n}\, x_0 + \left(1 - \frac{\chi_n \Delta_n}{\rho_n}\right) x_n - \frac{\tau_n \varsigma_n}{\rho_n}\frac{\Delta_n}{\tau_n}\, t_n^* \\
&= x_{n+1}.
\end{aligned} \tag{9.149}$$

(i): Let $n \in \mathbb{N}$. We derive from (9.145) that $\|x_{n+1} - x_0\| = \|\text{proj}_{H_n \cap G_n} x_0 - x_0\| \leqslant \|\text{proj}_Z x_0 -$

$x_0\|$. On the other hand, since $x_{n+1} \in G_n$ by virtue of (9.145), we have

$$\begin{aligned} \|x_n - x_0\|^2 + \|x_{n+1} - x_n\|^2 &\leqslant \|x_n - x_0\|^2 + \|x_{n+1} - x_n\|^2 + 2\langle x_{n+1} - x_n \mid x_n - x_0\rangle \\ &= \|x_{n+1} - x_0\|^2. \end{aligned} \tag{9.150}$$

(ii): Let $N \in \mathbb{N}$. In view of (9.150) and (i), $\sum_{n=0}^{N} \|x_{n+1} - x_n\|^2 \leqslant \sum_{n=0}^{N}(\|x_{n+1} - x_0\|^2 - \|x_n - x_0\|^2) = \|x_{N+1} - x_0\|^2 \leqslant \|\mathrm{proj}_Z x_0 - x_0\|^2$. Therefore, $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$. However, for every $n \in \mathbb{N}$, since (9.145) asserts that $x_{n+1} \in H_n$, we have $\|\mathrm{proj}_{H_n} x_n - x_n\| \leqslant \|x_{n+1} - x_n\|$. Thus $\sum_{n\in\mathbb{N}} \|\mathrm{proj}_{H_n} x_n - x_n\|^2 < +\infty$.

(iii): It results from (i) that $(x_n)_{n\in\mathbb{N}}$ is bounded. Now let $x \in \mathcal{K}$, let $(k_n)_{n\in\mathbb{N}}$ be a strictly increasing sequence in $\mathbb{N}$, and suppose that $x_{k_n} \rightharpoonup x$. Using [9, Lemma 2.42] and (i), we deduce that $\|x - x_0\| \leqslant \underline{\lim} \|x_{k_n} - x_0\| \leqslant \|\mathrm{proj}_Z x_0 - x_0\|$. Thus, since it results from our assumption that $x \in Z$, we have $x = \mathrm{proj}_Z x_0$, which implies that $x_n \rightharpoonup \mathrm{proj}_Z x_0$ [9, Lemma 2.46]. In turn, since $\overline{\lim} \|x_n - x_0\| \leqslant \|\mathrm{proj}_Z x_0 - x_0\|$ by (i), [9, Lemma 2.51(i)] forces $x_n \to \mathrm{proj}_Z x_0$. $\square$

References

[1] 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.

[2] 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.

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

[4] H. Attouch, 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.

[5] J.-F. Aujol and A. Chambolle, Dual norms and image decomposition models, *Int. J. Comput. Vision*, vol. 63, pp. 85–104, 2005.

[6] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, Optimization with sparsity-inducing penalties, *Found. Trends Machine Learn.*, vol. 4, pp. 1–106, 2012.

[7] S. Banert, A. Ringh, J. Adler, J. Karlsson, and O. Öktem, Data-driven nonsmooth optimization, *SIAM J. Optim.*, vol. 30, pp. 102–131, 2020.

[8] 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.

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

[10] 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.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] M. N. Bùi and P. L. Combettes, Warped proximal iterations for monotone inclusions, *J. Math. Anal. Appl.*, vol. 491, art. 124315, 21 pp., 2020.

[20] R. S. Burachik and V. Jeyakumar, A new geometric condition for Fenchel's duality in infinite dimensional spaces, *Math. Program.*, vol. B104, pp. 229–233, 2005.

[21] P. L. Combettes, Strong convergence of block-iterative outer approximation methods for convex optimization, *SIAM J. Control Optim.*, vol. 38, pp. 538–565, 2000.

[22] P. L. Combettes, Fejér-monotonicity in convex optimization, in: *Encyclopedia of Optimization*, (C. A. Floudas and P. M. Pardalos, Eds.), vol. 2, pp. 106–114. Springer-Verlag, New York, 2001. (Also available in 2nd ed., pp. 1016–1024, 2009.)

[23] P. L. Combettes, Systems of structured monotone inclusions: Duality, algorithms, and applications, *SIAM J. Optim.*, vol. 23, pp. 2420–2447, 2013.

[24] P. L. Combettes, Monotone operator theory in convex optimization, *Math. Program.*, vol. B170, pp. 177–206, 2018.

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

[26] P. L. Combettes and C. L. Müller, Perspective maximum likelihood-type estimation via proximal decomposition, *Electron. J. Stat.*, vol. 14, pp. 207–238, 2020.

[27] P. L. Combettes and J.-C. Pesquet, Deep neural network structures solving variational inequalities, *Set-Valued Var. Anal.*, vol. 28, pp. 491–518, 2020.

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

[29] 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.

[30] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems.* Springer-Verlag, New York, 2003.

[31] M. Fukushima, The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with applications to the traffic equilibrium problem, *Math. Program.*, vol. 72, pp. 1–15, 1996.

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

[33] M. Hintermüller and G. Stadler, An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration, *SIAM J. Sci. Comput.*, vol. 28, pp. 1–23, 2006.

[34] P. R. Johnstone and J. Eckstein, Projective splitting with forward steps, *Math. Program.*, published online 2020-09-30.

[35] P. R. Johnstone and J. Eckstein, Single-forward-step projective splitting: Exploiting cocoercivity, *Comput. Optim. Appl.*, vol. 78, pp. 125–166, 2021.

[36] M. Li and X. Yuan, A strictly contractive Peaceman-Rachford splitting method with logarithmic-quadratic proximal regularization for convex programming, *Math. Oper. Res.*, vol. 40, pp. 842–858, 2015.

[37] 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.

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

[39] T. Mizoguchi and I. Yamada, Hypercomplex tensor completion via convex optimization, *IEEE Trans. Signal Process.*, vol. 67, pp. 4078–4092, 2019.

[40] N. M. Nam, T. A. Nguyen, R. B. Rector, and J. Sun, Nonsmooth algorithms and Nesterov's smoothing technique for generalized Fermat-Torricelli problems, *SIAM J. Optim.*, vol. 24, pp. 1815–1839, 2014.

[41] S. Ono, T. Miyata, and I. Yamada, Cartoon-texture image decomposition using blockwise low-rank texture characterization, *IEEE Trans. Signal Process.*, vol. 23, pp. 1128–1142, 2014.

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

[43] 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.

[44] T. Pennanen, Dualization of generalized equations of maximal monotone type, *SIAM J. Optim.*, vol. 10, pp. 809–835, 2000.

[45] 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.

[46] S. M. Robinson, Composition duality and maximal monotonicity, *Math. Program.*, vol. 85, pp. 1–13, 1999.

[47] R. T. Rockafellar, Duality and stability in extremum problems involving convex functions, *Pacific J. Math.*, vol. 21, pp. 167–187, 1967.

[48] R. T. Rockafellar, *Convex Analysis.* Princeton University Press, Princeton, NJ, 1970.

[49] 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.

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

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

[52] L. Rosasco, S. Villa, and B. C. Vũ, A stochastic inertial forward-backward splitting algorithm for multivariate monotone inclusions, *Optimization*, vol. 65, pp. 1293–1314, 2016.

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

[54] 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.

[55] X. Yan and J. Bien, Rare feature selection in high dimensions, *J. Amer. Statist. Assoc.*, published online 2020-09-01.

[56] P. Yi and L. Pavel, An operator splitting approach for distributed generalized Nash equilibria computation, *Automatica*, vol. 102, pp. 111–121, 2019.

Chapter 10

CONCLUSION

10.1 Summary

We have addressed the important open questions **(Q1)**–**(Q8)** discussed in Chapter 1. In doing so, we have developed novel nonlinear analysis tools and methodologies to advance the field of monotone operator theory and its applications. More precisely:

- We have introduced the notion of a warped resolvent and developed a warped resolvent algorithmic framework for monotone inclusions. This framework brought together two seemingly different approaches: That of [4, 7, 9], which is based on Tseng's forward-backward-forward method [14], and that of [1, 10–12], which is based on the projective splitting framework.
- We have developed Bregman forward-backward algorithm for solving monotone inclusions in Banach spaces, as well as establishing its convergence.
- A saddle formalism was proposed for analyzing and solving highly structured system of monotone inclusions.
- Flexible algorithms for solving highly modular Nash equilibria, variational inequalities, and network flows were presented.
- We have shed more light on the implementation, the features, and the behavior of block-activated algorithms for solving multicomponent fully nonsmooth minimization.

10.2 Future work

Direction 10.1 While we have only emphasized the cutting plane methods in Section 3.2.4, we illustrated in [5, Section 3] that the proximal algorithm

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = J^{K_n}_{\gamma_n M} x_n, \tag{10.1}$$

based on direct applications of warped resolvents is of interest. In general, however, they do not conform to the format of (3.23), where the update equation is (we set $\lambda_n = 1$ for simplicity)

$$x_{n+1} = \mathrm{proj}_{H_n} x_n, \quad \text{where} \quad H_n = \Big\{ z \in \mathcal{H} \;\Big|\; \big\langle z - J^{K_n}_{\gamma_n M}\widetilde{x}_n \mid K_n\widetilde{x}_n - K_n J^{K_n}_{\gamma_n M}\widetilde{x}_n \big\rangle \leqslant 0 \Big\}. \tag{10.2}$$

These two approaches can be brought together as instantiations of the more general update

$$x_{n+1} = \mathrm{proj}^{Q_n}_{H_n} x_n, \quad \text{where} \quad H_n = \Big\{ z \in \mathcal{H} \;\Big|\; \big\langle z - J^{K_n}_{\gamma_n M}\widetilde{x}_n \mid K_n\widetilde{x}_n - K_n J^{K_n}_{\gamma_n M}\widetilde{x}_n \big\rangle \leqslant 0 \Big\}, \tag{10.3}$$

which involves the warped projector of Example 3.5 with respect to an operator $Q_n \colon \mathcal{H} \to \mathcal{H}$. If $Q_n = K_n$ and $\widetilde{x}_n = x_n$, then (10.3) yields (10.1). On the other hand, if $Q_n = \mathrm{Id}$, then (10.3) yields (10.2). Note also that if, $Q_n = K_n = \nabla f_n$, for some Legendre function f_n, and $\widetilde{x}_n = x_n$, then (10.3) gives the framework of [8]. Beyond this, replacing the standard projection $\mathrm{proj}_{H_n} x_n$ by a warped projection $\mathrm{proj}^{Q_n}_{H_n} x_n$ in Theorems 3.16 and 3.22 opens a vast field for algorithmic development.

Direction 10.2 In Theorems 3.16 and 3.22, the algorithms operate by using a single point (y_n, y_n^*) in gra M at iteration n. It may be advantageous to use a finite family $(y_{i,n}, y_{i,n}^*)_{i\in I_n}$ of points in gra M, say

$$(\forall i \in I_n) \quad (y_{i,n}, y_{i,n}^*) = \Big(J^{K_{i,n}}_{\gamma_{i,n} M}\widetilde{x}_{i,n}, \gamma_{i,n}^{-1}(K_{i,n}\widetilde{x}_{i,n} - K_{i,n}y_{i,n}) \Big). \tag{10.4}$$

By monotonicity of M, $(\forall i \in I_n)(\forall z \in \mathrm{zer}\, M)$ $\langle z \mid y_{i,n}^*\rangle \leqslant \langle y_{i,n} \mid y_{i,n}^*\rangle$. Therefore, using ideas found in the area of convex feasibility algorithms [6, 13], at every iteration n, given strictly positive weights $(\omega_{i,n})_{i\in I_n}$ adding up to 1, we average these inequalities to create a new half-space H_n containing zer M, namely

$$\mathrm{zer}\, M \subset H_n = \big\{ z \in \mathcal{X} \mid \langle z \mid y_n^*\rangle \leqslant \eta_n \big\}, \quad \text{where} \quad \begin{cases} y_n^* = \sum_{i\in I_n} \omega_{i,n} y_{i,n}^* \\ \eta_n = \sum_{i\in I_n} \omega_{i,n}\langle y_{i,n} \mid y_{i,n}^*\rangle. \end{cases} \tag{10.5}$$

Now set

$$\Lambda_n = \begin{cases} \dfrac{\sum_{i\in I_n} \omega_{i,n}\langle y_{i,n} - x_n \mid y_{i,n}^*\rangle}{\big\| \sum_{i\in I_n} \omega_{i,n} y_{i,n}^* \big\|^2}, & \text{if } \sum_{i\in I_n} \omega_{i,n}\langle x_n - y_{i,n} \mid y_{i,n}^*\rangle > 0; \\ 0, & \text{otherwise.} \end{cases} \tag{10.6}$$

Then, employing $\mathrm{proj}_{H_n} x_n = x_n + \Lambda_n \sum_{i\in I_n} \omega_{i,n} y_{i,n}^*$ as the point x_{n+1} in (3.23) and as the point $x_{n+1/2}$ in (3.38) results in multi-point extensions of Theorems 3.16 and 3.22.

Direction 10.3 The Yosida approximation plays an important role in monotone operator theory (see, e.g., [2, 3, 7]). As the warped resolvents can be much easier to compute than the standard ones, an interesting question is therefore to seek an extension of the Yosida approximation based on the warped resolvents, and analyze its properties.

Direction 10.4 We have proposed a full solution for the Bregman forward-backward splitting algorithm. It remains an open question what the Bregman version of the Douglas–Rachford algorithm is. This topic could be pursued by using tools from Chapters 3 and 4.

Raleigh, September 13, 2021

References

[1] 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.

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

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

[4] 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.

[5] M. N. Bùi and P. L. Combettes, Warped proximal iterations for monotone inclusions, 2019-08-19. `https://arxiv.org/pdf/1908.07077v1`

[6] P. L. Combettes, Construction d'un point fixe commun à une famille de contractions fermes, *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 320, pp. 1385–1390, 1995.

[7] P. L. Combettes, Systems of structured monotone inclusions: Duality, algorithms, and applications, *SIAM J. Optim.,* vol. 23, pp. 2420–2447, 2013.

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

[9] 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.

[10] Y. Dong, An LS-free splitting method for composite mappings, *Appl. Math. Lett.*, vol. 18, pp. 843–848, 2005.

[11] 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.

[12] 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.

[13] K. C. Kiwiel and B. Łopuch, Surrogate projection methods for finding fixed points of firmly nonexpansive mappings, *SIAM J. Optim.*, vol. 7, pp. 1084–1102, 1997.

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