SIAM J. OPTIM.
Vol. xx, pp. x
© xxxx Society for Industrial and Applied Mathematics
x–x

SOLVING COUPLED COMPOSITE MONOTONE INCLUSIONS BY SUCCESSIVE FEJÉR APPROXIMATIONS OF THEIR KUHN–TUCKER SET*

ABDULLAH ALOTAIBI†, PATRICK L. COMBETTES‡, and NASEER SHAHZAD§

Abstract. We propose a new class of primal-dual Fejér monotone algorithms for solving systems of composite monotone inclusions. Our construction is inspired by a framework used by Eckstein and Svaiter for the basic problem of finding a zero of the sum of two monotone operators. At each iteration, points in the graph of the monotone operators present in the model are used to construct a half-space containing the Kuhn–Tucker set associated with the system. The primal-dual update is then obtained via a relaxed projection of the current iterate onto this half-space. An important feature that distinguishes the resulting splitting algorithms from existing ones is that they do not require prior knowledge of bounds on the linear operators involved or the inversion of linear operators.

Key words. duality, Fejér monotonicity, monotone inclusion, monotone operator, primal-dual algorithm, splitting algorithm

AMS subject classifications. Primary 47H05; Secondary 65K05, 90C25, 94A08

1. Introduction.

The first monotone operator splitting methods arose in the late 1970s and were motivated by applications in mechanics and partial differential equations [32, 35, 39]. In recent years, the field of monotone operator splitting algorithms has benefited from a new impetus, fueled by emerging application areas such as signal and image processing, statistics, optimal transport, machine learning, and domain decomposition methods [3, 5, 24, 36, 41, 43, 46]. Three main algorithms dominate the field explicitly or implicitly: the forward-backward method [38], the Douglas–Rachford method [37], and the forward-backward-forward method [47]. These methods were originally designed to solve inclusions of the type $0 \in Ax + Bx$, where A and B are maximally monotone operators acting on a Hilbert space (via product space reformulations, they can also be extended to problems involving sums of more than two operators [9, 45]). Until recently, a significant challenge in the field was to design splitting techniques for inclusions involving linearly composed operators, say

$$0 \in Ax + L^*BLx, \tag{1.1}$$

where A and B are maximally monotone operators acting on Hilbert spaces $\mathcal{H}$ and $\mathcal{G}$, respectively, and L is a bounded linear operator from $\mathcal{H}$ to $\mathcal{G}$. In the case when A and B are subdifferentials, say $A = \partial f$ and $B = \partial g$, where $f\colon \mathcal{H} \to]-\infty, +\infty]$ and $g\colon \mathcal{G} \to]-\infty, +\infty]$ are lower semicontinuous convex functions satisfying a suitable

*Received by the editors December 20, 2013; accepted for publication (in revised form) October 01, 2014; published electronically DATE.
http://www.siam.org/journals/mms/x-x/XXXX.html.
†King Abdulaziz University, Department of Mathematics, Jeddah 21859, Saudi Arabia (aalotaibi@kau.edu.sa)
‡Sorbonne Universités – UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005, Paris, France (plc@ljll.math.upmc.fr)
§King Abdulaziz University, Department of Mathematics, Jeddah 21859, Saudi Arabia (nshahzad@kau.edu.sa)

© xxxx Society for Industrial and Applied Mathematics

constraint qualification, (1.1) corresponds to the minimization problem

$$\underset{x\in\mathcal{H}}{\text{minimize}}\ \ f(x)+g(Lx). \tag{1.2}$$

The Fenchel–Rockafellar dual of this problem is

$$\underset{v^*\in\mathcal{G}}{\text{minimize}}\ \ f^*(-L^*v^*)+g^*(v^*) \tag{1.3}$$

and the associated Kuhn–Tucker set is

$$\boldsymbol{Z}=\big\{(x,v^*)\in\mathcal{H}\oplus\mathcal{G}\ \big|\ -L^*v^*\in\partial f(x)\ \text{ and } Lx\in\partial g^*(v^*)\big\}. \tag{1.4}$$

The importance of this set is discussed extensively in [44], notably in connection with the fact that Kuhn–Tucker points provide solutions to (1.2) and (1.3). To the best of our knowledge, the first splitting method for composite problems of the form (1.1) is that proposed in [16], which was developed around the following formulation.

Problem 1.1. Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, and set $\boldsymbol{\mathcal{K}}=\mathcal{H}\oplus\mathcal{G}$. Let $A\colon\mathcal{H}\to 2^{\mathcal{H}}$ and $B\colon\mathcal{G}\to 2^{\mathcal{G}}$ be maximally monotone operators, and let $L\colon\mathcal{H}\to\mathcal{G}$ be a bounded linear operator. Consider the inclusion problem

$$\text{find}\ \ \overline{x}\in\mathcal{H}\ \ \text{such that}\ \ 0\in A\overline{x}+L^*BL\overline{x}, \tag{1.5}$$

the dual problem

$$\text{find}\ \ \overline{v}^*\in\mathcal{G}\ \ \text{such that}\ \ 0\in -LA^{-1}(-L^*\overline{v}^*)+B^{-1}\overline{v}^*, \tag{1.6}$$

and the associated Kuhn–Tucker set

$$\boldsymbol{Z}=\big\{(x,v^*)\in\boldsymbol{\mathcal{K}}\ \big|\ -L^*v^*\in Ax\ \text{ and } Lx\in B^{-1}v^*\big\}. \tag{1.7}$$

The problem is to find a point in $\boldsymbol{Z}$. The sets of solutions to (1.5) and (1.6) are denoted by $\mathscr{P}$ and $\mathscr{D}$, respectively.

The Kuhn–Tucker set (1.7) is a natural extension of (1.4) to general monotone operators. In [16], a point in $\boldsymbol{Z}$ was obtained by applying the forward-backward-forward method to a suitably decomposed inclusion in $\mathcal{H}\oplus\mathcal{G}$ (the use of Douglas–Rachford splitting was also discussed there). Subsequently, the idea of using traditional splitting techniques to find Kuhn–Tucker points was further exploited in a variety of settings, e.g., [1, 12, 14, 23, 25, 26, 48]. Despite their broad range of applicability, existing splitting methods suffer from two shortcomings that preclude their use in certain settings. Thus, a shortcoming of splitting methods based on the forward-backward-forward [16, 25] or the forward-backward algorithms [2, 26, 48] is that they require knowledge of $\|L\|$; this is also true for the Douglas–Rachford-based method of [14]. On the other hand, a shortcoming of splitting methods based on the Douglas–Rachford [16, Remark 2.9] or Spingarn [1] algorithms is that they require the inversion of linear operators, as does [12, Algorithm 3]. In some applications, however, $\|L\|$ cannot be evaluated reliably and the inversion of linear operators is not numerically feasible. As will be seen in Section 4, this issue becomes particularly acute when dealing with systems of coupled monotone inclusions, which constitute the main motivation for our investigation.

Our objective is to devise a new class of algorithms for solving Problem 1.1 that alleviate the above-mentioned shortcomings of existing methods. Our approach is

© xxxx Society for Industrial and Applied Mathematics

inspired by an original splitting framework proposed in [28] for solving the basic inclusion (see also [29] for the extension to the sum of several operators)

$$0 \in Ax + Bx. \tag{1.8}$$

The main idea of [28] is to use points in the graphs of A and B to construct a sequence of Fejér approximations to the so-called extended solution set

$$\big\{(x, v^*) \in \mathcal{H} \oplus \mathcal{H} \;\big|\; -v^* \in Ax \text{ and } v^* \in Bx\big\} \tag{1.9}$$

and to iterate by projection onto these successive approximations. This extended solution set is actually nothing but the specialization of the Kuhn–Tucker set (1.7) to the case when $\mathcal{G} = \mathcal{H}$ and $L = \mathrm{Id}$. This construction led to novel splitting methods for solving (1.8) that do not seem to derive from the traditional methods mentioned above. In the present paper, we extend it significantly beyond (1.8) in order to design new primal-dual splitting algorithms for Problem 1.1.

The paper is organized as follows. Preliminary results are established in Section 2 and algorithms for solving Problem 1.1 are developed in Section 3. These results are then used in Section 4 to solve systems of composite monotone inclusions in duality.

Notation. The scalar product of a Hilbert space is denoted by $\langle \cdot \mid \cdot \rangle$ and the associated norm by $\|\cdot\|$. The symbols $\rightharpoonup$ and $\rightarrow$ denote, respectively, weak and strong convergence, and Id denotes the identity operator. Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, let $2^{\mathcal{H}}$ be the power set of $\mathcal{H}$, and let $A\colon \mathcal{H} \rightarrow 2^{\mathcal{H}}$. We denote by $\operatorname{ran} A = \big\{u \in \mathcal{H} \;\big|\; (\exists\, x \in \mathcal{H})\; u \in Ax\big\}$ the range of A, by $\operatorname{gra} A = \big\{(x, u) \in \mathcal{H} \times \mathcal{H} \;\big|\; u \in Ax\big\}$ the graph of A, and by A^{-1} the inverse of A, which is defined through its graph $\big\{(u, x) \in \mathcal{H} \times \mathcal{H} \;\big|\; (x, u) \in \operatorname{gra} A\big\}$. The resolvent of A is $J_A = (\mathrm{Id} + A)^{-1}$. We say that A is monotone if

$$(\forall (x, u) \in \operatorname{gra} A)(\forall (y, v) \in \operatorname{gra} A) \quad \langle x - y \mid u - v \rangle \geqslant 0, \tag{1.10}$$

and maximally monotone if there does not exist any monotone operator $B\colon \mathcal{H} \rightarrow 2^{\mathcal{H}}$ such that $\operatorname{gra} A \subset \operatorname{gra} B \neq \operatorname{gra} A$. In this case, J_A is firmly nonexpansive and defined everywhere on $\mathcal{H}$. The Hilbert direct sum of $\mathcal{H}$ and $\mathcal{G}$ is denoted by $\mathcal{H} \oplus \mathcal{G}$. The projection operator onto a nonempty closed convex subset C of $\mathcal{H}$ is denoted by P_C. The necessary background on convex analysis and monotone operators will be found in [9].

2. Preliminary results.

We first investigate some basic properties of Problem 1.1, starting with the fact that Kuhn–Tucker points automatically provide primal and dual solutions.

Proposition 2.1. *In the setting of Problem 1.1, the following hold:*

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

(ii) $\mathscr{P} \neq \varnothing \Leftrightarrow \boldsymbol{Z} \neq \varnothing \Leftrightarrow \mathscr{D} \neq \varnothing$.

Proof. This is [16, Proposition 2.8] (see also [42] for (ii)). ■

A fundamental concept in algorithmic nonlinear analysis is that of Fejér monotonicity: a sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ in a Hilbert space $\boldsymbol{\mathcal{H}}$ is said to be Fejér monotone with respect to a set $\boldsymbol{C} \subset \boldsymbol{\mathcal{H}}$ if

$$(\forall \boldsymbol{z} \in \boldsymbol{C})(\forall n \in \mathbb{N}) \quad \|\boldsymbol{x}_{n+1} - \boldsymbol{z}\| \leqslant \|\boldsymbol{x}_n - \boldsymbol{z}\|. \tag{2.1}$$

Alternatively (see [8, Section 2]), $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\boldsymbol{C}$ if, for every $n \in \mathbb{N}$, $\boldsymbol{x}_{n+1}$ is a relaxed projection of $\boldsymbol{x}_n$ onto a closed affine half-space

© xxxx Society for Industrial and Applied Mathematics

$\boldsymbol{H}_n$ containing $\boldsymbol{C}$, i.e.,

(2.2)
$$(\forall n \in \mathbb{N}) \quad \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \lambda_n (P_{\boldsymbol{H}_n} \boldsymbol{x}_n - \boldsymbol{x}_n), \quad \text{where} \quad 0 \leqslant \lambda_n \leqslant 2 \quad \text{and} \quad \boldsymbol{C} \subset \boldsymbol{H}_n.$$

The half-spaces $(\boldsymbol{H}_n)_{n\in\mathbb{N}}$ in (2.2) are called Fejér approximations to $\boldsymbol{C}$. The Fejér monotonicity property (2.1) makes it possible to greatly simplify the analysis of the asymptotic behavior of a broad class of algorithms; see [7, 9, 21, 22, 30, 31] for background, examples, and historical notes.

In the following proposition, we consider the problem of constructing a Fejér approximation to the Kuhn–Tucker set (1.7).

Proposition 2.2. *In the setting of Problem 1.1, for every* $\mathsf{a} = (a, a^*) \in \operatorname{gra} A$ *and every* $\mathsf{b} = (b, b^*) \in \operatorname{gra} B$*, set*

(2.3)
$$\boldsymbol{H}_{\mathsf{a},\mathsf{b}} = \Big\{ \boldsymbol{x} \in \boldsymbol{\mathcal{K}} \;\Big|\; \langle \boldsymbol{x} \mid \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} \rangle \leqslant \eta_{\mathsf{a},\mathsf{b}} \Big\}, \quad \textit{where} \quad \begin{cases} \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} = (a^* + L^* b^*, b - La) \\ \eta_{\mathsf{a},\mathsf{b}} = \langle a \mid a^* \rangle + \langle b \mid b^* \rangle. \end{cases}$$

Then the following hold:

(i) *Let* $\mathsf{a} \in \operatorname{gra} A$ *and* $\mathsf{b} \in \operatorname{gra} B$*. Then* $\boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} = \boldsymbol{0} \;\Leftrightarrow\; \boldsymbol{H}_{\mathsf{a},\mathsf{b}} = \boldsymbol{\mathcal{K}} \;\Rightarrow\; (a, b^*) \in \boldsymbol{Z}$ *and* $\eta_{\mathsf{a},\mathsf{b}} = 0$*.*

(ii) *Let* $\mathsf{a} \in \operatorname{gra} A$ *and* $\mathsf{b} \in \operatorname{gra} B$*. Then* $\boldsymbol{Z} \subset \boldsymbol{H}_{\mathsf{a},\mathsf{b}}$*.*

(iii) $\boldsymbol{Z} = \bigcap_{\mathsf{a}\in\operatorname{gra} A} \bigcap_{\mathsf{b}\in\operatorname{gra} B} \boldsymbol{H}_{\mathsf{a},\mathsf{b}}$*.*

(iv) *Let* $(a, a^*) \in \operatorname{gra} A$*,* $(b, b^*) \in \operatorname{gra} B$*, and* $(x, v^*) \in \boldsymbol{\mathcal{K}}$*. Set* $s^* = a^* + L^* b^*$*,* $t = b - La$*, and* $\sigma = \sqrt{\|s^*\|^2 + \|t\|^2}$*; if* $\sigma > 0$*, set* $\Delta = (\langle x \mid s^* \rangle + \langle t \mid v^* \rangle - \langle a \mid a^* \rangle - \langle b \mid b^* \rangle)/\sigma$*. Then*

(2.4)
$$P_{\boldsymbol{H}_{\mathsf{a},\mathsf{b}}}(x, v^*) = \begin{cases} \big(x - (\Delta/\sigma) s^*, v^* - (\Delta/\sigma) t\big), & \textit{if } \sigma > 0 \textit{ and } \Delta > 0; \\ (x, v^*), & \textit{otherwise.} \end{cases}$$

Proof. (i): Suppose that $\boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} = \boldsymbol{0}$. Then $-L^* b^* = a^* \in Aa$ and $La = b \in B^{-1} b^*$. Hence, (1.7) implies that $(a, b^*) \in \boldsymbol{Z}$. In addition,

$$\eta_{\mathsf{a},\mathsf{b}} = \langle a \mid a^* \rangle + \langle b \mid b^* \rangle = \langle a \mid -L^* b^* \rangle + \langle La \mid b^* \rangle = -\langle La \mid b^* \rangle + \langle La \mid b^* \rangle = 0 \tag{2.5}$$

and therefore $\boldsymbol{H}_{\mathsf{a},\mathsf{b}} = \boldsymbol{\mathcal{K}}$. Conversely, $\boldsymbol{H}_{\mathsf{a},\mathsf{b}} = \boldsymbol{\mathcal{K}} \Rightarrow \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} = \boldsymbol{0}$ and $\eta_{\mathsf{a},\mathsf{b}} = 0$.

(ii): Suppose that $\boldsymbol{x} = (x, v^*) \in \boldsymbol{Z}$. Then $(x, -L^* v^*) \in \operatorname{gra} A$ and, by monotonicity of A,

$$\langle a - x \mid a^* + L^* v^* \rangle \geqslant 0. \tag{2.6}$$

Likewise, since $(Lx, v^*) \in \operatorname{gra} B$, we have

$$\langle b - Lx \mid b^* - v^* \rangle \geqslant 0. \tag{2.7}$$

Using (2.6) and (2.7), we obtain

$$\begin{aligned}
\big\langle \boldsymbol{x} \mid \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} \big\rangle &= \langle x \mid a^* + L^* b^* \rangle + \langle b - La \mid v^* \rangle \\
&= \langle x \mid a^* + L^* v^* \rangle + \langle Lx \mid b^* - v^* \rangle + \langle b - Lx \mid v^* \rangle + \langle x - a \mid L^* v^* \rangle \\
&= \langle x - a \mid a^* + L^* v^* \rangle + \langle a \mid a^* \rangle + \langle La \mid v^* \rangle \\
&\quad + \langle Lx - b \mid b^* - v^* \rangle + \langle b \mid b^* \rangle - \langle b \mid v^* \rangle + \langle b - Lx \mid v^* \rangle + \langle x - a \mid L^* v^* \rangle \\
&\leqslant \langle a \mid a^* \rangle + \langle La - b \mid v^* \rangle + \langle b \mid b^* \rangle + \langle b - Lx \mid v^* \rangle + \langle x - a \mid L^* v^* \rangle \\
&= \langle a \mid a^* \rangle + \langle b \mid b^* \rangle \\
&= \eta_{\mathsf{a},\mathsf{b}}. \qquad (2.8)
\end{aligned}$$

© xxxx Society for Industrial and Applied Mathematics

Thus, $\boldsymbol{x} \in \boldsymbol{H}_{\mathsf{a},\mathsf{b}}$.

(iii): By (ii), $\boldsymbol{Z} \subset \bigcap_{\mathsf{a}\in\operatorname{gra} A}\bigcap_{\mathsf{b}\in\operatorname{gra} B}\boldsymbol{H}_{\mathsf{a},\mathsf{b}}$. Conversely, fix $\mathsf{a} \in \operatorname{gra} A$ and $\mathsf{b} \in \operatorname{gra} B$, and let $\boldsymbol{x} = (x, v^*) \in \boldsymbol{H}_{\mathsf{a},\mathsf{b}}$. Then $\langle \boldsymbol{x} \mid \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}}\rangle \leqslant \eta_{\mathsf{a},\mathsf{b}}$ and therefore

$$\begin{aligned}\langle (a,b^*)-(x,v^*) \mid (a^*,b)-(-L^*v^*,Lx)\rangle &= \langle (a-x,b^*-v^*) \mid (a^*+L^*v^*,b-Lx)\rangle\\ &= \langle a-x \mid a^*+L^*v^*\rangle + \langle b-Lx \mid b^*-v^*\rangle\\ &= \eta_{\mathsf{a},\mathsf{b}} - \langle \boldsymbol{x} \mid \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}}\rangle\\ &\geqslant 0. \end{aligned} \tag{2.9}$$

Now set $\boldsymbol{M}\colon \boldsymbol{\mathcal{K}} \to 2^{\boldsymbol{\mathcal{K}}}\colon (z,w^*) \mapsto Az \times B^{-1}w^*$. Then, since $((a,b^*),(a^*,b))$ is an arbitrary point in $\operatorname{gra}\boldsymbol{M}$ and since [9, Propositions 20.22 and 20.23] imply that $\boldsymbol{M}$ is maximally monotone, we derive from (2.9) that $((x,v^*),(-L^*v^*,Lx)) \in \operatorname{gra}\boldsymbol{M}$, i.e., that $\boldsymbol{x} \in \boldsymbol{Z}$.

(iv): Let $\boldsymbol{x} \in \boldsymbol{\mathcal{K}}$. As seen in (i), if $\boldsymbol{s}^*_{\mathsf{a},\mathsf{b}} = \boldsymbol{0}$, then $\eta_{\mathsf{a},\mathsf{b}} = 0$ and $\boldsymbol{H}_{\mathsf{a},\mathsf{b}} = \boldsymbol{\mathcal{K}}$. Hence $\langle \boldsymbol{x} \mid \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}}\rangle = \eta_{\mathsf{a},\mathsf{b}}$ and $P_{\boldsymbol{H}_{\mathsf{a},\mathsf{b}}}\boldsymbol{x} = \boldsymbol{x}$. Otherwise, it follows from [9, Example 28.16] that

$$P_{\boldsymbol{H}_{\mathsf{a},\mathsf{b}}}\boldsymbol{x} = \begin{cases} \boldsymbol{x} - \dfrac{\langle \boldsymbol{x} \mid \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}}\rangle - \eta_{\mathsf{a},\mathsf{b}}}{\|\boldsymbol{s}^*_{\mathsf{a},\mathsf{b}}\|^2}\boldsymbol{s}^*_{\mathsf{a},\mathsf{b}}, & \text{if } \langle \boldsymbol{x} \mid \boldsymbol{s}^*_{\mathsf{a},\mathsf{b}}\rangle > \eta_{\mathsf{a},\mathsf{b}};\\ \boldsymbol{x}, & \text{otherwise.}\end{cases} \tag{2.10}$$

In view of (2.3), the proof is complete. ■

Remark 2.3.

(i) The fact that $\boldsymbol{Z}$ is closed and convex (Proposition 2.1(i)) is also apparent in Proposition 2.2(iii), which exhibits $\boldsymbol{Z}$ as an intersection of closed affine half-spaces.

(ii) The inclusion $\boldsymbol{Z} \subset \boldsymbol{H}_{\mathsf{a},\mathsf{b}}$ (Proposition 2.2(i)) will play a key role in the paper. This construction is inspired by that of [28, Lemma 3], where $\mathcal{G} = \mathcal{H}$ and $L = \operatorname{Id}$.

Our analysis will require the following asymptotic principle, which is of interest in its own right.

Proposition 2.4. *In the setting of Problem 1.1, let $(a_n, a_n^*)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} A$, let $(b_n, b_n^*)_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} B$, and let $(\overline{x},\overline{v}^*) \in \boldsymbol{\mathcal{K}}$. Suppose that $a_n \rightharpoonup \overline{x}$, $b_n^* \rightharpoonup \overline{v}^*$, $a_n^* + L^*b_n^* \to 0$, and $La_n - b_n \to 0$. Then $\langle a_n \mid a_n^*\rangle + \langle b_n \mid b_n^*\rangle \to 0$ and $(\overline{x},\overline{v}^*) \in \boldsymbol{Z}$.*

Proof. Define

$$\boldsymbol{V} = \big\{(x,y) \in \boldsymbol{\mathcal{K}} \;\big|\; Lx = y\big\}. \tag{2.11}$$

Then

$$\boldsymbol{V}^\perp = \big\{(u^*,v^*) \in \boldsymbol{\mathcal{K}} \;\big|\; u^* = -L^*v^*\big\}. \tag{2.12}$$

Now set

$$\boldsymbol{A}\colon \boldsymbol{\mathcal{K}} \to 2^{\boldsymbol{\mathcal{K}}}\colon (x,y) \mapsto Ax \times By. \tag{2.13}$$

We deduce from (1.7) that, for every $(x,v^*) \in \boldsymbol{\mathcal{K}}$,

$$\begin{aligned}(x,v^*) \in \boldsymbol{Z} &\Leftrightarrow \begin{cases}(x,-L^*v^*) \in \operatorname{gra} A\\ (Lx,v^*) \in \operatorname{gra} B\end{cases}\\ &\Leftrightarrow (\boldsymbol{x},\boldsymbol{u}) = \big((x,Lx),(-L^*v^*,v^*)\big) \in (\boldsymbol{V}\times\boldsymbol{V}^\perp)\cap\operatorname{gra}\boldsymbol{A}.\end{aligned} \tag{2.14}$$

© xxxx Society for Industrial and Applied Mathematics

On the other hand, [1, Lemma 3.1] asserts that

(2.15)
$$(\forall (x,y)\in\boldsymbol{\mathcal{K}})\quad \begin{cases} P_{\boldsymbol{V}}(x,y)=\big((\mathrm{Id}+L^*L)^{-1}(x+L^*y),L(\mathrm{Id}+L^*L)^{-1}(x+L^*y)\big)\\ P_{\boldsymbol{V}^\perp}(x,y)=\big(L^*(\mathrm{Id}+LL^*)^{-1}(Lx-y),-(\mathrm{Id}+LL^*)^{-1}(Lx-y)\big).\end{cases}$$

Now set

$$\overline{\boldsymbol{x}}=(\overline{x},L\overline{x}),\quad \overline{\boldsymbol{u}}=(-L^*\overline{v}^*,\overline{v}^*),\quad \text{and}\quad (\forall n\in\mathbb{N})\quad \begin{cases}\boldsymbol{x}_n=(a_n,b_n)\\ \boldsymbol{u}_n=(a_n^*,b_n^*).\end{cases}\tag{2.16}$$

Since $a_n^*+L^*b_n^*\to 0$ and $La_n-b_n\to 0$, we derive from (2.15) that $P_{\boldsymbol{V}}\boldsymbol{u}_n\to\boldsymbol{0}$ and $P_{\boldsymbol{V}^\perp}\boldsymbol{x}_n\to\boldsymbol{0}$. Altogether, since L and L^* are weakly continuous, the assumptions yield

$$\begin{cases}(\forall n\in\mathbb{N})\ \ (\boldsymbol{x}_n,\boldsymbol{u}_n)\in\operatorname{gra}\boldsymbol{A}\\ \boldsymbol{x}_n\rightharpoonup\overline{\boldsymbol{x}}\\ \boldsymbol{u}_n\rightharpoonup\overline{\boldsymbol{u}}\\ P_{\boldsymbol{V}^\perp}\boldsymbol{x}_n\to\boldsymbol{0}\\ P_{\boldsymbol{V}}\boldsymbol{u}_n\to\boldsymbol{0}.\end{cases}\tag{2.17}$$

However, (2.17) and [9, Proposition 25.3] imply that

$$\langle\boldsymbol{x}_n\mid\boldsymbol{u}_n\rangle\to 0\quad\text{and}\quad(\overline{\boldsymbol{x}},\overline{\boldsymbol{u}})\in(\boldsymbol{V}\times\boldsymbol{V}^\perp)\cap\operatorname{gra}\boldsymbol{A}.\tag{2.18}$$

In view of (2.14), the proof is complete. ■

Remark 2.5. In the special case when $\mathcal{G}=\mathcal{H}$ and $L=\mathrm{Id}$, Proposition 2.4 reduces to [6, Corollary 3] (see also [9, Corollary 25.5] for an alternate proof), where $m=2$. The decomposition $\boldsymbol{\mathcal{K}}=\boldsymbol{V}\oplus\boldsymbol{V}^\perp$, where $\boldsymbol{V}$ is as in (2.11), is used in [1] in a different context.

3. Finding Kuhn–Tucker points by Fejér approximations.

In view of Proposition 2.1(i), Problem 1.1 reduces to finding a point in a nonempty closed convex subset of a Hilbert space. This can be achieved via the following generic Fejér-monotone algorithm.

Proposition 3.1.[22] *Let $\boldsymbol{\mathcal{H}}$ be a real Hilbert space, let $\boldsymbol{C}$ be a nonempty closed convex subset of $\boldsymbol{\mathcal{H}}$, and let $\boldsymbol{x}_0\in\boldsymbol{\mathcal{H}}$. Iterate*

$$\begin{array}{l}\textit{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l}\boldsymbol{H}_n \textit{ is a closed affine half-space such that } \boldsymbol{C}\subset\boldsymbol{H}_n\\ \lambda_n\in\,]0,2[\\ \boldsymbol{x}_{n+1}=\boldsymbol{x}_n+\lambda_n(P_{\boldsymbol{H}_n}\boldsymbol{x}_n-\boldsymbol{x}_n).\end{array}\right.\end{array}\tag{3.1}$$

Then the following hold:

(i) *$(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\boldsymbol{C}$: $(\forall\boldsymbol{z}\in\boldsymbol{C})(\forall n\in\mathbb{N})\ \|\boldsymbol{x}_{n+1}-\boldsymbol{z}\|\leqslant\|\boldsymbol{x}_n-\boldsymbol{z}\|$.*

(ii) *$\sum_{n\in\mathbb{N}}\lambda_n(2-\lambda_n)\|P_{\boldsymbol{H}_n}\boldsymbol{x}_n-\boldsymbol{x}_n\|^2<+\infty$.*

(iii) *Suppose that, for every $\boldsymbol{x}\in\boldsymbol{\mathcal{H}}$ and every strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$, $\boldsymbol{x}_{k_n}\rightharpoonup\boldsymbol{x}\Rightarrow\boldsymbol{x}\in\boldsymbol{C}$. Then $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\boldsymbol{C}$.*

© xxxx Society for Industrial and Applied Mathematics

We now derive from the above convergence principle a conceptual primal-dual splitting framework.

Proposition 3.2. *Consider the setting of Problem 1.1. Suppose that $\mathscr{P} \neq \varnothing$, let $x_0 \in \mathcal{H}$, let $v_0^* \in \mathcal{G}$, and iterate*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\quad (a_n, a_n^*) \in \operatorname{gra} A \\
\quad (b_n, b_n^*) \in \operatorname{gra} B \\
\quad s_n^* = a_n^* + L^* b_n^* \\
\quad t_n = b_n - L a_n \\
\quad \sigma_n = \sqrt{\|s_n^*\|^2 + \|t_n\|^2} \\
\quad \textit{if } \sigma_n = 0 \\
\qquad \overline{x} = a_n \\
\qquad \overline{v}^* = b_n^* \\
\qquad \textit{terminate.} \\
\quad \textit{if } \sigma_n > 0 \\
\qquad \lambda_n \in \left]0, 2\right[\\
\qquad \Delta_n = \max\big\{0, (\langle x_n \mid s_n^* \rangle + \langle t_n \mid v_n^* \rangle - \langle a_n \mid a_n^* \rangle - \langle b_n \mid b_n^* \rangle)/\sigma_n\big\} \\
\qquad \theta_n = \lambda_n \Delta_n / \sigma_n \\
\qquad x_{n+1} = x_n - \theta_n s_n^* \\
\qquad v_{n+1}^* = v_n^* - \theta_n t_n.
\end{array}
\tag{3.2}
$$

Then either (3.2) terminates at a solution $(\overline{x}, \overline{v}^*) \in \boldsymbol{Z}$ in a finite number of iterations or it generates infinite sequences $(x_n)_{n\in\mathbb{N}}$ and $(v_n^*)_{n\in\mathbb{N}}$ such that the following hold:

(i) *$(x_n, v_n^*)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $\boldsymbol{Z}$.*

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

(iii) *Suppose that for every $x \in \mathcal{H}$, every $v^* \in \mathcal{G}$, and every strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$,*

$$
\big[\, x_{k_n} \rightharpoonup x \ \textit{ and } \ v_{k_n}^* \rightharpoonup v^* \,\big] \quad \Rightarrow \quad (x, v^*) \in \boldsymbol{Z}. \tag{3.3}
$$

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 $(\overline{x}, \overline{v}^*) \in \boldsymbol{Z}$.

Proof. We first observe that, by Proposition 2.1, $\boldsymbol{Z}$ is nonempty, closed, and convex. Two alternatives are possible. First, suppose that, for some $n \in \mathbb{N}$, $\sigma_n = 0$. Then Proposition 2.2(i) asserts that the algorithm terminates at $(\overline{x}, \overline{v}^*) = (a_n, b_n^*) \in \boldsymbol{Z}$. Now suppose that $(\forall n \in \mathbb{N})\ \sigma_n > 0$. For every $n \in \mathbb{N}$, set

$$
\boldsymbol{x}_n = (x_n, v_n^*), \quad \boldsymbol{s}_n^* = (s_n^*, t_n), \quad \text{and} \quad \eta_n = \langle a_n \mid a_n^* \rangle + \langle b_n \mid b_n^* \rangle, \tag{3.4}
$$

and define

$$
\boldsymbol{H}_n = \big\{ \boldsymbol{x} \in \boldsymbol{\mathcal{K}} \;\big|\; \langle \boldsymbol{x} \mid \boldsymbol{s}_n^* \rangle \leqslant \eta_n \big\}. \tag{3.5}
$$

Then we derive from (3.2) and Proposition 2.2(ii) that $(\forall n \in \mathbb{N})\ \boldsymbol{Z} \subset \boldsymbol{H}_n$. On the other hand, Proposition 2.2(iv) implies that

$$
(\forall n \in \mathbb{N}) \quad \Delta_n = \|P_{\boldsymbol{H}_n} \boldsymbol{x}_n - \boldsymbol{x}_n\| \quad \text{and} \quad \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \lambda_n (P_{\boldsymbol{H}_n} \boldsymbol{x}_n - \boldsymbol{x}_n). \tag{3.6}
$$

Thus, the conclusions follow from Proposition 2.1(i) and Proposition 3.1. ■

At the nth iteration of algorithm (3.2), one picks the quadruple (a_n, a_n^*, b_n, b_n^*) in $\operatorname{gra} A \times \operatorname{gra} B$. In the following corollary, this quadruple is taken in a more restricted

© xxxx Society for Industrial and Applied Mathematics

set adapted to the current primal-dual iterate (x_n, v_n^*), which leads to more explicit convergence conditions.

Corollary 3.3. *Consider the setting of Problem 1.1. Suppose that $\mathscr{P} \neq \varnothing$, let $\varepsilon \in \left]0,1\right[$, let $\alpha \in \left]0,+\infty\right[$, let $x_0 \in \mathcal{H}$, and let $v_0^* \in \mathcal{G}$. For every $(x, v^*) \in \boldsymbol{\mathcal{K}}$, set*

$$\boldsymbol{G}_\alpha(x,v^*) = \Big\{(a,b,a^*,b^*) \in \boldsymbol{\mathcal{K}} \times \boldsymbol{\mathcal{K}} \;\Big|\; (a,a^*) \in \operatorname{gra} A,\ (b,b^*) \in \operatorname{gra} B,\ \text{and} \\ \langle x-a \mid a^* + L^*v^*\rangle + \langle Lx - b \mid b^* - v^*\rangle \geqslant \alpha\big(\|a^* + L^*b^*\|^2 + \|La - b\|^2\big)\Big\}. \tag{3.7}$$

Iterate

$$\begin{array}{l}
\textit{for } n = 0,1,\ldots \\
\quad (a_n, b_n, a_n^*, b_n^*) \in \boldsymbol{G}_\alpha(x_n, v_n^*) \\
\quad s_n^* = a_n^* + L^*b_n^* \\
\quad t_n = b_n - La_n \\
\quad \tau_n = \|s_n^*\|^2 + \|t_n\|^2 \\
\quad \textit{if } \tau_n = 0 \\
\qquad \overline{x} = a_n \\
\qquad \overline{v}^* = b_n^* \\
\qquad \textit{terminate.} \\
\quad \textit{if } \tau_n > 0 \\
\qquad \lambda_n \in [\varepsilon, 2-\varepsilon] \\
\qquad \theta_n = \lambda_n\big(\langle x_n \mid s_n^*\rangle + \langle t_n \mid v_n^*\rangle - \langle a_n \mid a_n^*\rangle - \langle b_n \mid b_n^*\rangle\big)/\tau_n \\
\qquad x_{n+1} = x_n - \theta_n s_n^* \\
\qquad v_{n+1}^* = v_n^* - \theta_n t_n.
\end{array} \tag{3.8}$$

Then either (3.8) terminates at a solution $(\overline{x}, \overline{v}^*) \in \boldsymbol{Z}$ in a finite number of iterations or it generates infinite sequences $(x_n)_{n\in\mathbb{N}}$ and $(v_n^*)_{n\in\mathbb{N}}$ such that the following hold:

(i) $\sum_{n\in\mathbb{N}} \|s_n^*\|^2 < +\infty$ *and* $\sum_{n\in\mathbb{N}} \|t_n\|^2 < +\infty$.

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

(iii) *Suppose that*

$$x_n - a_n \rightharpoonup 0 \quad \text{and} \quad v_n^* - b_n^* \rightharpoonup 0. \tag{3.9}$$

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 $(\overline{x}, \overline{v}^*) \in \boldsymbol{Z}$.

Proof. This corollary is an application of Proposition 3.2. To see this, let $(x, v^*) \in \boldsymbol{\mathcal{K}}$. First, to show that the algorithm is well defined, we must prove that $\boldsymbol{G}_\alpha(x, v^*) \neq \varnothing$. Since $\mathscr{P} \neq \varnothing$, it follows from Proposition 2.1(ii) that $\boldsymbol{Z} \neq \varnothing$. Now let $(a, b^*) \in \boldsymbol{Z}$, and set $a^* = -L^*b^*$ and $b = La$. Then (1.7) yields $(a, a^*) \in \operatorname{gra} A$ and $(b, b^*) \in \operatorname{gra} B$. Moreover,

$$\begin{aligned}
&\langle x-a \mid a^* + L^*v^*\rangle + \langle Lx - b \mid b^* - v^*\rangle \\
&\qquad = -\langle x-a \mid L^*(b^* - v^*)\rangle + \langle L(x-a) \mid b^* - v^*\rangle \\
&\qquad = 0 \\
&\qquad = \alpha\big(\|a^* + L^*b^*\|^2 + \|La - b\|^2\big).
\end{aligned} \tag{3.10}$$

Hence $(a, b, a^*, b^*) \in \boldsymbol{G}_\alpha(x, v^*)$ and (3.8) is well defined. Next, to show that (3.8) is a special case of (3.2) it is enough to consider the case when $(\forall n \in \mathbb{N})\ \tau_n > 0$. Note

© xxxx Society for Industrial and Applied Mathematics

that (3.8) yields

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad \langle x_n \mid s_n^* \rangle + \langle t_n \mid v_n^* \rangle - \langle a_n \mid a_n^* \rangle - \langle b_n \mid b_n^* \rangle & \\
&= \langle x_n - a_n \mid a_n^* + L^* v_n^* \rangle + \langle L x_n - b_n \mid b_n^* - v_n^* \rangle \\
&\geqslant \alpha \big(\|a_n^* + L^* b_n^*\|^2 + \|L a_n - b_n\|^2 \big) \\
&= \alpha \tau_n \\
&> 0. \qquad (3.11)
\end{aligned}
$$

In turn, if we define $(\Delta_n)_{n\in\mathbb{N}}$ as in (3.2), we obtain

$$
(\forall n \in \mathbb{N}) \quad \Delta_n = \frac{\langle x_n \mid s_n^* \rangle + \langle t_n \mid v_n^* \rangle - \langle a_n \mid a_n^* \rangle - \langle b_n \mid b_n^* \rangle}{\sqrt{\tau_n}} \geqslant \alpha \sqrt{\tau_n} > 0. \qquad (3.12)
$$

Hence (3.8) is a special case of (3.2). Moreover, it follows from (3.12) and Proposition 3.2(ii) that

$$
\sum_{n\in\mathbb{N}} \big(\|s_n^*\|^2 + \|t_n\|^2 \big) = \sum_{n\in\mathbb{N}} \tau_n \leqslant \frac{1}{\alpha^2} \sum_{n\in\mathbb{N}} \Delta_n^2 \leqslant \frac{1}{(\alpha\varepsilon)^2} \sum_{n\in\mathbb{N}} \lambda_n (2-\lambda_n) \Delta_n^2 < +\infty, \qquad (3.13)
$$

which establishes (i). On the other hand, (ii) results from (3.8) and (3.13) since

$$
\begin{aligned}
\sum_{n\in\mathbb{N}} \big(\|x_{n+1} - x_n\|^2 + \|v_{n+1}^* - v_n^*\|^2 \big) &= \sum_{n\in\mathbb{N}} \theta_n^2 \tau_n \\
&= \sum_{n\in\mathbb{N}} \lambda_n^2 \Delta_n^2 \\
&\leqslant (2-\varepsilon)^2 \sum_{n\in\mathbb{N}} \Delta_n^2 \\
&< +\infty. \qquad (3.14)
\end{aligned}
$$

Finally, to prove (iii), it remains to check (3.3). Take $x \in \mathcal{H}$, $v^* \in \mathcal{G}$, and a strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $x_{k_n} \rightharpoonup x$ and $v_{k_n}^* \rightharpoonup v^*$. Then it follows from (3.9) and (i) that

$$
a_{k_n} \rightharpoonup x, \quad b_{k_n}^* \rightharpoonup v^*, \quad a_{k_n}^* + L^* b_{k_n}^* \to 0, \quad \text{and} \quad L a_{k_n} - b_{k_n} \to 0, \qquad (3.15)
$$

and from (3.8) that $(\forall n \in \mathbb{N})$ $(a_n, a_n^*) \in \operatorname{gra} A$ and $(b_n, b_n^*) \in \operatorname{gra} B$. We therefore appeal to Proposition 2.4 to conclude that $(x, v^*) \in \boldsymbol{Z}$. ■

Remark 3.4. In the special case when $\mathcal{G} = \mathcal{H}$ and $L = \operatorname{Id}$, Corollary 3.3(iii) was established in [28, Proposition 2] under the following additional assumptions: $A + B$ is maximally monotone or $\mathcal{H}$ is finite-dimensional, $x_n - a_n \to 0$, and $v_n^* - b_n^* \to 0$.

Corollary 3.3 is conceptual in that it does not specify a rule for selecting the quadruple (a_n, b_n, a_n^*, b_n^*) in $\boldsymbol{G}_\alpha(x_n, v_n^*)$ at iteration n. We now provide an example of a concrete selection rule.

Proposition 3.5. *Consider the setting of Problem 1.1. Suppose that $\mathscr{P} \neq \varnothing$, let*

© xxxx Society for Industrial and Applied Mathematics

$\varepsilon \in \left]0,1\right[$, *let* $x_0 \in \mathcal{H}$, *let* $v_0^* \in \mathcal{G}$, *and iterate*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\quad (\gamma_n, \mu_n) \in [\varepsilon, 1/\varepsilon]^2 \\
\quad a_n = J_{\gamma_n A}(x_n - \gamma_n L^* v_n^*) \\
\quad l_n = L x_n \\
\quad b_n = J_{\mu_n B}(l_n + \mu_n v_n^*) \\
\quad s_n^* = \gamma_n^{-1}(x_n - a_n) + \mu_n^{-1} L^*(l_n - b_n) \\
\quad t_n = b_n - L a_n \\
\quad \tau_n = \|s_n^*\|^2 + \|t_n\|^2 \\
\quad \textit{if } \tau_n = 0 \\
\qquad \overline{x} = a_n \\
\qquad \overline{v}^* = v_n^* + \mu_n^{-1}(l_n - b_n) \\
\qquad \textit{terminate.} \\
\quad \textit{if } \tau_n > 0 \\
\qquad \lambda_n \in [\varepsilon, 2 - \varepsilon] \\
\qquad \theta_n = \lambda_n \big(\gamma_n^{-1}\|x_n - a_n\|^2 + \mu_n^{-1}\|l_n - b_n\|^2\big)/\tau_n \\
\qquad x_{n+1} = x_n - \theta_n s_n^* \\
\qquad v_{n+1}^* = v_n^* - \theta_n t_n.
\end{array}
\tag{3.16}
$$

Then either (3.16) *terminates at a solution* $(\overline{x}, \overline{v}^*) \in \boldsymbol{Z}$ *in a finite number of iterations or it generates infinite sequences* $(x_n)_{n\in\mathbb{N}}$ *and* $(v_n^*)_{n\in\mathbb{N}}$ *such that the following hold:*

(i) $\sum_{n\in\mathbb{N}} \|s_n^*\|^2 < +\infty$ *and* $\sum_{n\in\mathbb{N}} \|t_n\|^2 < +\infty$.
(ii) $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$ *and* $\sum_{n\in\mathbb{N}} \|v_{n+1}^* - v_n^*\|^2 < +\infty$.
(iii) $\sum_{n\in\mathbb{N}} \|x_n - a_n\|^2 < +\infty$ *and* $\sum_{n\in\mathbb{N}} \|Lx_n - b_n\|^2 < +\infty$.
(iv) $(x_n)_{n\in\mathbb{N}}$ *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* $(\overline{x}, \overline{v}^*) \in \boldsymbol{Z}$.

Proof. We are going to derive the results from Corollary 3.3. To this end, let us set

$$
(\forall n \in \mathbb{N}) \quad a_n^* = \gamma_n^{-1}(x_n - a_n) - L^* v_n^* \quad \text{and} \quad b_n^* = \mu_n^{-1}(Lx_n - b_n) + v_n^*. \tag{3.17}
$$

Now let $n \in \mathbb{N}$. To show that (3.16) is an instantiation of (3.8), let us check that there exists $\alpha \in \left]0, +\infty\right[$ such that $(a_n, b_n, a_n^*, b_n^*) \in \boldsymbol{G}_\alpha(x_n, v_n^*)$. By construction, we have

$$
(a_n, a_n^*) \in \operatorname{gra} A \quad \text{and} \quad (b_n, b_n^*) \in \operatorname{gra} B. \tag{3.18}
$$

In view of (3.7), we must find $\alpha \in \left]0, +\infty\right[$ such that

$$
\langle x_n - a_n \mid a_n^* + L^* v_n^* \rangle + \langle Lx_n - b_n \mid b_n^* - v_n^* \rangle \geqslant \alpha\big(\|a_n^* + L^* b_n^*\|^2 + \|La_n - b_n\|^2\big). \tag{3.19}
$$

By (3.17),

$$
\langle x_n - a_n \mid a_n^* + L^* v_n^* \rangle + \langle Lx_n - b_n \mid b_n^* - v_n^* \rangle = \gamma_n^{-1}\|x_n - a_n\|^2 + \mu_n^{-1}\|Lx_n - b_n\|^2 \tag{3.20}
$$

and

$$
\begin{aligned}
\|a_n^* + L^* b_n^*\|^2 + \|La_n - b_n\|^2 \\
= \|(\gamma_n^{-1}\operatorname{Id} + \mu_n^{-1} L^* L)x_n - (\gamma_n^{-1} a_n + \mu_n^{-1} L^* b_n)\|^2 + \|La_n - b_n\|^2.
\end{aligned} \tag{3.21}
$$

© xxxx Society for Industrial and Applied Mathematics

On the other hand,

$$\|La_n - b_n\|^2 = \|La_n\|^2 - 2\langle La_n \mid b_n\rangle + \|b_n\|^2 \tag{3.22}$$

and

$$\begin{aligned}
&\|(\gamma_n^{-1}\operatorname{Id} + \mu_n^{-1}L^*L)x_n - (\gamma_n^{-1}a_n + \mu_n^{-1}L^*b_n)\|^2 \\
&\quad = \|\gamma_n^{-1}x_n + \mu_n^{-1}L^*Lx_n\|^2 - 2\langle \gamma_n^{-1}x_n + \mu_n^{-1}L^*Lx_n \mid \gamma_n^{-1}a_n + \mu_n^{-1}L^*b_n\rangle \\
&\qquad + \|\gamma_n^{-1}a_n + \mu_n^{-1}L^*b_n\|^2 \\
&\quad = \gamma_n^{-2}\|x_n\|^2 + 2\gamma_n^{-1}\mu_n^{-1}\|Lx_n\|^2 + \mu_n^{-2}\|L^*Lx_n\|^2 - 2\gamma_n^{-2}\langle x_n \mid a_n\rangle \\
&\qquad - 2\gamma_n^{-1}\mu_n^{-1}\langle Lx_n \mid b_n\rangle - 2\gamma_n^{-1}\mu_n^{-1}\langle Lx_n \mid La_n\rangle - 2\mu_n^{-2}\langle L^*Lx_n \mid L^*b_n\rangle \\
&\qquad + \gamma_n^{-2}\|a_n\|^2 + 2\gamma_n^{-1}\mu_n^{-1}\langle La_n \mid b_n\rangle + \mu_n^{-2}\|L^*b_n\|^2 \\
&\quad = \gamma_n^{-2}\|x_n - a_n\|^2 + \mu_n^{-2}\|L^*(Lx_n - b_n)\|^2 + 2\gamma_n^{-1}\mu_n^{-1}\|Lx_n\|^2 \\
&\qquad - 2\gamma_n^{-1}\mu_n^{-1}\langle Lx_n \mid b_n\rangle - 2\gamma_n^{-1}\mu_n^{-1}\langle Lx_n \mid La_n\rangle + 2\gamma_n^{-1}\mu_n^{-1}\langle La_n \mid b_n\rangle \\
&\quad = \gamma_n^{-2}\|x_n - a_n\|^2 + \mu_n^{-2}\|L^*(Lx_n - b_n)\|^2 + \gamma_n^{-1}\mu_n^{-1}\|L(x_n - a_n)\|^2 \\
&\qquad - \gamma_n^{-1}\mu_n^{-1}\|La_n\|^2 + \gamma_n^{-1}\mu_n^{-1}\|Lx_n - b_n\|^2 - \gamma_n^{-1}\mu_n^{-1}\|b_n\|^2 \\
&\qquad + 2\gamma_n^{-1}\mu_n^{-1}\langle La_n \mid b_n\rangle. \qquad (3.23)
\end{aligned}$$

Combining (3.21), (3.22), and (3.23), and recalling that $\{\gamma_n, \mu_n\} \subset [\varepsilon, \varepsilon^{-1}]$, we obtain

$$\begin{aligned}
&\|a_n^* + L^*b_n^*\|^2 + \|La_n - b_n\|^2 \\
&\quad = \gamma_n^{-2}\|x_n - a_n\|^2 + \mu_n^{-2}\|L^*(Lx_n - b_n)\|^2 \\
&\qquad + \gamma_n^{-1}\mu_n^{-1}\|L(x_n - a_n)\|^2 + \gamma_n^{-1}\mu_n^{-1}\|Lx_n - b_n\|^2 \\
&\qquad + \big(1 - \gamma_n^{-1}\mu_n^{-1}\big)\big(\|La_n\|^2 - 2\langle La_n \mid b_n\rangle + \|b_n\|^2\big) \\
&\quad = \gamma_n^{-2}\|x_n - a_n\|^2 + \mu_n^{-2}\|L^*(Lx_n - b_n)\|^2 + \gamma_n^{-1}\mu_n^{-1}\|L(x_n - a_n)\|^2 \\
&\qquad + \gamma_n^{-1}\mu_n^{-1}\|Lx_n - b_n\|^2 + \big(1 - \gamma_n^{-1}\mu_n^{-1}\big)\|La_n - b_n\|^2 \\
&\quad \leqslant \varepsilon^{-1}\Big(\gamma_n^{-1}\|x_n - a_n\|^2 + \mu_n^{-1}\|L^*(Lx_n - b_n)\|^2 + \gamma_n^{-1}\|L(x_n - a_n)\|^2 \\
&\qquad + \mu_n^{-1}\|Lx_n - b_n\|^2\Big) + 2\big(1 - \gamma_n^{-1}\mu_n^{-1}\big)\big(\|L(a_n - x_n)\|^2 + \|Lx_n - b_n\|^2\big) \\
&\quad \leqslant \varepsilon^{-1}\big(1 + \|L\|^2\big)\Big(\gamma_n^{-1}\|x_n - a_n\|^2 + \mu_n^{-1}\|Lx_n - b_n\|^2\Big) \\
&\qquad + 2\big(\gamma_n - \mu_n^{-1}\big)\gamma_n^{-1}\|L\|^2\,\|x_n - a_n\|^2 + 2\big(\mu_n - \gamma_n^{-1}\big)\mu_n^{-1}\|Lx_n - b_n\|^2 \\
&\quad \leqslant \varepsilon^{-1}\big(1 + \|L\|^2 + 2(1 - \varepsilon^2)\max\{1, \|L\|^2\}\big) \\
&\qquad \times \big(\gamma_n^{-1}\|x_n - a_n\|^2 + \mu_n^{-1}\|Lx_n - b_n\|^2\big). \qquad (3.24)
\end{aligned}$$

Therefore, (3.20) implies that (3.19) is satisfied with

$$\alpha = \frac{\varepsilon}{1 + \|L\|^2 + 2(1 - \varepsilon^2)\max\{1, \|L\|^2\}}. \tag{3.25}$$

We thus obtain (i) and (ii). To prove (iii), note that it follows from (3.17) that

$$\begin{aligned}
x_n - a_n &= (\gamma_n^{-1}\operatorname{Id} + \mu_n^{-1}L^*L)^{-1}\big(\gamma_n^{-1}(x_n - a_n) + \mu_n^{-1}L^*L(x_n - a_n)\big) \\
&= (\gamma_n^{-1}\operatorname{Id} + \mu_n^{-1}L^*L)^{-1}\big(\gamma_n^{-1}(x_n - a_n) + \mu_n^{-1}L^*(Lx_n - b_n) \\
&\qquad + \mu_n^{-1}L^*(b_n - La_n)\big) \\
&= \gamma_n(\operatorname{Id} + (\gamma_n/\mu_n)L^*L)^{-1}\big((a_n^* + L^*b_n^*) + \mu_n^{-1}L^*(b_n - La_n)\big). \qquad (3.26)
\end{aligned}$$

© xxxx Society for Industrial and Applied Mathematics

Thus, since $\|(\mathrm{Id}+(\gamma_n/\mu_n)L^*L)^{-1}\| \leqslant 1$ and since $\max\{\gamma_n, \mu_n^{-1}\} \leqslant \varepsilon^{-1}$, we have

$$\begin{aligned}\|x_n - a_n\|^2 &\leqslant \gamma_n^2 \big\|(\mathrm{Id}+(\gamma_n/\mu_n)L^*L)^{-1}\big\|^2 \big(\|a_n^* + L^*b_n^*\| + \mu_n^{-1}\|L^*(La_n - b_n)\|\big)^2 \\ &\leqslant 2\gamma_n^2\big(\|a_n^* + L^*b_n^*\|^2 + \mu_n^{-2}\|L^*(La_n - b_n)\|^2\big) \\ &\leqslant 2\varepsilon^{-2}\big(\|s_n^*\|^2 + \varepsilon^{-2}\|L\|^2\,\|t_n\|^2\big). \end{aligned} \tag{3.27}$$

Hence, (i) yields $\sum_{n\in\mathbb{N}} \|x_n - a_n\|^2 < +\infty$. In turn, since

$$\|Lx_n - b_n\|^2 = \|L(x_n - a_n) + La_n - b_n\|^2 \leqslant 2\big(\|L\|^2\,\|x_n - a_n\|^2 + \|t_n\|^2\big), \tag{3.28}$$

we obtain $\sum_{n\in\mathbb{N}} \|Lx_n - b_n\|^2 < +\infty$. Therefore

$$\sum_{n\in\mathbb{N}} \|v_n^* - b_n^*\|^2 = \sum_{n\in\mathbb{N}} \mu_n^{-2}\|Lx_n - b_n\|^2 \leqslant \varepsilon^{-2}\sum_{n\in\mathbb{N}} \|Lx_n - b_n\|^2 < +\infty. \tag{3.29}$$

Thus, (3.9) is satisfied and (iv) follows. ■

Remark 3.6. As mentioned in the Introduction, existing methods for solving Problem 1.1 either require knowledge of $\|L\|$ or necessitate potentially hard to implement inversions of linear operators. For instance, the method of [16], which hinges on a reformulation that employs Tseng's forward-backward-forward algorithm [47], imposes the same scaling coefficients on A and B at each iteration and they must be bounded by a specific constant which depends on $\|L\|$; more precisely, $(\forall n \in \mathbb{N})$ $\gamma_n = \mu_n \in \,]0, 1/\|L\|[$. These restrictions are lifted in (3.16), where the parameters $(\gamma_n)_{n\in\mathbb{N}}$ and $(\mu_n)_{n\in\mathbb{N}}$ can evolve freely in an arbitrarily large interval of $]0, +\infty[$ independent from L.

We now highlight two particular instances of interest.

Example 3.7. Consider the setting of Problem 1.1 with $A = 0$. Then the primal problem (1.5) reduces to

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } 0 \in L^*BL\overline{x}. \tag{3.30}$$

Assume that it has at least one solution, let $\lambda \in \,]0, 2[$, let $x_0 \in \mathcal{H}$, let $v_0^* \in \mathcal{G}$, and iterate

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\quad a_n = x_n - L^*v_n^* \\
\quad l_n = Lx_n \\
\quad b_n = J_B(l_n + v_n^*) \\
\quad s_n^* = x_n - a_n + L^*(l_n - b_n) \\
\quad t_n = b_n - La_n \\
\quad \tau_n = \|s_n^*\|^2 + \|t_n\|^2 \\
\quad \text{if } \tau_n = 0 \\
\qquad \overline{x} = a_n \\
\qquad \overline{v}^* = v_n^* + l_n - b_n \\
\qquad \text{terminate.} \\
\quad \text{if } \tau_n > 0 \\
\qquad \theta_n = \lambda\big(\|x_n - a_n\|^2 + \|l_n - b_n\|^2\big)/\tau_n \\
\qquad x_{n+1} = x_n - \theta_n s_n^* \\
\qquad v_{n+1}^* = v_n^* - \theta_n t_n.
\end{array} \tag{3.31}$$

© xxxx Society for Industrial and Applied Mathematics

This algorithm is the instance of (3.16) in which $A = 0$ and $(\forall n \in \mathbb{N})\ \gamma_n = \mu_n = 1$ and $\lambda_n = \lambda$. It follows from Proposition 3.5 that, if it does not terminate, it produces a sequence $(x_n)_{n\in\mathbb{N}}$ that converges weakly to a solution to (3.30). In the special case when $\mathcal{H} = \mathbb{R}^N$ and $\mathcal{G} = \mathbb{R}^M$ this result was established in [27] (the fact that [27, Algorithm 3.1] is equivalent to (3.31) follows from elementary manipulations). Interestingly, the analysis of [27] is quite different from ours and it does not employ a geometric construction.

Example 3.8. Consider the setting of Problem 1.1 with $\mathcal{G} = \mathcal{H}$ and $L = \mathrm{Id}$. Then the primal problem (1.5) reduces to

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } 0 \in A\overline{x} + B\overline{x}. \tag{3.32}$$

Assume that it has at least one solution, let $\varepsilon \in \,]0,1[$, let $x_0 \in \mathcal{H}$, let $v_0^* \in \mathcal{H}$, and iterate

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\quad (\gamma_n, \mu_n) \in [\varepsilon, 1/\varepsilon]^2 \\
\quad a_n = J_{\gamma_n A}(x_n - \gamma_n v_n^*) \\
\quad b_n = J_{\mu_n B}(x_n + \mu_n v_n^*) \\
\quad s_n^* = \gamma_n^{-1}(x_n - a_n) + \mu_n^{-1}(x_n - b_n) \\
\quad t_n = b_n - a_n \\
\quad \tau_n = \|s_n^*\|^2 + \|t_n\|^2 \\
\quad \text{if } \tau_n = 0 \\
\qquad \overline{x} = a_n \\
\qquad \overline{v}^* = v_n^* + \mu_n^{-1}(x_n - b_n) \\
\qquad \text{terminate.} \\
\quad \text{if } \tau_n > 0 \\
\qquad \lambda_n \in [\varepsilon, 2 - \varepsilon] \\
\qquad \theta_n = \lambda_n\big(\gamma_n^{-1}\|x_n - a_n\|^2 + \mu_n^{-1}\|x_n - b_n\|^2\big)/\tau_n \\
\qquad x_{n+1} = x_n - \theta_n s_n^* \\
\qquad v_{n+1}^* = v_n^* - \theta_n t_n.
\end{array} \tag{3.33}$$

Then Proposition 3.5 asserts that, if the algorithm does not terminate, it produces a sequence $(x_n, v_n^*)_{n\in\mathbb{N}}$ that converges weakly to a point $(\overline{x}, \overline{v}^*)$ such that $-\overline{v}^* \in A\overline{x}$ and $\overline{v}^* \in B\overline{x}$, so that $\overline{x}$ solves (3.32). Under the additional assumptions that $A + B$ is maximally monotone or that $\mathcal{H}$ is finite-dimensional, this result was established in [28, Proposition 3] for a version of (3.33) in which an additional relaxation parameter is allowed in the definition of a_n.

4. A Fejér monotone algorithm for coupled monotone inclusions.

Many complex systems feature interactions between several variables can be modeled in terms of equilibria involving composite monotone operators. A mathematical formulation of such problems in duality is the following.

Problem 4.1. Let m and K be strictly positive integers, let $(\mathcal{H}_i)_{1\leqslant i\leqslant m}$ and $(\mathcal{G}_k)_{1\leqslant k\leqslant K}$ be real Hilbert spaces, and set $\boldsymbol{\mathcal{K}} = \mathcal{H}_1 \oplus \cdots \mathcal{H}_m \oplus \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_K$. For every $i \in \{1, \ldots, m\}$ and every $k \in \{1, \ldots, K\}$, 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 $L_{ki} \colon \mathcal{H}_i \to \mathcal{G}_k$ be linear

© xxxx Society for Industrial and Applied Mathematics

and bounded. Consider the coupled inclusions problem

$$\text{(4.1)}\quad \text{find } \overline{x}_1 \in \mathcal{H}_1, \ldots, \overline{x}_m \in \mathcal{H}_m \text{ such that}$$
$$(\forall i \in \{1,\ldots,m\}) \quad z_i \in A_i\overline{x}_i + \sum_{k=1}^{K} L_{ki}^*\Big(B_k\Big(\sum_{j=1}^{m} L_{kj}\overline{x}_j - r_k\Big)\Big),$$

the dual problem

$$\text{(4.2)}\quad \text{find } \overline{v}_1^* \in \mathcal{G}_1, \ldots, \overline{v}_K^* \in \mathcal{G}_K \text{ such that}$$
$$(\forall k \in \{1,\ldots,K\}) \quad -r_k \in -\sum_{i=1}^{m} L_{ki}\Big(A_i^{-1}\Big(z_i - \sum_{l=1}^{K} L_{li}^*\overline{v}_l^*\Big)\Big) + B_k^{-1}\overline{v}_k^*,$$

and the associated Kuhn–Tucker set

(4.3)
$$\boldsymbol{Z} = \Bigg\{(x_1,\ldots,x_m,v_1^*,\ldots,v_K^*) \in \boldsymbol{\mathcal{K}} \;\Bigg|\; (\forall i \in \{1,\ldots,m\})\;\; z_i - \sum_{k=1}^{K} L_{ki}^* v_k^* \in A_i x_i$$
$$\text{and } (\forall k \in \{1,\ldots,K\})\;\; \sum_{i=1}^{m} L_{ki}x_i - r_k \in B_k^{-1}v_k^*\Bigg\}.$$

The problem is to find a point in $\boldsymbol{Z}$. The sets of solutions to (4.1) and (4.2) are denoted by $\mathscr{P}$ and $\mathscr{D}$, respectively.

Such formulations, at least in their primal form (4.2), have been investigated at various levels of generality in [2, 3, 5, 10, 13, 15, 17, 20, 18, 23, 33, 34] to model problems arising in areas such as game theory, evolution equations, machine learning, signal and image processing, mechanics, the cognitive sciences, and domain decomposition methods in partial differential equations. As shown in [11] and [23, Section 3], another important motivation for studying such systems is the fact that single-variable inclusion problems involving various types of parallel sums of monotone operators, can be recast in the multivariate format (4.1) via the introduction of auxiliary variables.

In this section, we shall use the following result, which establishes a bridge between Problem 1.1 and Problem 4.1, to devise a splitting method for the latter based on Proposition 3.5.

Proposition 4.2. *Consider the setting of Problem 4.1 and set*

$$\text{(4.4)}\quad \begin{cases} \mathcal{H} = \bigoplus_{i=1}^{m} \mathcal{H}_i \\ \mathcal{G} = \bigoplus_{k=1}^{K} \mathcal{G}_k \\ A\colon \mathcal{H} \to 2^{\mathcal{H}}\colon (x_i)_{1\leqslant i\leqslant m} \mapsto \bigtimes_{i=1}^{m}(-z_i + A_i x_i) \\ B\colon \mathcal{G} \to 2^{\mathcal{G}}\colon (y_k)_{1\leqslant k\leqslant K} \mapsto \bigtimes_{k=1}^{K} B_k(y_k - r_k) \\ L\colon \mathcal{H} \to \mathcal{G}\colon (x_i)_{1\leqslant i\leqslant m} \mapsto \big(\sum_{i=1}^{m} L_{ki}x_i\big)_{1\leqslant k\leqslant K} \end{cases}$$

in Problem 1.1. Then the following hold:

(i) *Problem 1.1 coincides with Problem 4.1.*

(ii) *Let $\gamma \in \,]0,+\infty[$, let $(x_i)_{1\leqslant i\leqslant m} \in \mathcal{H}$, and let $(y_k)_{1\leqslant k\leqslant K} \in \mathcal{G}$. Then*

$$\text{(4.5)}\quad J_{\gamma A}(x_i)_{1\leqslant i\leqslant m} = \big(J_{\gamma A_i}(x_i + \gamma z_i)\big)_{1\leqslant i\leqslant m}$$
$$\text{and}\quad J_{\gamma B}(y_k)_{1\leqslant k\leqslant K} = \big(r_k + J_{\gamma B_k}(y_k - r_k)\big)_{1\leqslant k\leqslant K}.$$

© xxxx Society for Industrial and Applied Mathematics

Proof. (i): This follows from (4.4) and the fact that $L^*\colon \mathcal{G} \to \mathcal{H}\colon (y_k)_{1\leqslant k\leqslant K} \mapsto (\sum_{k=1}^K L_{ki}^* y_k)_{1\leqslant i\leqslant m}$.

(ii): [9, Propositions 23.15 and 23.16]. ■

To find a Kuhn–Tucker point in Problem 4.1 we can invoke Proposition 4.2 and apply the algorithms devised in Section 3 in the setting of (4.4). Thus, Proposition 3.5 leads to the following result.

Theorem 4.3. *Consider the setting of Problem 4.1. Suppose that $\mathscr{P} \neq \varnothing$, let $\varepsilon \in \left]0,1\right[$, let $x_{1,0} \in \mathcal{H}_1, \ldots, x_{m,0} \in \mathcal{H}_m$, $v_{1,0}^* \in \mathcal{G}_1, \ldots, v_{K,0}^* \in \mathcal{G}_K$, and iterate*

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\quad (\gamma_n, \mu_n) \in [\varepsilon, 1/\varepsilon]^2 \\
\quad \textit{for } i = 1, \ldots, m \\
\quad\quad a_{i,n} = J_{\gamma_n A_i}\Big(x_{i,n} + \gamma_n\big(z_i - \sum_{k=1}^K L_{ki}^* v_{k,n}^*\big)\Big) \\
\quad \textit{for } k = 1, \ldots, K \\
\quad\quad l_{k,n} = \sum_{i=1}^m L_{ki} x_{i,n} \\
\quad\quad b_{k,n} = r_k + J_{\mu_n B_k}\big(l_{k,n} + \mu_n v_{k,n}^* - r_k\big) \\
\quad\quad t_{k,n} = b_{k,n} - \sum_{i=1}^m L_{ki} a_{i,n} \\
\quad \textit{for } i = 1, \ldots, m \\
\quad\quad s_{i,n}^* = \gamma_n^{-1}(x_{i,n} - a_{i,n}) + \mu_n^{-1} \sum_{k=1}^K L_{ki}^*(l_{k,n} - b_{k,n}) \\
\quad \tau_n = \sum_{i=1}^m \|s_{i,n}^*\|^2 + \sum_{k=1}^K \|t_{k,n}\|^2 \\
\quad \textit{if } \tau_n = 0 \\
\quad\quad \textit{for } i = 1, \ldots, m \\
\quad\quad\quad \overline{x}_i = a_{i,n} \\
\quad\quad \textit{for } k = 1, \ldots, K \\
\quad\quad\quad \overline{v}_k^* = v_{k,n}^* + \mu_n^{-1}(l_{k,n} - b_{k,n}) \\
\quad\quad \textit{terminate.} \\
\quad \textit{if } \tau_n > 0 \\
\quad\quad \lambda_n \in [\varepsilon, 2 - \varepsilon] \\
\quad\quad \theta_n = \lambda_n\big(\gamma_n^{-1} \sum_{i=1}^m \|x_{i,n} - a_{i,n}\|^2 + \mu_n^{-1} \sum_{k=1}^K \|l_{k,n} - b_{k,n}\|^2\big)/\tau_n \\
\quad\quad \textit{for } i = 1, \ldots, m \\
\quad\quad\quad x_{i,n+1} = x_{i,n} - \theta_n s_{i,n}^* \\
\quad\quad \textit{for } k = 1, \ldots, K \\
\quad\quad\quad v_{k,n+1}^* = v_{k,n}^* - \theta_n t_{k,n}.
\end{array}
\tag{4.6}
$$

Then either (4.6) terminates at a solution $(\overline{x}_1, \ldots, \overline{x}_m, \overline{v}_1^*, \ldots, \overline{v}_K^*) \in \boldsymbol{Z}$ in a finite number of iterations or it generates infinite sequences $(x_{1,n})_{n\in\mathbb{N}}, \ldots, (x_{m,n})_{n\in\mathbb{N}}$, $(v_{1,n}^*)_{n\in\mathbb{N}}, \ldots, (v_{K,n}^*)_{n\in\mathbb{N}}$ such that the following hold:

(i) $(\forall i \in \{1, \ldots, m\})$ $\sum_{n\in\mathbb{N}} \|s_{i,n}^*\|^2 < +\infty$, $\sum_{n\in\mathbb{N}} \|x_{i,n+1} - x_{i,n}\|^2 < +\infty$, *and* $\sum_{n\in\mathbb{N}} \|x_{i,n} - a_{i,n}\|^2 < +\infty$.

(ii) $(\forall k \in \{1, \ldots, K\})$ $\sum_{n\in\mathbb{N}} \|t_{k,n}\|^2 < +\infty$, $\sum_{n\in\mathbb{N}} \|v_{k,n+1}^* - v_{k,n}^*\|^2 < +\infty$, *and* $\sum_{n\in\mathbb{N}} \|\sum_{i=1}^m L_{ki} x_{i,n} - b_{k,n}\|^2 < +\infty$.

(iii) *For every $i \in \{1, \ldots, m\}$ $(x_{i,n})_{n\in\mathbb{N}}$ converges weakly to a point $\overline{x}_i$, for every $k \in \{1, \ldots, K\}$ $(v_{k,n}^*)_{n\in\mathbb{N}}$ converges weakly to a point $\overline{v}_k^*$, $(\overline{x}_1, \ldots, \overline{x}_m) \in \mathscr{P}$, $(\overline{v}_1^*, \ldots, \overline{v}_K^*) \in \mathscr{D}$, and $(\overline{x}_1, \ldots, \overline{x}_m, \overline{v}_1^*, \ldots, \overline{v}_K^*) \in \boldsymbol{Z}$.*

Proof. Define $\mathcal{H}$, $\mathcal{G}$, A, B, and L as in (4.4). Then, as seen in Proposition 4.2(i),

© xxxx Society for Industrial and Applied Mathematics

Problem 1.1 coincides with Problem 4.1. Now set

$$(\forall n \in \mathbb{N}) \quad \begin{cases} a_n = (a_{i,n})_{1 \leqslant i \leqslant m} \\ s_n^* = (s_{i,n}^*)_{1 \leqslant i \leqslant m} \\ x_n = (x_{i,n})_{1 \leqslant i \leqslant m} \\ b_n = (b_{k,n})_{1 \leqslant k \leqslant K} \\ l_n = (l_{k,n})_{1 \leqslant k \leqslant K} \\ t_n = (t_{k,n})_{1 \leqslant k \leqslant K} \\ v_n^* = (v_{k,n}^*)_{1 \leqslant k \leqslant K}. \end{cases} \tag{4.7}$$

Then we derive from Proposition 4.2(ii) that (4.6) coincides with (3.16). The assertions therefore follow from Proposition 3.5. ■

Remark 4.4. In the special case when $m = 1$, $A_1 = 0$, $z_1 = 0$, and, for every $k \in \{1, \ldots, K\}$, $\mathcal{G}_k = \mathcal{H}$, $L_{k1} = \mathrm{Id}$, and $r_k = 0$, the primal problem (4.1) becomes

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } 0 \in \sum_{k=1}^{K} B_k \overline{x}, \tag{4.8}$$

and the associated Kuhn–Tucker set of (4.3) becomes

$$\boldsymbol{Z} = \left\{ (x, v_1^*, \ldots, v_K^*) \in \mathcal{H}^{K+1} \;\middle|\; \sum_{k=1}^{K} v_k^* = 0 \text{ and } (\forall k \in \{1, \ldots, K\}) \;\; v_k^* \in B_k x \right\}. \tag{4.9}$$

In this setting, (4.6) reduces to an algorithm which is similar to that of [29, Section 4]. The convergence of the latter was established under the additional assumption that $\sum_{k=1}^{K} B_k$ is maximally monotone or that $\mathcal{H}$ is finite-dimensional [29, Proposition 4.2], but these assumptions were subsequently shown not to be necessary [6]. Let us note that in this special case, (4.6) is different from the algorithm of [29] as it has a parallel structure (all the operators $(B_k)_{1 \leqslant k \leqslant K}$ are used simultaneously), whereas that of [29] allows for more flexibility (e.g., sequential activation) and it assigns to each monotone operator its own scaling parameter. It is natural to ask whether, in our general setting, (4.6) could be extended to include such features by using Corollary 3.3 directly instead of Proposition 3.5. We have not been successful in bringing an affirmative answer to this question.

Remark 4.5. Using Proposition 4.2, any algorithm for solving Problem 1.1 can in principle be used to solve Problem 4.1. However, methods which require the computation of the norm of the operator L of (4.4) face the difficulty of expressing it tightly in terms of those of the individual coupling operators $(L_{ki})_{\substack{1 \leqslant k \leqslant K \\ 1 \leqslant i \leqslant m}}$; see, for instance, [11, 23] for examples of such approximations. This task is further complicated by the fact that in some situations the norms of the individual coupling operators may not even be computable precisely. For instance, in domain decomposition methods, L_{ki} is the trace operator relative to the interface between two subdomains and, depending on the underlying assumptions, its norm may not be easy to estimate. Likewise, inverting linear operators based on various combinations of the individual coupling operators is typically unfeasible in such applications, which renders inoperative those methods of [1, 14, 16] using such computations. These shortcomings of existing methods are circumvented by (4.6), which makes it particularly attractive for solving Problem 4.1.

© xxxx Society for Industrial and Applied Mathematics

Remark 4.6. An alternative method to solve Problem 4.1 is that proposed in [23]. In terms of complexity per iteration and parallelizability, both algorithms are quite comparable. However, as noted in Remark 4.5, the method of [23] implicitly requires a tight bound on the norm of the global operator L of (4.4), which can be a serious drawback. Another difference between the method of [23] and (4.6), is that the latter features two sequences of scaling parameters $(\gamma_n)_{n\in\mathbb{N}}$ and $(\mu_n)_{n\in\mathbb{N}}$ which, furthermore, can be arbitrarily large or small. The proposed algorithm (4.6) also incorporates relaxation parameters $(\lambda_n)_{n\in\mathbb{N}}$ that can induce large step sizes through overrelaxations up to almost 2, whereas the method of [23] is unrelaxed.

An important area of application of Problem 4.1 is multivariate convex minimization problems. We denote by $\Gamma_0(\mathcal{H})$ the class of lower semicontinuous convex proper functions from $\mathcal{H}$ to $]-\infty,+\infty]$. Let $f \in \Gamma_0(\mathcal{H})$. The conjugate of f is $\Gamma_0(\mathcal{H}) \ni f^*\colon u \mapsto \sup_{x\in\mathcal{H}}(\langle x \mid u\rangle - f(x))$. For every $x \in \mathcal{H}$, $f + \|x - \cdot\|^2/2$ has a unique minimizer, which is denoted by $\operatorname{prox}_f x$ [40]. We have

$$\operatorname{prox}_f = J_{\partial f}, \tag{4.10}$$

where

$$\partial f\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \big\{u^* \in \mathcal{H} \;\big|\; (\forall y \in \mathcal{H})\;\; \langle y - x \mid u^*\rangle + f(x) \leqslant f(y)\big\} \tag{4.11}$$

is the Moreau subdifferential of f. The following formulation captures a variety of multivariate minimization problems, e.g., [3, 4, 5, 15, 18, 19, 20, 33, 34].

Problem 4.7. Let m and K be strictly positive integers, let $(\mathcal{H}_i)_{1\leqslant i\leqslant m}$ and $(\mathcal{G}_k)_{1\leqslant k\leqslant K}$ be real Hilbert spaces, and set $\boldsymbol{\mathcal{K}} = \mathcal{H}_1 \oplus \cdots \mathcal{H}_m \oplus \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_K$. For every $i \in \{1,\ldots,m\}$ and every $k \in \{1,\ldots,K\}$, let $f_i \in \Gamma_0(\mathcal{H}_i)$, let $g_k \in \Gamma_0(\mathcal{G}_k)$, 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. Let $\mathscr{P}$ be the set of solutions to the primal problem

$$\underset{x_1\in\mathcal{H}_1,\ldots,\,x_m\in\mathcal{H}_m}{\text{minimize}} \;\; \sum_{i=1}^{m}\big(f_i(x_i) - \langle x_i \mid z_i\rangle\big) + \sum_{k=1}^{K} g_k\bigg(\sum_{i=1}^{m} L_{ki}x_i - r_k\bigg), \tag{4.12}$$

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

$$\underset{v_1^*\in\mathcal{G}_1,\ldots,\,v_K^*\in\mathcal{G}_K}{\text{minimize}} \;\; \sum_{i=1}^{m} f_i^*\bigg(z_i - \sum_{k=1}^{K} L_{ki}^* v_k^*\bigg) + \sum_{k=1}^{K}\big(g_k^*(v_k^*) + \langle v_k^* \mid r_k\rangle\big). \tag{4.13}$$

The problem is to find a point in the associated Kuhn–Tucker set

$$\begin{aligned}\boldsymbol{Z} = \bigg\{(x_1,\ldots,x_m,v_1^*,\ldots,v_K^*) \in \boldsymbol{\mathcal{K}} \;\bigg|\; (\forall i \in \{1,\ldots,m\})\;\; z_i - \sum_{k=1}^{K} L_{ki}^* v_k^* \in \partial f_i x_i \\ \text{and}\;\; (\forall k \in \{1,\ldots,K\})\;\; \sum_{i=1}^{m} L_{ki}x_i - r_k \in \partial g_k^* v_k^*\bigg\}.\end{aligned} \tag{4.14}$$

Corollary 4.8. *Consider the setting of Problem 4.7. Suppose that*

$$(\forall i \in \{1,\ldots,m\}) \quad z_i \in \operatorname{ran}\bigg(\partial f_i + \sum_{k=1}^{K} L_{ki}^* \circ \partial g_k \circ \bigg(\sum_{j=1}^{m} L_{kj}\cdot - r_k\bigg)\bigg), \tag{4.15}$$

© xxxx Society for Industrial and Applied Mathematics

let $\varepsilon \in \,]0,1[$, *let* $x_{1,0} \in \mathcal{H}_1, \ldots, x_{m,0} \in \mathcal{H}_m$, *let* $v^*_{1,0} \in \mathcal{G}_1, \ldots, v^*_{K,0} \in \mathcal{G}_K$, *and iterate* (4.6), *where the only modification is that we now set*

$$a_{i,n} = \operatorname{prox}_{\gamma_n f_i}\left(x_{i,n} + \gamma_n\left(z_i - \sum_{k=1}^{K} L^*_{ki} v^*_{k,n}\right)\right) \tag{4.16}$$

and

$$b_{k,n} = r_k + \operatorname{prox}_{\mu_n g_k}\big(l_{k,n} + \mu_n v^*_{k,n} - r_k\big). \tag{4.17}$$

Then the conclusions of Theorem 4.3 hold true.

Proof. Set $(\forall i \in \{1,\ldots,m\})\ A_i = \partial f_i$ and $(\forall k \in \{1,\ldots,K\})\ B_k = \partial g_k$. Then, using the same arguments as in [23, Proposition 5.4], we obtain that (4.12) and (4.13) are instances of (4.1) and (4.2), respectively. ■

Sufficient conditions for this constraint qualification (4.15) to hold can be found in [23, Proposition 5.3]. In particular, if $(\mathcal{H}_i)_{1\leqslant i\leqslant m}$ and $(\mathcal{G}_k)_{1\leqslant k\leqslant K}$ are finite-dimensional and if $\mathscr{P} \neq \varnothing$, then (4.15) is satisfied if $(\forall i \in \{1,\ldots,m\})(\exists\, x_i \in \operatorname{ri}\operatorname{dom} f_i)(\forall k \in \{1,\ldots,K\})\ \sum_{i=1}^m L_{ki}x_i - r_k \in \operatorname{ri}\operatorname{dom} g_k$.

REFERENCES

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

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

[3] H. Attouch, A. Cabot, P. Frankel, and J. Peypouquet, *Alternating proximal algorithms for linearly constrained variational inequalities: Application to domain decomposition for PDE's,* Nonlinear Anal., 74 (2011), pp. 7455–7473.

[4] J.-F. Aujol and S. H. Kang, *Color image decomposition and restoration*, J. Vis. Comm. Image Represent., 17 (2006), pp. 916–928.

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

[6] H. H. Bauschke, *A note on the paper by Eckstein and Svaiter on "General projective splitting methods for sums of maximal monotone operators,"* SIAM J. Control Optim., 48 (2009), pp. 2513–2515.

[7] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev., 38 (1996), pp. 367-426.

[8] H. H. Bauschke and P. L. Combettes, *A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces,* Math. Oper. Res., 26 (2001), pp. 248–264.

[9] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces,* Springer, New York, 2011.

[10] H. H. Bauschke, P. L. Combettes, and S. Reich, *The asymptotic behavior of the composition of two resolvents,* Nonlinear Anal., 60 (2005), pp. 283–301.

[11] S. R. Becker and P. L. Combettes, *An algorithm for splitting parallel sums of linearly composed monotone operators, with applications to signal recovery,* J. Nonlinear Convex Anal., 15 (2014), pp. 137–159.

[12] 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., 23 (2013), pp. 2011–2036.

[13] R. I. Boţ, E. R. Csetnek, and E. Nagy, *Solving systems of monotone inclusions via primal-dual splitting techniques,* Taiwanese J. Math., 17 (2013), pp. 1983–2009.

[14] 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., 23 (2013), pp. 2541–2565.

[15] 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., 2 (2009), pp. 485–508.

© xxxx Society for Industrial and Applied Mathematics

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

[17] 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, H. H. Bauschke, P. Borwein, F. Garvan, M. Théra, J. Vanderwerff, and H. Wolkowicz, eds.), Springer, New York, 2013, pp. 143–159.

[18] L. M. Briceño-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik, *Proximal algorithms for multicomponent image recovery problems,* J. Math. Imaging Vision, 41 (2011), pp. 3–22.

[19] E. J. Candès, X. Li, Y. Ma, and J. Wright, *Robust principal component analysis?,* J. ACM, 58 (2011), art. 11, 37 pp.

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

[21] P. L. Combettes, *The convex feasibility problem in image recovery,* in: Advances in Imaging and Electron Physics (P. Hawkes, Ed.), vol. 95, Academic Press, New York, 1996, pp. 155–270.

[22] 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.)

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

[24] 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., Springer, New York, 2011, pp. 185–212.

[25] 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., 20 (2012), pp. 307–330.

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

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

[28] J. Eckstein and B. F. Svaiter, *A family of projective splitting methods for the sum of two maximal monotone operators,* Math. Program., 111 (2008), pp. 173–199.

[29] J. Eckstein and B. F. Svaiter, *General projective splitting methods for sums of maximal monotone operators,* SIAM J. Control Optim., 48 (2009), pp. 787–811.

[30] I. I. Eremin, *Methods of Fejér approximations in convex programming,* Math. Notes, 3 (1968), pp. 139–149.

[31] I. I. Eremin and L. D. Popov, *Fejér processes in theory and practice: Recent results,* Russian Math. (Iz. VUZ), 53 (2009), pp. 36–55.

[32] M. Fortin and R. Glowinski, Eds., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems.* North-Holland, Amsterdam, 1983.

[33] P. Frankel, *Alternating proximal algorithm with costs-to-move, dual description and application to PDE's,* Discrete Contin. Dyn. Syst. Ser. S, 5 (2012), pp. 545–557.

[34] P. Frankel and J. Peypouquet, *Lagrangian-penalization algorithm for constrained optimization and variational inequalities,* Set-Valued Var. Anal., 20 (2012), pp. 169–185.

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

[36] D. Goldfarb and S. Ma, *Fast multiple-splitting algorithms for convex optimization,* SIAM J. Optim., 22 (2012), pp. 533–556.

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

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

[39] B. Mercier, *Inéquations Variationnelles de la Mécanique* (Publications Mathématiques d'Orsay, no. 80.01), Université de Paris-XI, Orsay, France, 1980.

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

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

© xxxx Society for Industrial and Applied Mathematics

[42] T. PENNANEN, *Dualization of generalized equations of maximal monotone type,* SIAM J. Optim., 10 (2000), pp. 809–835.

[43] H. RAGUET, J. FADILI, AND G. PEYRÉ, *A generalized forward-backward splitting,* SIAM J. Imaging Sci., 6 (2013), pp. 1199–1226.

[44] R. T. ROCKAFELLAR, *Conjugate Duality and Optimization,* SIAM, Philadelphia, PA, 1974.

[45] J. E. SPINGARN, *Partial inverse of a monotone operator,* Appl. Math. Optim., 10 (1983), pp. 247–265.

[46] S. SRA, S. NOWOZIN, AND S. J. WRIGHT, *Optimization for Machine Learning,* MIT Press, Cambridge, MA, 2012.

[47] P. TSENG, *A modified forward-backward splitting method for maximal monotone mappings,* SIAM J. Control Optim., 38 (2000), pp. 431–446.

[48] B. C. VŨ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators,* Adv. Comput. Math., 38 (2013), pp. 667–681.