A PARALLEL SPLITTING METHOD FOR COUPLED MONOTONE INCLUSIONS*

Hédy Attouch,[1] Luis M. Briceño-Arias,[2] and Patrick L. Combettes[3]

[1]Université Montpellier II
Institut de Mathématiques et de Modélisation de Montpellier – UMR CNRS 5149
34095 Montpellier Cedex 5, France (attouch@math.univ-montp2.fr)

[2]UPMC Université Paris 06
Laboratoire Jacques-Louis Lions – UMR CNRS 7598
and Équipe Combinatoire et Optimisation – UMR CNRS 7090
75005 Paris, France (lbriceno@math.jussieu.fr)

[3]UPMC Université Paris 06
Laboratoire Jacques-Louis Lions – UMR CNRS 7598, 75005 Paris, France (plc@math.jussieu.fr)

Abstract

A parallel splitting method is proposed for solving systems of coupled monotone inclusions in Hilbert spaces, and its convergence is established under the assumption that solutions exist. Unlike existing alternating algorithms, which are limited to two variables and linear coupling, our parallel method can handle an arbitrary number of variables as well as nonlinear coupling schemes. The breadth and flexibility of the proposed framework is illustrated through applications in the areas of evolution inclusions, variational problems, best approximation, and network flows.

Keywords: coupled systems, demiregular operator, evolution inclusion, forward-backward algorithm, maximal monotone operator, operator splitting, parallel algorithm, weak convergence.

*Contact author: P. L. Combettes, `plc@math.jussieu.fr`, phone: +33 1 4427 6319, fax: +33 1 4427 7200. This work was supported by the Agence Nationale de la Recherche under grant ANR-08-BLAN-0294-02.

1 Problem statement

This paper is concerned with the numerical solution of systems of coupled monotone inclusions in Hilbert spaces. A simple instance of this problem is to

$$\text{find } x_1 \in \mathcal{H},\ x_2 \in \mathcal{H} \quad \text{such that} \quad \begin{cases} 0 \in A_1x_1 + x_1 - x_2 \\ 0 \in A_2x_2 - x_1 + x_2, \end{cases} \tag{1.1}$$

where $(\mathcal{H}, \|\cdot\|)$ is a real Hilbert space, and where A_1 and A_2 are maximal monotone operators acting on $\mathcal{H}$. This formulation arises in various areas of nonlinear analysis [12]. For example, if $A_1 = \partial f_1$ and $A_2 = \partial f_2$ are the subdifferentials of proper lower semicontinuous convex functions f_1 and f_2 from $\mathcal{H}$ to $]-\infty, +\infty]$, (1.1) is equivalent to

$$\underset{x_1\in\mathcal{H},\, x_2\in\mathcal{H}}{\text{minimize}} \quad f_1(x_1) + f_2(x_2) + \frac{1}{2}\|x_1 - x_2\|^2. \tag{1.2}$$

This joint minimization problem, which was first investigated in [1], models problems in disciplines such as the cognitive sciences [4], image processing [26], and signal processing [28] (see also the references therein for further applications in mechanics, filter design, and dynamical games). In particular, if f_1 and f_2 are the indicator functions of closed convex subsets C_1 and C_2 of $\mathcal{H}$, (1.2) reduces to the classical best approximation pair problem [8, 11, 18, 29]

$$\underset{x_1\in C_1,\, x_2\in C_2}{\text{minimize}} \quad \|x_1 - x_2\|. \tag{1.3}$$

On the numerical side, a simple algorithm is available to solve (1.1), namely,

$$x_{1,0} \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \begin{cases} x_{2,n} &= (\mathrm{Id} + A_2)^{-1}x_{1,n} \\ x_{1,n+1} &= (\mathrm{Id} + A_1)^{-1}x_{2,n}. \end{cases} \tag{1.4}$$

This alternating resolvent method produces sequences $(x_{1,n})_{n\in\mathbb{N}}$ and $(x_{2,n})_{n\in\mathbb{N}}$ that converge weakly to points x_1 and x_2, respectively, such that (x_1, x_2) solves (1.1) if solutions exist [12, Theorem 3.3]. In [3], the variational formulation (1.2) was extended to

$$\underset{x_1\in\mathcal{H}_1,\, x_2\in\mathcal{H}_2}{\text{minimize}} \quad f_1(x_1) + f_2(x_2) + \frac{1}{2}\|L_1x_1 - L_2x_2\|_{\mathcal{G}}^2, \tag{1.5}$$

where $\mathcal{H}_1$, $\mathcal{H}_2$, and $\mathcal{G}$ are Hilbert spaces, $f_1 \colon \mathcal{H}_1 \to]-\infty, +\infty]$ and $f_2 \colon \mathcal{H}_2 \to]-\infty, +\infty]$ are proper lower semicontinuous convex functions, and $L_1 \colon \mathcal{H}_1 \to \mathcal{G}$ and $L_2 \colon \mathcal{H}_2 \to \mathcal{G}$ are linear and bounded. This problem was solved in [3] via an inertial alternating minimization procedure first proposed in [4] for (1.2).

The above problems and their solution algorithms are limited to two variables which, in addition, must be linearly coupled. These are serious restrictions since models featuring more than two variables and/or nonlinear coupling schemes arise naturally in applications. The purpose of this paper is to address simultaneously these restrictions by proposing a parallel algorithm for solving systems of monotone inclusions involving an arbitrary number of variables and nonlinear coupling. The breadth and flexibility of this framework will be illustrated through applications in the areas of evolution inclusions, variational problems, best approximation, and network flows.

We now state our problem formulation and our standing assumptions.

Problem 1.1 Let $(\mathcal{H}_i)_{1\leq i\leq m}$ be real Hilbert spaces, where $m \geq 2$. For every $i \in \{1,\ldots,m\}$, let $A_i\colon \mathcal{H}_i \to 2^{\mathcal{H}_i}$ be maximal monotone and let $B_i\colon \mathcal{H}_1\times\cdots\times\mathcal{H}_m \to \mathcal{H}_i$. It is assumed that there exists $\beta \in \left]0,+\infty\right[$ such that

$$(\forall (x_1,\ldots,x_m) \in \mathcal{H}_1\times\cdots\times\mathcal{H}_m)(\forall (y_1,\ldots,y_m) \in \mathcal{H}_1\times\cdots\times\mathcal{H}_m)$$
$$\sum_{i=1}^m \langle x_i - y_i \mid B_i(x_1,\ldots,x_m) - B_i(y_1,\ldots,y_m)\rangle \geq \beta \sum_{i=1}^m \big\|B_i(x_1,\ldots,x_m) - B_i(y_1,\ldots,y_m)\big\|^2. \quad (1.6)$$

The problem is to

$$\text{find } x_1 \in \mathcal{H}_1,\ldots,x_m \in \mathcal{H}_m \quad \text{such that} \quad \begin{cases} 0 \in A_1x_1 + B_1(x_1,\ldots,x_m)\\ \quad\vdots\\ 0 \in A_mx_m + B_m(x_1,\ldots,x_m),\end{cases} \quad (1.7)$$

under the assumption that such points exist.

In abstract terms, the system of inclusions in (1.7) models an equilibrium involving m variables in different Hilbert spaces. The ith inclusion in this system is a perturbation of the basic inclusion $0 \in A_ix_i$ by addition of the coupling term $B_i(x_1,\ldots,x_m)$. Our analysis captures various linear and nonlinear coupling schemes. If

$$(\forall i \in \{1,\ldots,m\}) \quad \mathcal{H}_i = \mathcal{H} \quad \text{and} \quad (\forall x \in \mathcal{H}) \quad B_i(x,\ldots,x) = 0, \quad (1.8)$$

then Problem 1.1 is a relaxation of the standard problem [20, 33] of finding a common zero of the operators $(A_i)_{1\leq i\leq m}$, i.e., of solving the inclusion $0 \in \bigcap_{i=1}^m A_ix$. In particular, if $m = 2$, $\mathcal{H}_1 = \mathcal{H}_2 = \mathcal{H}$, $B_1 = -B_2\colon (x_1,x_2) \mapsto x_1 - x_2$, and $\beta = 1/2$, then Problem 1.1 reverts to (1.1). On the other hand, if $m = 2$, $A_1 = \partial f_1$, $A_2 = \partial f_2$, $B_1\colon (x_1,x_2) \mapsto L_1^*(L_1x_1 - L_2x_2)$, $B_2\colon (x_1,x_2) \mapsto -L_2^*(L_1x_1 - L_2x_2)$, and $\beta = (\|L_1\|^2 + \|L_2\|^2)^{-1}$, then Problem 1.1 reverts to (1.5). Generally speaking, (1.7) covers coupled problems involving minimizations, variational inequalities, saddle points, or evolution inclusions, depending on the type of the maximal monotone operators $(A_i)_{1\leq i\leq m}$.

The paper is organized as follows. In Section 2, we present our algorithm for solving Problem 1.1 and prove its convergence. Applications to systems of evolution inclusions are treated in Section 3. Finally, Section 4 is devoted to variational formulations deriving from Problem 1.1 and features applications to best approximation and network flows.

Notation. Throughout, $\mathcal{H}$ and $(\mathcal{H}_i)_{1\leq i\leq m}$ are real Hilbert spaces. For convenience, their scalar products are all denoted by $\langle\cdot\mid\cdot\rangle$ and the associated norms by $\|\cdot\|$. The symbols $\rightharpoonup$ and $\to$ denote, respectively, weak and strong convergence, Id denotes the identity operator, and L^* denotes the adjoint of a bounded linear operator L. The indicator function of a subset C of $\mathcal{H}$ is

$$\iota_C\colon x \mapsto \begin{cases} 0, & \text{if } x \in C;\\ +\infty, & \text{if } x \notin C,\end{cases} \quad (1.9)$$

and the distance from $x \in \mathcal{H}$ to C is $d_C(x) = \inf_{y\in C}\|x - y\|$; if C is nonempty closed and convex, the projection of x onto C is the unique point P_Cx in C such that $\|x - P_Cx\| = d_C(x)$. We denote by $\Gamma_0(\mathcal{H})$ the class of lower semicontinuous convex functions $f\colon \mathcal{H} \to \left]-\infty,+\infty\right]$ which are proper in the sense that $\operatorname{dom} f = \{x \in \mathcal{H} \mid f(x) < +\infty\} \neq \varnothing$. The subdifferential of $f \in \Gamma_0(\mathcal{H})$ is the maximal monotone operator

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

We denote by $\operatorname{gra} A = \big\{(x,u) \in \mathcal{H} \times \mathcal{H} \bigm| u \in Ax\big\}$ the graph of a set-valued operator $A \colon \mathcal{H} \to 2^{\mathcal{H}}$, by $\operatorname{dom} A = \big\{x \in \mathcal{H} \bigm| Ax \neq \varnothing\big\}$ its domain, and by $J_A = (\mathrm{Id} + A)^{-1}$ its resolvent. If A is monotone, then J_A is single-valued and nonexpansive and, furthermore, if A is maximal monotone, then $\operatorname{dom} J_A = \mathcal{H}$. For complements and further background on convex analysis and monotone operator theory, see [5, 15, 44, 46, 48].

2 Algorithm

Let us start with a characterization of the solutions to Problem 1.1.

Proposition 2.1 *Let $(x_i)_{1 \leq i \leq m} \in \mathcal{H}_1 \times \cdots \times \mathcal{H}_m$, let $(\lambda_i)_{1 \leq i \leq m} \in [0,1[^m$, and let $\gamma \in \left]0,+\infty\right[$. Then $(x_i)_{1 \leq i \leq m}$ solves Problem 1.1 if and only if*

$$(\forall i \in \{1,\ldots,m\}) \quad x_i = \lambda_i x_i + (1-\lambda_i) J_{\gamma A_i}\big(x_i - \gamma B_i(x_1,\ldots,x_m)\big). \tag{2.1}$$

Proof. Let $i \in \{1,\ldots,m\}$. Then, since B_i is single-valued,

$$\begin{aligned} 0 \in A_i x_i + B_i(x_1,\ldots,x_m) \quad &\Leftrightarrow \quad x_i - \gamma B_i(x_1,\ldots,x_m) \in x_i + \gamma A_i x_i \\ &\Leftrightarrow \quad x_i = J_{\gamma A_i}\big(x_i - \gamma B_i(x_1,\ldots,x_m)\big) \\ &\Leftrightarrow \quad x_i = x_i + (1-\lambda_i)\Big(J_{\gamma A_i}\big(x_i - \gamma B_i(x_1,\ldots,x_m)\big) - x_i\Big), \end{aligned} \tag{2.2}$$

and we obtain (2.1). ☐

The above characterization suggests the following algorithm, which constructs m sequences $((x_{i,n})_{n\in\mathbb{N}})_{1 \leq i \leq m}$. Recall that β is the constant appearing in (1.6).

Algorithm 2.2 Fix $\varepsilon \in \left]0, \min\{1,\beta\}\right[$, $(\gamma_n)_{n\in\mathbb{N}}$ in $[\varepsilon, 2\beta - \varepsilon]$, $(\lambda_n)_{n\in\mathbb{N}}$ in $[0, 1-\varepsilon]$, and $(x_{i,0})_{1 \leq i \leq m} \in \mathcal{H}_1 \times \cdots \times \mathcal{H}_m$. Set, for every $n \in \mathbb{N}$,

$$\begin{cases} x_{1,n+1} = \lambda_{1,n} x_{1,n} + (1-\lambda_{1,n})\Big(J_{\gamma_n A_{1,n}}\big(x_{1,n} - \gamma_n(B_{1,n}(x_{1,n},\ldots,x_{m,n}) + b_{1,n})\big) + a_{1,n}\Big) \\ \quad\vdots \\ x_{m,n+1} = \lambda_{m,n} x_{m,n} + (1-\lambda_{m,n})\Big(J_{\gamma_n A_{m,n}}\big(x_{m,n} - \gamma_n(B_{m,n}(x_{1,n},\ldots,x_{m,n}) + b_{m,n})\big) + a_{m,n}\Big), \end{cases} \tag{2.3}$$

where, for every $i \in \{1,\ldots,m\}$, the following hold.

(i) $(A_{i,n})_{n\in\mathbb{N}}$ are maximal monotone operators from $\mathcal{H}_i$ to $2^{\mathcal{H}_i}$ such that

$$(\forall \rho \in \left]0,+\infty\right[) \ \sum_{n\in\mathbb{N}} \sup_{\|y\| \leq \rho} \|J_{\gamma_n A_{i,n}} y - J_{\gamma_n A_i} y\| < +\infty. \tag{2.4}$$

(ii) $(B_{i,n})_{n\in\mathbb{N}}$ are operators from $\mathcal{H}_1 \times \cdots \times \mathcal{H}_m$ to $\mathcal{H}_i$ such that

(a) the operators $(B_{i,n} - B_i)_{n\in\mathbb{N}}$ are Lipschitz-continuous with respective constants $(\kappa_{i,n})_{n\in\mathbb{N}}$ in $]0,+\infty[$ satisfying $\sum_{n\in\mathbb{N}} \kappa_{i,n} < +\infty$; and

(b) there exists $\boldsymbol{z} \in \mathcal{H}_1 \times \cdots \times \mathcal{H}_m$, independent of i, such that $(\forall n \in \mathbb{N})$ $B_{i,n}\boldsymbol{z} = B_i\boldsymbol{z}$.

(iii) $(a_{i,n})_{n\in\mathbb{N}}$ and $(b_{i,n})_{n\in\mathbb{N}}$ are sequences in $\mathcal{H}_i$ such that $\sum_{n\in\mathbb{N}} \|a_{i,n}\| < +\infty$ and $\sum_{n\in\mathbb{N}} \|b_{i,n}\| < +\infty$.

(iv) $(\lambda_{i,n})_{n\in\mathbb{N}}$ is a sequence in $[0,1[$ such that $\sum_{n\in\mathbb{N}} |\lambda_{i,n} - \lambda_n| < +\infty$.

Conditions (i) and (ii) describe the types of approximations to the original operators $(A_i)_{1\leq i\leq m}$ and $(B_i)_{1\leq i\leq m}$ which can be utilized. Condition (iii) quantifies the tolerance which is allowed in the implementation of these approximations (see [25, 31, 32] for specific examples), while (iv) quantifies that allowed in the departure from the global relaxation scheme. The parallel nature of Algorithm 2.2 stems from the fact that the m evaluations of the resolvent operators in (2.3) can be performed independently and, therefore, simultaneously on concurrent processors.

Our asymptotic analysis of Algorithm 2.2 will be based on Theorem 2.8 below on the convergence of the forward-backward algorithm. First, we need to introduce the notion of demiregularity. This notion captures various properties typically used to establish the strong convergence of dynamical systems, e.g., compactness [18], bounded compactness [8, 21, 22], uniform monotonicity [22, 24, 48], uniform convexity [26, 29, 34, 46], compactness of resolvents [30], and demicompactness [38, 47]. In the case of at most single-valued operators, demiregularity captures standard regularity properties used in nonlinear analysis [48, Definition 27.1].

Definition 2.3 An operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is demiregular at $y \in \operatorname{dom} A$ if, for every sequence $((y_n, v_n))_{n\in\mathbb{N}}$ in $\operatorname{gra} A$ and every $v \in Ay$, we have

$$\begin{cases} y_n \rightharpoonup y \\ v_n \to v \end{cases} \quad \Rightarrow \quad y_n \to y. \tag{2.5}$$

Proposition 2.4 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, let $y \in \operatorname{dom} A$, and let $\mathcal{M}$ be the class of nondecreasing functions from $[0,+\infty[$ to $[0,+\infty]$ that vanish only at 0. Suppose that one of the following holds.*

(i) *A is uniformly monotone at y, i.e., there exists $\phi \in \mathcal{M}$ such that*

$$(\forall v \in Ay)(\forall (x,u) \in \operatorname{gra} A) \quad \langle x - y \mid u - v\rangle \geq \phi(\|x - y\|). \tag{2.6}$$

(ii) *A is uniformly monotone, i.e., there exists $\phi \in \mathcal{M}$ such that (2.6) holds for every $y \in \operatorname{dom} A$.*

(iii) *A is strongly monotone, i.e., there exists $\rho \in \,]0,+\infty[$ such that $A - \rho\,\mathrm{Id}$ is monotone.*

(iv) *$A = \partial f$, where $f \in \Gamma_0(\mathcal{H})$ is uniformly convex at y [46, Section 3.4], i.e., there exists $\phi \in \mathcal{M}$ such that*

$$(\forall \alpha \in \,]0,1[)(\forall x \in \operatorname{dom} f) \\ f\big(\alpha x + (1-\alpha)y\big) + \alpha(1-\alpha)\phi(\|x - y\|) \leq \alpha f(x) + (1-\alpha) f(y). \tag{2.7}$$

(v) *$A = \partial f$, where $f \in \Gamma_0(\mathcal{H})$ is uniformly convex, i.e., there exists $\phi \in \mathcal{M}$ such that (2.7) holds for every $y \in \operatorname{dom} f$.*

(vi) *$A=\partial f$, where $f\in\Gamma_0(\mathcal{H})$ is strongly convex, i.e., there exists $\rho\in\,]0,+\infty[$ such that $f-\rho\|\cdot\|^2/2$ is convex.*

(vii) *$A=\partial f$, where $f\in\Gamma_0(\mathcal{H})$ and the lower level sets of f are boundedly compact.*

(viii) *J_A is compact, i.e., for every bounded set $C\subset\mathcal{H}$, the closure of $J_A(C)$ is compact.*

(ix) *$\operatorname{dom} A$ is boundedly relatively compact, i.e., the intersection of its closure with every closed ball is compact.*

(x) *$\mathcal{H}$ is finite-dimensional.*

(xi) *$A\colon\mathcal{H}\to\mathcal{H}$ is single-valued with a single-valued continuous inverse.*

(xii) *A is single-valued on $\operatorname{dom} A$ and $\operatorname{Id}-A$ demicompact [38], [47, Section 10.4], i.e., for every bounded sequence $(x_n)_{n\in\mathbb{N}}$ in $\operatorname{dom} A$ such that $(Ax_n)_{n\in\mathbb{N}}$ converges strongly, $(x_n)_{n\in\mathbb{N}}$ admits a strong cluster point.*

Then A is demiregular at y.

Proof. Let $((y_n,v_n))_{n\in\mathbb{N}}$ be a sequence in $\operatorname{gra} A$ and let $v\in Ay$ be such that $y_n\rightharpoonup y$ and $v_n\to v$. We must show that $y_n\to y$.

(i): By (2.6), there exists $\phi\in\mathcal{M}$ such that $(\forall n\in\mathbb{N})$ $\langle y_n-y\mid v_n-v\rangle\geq\phi(\|y_n-y\|)$. However, since $y_n\rightharpoonup y$ and $v_n\to v$, we have $\langle y_n-y\mid v_n-v\rangle\to 0$. Therefore, appealing to the properties of ϕ, we conclude that $\|y_n-y\|\to 0$.

(ii)⇒(i): Clear.

(iii)⇒(ii): Indeed, A is uniformly monotone with $\phi\colon t\mapsto\rho t^2$.

(iv)⇒(i): See [46, Section 3.4].

(v)⇒(iv): Clear.

(vi)⇒(v): Indeed, f is uniformly convex with $\phi\colon t\mapsto\rho t^2/2$.

(vii): Since $\langle y_n-y\mid v_n\rangle\to 0$, there exists $\rho\in\,]0,+\infty[$ such that $\sup_{n\in\mathbb{N}}\langle y_n-y\mid v_n\rangle\leq\rho$. Hence, since $y\in\operatorname{dom}\partial f\subset\operatorname{dom} f$, (1.10) yields

$$(\forall n\in\mathbb{N})\quad f(y_n)\leq f(y)+\langle y_n-y\mid v_n\rangle\leq f(y)+\rho<+\infty,\tag{2.8}$$

which shows that $(y_n)_{n\in\mathbb{N}}$ lies in a lower level set of f. Since $(y_n)_{n\in\mathbb{N}}$ is bounded, it therefore lies in a compact set. However, since weak convergence and strong convergence coincide for sequences in compact sets, we conclude that $y_n\to y$.

(viii): We have $(\forall n\in\mathbb{N})$ $(y_n,v_n)\in\operatorname{gra} A\Rightarrow(v_n+y_n)-y_n\in Ay_n\Rightarrow y_n=J_A(v_n+y_n)$. Since $(v_n+y_n)_{n\in\mathbb{N}}$ converge weakly, it lies in a bounded set C. Thus, $(y_n)_{n\in\mathbb{N}}$ lies in $J_A(C)$, which has compact closure. Hence $y_n\rightharpoonup y\Rightarrow y_n\to y$.

(ix)⇒(viii): Let $C\subset\mathcal{H}$ be bounded. Then $J_A(C)\subset J_A(\mathcal{H})=\operatorname{dom} A$ and, by nonexpansivity of J_A, $J_A(C)$ is bounded. Altogether, $J_A(C)$ has compact closure.

(x)⇒(ix): Clear.

(xi): Since $Ay_n = v_n \to v = Ay$, we have $y_n = A^{-1}v_n \to A^{-1}v = y$.

(xii): Since $(y_n)_{n\in\mathbb{N}}$ converges weakly, it is bounded. In addition, $(Ay_n)_{n\in\mathbb{N}} = (v_n)_{n\in\mathbb{N}}$ converges strongly. Hence, by demicompactness of $\mathrm{Id} - A$, $(y_n)_{n\in\mathbb{N}}$ has a strong cluster point x and, since $y_n \rightharpoonup y$, we must have $x = y$. Now suppose that $y_n \not\to y$. Then, there exist $\varepsilon \in \left]0,+\infty\right[$ and a subsequence $(y_{k_n})_{n\in\mathbb{N}}$ such that

$$(\forall n \in \mathbb{N}) \quad \|y_{k_n} - y\| \geq \varepsilon. \tag{2.9}$$

However, since $y_{k_n} \rightharpoonup y$ and $(Ay_{k_n})_{n\in\mathbb{N}}$ converges strongly, arguing as above, we can extract a further subsequence $(y_{l_{k_n}})_{n\in\mathbb{N}}$ such that $y_{l_{k_n}} \to y$, which contradicts (2.9). Therefore, $y_n \to y$. □

Next, we recall the notion of cocoercivity.

Definition 2.5 Let $\chi \in \left]0,+\infty\right[$. An operator $B\colon \mathcal{H} \to \mathcal{H}$ is χ-cocoercive if χB is firmly nonexpansive, i.e.,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Bx - By\rangle \geq \chi\|Bx - By\|^2. \tag{2.10}$$

Firmly nonexpansive operators include resolvents of maximal monotone operators, proximity operators, and projectors onto nonempty closed convex sets. In addition, the Yosida approximation of a maximal monotone operator of index χ is χ-cocoercive [2] (further examples of cocoercive operators can be found in [49]). It is clear from (2.10) that, if B is χ-cocoercive, then it is χ^{-1}-Lipschitz continuous. The next lemma, which provides a converse implication, supplies us with another important instance of cocoercive operator (see also [27]).

Lemma 2.6 [7, Corollaire 10] *Let $\varphi\colon \mathcal{H} \to \mathbb{R}$ be a differentiable convex function and let $\tau \in \left]0,+\infty\right[$. Suppose that $\nabla\varphi$ is τ-Lipschitz continuous. Then $\nabla\varphi$ is τ^{-1}-cocoercive.*

We shall also use the following fact.

Lemma 2.7 [22, Lemma 2.3] *Let $\chi \in \left]0,+\infty\right[$, let $B\colon \mathcal{H} \to \mathcal{H}$ be a χ-cocoercive operator, and let $\gamma \in \left]0,2\chi\right[$. Then $\mathrm{Id} - \gamma B$ is nonexpansive.*

We are now ready to record some convergence properties of the forward-backward algorithm, which are of interest in their own right. The forward-backward algorithm finds its roots in the projected gradient method [34] and certain methods for solving variational inequalities [6, 16, 35, 43] (see also the bibliography of [22] for more recent developments).

Theorem 2.8 *Let $(\boldsymbol{\mathcal{H}}, |||\cdot|||)$ be a real Hilbert space, let $\chi \in \left]0,+\infty\right[$, let $\boldsymbol{A}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}$ be a maximal monotone operator, and let $\boldsymbol{B}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ be a χ-cocoercive operator such that*

$$\boldsymbol{Z} = (\boldsymbol{A} + \boldsymbol{B})^{-1}(\boldsymbol{0}) \neq \varnothing. \tag{2.11}$$

Fix $\varepsilon \in \left]0,\min\{1,\chi\}\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon, 2\chi-\varepsilon]$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $[0, 1-\varepsilon]$, and let $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$ and $(\boldsymbol{b}_n)_{n\in\mathbb{N}}$ be sequences in $\boldsymbol{\mathcal{H}}$ such that $\sum_{n\in\mathbb{N}} |||\boldsymbol{a}_n||| < +\infty$ and $\sum_{n\in\mathbb{N}} |||\boldsymbol{b}_n||| < +\infty$. Finally, fix $\boldsymbol{x}_0 \in \boldsymbol{\mathcal{H}}$ and, for every $n \in \mathbb{N}$, set

$$\boldsymbol{x}_{n+1} = \lambda_n \boldsymbol{x}_n + (1-\lambda_n)\big(J_{\gamma_n \boldsymbol{A}}(\boldsymbol{x}_n - \gamma_n(\boldsymbol{B}\boldsymbol{x}_n + \boldsymbol{b}_n)) + \boldsymbol{a}_n\big). \tag{2.12}$$

Then the following hold for some $\boldsymbol{x} \in \boldsymbol{Z}$.

(i) $\boldsymbol{x}_n \rightharpoonup \boldsymbol{x}$.

(ii) $\boldsymbol{B}\boldsymbol{x}_n \to \boldsymbol{B}\boldsymbol{x}$.

(iii) $\boldsymbol{x}_n - J_{\gamma_n \boldsymbol{A}}(\boldsymbol{x}_n - \gamma_n \boldsymbol{B}\boldsymbol{x}_n) \to \mathbf{0}$.

(iv) *Suppose that one of the following is satisfied.*

(a) $\boldsymbol{A}$ *is demiregular at* $\boldsymbol{x}$ *(see Proposition 2.4 for special cases).*

(b) $\boldsymbol{B}$ *is demiregular at* $\boldsymbol{x}$ *(see Proposition 2.4 for special cases).*

(c) $\operatorname{int} \boldsymbol{Z} \neq \varnothing$.

Then $\boldsymbol{x}_n \to \boldsymbol{x}$.

Proof. For every $n \in \mathbb{N}$, set

$$\boldsymbol{T}_{1,n} = J_{\gamma_n \boldsymbol{A}},\ \boldsymbol{T}_{2,n} = \mathbf{Id} - \gamma_n \boldsymbol{B},$$
$$\boldsymbol{e}_{1,n} = \boldsymbol{a}_n,\ \boldsymbol{e}_{2,n} = -\gamma_n \boldsymbol{b}_n,\ \mu_n = 1 - \lambda_n,\ \beta_{1,n} = 2,\ \text{ and }\ \beta_{2,n} = \frac{2\chi}{\gamma_n}. \tag{2.13}$$

Then $\sum_{n\in\mathbb{N}} \mu_n |||\boldsymbol{e}_{1,n}||| < +\infty$, $\sum_{n\in\mathbb{N}} \mu_n |||\boldsymbol{e}_{2,n}||| < +\infty$, and, by [22, Equation (6.5)], $\boldsymbol{Z} = \bigcap_{n\in\mathbb{N}} \operatorname{Fix} \boldsymbol{T}_{1,n}\boldsymbol{T}_{2,n}$. Moreover, as seen in [22, Section 6], $(1-\beta_{1,n})\mathbf{Id} + \beta_{1,n}\boldsymbol{T}_{1,n}$ and $(1-\beta_{2,n})\mathbf{Id} + \beta_{2,n}\boldsymbol{T}_{2,n}$ are nonexpansive, and (2.12) can be rewritten as

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \mu_n\big(\boldsymbol{T}_{1,n}\big(\boldsymbol{T}_{2,n}\boldsymbol{x}_n + \boldsymbol{e}_{2,n}\big) + \boldsymbol{e}_{1,n} - \boldsymbol{x}_n\big), \tag{2.14}$$

which is precisely the iteration governing [22, Algorithm 1.2], where $m = 2$.

(i): [22, Corollary 6.5].

(ii)&(iii): We derive from (2.14), [22, Remark 3.4], and our assumptions on $(\lambda_n)_{n\in\mathbb{N}}$ and $(\gamma_n)_{n\in\mathbb{N}}$ that $(\mathbf{Id} - \boldsymbol{T}_{2,n})\boldsymbol{x}_n - (\mathbf{Id} - \boldsymbol{T}_{2,n})\boldsymbol{x} \to \mathbf{0}$ and, in turn, that $\boldsymbol{B}\boldsymbol{x}_n \to \boldsymbol{B}\boldsymbol{x}$. Likewise, [22, Remark 3.4] yields $\boldsymbol{x}_n - \boldsymbol{T}_{1,n}\boldsymbol{T}_{2,n}\boldsymbol{x}_n \to \mathbf{0}$ and, therefore, $\boldsymbol{x}_n - J_{\gamma_n \boldsymbol{A}}(\boldsymbol{x}_n - \gamma_n \boldsymbol{B}\boldsymbol{x}_n) \to \mathbf{0}$.

(iv)(a): Set $\boldsymbol{v} = -\boldsymbol{B}\boldsymbol{x}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \boldsymbol{y}_n = J_{\gamma_n \boldsymbol{A}}\big(\boldsymbol{x}_n - \gamma_n \boldsymbol{B}\boldsymbol{x}_n\big) \\ \boldsymbol{v}_n = \gamma_n^{-1}(\boldsymbol{x}_n - \boldsymbol{y}_n) - \boldsymbol{B}\boldsymbol{x}_n. \end{cases} \tag{2.15}$$

On the one hand, we have $\boldsymbol{v} = -\boldsymbol{B}\boldsymbol{x} \in \boldsymbol{A}\boldsymbol{x}$ and $(\forall n \in \mathbb{N})$ $(\boldsymbol{y}_n, \boldsymbol{v}_n) \in \operatorname{gra} \boldsymbol{A}$. On the other hand, we derive from (i) and (iii) that $\boldsymbol{y}_n \rightharpoonup \boldsymbol{x}$. Furthermore, since

$$(\forall n \in \mathbb{N}) \ \ |||\boldsymbol{v}_n - \boldsymbol{v}||| \leq \frac{|||\boldsymbol{x}_n - \boldsymbol{y}_n|||}{\gamma_n} + |||\boldsymbol{B}\boldsymbol{x}_n - \boldsymbol{B}\boldsymbol{x}|||, \tag{2.16}$$

it follows from (ii), (iii), and the condition $\inf_{n\in\mathbb{N}} \gamma_n > 0$ that $\boldsymbol{v}_n \to \boldsymbol{v}$. It then results from Definition 2.3 that $\boldsymbol{y}_n \to \boldsymbol{x}$ and, in turn, from (iii) that $\boldsymbol{x}_n \to \boldsymbol{x}$.

(iv)(b): Set $\boldsymbol{v} = \boldsymbol{B}\boldsymbol{x}$ and $(\forall n \in \mathbb{N})$ $\boldsymbol{v}_n = \boldsymbol{B}\boldsymbol{x}_n$. Then (i) yields $\boldsymbol{x}_n \rightharpoonup \boldsymbol{x}$ and (ii) yields $\boldsymbol{v}_n \to \boldsymbol{v}$. It thus follows from Definition 2.3 that $\boldsymbol{x}_n \to \boldsymbol{x}$.

(iv)(c): This follows from (i) and [22, Theorem 3.3(i) & Lemma 2.8(iv)]. ☐

The main results of this section are the following theorems. Let us start with weak convergence.

Theorem 2.9 *Let $((x_{i,n})_{n\in\mathbb{N}})_{1\leq i\leq m}$ be sequences generated by Algorithm 2.2. Then, for every $i \in \{1,\ldots,m\}$, $(x_{i,n})_{n\in\mathbb{N}}$ converges weakly to a point $x_i \in \mathcal{H}_i$, and $(x_i)_{1\leq i\leq m}$ is a solution to Problem 1.1.*

Proof. Throughout the proof, a generic element $\boldsymbol{x}$ in the Cartesian product $\mathcal{H}_1 \times \cdots \times \mathcal{H}_m$ will be expressed in terms of its components as $\boldsymbol{x} = (x_i)_{1\leq i\leq m}$. We shall show that our algorithmic setting reduces to the situation described in Theorem 2.8(i) in the Hilbert direct sum $\boldsymbol{\mathcal{H}} = \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_m$ obtained by endowing $\mathcal{H}_1 \times \cdots \times \mathcal{H}_m$ with the scalar product

$$\langle\langle \cdot \mid \cdot \rangle\rangle \colon (\boldsymbol{x}, \boldsymbol{y}) \mapsto \sum_{i=1}^{m} \langle x_i \mid y_i \rangle, \tag{2.17}$$

with associated norm

$$|||\cdot||| \colon \boldsymbol{x} \mapsto \sqrt{\sum_{i=1}^{m} \|x_i\|^2}. \tag{2.18}$$

To this end, we shall show that the iterations (2.3) can be cast in the form of (2.12). First, define

$$\boldsymbol{A} \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}} \colon \boldsymbol{x} \mapsto \bigtimes_{i=1}^{m} A_i x_i \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \boldsymbol{A}_n \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}} \colon \boldsymbol{x} \mapsto \bigtimes_{i=1}^{m} A_{i,n} x_i. \tag{2.19}$$

It follows from the maximal monotonicity of the operators $(A_i)_{1\leq i\leq m}$, condition (i) in Algorithm 2.2, (2.17), and (2.19) that

$$\boldsymbol{A} \text{ and } (\boldsymbol{A}_n)_{n\in\mathbb{N}} \text{ are maximal monotone,} \tag{2.20}$$

with resolvents

$$J_{\boldsymbol{A}} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{x} \mapsto (J_{A_i} x_i)_{1\leq i\leq m} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad J_{\boldsymbol{A}_n} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{x} \mapsto (J_{A_{i,n}} x_i)_{1\leq i\leq m}, \tag{2.21}$$

respectively. Moreover, for every $\rho \in \left]0,+\infty\right[$, we derive from (2.18), (2.21), and condition (i) in Algorithm 2.2 that

$$\begin{aligned}
\sum_{n\in\mathbb{N}} \sup_{|||\boldsymbol{y}|||\leq\rho} |||J_{\gamma_n \boldsymbol{A}_n} \boldsymbol{y} - J_{\gamma_n \boldsymbol{A}} \boldsymbol{y}||| &= \sum_{n\in\mathbb{N}} \sup_{|||\boldsymbol{y}|||\leq\rho} \sqrt{\sum_{i=1}^{m} \|J_{\gamma_n A_{i,n}} y_i - J_{\gamma_n A_i} y_i\|^2} \\
&\leq \sum_{n\in\mathbb{N}} \sup_{|||\boldsymbol{y}|||\leq\rho} \sum_{i=1}^{m} \|J_{\gamma_n A_{i,n}} y_i - J_{\gamma_n A_i} y_i\| \\
&\leq \sum_{i=1}^{m} \sum_{n\in\mathbb{N}} \sup_{\|y_i\|\leq\rho} \|J_{\gamma_n A_{i,n}} y_i - J_{\gamma_n A_i} y_i\| \\
&< +\infty.
\end{aligned} \tag{2.22}$$

Now define

$$\boldsymbol{B} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{x} \mapsto (B_i \boldsymbol{x})_{1\leq i\leq m} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \boldsymbol{B}_n \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{x} \mapsto (B_{i,n} \boldsymbol{x})_{1\leq i\leq m}. \tag{2.23}$$

Then (1.7) is equivalent to

$$\text{find } \boldsymbol{x} \in \boldsymbol{Z} = (\boldsymbol{A} + \boldsymbol{B})^{-1}(\boldsymbol{0}). \tag{2.24}$$

Moreover, in the light of (2.17), (2.18), and (2.23), (1.6) becomes

$$(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}})(\forall \boldsymbol{y} \in \boldsymbol{\mathcal{H}}) \quad \langle\langle \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{B}\boldsymbol{x} - \boldsymbol{B}\boldsymbol{y} \rangle\rangle \geq \beta |||\boldsymbol{B}\boldsymbol{x} - \boldsymbol{B}\boldsymbol{y}|||^2. \tag{2.25}$$

In other words, $\boldsymbol{B}$ is β-cocoercive. Next, let $n \in \mathbb{N}$ and set

$$\boldsymbol{c}_n = (a_{i,n})_{1\leq i\leq m} \quad \text{and} \quad \boldsymbol{d}_n = (b_{i,n})_{1\leq i\leq m}. \tag{2.26}$$

We deduce from (2.18) and condition (iii) in Algorithm 2.2 that

$$\sum_{k\in\mathbb{N}} |||\boldsymbol{c}_k||| \leq \sum_{k\in\mathbb{N}} \sqrt{\sum_{i=1}^m \|a_{i,k}\|^2} \leq \sum_{i=1}^m \sum_{k\in\mathbb{N}} \|a_{i,k}\| < +\infty \tag{2.27}$$

and, likewise, that

$$\sum_{k\in\mathbb{N}} |||\boldsymbol{d}_k||| < +\infty. \tag{2.28}$$

Now set

$$\boldsymbol{x}_n = (x_{i,n})_{1\leq i\leq m} \quad \text{and} \quad \boldsymbol{\Lambda}_n \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{x} \mapsto (\lambda_{i,n} x_i)_{1\leq i\leq m}. \tag{2.29}$$

It follows from (2.18) and condition (iv) in Algorithm 2.2 that

$$|||\boldsymbol{\Lambda}_n||| = \max_{1\leq i\leq m} \lambda_{i,n} \leq 1 \quad \text{and} \quad |||\mathbf{Id} - \boldsymbol{\Lambda}_n||| = 1 - \min_{1\leq i\leq m} \lambda_{i,n} \leq 1. \tag{2.30}$$

Hence,

$$|||\boldsymbol{\Lambda}_n||| + |||\mathbf{Id} - \boldsymbol{\Lambda}_n||| = 1 + \max_{1\leq i\leq m} (\lambda_{i,n} - \lambda_n) - \min_{1\leq i\leq m} (\lambda_{i,n} - \lambda_n) \leq 1 + \tau_n, \tag{2.31}$$

where

$$\tau_n = 2 \max_{1\leq i\leq m} |\lambda_{i,n} - \lambda_n|. \tag{2.32}$$

We observe that, by virtue of condition (iv) in Algorithm 2.2,

$$\sum_{k\in\mathbb{N}} \tau_k = 2 \sum_{k\in\mathbb{N}} \max_{1\leq i\leq m} |\lambda_{i,k} - \lambda_k| \leq 2 \sum_{i=1}^m \sum_{k\in\mathbb{N}} |\lambda_{i,k} - \lambda_k| < +\infty. \tag{2.33}$$

Moreover, in view of (2.21), (2.23), (2.26), and (2.29), the iterations (2.3) are equivalent to

$$\boldsymbol{x}_{n+1} = \boldsymbol{\Lambda}_n \boldsymbol{x}_n + (\mathbf{Id} - \boldsymbol{\Lambda}_n)\big(J_{\gamma_n \boldsymbol{A}_n}\big(\boldsymbol{x}_n - \gamma_n(\boldsymbol{B}_n \boldsymbol{x}_n + \boldsymbol{d}_n)\big) + \boldsymbol{c}_n\big). \tag{2.34}$$

Now define

$$\boldsymbol{D}_n = \boldsymbol{B}_n - \boldsymbol{B}. \tag{2.35}$$

It follows from condition (ii)(a) in Algorithm 2.2, (2.18), and (2.23) that $\boldsymbol{D}_n$ is Lipschitz continuous with constant $\kappa_n = \sqrt{\sum_{i=1}^m \kappa_{i,n}^2}$ and that

$$\sum_{k\in\mathbb{N}} \kappa_k = \sum_{k\in\mathbb{N}} \sqrt{\sum_{i=1}^m \kappa_{i,k}^2} \leq \sum_{i=1}^m \sum_{k\in\mathbb{N}} \kappa_{i,k} < +\infty. \tag{2.36}$$

Furthermore, set

$$\boldsymbol{b}_n = \boldsymbol{D}_n \boldsymbol{x}_n + \boldsymbol{d}_n \tag{2.37}$$

and let $\boldsymbol{x} \in \boldsymbol{Z}$. Then

$$\begin{aligned} |||\boldsymbol{b}_n||| &\leq |||\boldsymbol{D}_n \boldsymbol{x}_n||| + |||\boldsymbol{d}_n||| \\ &\leq |||\boldsymbol{D}_n \boldsymbol{x}_n - \boldsymbol{D}_n \boldsymbol{x}||| + |||\boldsymbol{D}_n \boldsymbol{x} - \boldsymbol{D}_n \boldsymbol{z}||| + |||\boldsymbol{d}_n||| \\ &\leq \kappa_n(|||\boldsymbol{x}_n - \boldsymbol{x}||| + |||\boldsymbol{x} - \boldsymbol{z}|||) + |||\boldsymbol{d}_n|||, \end{aligned} \tag{2.38}$$

where $\boldsymbol{z}$ is provided by assumption (ii)(b) in Algorithm 2.2. We now set

$$\boldsymbol{T}_n = \mathbf{Id} - \gamma_n \boldsymbol{B} \quad \text{and} \quad \boldsymbol{e}_n = J_{\gamma_n \boldsymbol{A}_n}\big(\boldsymbol{T}_n \boldsymbol{x}\big) - \boldsymbol{x}. \tag{2.39}$$

On the one hand, the inequality $\sup_{k\in\mathbb{N}} \gamma_k \leq 2\beta$ yields

$$|||\boldsymbol{T}_n \boldsymbol{x}||| \leq \rho, \quad \text{where} \quad \rho = |||\boldsymbol{x}||| + 2\beta |||\boldsymbol{B}\boldsymbol{x}|||. \tag{2.40}$$

On the other hand, since $\boldsymbol{x}$ is a solution to Problem 1.1, Proposition 2.1, (2.21), and (2.23) supply

$$\boldsymbol{x} = J_{\gamma_n \boldsymbol{A}}(\boldsymbol{T}_n \boldsymbol{x}). \tag{2.41}$$

Therefore, (2.39), (2.40), and (2.22) imply that

$$\sum_{k\in\mathbb{N}} |||\boldsymbol{e}_k||| = \sum_{k\in\mathbb{N}} |||J_{\gamma_k \boldsymbol{A}_k}(\boldsymbol{T}_k \boldsymbol{x}) - \boldsymbol{x}||| = \sum_{k\in\mathbb{N}} |||J_{\gamma_k \boldsymbol{A}_k}(\boldsymbol{T}_k \boldsymbol{x}) - J_{\gamma_k \boldsymbol{A}}(\boldsymbol{T}_k \boldsymbol{x})||| < +\infty. \tag{2.42}$$

In addition, (2.35), (2.37), and (2.39) yield

$$J_{\gamma_n \boldsymbol{A}_n}\big(\boldsymbol{x}_n - \gamma_n(\boldsymbol{B}_n \boldsymbol{x}_n + \boldsymbol{d}_n)\big) - \boldsymbol{x} = J_{\gamma_n \boldsymbol{A}_n}\big(\boldsymbol{T}_n \boldsymbol{x}_n - \gamma_n \boldsymbol{b}_n\big) - J_{\gamma_n \boldsymbol{A}_n}(\boldsymbol{T}_n \boldsymbol{x}) + \boldsymbol{e}_n. \tag{2.43}$$

Since $J_{\gamma_n \boldsymbol{A}}$ and, by Lemma 2.7, $\boldsymbol{T}_n$ are nonexpansive, we derive from (2.43) and (2.38) that

$$\begin{aligned}
|||J_{\gamma_n \boldsymbol{A}_n}\big(\boldsymbol{x}_n - \gamma_n(\boldsymbol{B}_n \boldsymbol{x}_n + \boldsymbol{d}_n)\big) - \boldsymbol{x}||| &\leq |||J_{\gamma_n \boldsymbol{A}_n}\big(\boldsymbol{T}_n \boldsymbol{x}_n - \gamma_n \boldsymbol{b}_n\big) - J_{\gamma_n \boldsymbol{A}_n}(\boldsymbol{T}_n \boldsymbol{x})||| + |||\boldsymbol{e}_n||| \\
&\leq |||\boldsymbol{T}_n \boldsymbol{x}_n - \gamma_n \boldsymbol{b}_n - \boldsymbol{T}_n \boldsymbol{x}||| + |||\boldsymbol{e}_n||| \\
&\leq |||\boldsymbol{x}_n - \boldsymbol{x}||| + \gamma_n |||\boldsymbol{b}_n||| + |||\boldsymbol{e}_n||| \\
&\leq |||\boldsymbol{x}_n - \boldsymbol{x}||| + 2\beta |||\boldsymbol{b}_n||| + |||\boldsymbol{e}_n||| \\
&\leq (1 + 2\beta\kappa_n)|||\boldsymbol{x}_n - \boldsymbol{x}||| + 2\beta\kappa_n |||\boldsymbol{x} - \boldsymbol{z}||| \\
&\quad + 2\beta |||\boldsymbol{d}_n||| + |||\boldsymbol{e}_n|||.
\end{aligned} \tag{2.44}$$

Thus, it results from (2.34), (2.44), (2.31), and (2.30) that

$$\begin{aligned}
|||\boldsymbol{x}_{n+1} - \boldsymbol{x}||| &= |||\boldsymbol{\Lambda}_n(\boldsymbol{x}_n - \boldsymbol{x}) + (\mathbf{Id} - \boldsymbol{\Lambda}_n)\big(J_{\gamma_n \boldsymbol{A}_n}\big(\boldsymbol{x}_n - \gamma_n(\boldsymbol{B}_n \boldsymbol{x}_n + \boldsymbol{d}_n)\big) - \boldsymbol{x} + \boldsymbol{c}_n\big)||| \\
&\leq |||\boldsymbol{\Lambda}_n|||\,|||\boldsymbol{x}_n - \boldsymbol{x}||| + |||\mathbf{Id} - \boldsymbol{\Lambda}_n|||\,|||\boldsymbol{c}_n||| \\
&\quad + |||\mathbf{Id} - \boldsymbol{\Lambda}_n|||\,|||J_{\gamma_n \boldsymbol{A}_n}\big(\boldsymbol{x}_n - \gamma_n(\boldsymbol{B}_n \boldsymbol{x}_n + \boldsymbol{d}_n)\big) - \boldsymbol{x}||| \\
&\leq |||\boldsymbol{\Lambda}_n|||\,|||\boldsymbol{x}_n - \boldsymbol{x}||| + |||\mathbf{Id} - \boldsymbol{\Lambda}_n|||\,|||\boldsymbol{c}_n||| \\
&\quad + |||\mathbf{Id} - \boldsymbol{\Lambda}_n|||\,\big((1 + 2\beta\kappa_n)|||\boldsymbol{x}_n - \boldsymbol{x}||| + 2\beta\kappa_n |||\boldsymbol{x} - \boldsymbol{z}||| \\
&\quad + 2\beta |||\boldsymbol{d}_n||| + |||\boldsymbol{e}_n|||\big) \\
&\leq \big(|||\boldsymbol{\Lambda}_n||| + |||\mathbf{Id} - \boldsymbol{\Lambda}_n|||\big)|||\boldsymbol{x}_n - \boldsymbol{x}||| + |||\mathbf{Id} - \boldsymbol{\Lambda}_n|||\,\big(|||\boldsymbol{c}_n||| + 2\beta\kappa_n |||\boldsymbol{x}_n - \boldsymbol{x}||| \\
&\quad + 2\beta\kappa_n |||\boldsymbol{x} - \boldsymbol{z}||| + 2\beta |||\boldsymbol{d}_n||| + |||\boldsymbol{e}_n|||\big) \\
&\leq (1 + \tau_n)|||\boldsymbol{x}_n - \boldsymbol{x}||| + |||\boldsymbol{c}_n||| + 2\beta\kappa_n |||\boldsymbol{x}_n - \boldsymbol{x}||| \\
&\quad + 2\beta\kappa_n |||\boldsymbol{x} - \boldsymbol{z}||| + 2\beta |||\boldsymbol{d}_n||| + |||\boldsymbol{e}_n||| \\
&\leq (1 + \alpha_n)|||\boldsymbol{x}_n - \boldsymbol{x}||| + \delta_n,
\end{aligned} \tag{2.45}$$

where

$$\alpha_n = \tau_n + 2\beta\kappa_n \quad \text{and} \quad \delta_n = |||\boldsymbol{c}_n||| + 2\beta\kappa_n |||\boldsymbol{x} - \boldsymbol{z}||| + 2\beta |||\boldsymbol{d}_n||| + |||\boldsymbol{e}_n|||. \tag{2.46}$$

In turn, it follows from (2.33), (2.36), (2.27), (2.28), and (2.42) that $\sum_{k\in\mathbb{N}} \alpha_k < +\infty$ and $\sum_{k\in\mathbb{N}} \delta_k < +\infty$. Thus, (2.45) and [39, Lemma 2.2.2] yield

$$\sup_{k\in\mathbb{N}} |||\boldsymbol{x}_k - \boldsymbol{x}||| < +\infty \tag{2.47}$$

and, using (2.36) and (2.28), we derive from (2.38) that

$$\sum_{k\in\mathbb{N}} |||\boldsymbol{b}_k||| < +\infty. \tag{2.48}$$

In view of (2.37), (2.35), and (2.39), (2.34) is equivalent to

$$\boldsymbol{x}_{n+1} = \boldsymbol{\Lambda}_n\boldsymbol{x}_n + (\mathbf{Id} - \boldsymbol{\Lambda}_n)\big(J_{\gamma_n\boldsymbol{A}}(\boldsymbol{T}_n\boldsymbol{x}_n - \gamma_n\boldsymbol{b}_n) + \boldsymbol{h}_n\big), \tag{2.49}$$

where

$$\boldsymbol{h}_n = J_{\gamma_n\boldsymbol{A}_n}(\boldsymbol{T}_n\boldsymbol{x}_n - \gamma_n\boldsymbol{b}_n) - J_{\gamma_n\boldsymbol{A}}(\boldsymbol{T}_n\boldsymbol{x}_n - \gamma_n\boldsymbol{b}_n) + \boldsymbol{c}_n. \tag{2.50}$$

Now set $\mu = \sup_{k\in\mathbb{N}} |||\boldsymbol{x}_k - \boldsymbol{x}||| + \rho + 2\beta\sup_{k\in\mathbb{N}} |||\boldsymbol{b}_k|||$. Then it follows from (2.47) and (2.48) that $\mu < +\infty$. Moreover, we deduce from the nonexpansivity of $\boldsymbol{T}_n$ and (2.40) that

$$\begin{aligned} |||\boldsymbol{T}_n\boldsymbol{x}_n - \gamma_n\boldsymbol{b}_n||| &\leq |||\boldsymbol{T}_n\boldsymbol{x}_n - \boldsymbol{T}_n\boldsymbol{x}||| + |||\boldsymbol{T}_n\boldsymbol{x}||| + 2\beta|||\boldsymbol{b}_n||| \\ &\leq |||\boldsymbol{x}_n - \boldsymbol{x}||| + \rho + 2\beta|||\boldsymbol{b}_n||| \\ &\leq \mu. \end{aligned} \tag{2.51}$$

Hence, appealing to (2.22) and (2.27), we infer from (2.50) that

$$\sum_{k\in\mathbb{N}} |||\boldsymbol{h}_k||| < +\infty. \tag{2.52}$$

Note that, upon introducing

$$\boldsymbol{a}_n = \boldsymbol{h}_n + \frac{1}{1-\lambda_n}(\boldsymbol{\Lambda}_n - \lambda_n\mathbf{Id}\,)\big(\boldsymbol{x}_n - J_{\gamma_n\boldsymbol{A}}(\boldsymbol{T}_n\boldsymbol{x}_n - \gamma_n\boldsymbol{b}_n) - \boldsymbol{h}_n\big) \tag{2.53}$$

and using (2.39), we can rewrite (2.49) in the form of (2.12), namely,

$$\boldsymbol{x}_{n+1} = \lambda_n\boldsymbol{x}_n + (1-\lambda_n)\big(J_{\gamma_n\boldsymbol{A}}(\boldsymbol{x}_n - \gamma_n(\boldsymbol{B}\boldsymbol{x}_n + \boldsymbol{b}_n)) + \boldsymbol{a}_n\big). \tag{2.54}$$

On the other hand, using (2.41) and the nonexpansivity of $J_{\gamma_n\boldsymbol{A}}$ and $\boldsymbol{T}_n$, we get

$$\begin{aligned} |||\boldsymbol{x}_n - J_{\gamma_n\boldsymbol{A}}(\boldsymbol{T}_n\boldsymbol{x}_n - \gamma_n\boldsymbol{b}_n) - \boldsymbol{h}_n||| &\leq |||\boldsymbol{x}_n - \boldsymbol{x}||| + |||J_{\gamma_n\boldsymbol{A}}(\boldsymbol{T}_n\boldsymbol{x}) - J_{\gamma_n\boldsymbol{A}}(\boldsymbol{T}_n\boldsymbol{x}_n - \gamma_n\boldsymbol{b}_n)||| \\ &\quad + |||\boldsymbol{h}_n||| \\ &\leq 2|||\boldsymbol{x}_n - \boldsymbol{x}||| + 2\beta|||\boldsymbol{b}_n||| + |||\boldsymbol{h}_n|||. \end{aligned} \tag{2.55}$$

Therefore, we derive from (2.47), (2.48), and (2.52) that

$$\nu = \sup_{k\in\mathbb{N}} |||\boldsymbol{x}_k - J_{\gamma_k\boldsymbol{A}}(\boldsymbol{T}_k\boldsymbol{x}_k - \gamma_k\boldsymbol{b}_k) - \boldsymbol{h}_k||| < +\infty, \tag{2.56}$$

and hence, from (2.53), (2.29) and the inequality $\lambda_n \leq 1-\varepsilon$, that

$$\begin{aligned} |||\boldsymbol{a}_n||| &\leq |||\boldsymbol{h}_n||| + \frac{1}{1-\lambda_n}|||\boldsymbol{\Lambda}_n - \lambda_n\mathbf{Id}\,|||\;|||\boldsymbol{x}_n - J_{\gamma_n\boldsymbol{A}}(\boldsymbol{T}_n\boldsymbol{x}_n - \gamma_n\boldsymbol{b}_n) - \boldsymbol{h}_n||| \\ &\leq |||\boldsymbol{h}_n||| + \frac{\nu}{\varepsilon}\max_{1\leq i\leq m}|\lambda_{i,n} - \lambda_n|. \end{aligned} \tag{2.57}$$

Thus, using (2.52) and arguing as in (2.33), we get

$$\sum_{k\in\mathbb{N}} |||\boldsymbol{a}_k||| < +\infty. \tag{2.58}$$

However, Theorem 2.8(i) asserts that, with properties (2.20), (2.25), (2.48), (2.58), and under the hypotheses on $(\gamma_n)_{n\in\mathbb{N}}$ and $(\lambda_n)_{n\in\mathbb{N}}$ stated in Algorithm 2.2, the sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ generated by (2.54) converges weakly to a point in $\boldsymbol{Z}$. Since (2.54) is equivalent to (2.3) and (2.24) is equivalent to (1.7), the proof is complete. □

We conclude this section with the following theorem, in which we describe instances of strong convergence derived from Theorem 2.8.

Theorem 2.10 *Let $((x_{i,n})_{n\in\mathbb{N}})_{1\leq i\leq m}$ and $(x_i)_{1\leq i\leq m}$ be as in Theorem 2.9. Then the following hold.*

(i) *Suppose that, for some $i\in\{1,\ldots,m\}$, A_i is demiregular at x_i (see Proposition 2.4 for special cases). Then $x_{i,n}\to x_i$.*

(ii) *Suppose that the operator $(y_j)_{1\leq j\leq m}\mapsto \big(B_i(y_j)_{1\leq j\leq m}\big)_{1\leq i\leq m}$ is demiregular at $(x_i)_{1\leq i\leq m}$ (see Proposition 2.4 for special cases). Then, for every $i\in\{1,\ldots,m\}$, $x_{i,n}\to x_i$.*

(iii) *Suppose that the set of solutions to Problem 1.1 has a nonempty interior. Then, for every $i\in\{1,\ldots,m\}$, $x_{i,n}\to x_i$.*

Proof. We use the same product space setting and notation as in the proof of Theorem 2.9. In particular, we set $\boldsymbol{x}=(x_1,\ldots,x_m)$ and $\boldsymbol{\mathcal{H}}=\mathcal{H}_1\oplus\cdots\oplus\mathcal{H}_m$, and we define

$$\boldsymbol{A}\colon\boldsymbol{\mathcal{H}}\to 2^{\boldsymbol{\mathcal{H}}}\colon\boldsymbol{y}\mapsto \underset{i=1}{\overset{m}{\times}} A_iy_i \quad\text{and}\quad \boldsymbol{B}\colon\boldsymbol{\mathcal{H}}\to\boldsymbol{\mathcal{H}}\colon\boldsymbol{y}\mapsto (B_i\boldsymbol{y})_{1\leq i\leq m}. \tag{2.59}$$

As seen in the proof of Theorem 2.9, the convergence properties of $(\boldsymbol{x}_n)_{n\in\mathbb{N}}=((x_{i,n})_{1\leq i\leq m})_{n\in\mathbb{N}}$ follow from those listed in Theorem 2.8 and applied to the operators defined in (2.59); moreover, the set of solutions to Problem 1.1 is $\boldsymbol{Z}=(\boldsymbol{A}+\boldsymbol{B})^{-1}(\boldsymbol{0})$.

(i): Set $v_i=-B_i(x_1,\ldots,x_m)$ and

$$(\forall n\in\mathbb{N})\quad \begin{cases} y_{i,n}=J_{\gamma_nA_i}\big(x_{i,n}-\gamma_nB_i(x_{1,n},\ldots,x_{m,n})\big)\\ v_{i,n}=\gamma_n^{-1}(x_{i,n}-y_{i,n})-B_i(x_{1,n},\ldots,x_{m,n}). \end{cases} \tag{2.60}$$

We first derive from (1.7) that

$$v_i=-B_i(x_1,\ldots,x_m)\in A_ix_i. \tag{2.61}$$

Moreover, it follows from Theorem 2.8(i) that

$$x_{i,n}\rightharpoonup x_i, \tag{2.62}$$

from Theorem 2.8(ii) that

$$\|B_i(x_{1,n},\ldots,x_{m,n})-B_i(x_1,\ldots,x_m)\|=\|B_i\boldsymbol{x}_n-B_i\boldsymbol{x}\|\leq|||\boldsymbol{B}\boldsymbol{x}_n-\boldsymbol{B}\boldsymbol{x}|||\to 0, \tag{2.63}$$

and from Theorem 2.8(iii) and (2.21) that

$$\|x_{i,n}-y_{i,n}\|\leq|||\boldsymbol{x}_n-J_{\gamma_n\boldsymbol{A}}(\boldsymbol{x}_n-\gamma_n\boldsymbol{B}\boldsymbol{x}_n)|||\to 0. \tag{2.64}$$

Combining (2.62) and (2.64), we obtain

$$y_{i,n}\rightharpoonup x_i. \tag{2.65}$$

Next, we derive from (2.60) that

$$(\forall n \in \mathbb{N}) \quad (y_{i,n}, v_{i,n}) \in \operatorname{gra} A_i \tag{2.66}$$

and that

$$(\forall n \in \mathbb{N}) \ \ \|v_{i,n} - v_i\| \leq \frac{\|x_{i,n} - y_{i,n}\|}{\gamma_n} + \|B_i(x_{1,n}, \dots, x_{m,n}) - B_i(x_1, \dots, x_m)\|. \tag{2.67}$$

Hence, it follows from (2.64), the condition $\inf_{n\in\mathbb{N}} \gamma_n > 0$, and (2.63), that

$$v_{i,n} \to v_i. \tag{2.68}$$

Altogether, (2.61), (2.65), (2.66), (2.68), and Definition 2.3 yield $y_{i,n} \to x_i$. In turn, appealing to (2.64), we conclude that $x_{i,n} \to x_i$.

(ii): This follows Theorem 2.8(iv)(b).

(iii): This follows Theorem 2.8(iv)(c). □

3 Coupling evolution inclusions

Evolution inclusions arise in various fields of applied mathematics [30, 42]. In this section, we address the problem of solving systems of coupled evolution inclusions with periodicity conditions.

Let us recall some standard notation [15, 48]. Fix $T \in \,]0, +\infty[$ and $p \in [1, +\infty[$. Then $\mathcal{D}(]0, T[)$ is the set of infinitely differentiable functions from $]0, T[$ to $\mathbb{R}$ with compact support in $]0, T[$. Given a real Hilbert space H, $\mathcal{C}([0, T]; \mathsf{H})$ is the space of continuous functions from $[0, T]$ to H and $L^p([0, T]; \mathsf{H})$ is the space of classes of equivalences of Borel measurable functions $x\colon [0, T] \to \mathsf{H}$ such that $\int_0^T \|x(t)\|_{\mathsf{H}}^p dt < +\infty$. $L^2([0, T]; \mathsf{H})$ is a Hilbert space with scalar product $(x, y) \mapsto \int_0^T \langle x(t) \mid y(t)\rangle_{\mathsf{H}} dt$. Now take x and y in $L^1([0, T]; \mathsf{H})$. Then y is the weak derivative of x if $\int_0^T \phi(t)y(t)dt = -\int_0^T (d\phi(t)/dt)x(t)dt$ for every $\phi \in \mathcal{D}(]0, T[)$, in which case we use the notation $y = x'$. Moreover,

$$W^{1,2}([0, T]; \mathsf{H}) = \big\{x \in L^2([0, T]; \mathsf{H}) \;\big|\; x' \in L^2([0, T]; \mathsf{H})\big\}, \tag{3.1}$$

equipped with the scalar product $(x, y) \mapsto \int_0^T \langle x(t) \mid y(t)\rangle_{\mathsf{H}} dt + \int_0^T \langle x'(t) \mid y'(t)\rangle_{\mathsf{H}} dt$, is a Hilbert space.

Problem 3.1 Let $(\mathsf{H}_i)_{1\leq i\leq m}$ be real Hilbert spaces and let $T \in \,]0, +\infty[$. For every $i \in \{1, \dots, m\}$, set

$$\mathcal{W}_i = \big\{x \in \mathcal{C}([0, T]; \mathsf{H}_i) \cap W^{1,2}([0, T]; \mathsf{H}_i) \;\big|\; x(T) = x(0)\big\}, \tag{3.2}$$

let $\mathsf{f}_i \in \Gamma_0(\mathsf{H}_i)$, and let $\mathsf{B}_i\colon \mathsf{H}_1 \times \cdots \times \mathsf{H}_m \to \mathsf{H}_i$. It is assumed that there exists $\beta \in \,]0, +\infty[$ such that

$$(\forall(\mathsf{x}_1, \dots, \mathsf{x}_m) \in \mathsf{H}_1 \times \cdots \times \mathsf{H}_m)(\forall(\mathsf{y}_1, \dots, \mathsf{y}_m) \in \mathsf{H}_1 \times \cdots \times \mathsf{H}_m)$$
$$\sum_{i=1}^m \langle \mathsf{x}_i - \mathsf{y}_i \mid \mathsf{B}_i(\mathsf{x}_1, \dots, \mathsf{x}_m) - \mathsf{B}_i(\mathsf{y}_1, \dots, \mathsf{y}_m)\rangle_{\mathsf{H}_i} \geq \beta \sum_{i=1}^m \big\|\mathsf{B}_i(\mathsf{x}_1, \dots, \mathsf{x}_m) - \mathsf{B}_i(\mathsf{y}_1, \dots, \mathsf{y}_m)\big\|_{\mathsf{H}_i}^2. \tag{3.3}$$

The problem is to

find $x_1 \in \mathcal{W}_1, \dots, x_m \in \mathcal{W}_m$ such that

$$(\forall i \in \{1,\dots,m\}) \quad 0 \in x_i'(t) + \partial \mathsf{f}_i(x_i(t)) + \mathsf{B}_i(x_1(t),\dots,x_m(t)) \ \text{ a.e. on } \left]0,T\right[, \tag{3.4}$$

under the assumption that such functions exist.

Algorithm 3.2 Fix $\varepsilon \in \left]0,\min\{1,\beta\}\right[$, $(\gamma_n)_{n\in\mathbb{N}}$ in $[\varepsilon, 2\beta-\varepsilon]$, and $(\lambda_n)_{n\in\mathbb{N}}$ in $[0,1-\varepsilon]$. Let, for every $n \in \mathbb{N}$ and every $i \in \{1,\dots,m\}$, $y_{i,n}$ be the unique solution in $\mathcal{W}_i$ to the inclusion

$$\frac{x_{i,n}(t) - y_{i,n}(t)}{\gamma_n} - \big(\mathsf{B}_i(x_{1,n}(t),\dots,x_{m,n}(t)) + b_{i,n}(t)\big) \in y_{i,n}'(t) + \partial \mathsf{f}_i(y_{i,n}(t)) + e_{i,n}(t) \ \text{ a.e. on } \left]0,T\right[\tag{3.5}$$

and set

$$x_{i,n+1} = \lambda_{i,n} x_{i,n} + (1-\lambda_{i,n}) y_{i,n} \tag{3.6}$$

where, for every $i \in \{1,\dots,m\}$, the following hold.

(i) $x_{i,0} \in W^{1,2}([0,T];\mathsf{H}_i)$.

(ii) $(b_{i,n})_{n\in\mathbb{N}}$ and $(e_{i,n})_{n\in\mathbb{N}}$ are sequences in $L^2([0,T];\mathsf{H}_i)$ such that

$$\sum_{n\in\mathbb{N}} \sqrt{\int_0^T \|b_{i,n}(t)\|^2_{\mathsf{H}_i} dt} < +\infty \quad \text{and} \quad \sum_{n\in\mathbb{N}} \sqrt{\int_0^T \|e_{i,n}(t)\|^2_{\mathsf{H}_i} dt} < +\infty. \tag{3.7}$$

(iii) $(\lambda_{i,n})_{n\in\mathbb{N}}$ is a sequence in $[0,1[$ such that $\sum_{n\in\mathbb{N}} |\lambda_{i,n} - \lambda_n| < +\infty$.

In (3.5), $b_{i,n}(t)$ models the error tolerated in computing $\mathsf{B}_i(x_{1,n}(t),\dots,x_{m,n}(t))$, while $e_{i,n}(t)$ models the error tolerated in solving the inclusion with respect to $\partial \mathsf{f}_i(y_{i,n}(t))$.

We now examine the weak convergence properties of Algorithm 3.2 (strong convergence conditions can be derived from Theorem 2.10).

Theorem 3.3 *Let $((x_{i,n})_{n\in\mathbb{N}})_{1\le i\le m}$ be sequences generated by Algorithm 3.2. Then, for every $i \in \{1,\dots,m\}$, $(x_{i,n})_{n\in\mathbb{N}}$ converges weakly in $W^{1,2}([0,T];\mathsf{H}_i)$ to a point $x_i \in \mathcal{W}_i$, and $(x_i)_{1\le i\le m}$ is a solution to Problem 3.1.*

Proof. For every $i \in \{1,\dots,m\}$, set $\mathcal{H}_i = L^2([0,T];\mathsf{H}_i)$ and

$$\begin{aligned} A_i \colon \mathcal{H}_i &\to 2^{\mathcal{H}_i} \\ x &\mapsto \begin{cases} \Big\{ u \in \mathcal{H}_i \;\Big|\; u(t) \in x'(t) + \partial \mathsf{f}_i(x(t)) \ \text{ a.e. in } \left]0,T\right[\Big\}, & \text{if } x \in \mathcal{W}_i; \\ \varnothing, & \text{otherwise.} \end{cases} \end{aligned} \tag{3.8}$$

Let us first show that the operators $(A_i)_{1\le i\le m}$ are maximal monotone. For this purpose, let $i \in \{1,\dots,m\}$, and take $(x,u) \in \operatorname{gra} A_i$ and $(y,v) \in \operatorname{gra} A_i$. It follows from (3.8) that, almost everywhere on $]0,T[$, $u(t)-x'(t)\in\partial\mathsf{f}_i(x(t))$ and $v(t)-y'(t)\in\partial\mathsf{f}_i(y(t))$. Therefore, by monotonicity of $\partial\mathsf{f}_i$, we have

$$\int_0^T \big\langle x(t)-y(t)\mid \big(u(t)-x'(t)\big)-\big(v(t)-y'(t)\big)\big\rangle_{\mathsf{H}_i} dt \geq 0. \tag{3.9}$$

Hence,

$$\begin{aligned}
\langle x-y\mid u-v\rangle &= \int_0^T \langle x(t)-y(t)\mid u(t)-v(t)\rangle_{\mathsf{H}_i} dt\\
&= \int_0^T \big\langle x(t)-y(t)\mid \big(u(t)-x'(t)\big)-\big(v(t)-y'(t)\big)\big\rangle_{\mathsf{H}_i} dt\\
&\quad+\int_0^T \big\langle x(t)-y(t)\mid x'(t)-y'(t)\big\rangle_{\mathsf{H}_i} dt\\
&\geq \frac{1}{2}\int_0^T \frac{d\,\|x(t)-y(t)\|^2_{\mathsf{H}_i}}{dt}dt\\
&= \frac{1}{2}\big(\|x(T)-y(T)\|^2_{\mathsf{H}_i}-\|x(0)-y(0)\|^2_{\mathsf{H}_i}\big)\\
&= 0.
\end{aligned} \tag{3.10}$$

Thus, A_i is monotone. To prove maximality, set $\mathsf{g}_i=(1/2)\|\cdot\|^2_{\mathsf{H}_i}+\mathsf{f}_i$. Then $\mathsf{g}_i\in\Gamma_0(\mathsf{H}_i)$ and $\partial\mathsf{g}_i=\mathrm{Id}+\partial\mathsf{f}_i$. Moreover, since $\mathsf{f}_i\in\Gamma_0(\mathsf{H}_i)$, it follows from the Fenchel-Moreau theorem that it is minorized by a continuous affine functional, say $\mathsf{f}_i\geq\langle\cdot\mid\mathsf{v}\rangle_{\mathsf{H}_i}+\eta$ for some $\mathsf{v}\in\mathsf{H}_i$ and $\eta\in\mathbb{R}$. Now, let $\mathsf{y}\in\operatorname{dom}\mathsf{f}_i=\operatorname{dom}\mathsf{g}_i$ and take $(\mathsf{x},\mathsf{u})\in\operatorname{gra}\partial\mathsf{g}_i$. Then (1.10) and Cauchy-Schwarz imply the coercivity property

$$\begin{aligned}
\frac{\langle\mathsf{x}-\mathsf{y}\mid\mathsf{u}\rangle_{\mathsf{H}_i}}{\|\mathsf{x}\|_{\mathsf{H}_i}} &\geq \frac{\mathsf{g}_i(\mathsf{x})-\mathsf{g}_i(\mathsf{y})}{\|\mathsf{x}\|_{\mathsf{H}_i}}\\
&= \frac{\|\mathsf{x}\|_{\mathsf{H}_i}}{2}+\frac{\mathsf{f}_i(\mathsf{x})-\mathsf{g}_i(\mathsf{y})}{\|\mathsf{x}\|_{\mathsf{H}_i}}\\
&\geq \frac{\|\mathsf{x}\|_{\mathsf{H}_i}}{2}-\|\mathsf{v}\|_{\mathsf{H}_i}+\frac{\eta-\mathsf{g}_i(\mathsf{y})}{\|\mathsf{x}\|_{\mathsf{H}_i}}\\
&\to+\infty\quad\text{as}\quad \|\mathsf{x}\|_{\mathsf{H}_i}\to+\infty.
\end{aligned} \tag{3.11}$$

Therefore, [15, Corollaire 3.4] asserts that for every $w\in\mathcal{H}_i$ there exists $z\in\mathcal{W}_i$ such that

$$w(t)\in z'(t)+\partial\mathsf{g}_i(z(t))=z'(t)+z(t)+\partial\mathsf{f}_i(z(t))\ \text{ a.e. on }\ ]0,T[\,, \tag{3.12}$$

i.e., by (3.8), such that $w-z\in A_iz$. This shows that the range of $\mathrm{Id}+A_i$ is $\mathcal{H}_i$ and hence, by Minty's theorem [5, Theorem 3.5.8], that A_i is maximal monotone.

Next, for every $i\in\{1,\ldots,m\}$ and every $(x_1,\ldots,x_m)\in\mathcal{H}_1\times\cdots\times\mathcal{H}_m$, define almost everywhere

$$B_i(x_1,\ldots,x_m)\colon [0,T]\to\mathsf{H}_i\colon t\mapsto\mathsf{B}_i(x_1(t),\ldots,x_m(t)). \tag{3.13}$$

Now let $(x_1,\ldots,x_m)\in\mathcal{H}_1\times\cdots\times\mathcal{H}_m$ and set $(\forall i\in\{1,\ldots,m\})$ $\mathsf{b}_i=\mathsf{B}_i(0,\ldots,0)$. Then it follows from (3.3) and Cauchy-Schwarz that, almost everywhere on $[0,T]$,

$$\begin{aligned}
\beta\sum_{j=1}^m \|\mathsf{B}_j(x_1(t),\ldots,x_m(t))-\mathsf{b}_j\|^2_{\mathsf{H}_j} &\leq \sum_{j=1}^m \langle x_j(t)-\mathsf{0} \mid \mathsf{B}_j(x_1(t),\ldots,x_m(t))-\mathsf{b}_j\rangle_{\mathsf{H}_j}\\
&\leq \sum_{j=1}^m \|x_j(t)\|_{\mathsf{H}_j}\,\|\mathsf{B}_j(x_1(t),\ldots,x_m(t))-\mathsf{b}_j\|_{\mathsf{H}_j}\\
&\leq \sqrt{\sum_{j=1}^m \|x_j(t)\|^2_{\mathsf{H}_j}}\sqrt{\sum_{j=1}^m \|\mathsf{B}_j(x_1(t),\ldots,x_m(t))-\mathsf{b}_j\|^2_{\mathsf{H}_j}}. \qquad (3.14)
\end{aligned}$$

Therefore, for every $i\in\{1,\ldots,m\}$,

$$\begin{aligned}
\|B_i(x_1,\ldots,x_m)(t)\|^2_{\mathsf{H}_i} &\leq 2\big(\|\mathsf{b}_i\|^2_{\mathsf{H}_i}+\|B_i(x_1,\ldots,x_m)(t)-\mathsf{b}_i\|^2_{\mathsf{H}_i}\big)\\
&\leq 2\Bigg(\|\mathsf{b}_i\|^2_{\mathsf{H}_i}+\sum_{j=1}^m\|\mathsf{B}_j(x_1(t),\ldots,x_m(t))-\mathsf{b}_j\|^2_{\mathsf{H}_j}\Bigg)\\
&\leq 2\Bigg(\|\mathsf{b}_i\|^2_{\mathsf{H}_i}+\frac{1}{\beta^2}\sum_{j=1}^m\|x_j(t)\|^2_{\mathsf{H}_j}\Bigg)\ \text{ a.e. on }]0,T[\,, \qquad (3.15)
\end{aligned}$$

which yields

$$\int_0^T \|B_i(x_1,\ldots,x_m)(t)\|^2_{\mathsf{H}_i}dt \leq 2T\|\mathsf{b}_i\|^2_{\mathsf{H}_i}+\frac{2}{\beta^2}\sum_{j=1}^m\|x_j\|^2, \qquad (3.16)$$

so that we can now claim that $B_i\colon \mathcal{H}_1\times\cdots\times\mathcal{H}_m\to L^2([0,T];\mathsf{H}_i)=\mathcal{H}_i$. In addition, upon integrating, we derive from (3.3) and (3.13) that, for every $(y_1,\ldots,y_m)\in\mathcal{H}_1\times\cdots\times\mathcal{H}_m$,

$$\sum_{i=1}^m\langle x_i-y_i \mid B_i(x_1,\ldots,x_m)-B_i(y_1,\ldots,y_m)\rangle \geq \beta\sum_{i=1}^m\big\|B_i(x_1,\ldots,x_m)-B_i(y_1,\ldots,y_m)\big\|^2. \qquad (3.17)$$

We have thus established (1.6).

Let us now make the connection between Algorithm 3.2 and Algorithm 2.2. For every $n\in\mathbb{N}$ and every $i\in\{1,\ldots,m\}$, it follows from (3.5), (3.8), (3.13), and the maximal monotonicity of A_i that $y_{i,n}$ is uniquely defined and can be expressed as

$$y_{i,n}=J_{\gamma_n A_i}\Big(x_{i,n}-\gamma_n\big(B_i(x_{1,n},\ldots,x_{m,n})+b_{i,n}\big)\Big)+a_{i,n}\,, \qquad (3.18)$$

where

$$\begin{aligned}
a_{i,n}=J_{\gamma_n A_i}&\Big(-\gamma_n e_{i,n}+x_{i,n}-\gamma_n\big(B_i(x_{1,n},\ldots,x_{m,n})+b_{i,n}\big)\Big)\\
&-J_{\gamma_n A_i}\Big(x_{i,n}-\gamma_n\big(B_i(x_{1,n},\ldots,x_{m,n})+b_{i,n}\big)\Big). \qquad (3.19)
\end{aligned}$$

We therefore deduce from (3.6) that

$$x_{i,n+1}=\lambda_{i,n}x_{i,n}+(1-\lambda_{i,n})\Big(J_{\gamma_n A_i}\Big(x_{i,n}-\gamma_n\big(B_i(x_{1,n},\ldots,x_{m,n})+b_{i,n}\big)\Big)+a_{i,n}\Big). \qquad (3.20)$$

Thus, (3.20) derives from (2.3) with $A_{i,n} \equiv A_i$ and $B_{i,n} \equiv B_i$. On the other hand, for every $i \in \{1,\dots,m\}$, by nonexpansivity of the operators $(J_{\gamma_n A_i})_{n\in\mathbb{N}}$, we deduce from (3.19) and (3.7) that

$$\sum_{n\in\mathbb{N}} \|a_{i,n}\| \leq \sum_{n\in\mathbb{N}} \gamma_n \|e_{i,n}\| \leq 2\beta \sum_{n\in\mathbb{N}} \|e_{i,n}\| < +\infty. \tag{3.21}$$

As a result, all the hypotheses of Algorithm 2.2 are satisfied and hence Theorem 2.9 asserts that, for every $i \in \{1,\dots,m\}$, $(x_{i,n})_{n\in\mathbb{N}}$ converges weakly in $\mathcal{H}_i = L^2([0,T];\mathsf{H}_i)$ to a point x_i, and $(x_i)_{1\leq i\leq m}$ satisfies

$$(\forall i \in \{1,\dots,m\}) \quad 0 \in A_i x_i + B_i(x_1,\dots,x_m). \tag{3.22}$$

Accordingly,

$$\sigma = \max_{1\leq i\leq m} \sup_{n\in\mathbb{N}} \|x_{i,n}\| < +\infty \tag{3.23}$$

and $(\forall i \in \{1,\dots,m\})\ x_i \in \operatorname{dom} A_i \subset \mathcal{W}_i$. Moreover since, in view of (3.8) and (3.13), (3.22) reduces to (3.4), $(x_i)_{1\leq i\leq m}$ is a solution to Problem 3.1.

To complete the proof, let $i \in \{1,\dots,m\}$. To show that $(x_{i,n})_{n\in\mathbb{N}}$ converges weakly to x_i in $W^{1,2}([0,T];\mathsf{H}_i)$, it remains to show that $(x'_{i,n})_{n\in\mathbb{N}}$ converges weakly to x'_i in $L^2([0,T];\mathsf{H}_i)$. We first observe that $(x_{i,n})_{n\in\mathbb{N}}$ lies in $W^{1,2}([0,T];\mathsf{H}_i)$. Indeed, it follows from (3.8) that

$$(\forall n \in \mathbb{N})(\forall z \in \mathcal{H}_i) \quad J_{\gamma_n A_i} z \in \operatorname{dom}(\gamma_n A_i) \subset \mathcal{W}_i \subset W^{1,2}([0,T];\mathsf{H}_i). \tag{3.24}$$

As a result, we deduce from (3.19) that $(a_{i,n})_{n\in\mathbb{N}}$ lies in $W^{1,2}([0,T];\mathsf{H}_i)$. On the other hand, by construction, $(y_{i,n})_{n\in\mathbb{N}}$ lies in $\mathcal{W}_i \subset W^{1,2}([0,T];\mathsf{H}_i)$. In view of (3.6) and (i) in Algorithm 3.2, $(x_{i,n})_{n\in\mathbb{N}}$ is therefore in $W^{1,2}([0,T];\mathsf{H}_i)$. Next, let us show that $(x'_{i,n})_{n\in\mathbb{N}}$ is bounded in $L^2([0,T];\mathsf{H}_i)$. To this end, let $n \in \mathbb{N}$ and set

$$w_{i,n}(t) = \frac{x_{i,n}(t) - y_{i,n}(t)}{\gamma_n} - \mathsf{B}_i(x_{1,n}(t),\dots,x_{m,n}(t)) - b_{i,n}(t) - y'_{i,n}(t) - e_{i,n}(t) \ \text{ a.e. on }]0,T[\,. \tag{3.25}$$

Then we derive from (3.5) that

$$w_{i,n}(t) \in \partial \mathsf{f}_i(y_{i,n}(t)) \ \text{ a.e. on }]0,T[\,. \tag{3.26}$$

Hence, since $w_{i,n} \in \mathcal{H}_i$, it follows from [15, Lemme 3.3] that

$$\frac{d(\mathsf{f}_i \circ y_{i,n})(t)}{dt} = \langle w_{i,n}(t) \mid y'_{i,n}(t)\rangle_{\mathsf{H}_i} \ \text{ a.e. on }]0,T[\,. \tag{3.27}$$

On the other hand, since $y_{i,n} \in \mathcal{W}_i$, we have $y_{i,n}(T) = y_{i,n}(0)$. Therefore

$$\int_0^T \langle w_{i,n}(t) \mid y'_{i,n}(t)\rangle_{\mathsf{H}_i} dt = \int_0^T \frac{d(\mathsf{f}_i \circ y_{i,n})(t)}{dt} dt = \mathsf{f}_i(y_{i,n}(T)) - \mathsf{f}_i(y_{i,n}(0)) = 0 \tag{3.28}$$

and, furthermore,

$$\int_0^T \langle y_{i,n}(t) \mid y'_{i,n}(t)\rangle_{\mathsf{H}_i} dt = \frac{1}{2}\int_0^T \frac{d\|y_{i,n}(t)\|^2_{\mathsf{H}_i}}{dt} dt = \frac{\|y_{i,n}(T)\|^2_{\mathsf{H}_i} - \|y_{i,n}(0)\|^2_{\mathsf{H}_i}}{2} = 0. \tag{3.29}$$

We deduce from (3.28), (3.25), and (3.29) that

$$\begin{aligned} 0 = &\int_0^T \left\langle \frac{x_{i,n}(t)}{\gamma_n} \,\middle|\, y'_{i,n}(t)\right\rangle_{\mathsf{H}_i} dt - \int_0^T \langle \mathsf{B}_i(x_{1,n}(t),\dots,x_{m,n}(t)) \mid y'_{i,n}(t)\rangle_{\mathsf{H}_i} dt \\ &- \int_0^T \langle b_{i,n}(t) \mid y'_{i,n}(t)\rangle_{\mathsf{H}_i} dt - \int_0^T \|y'_{i,n}(t)\|^2_{\mathsf{H}_i} dt - \int_0^T \langle e_{i,n}(t) \mid y'_{i,n}(t)\rangle_{\mathsf{H}_i} dt. \end{aligned} \tag{3.30}$$

Thus, using Cauchy-Schwarz, the inequality $\gamma_n \geq \varepsilon$, and (3.13), we obtain

$$\|y'_{i,n}\|^2 \leq \left(\frac{1}{\varepsilon}\|x_{i,n}\| + \|B_i(x_{1,n},\ldots,x_{m,n})\| + \|b_{i,n}\| + \|e_{i,n}\|\right)\|y'_{i,n}\|. \tag{3.31}$$

In turn, it follows from (3.6) that

$$\|x'_{i,n+1}\| \leq \lambda_{i,n}\|x'_{i,n}\| + (1-\lambda_{i,n})\left(\frac{1}{\varepsilon}\|x_{i,n}\| + \|B_i(x_{1,n},\ldots,x_{m,n})\| + \|b_{i,n}\| + \|e_{i,n}\|\right). \tag{3.32}$$

On the other hand, arguing as in (3.16), we derive from (3.23) that

$$\|B_i(x_{1,n},\ldots,x_{m,n})\| \leq \sqrt{2T\|\mathsf{b}_i\|^2_{\mathsf{H}_i} + \frac{2m\sigma^2}{\beta^2}} \leq \sqrt{2T}\|\mathsf{b}_i\|_{\mathsf{H}_i} + \sqrt{2m}\,\frac{\sigma}{\beta}. \tag{3.33}$$

Hence, using (ii) in Algorithm 3.2, we derive by induction from (3.32) that

$$\|x'_{i,n}\| \leq \max\left\{\|x'_{i,0}\|, \frac{\sigma}{\varepsilon} + \sqrt{2T}\|\mathsf{b}_i\|_{\mathsf{H}_i} + \sqrt{2m}\,\frac{\sigma}{\beta} + \sup_{k\in\mathbb{N}}\big(\|b_{i,k}\| + \|e_{i,k}\|\big)\right\}. \tag{3.34}$$

This shows the boundedness of $(x'_{i,n})_{n\in\mathbb{N}}$ in $L^2([0,T];\mathsf{H}_i)$. Now let z be the weak limit in $L^2([0,T];\mathsf{H}_i)$ of an arbitrary weakly convergent subsequence of $(x'_{i,n})_{n\in\mathbb{N}}$. Since $(x_{i,n})_{n\in\mathbb{N}}$ converges weakly in $L^2([0,T];\mathsf{H}_i)$ to x_i, it therefore follows from [48, Proposition 23.19] that $z = x'_i$. In turn, this shows that $(x'_{i,n})_{n\in\mathbb{N}}$ converges weakly in $L^2([0,T];\mathsf{H}_i)$ to x'_i. □

4 The variational case

We study a special case of Problem 1.1 which yields a variational formulation that extends (1.5).

Recall that, for every $f \in \Gamma_0(\mathcal{H})$ and every $x \in \mathcal{H}$, the function $y \mapsto f(y) + \|x-y\|^2/2$ admits a unique minimizer, which is denoted by $\operatorname{prox}_f x$. The proximity operator thus defined can be expressed as $\operatorname{prox}_f = J_{\partial f}$ [36].

Problem 4.1 Let $(\mathcal{H}_i)_{1\leq i\leq m}$ and $(\mathcal{G}_k)_{1\leq k\leq p}$ be real Hilbert spaces. For every $i \in \{1,\ldots,m\}$, let $f_i \in \Gamma_0(\mathcal{H}_i)$ and, for every $k \in \{1,\ldots,p\}$, let $\tau_k \in \left]0,+\infty\right[$, let $\varphi_k \colon \mathcal{G}_k \to \mathbb{R}$ be a differentiable convex function with a τ_k-Lipschitz-continuous gradient, and let $L_{ki} \colon \mathcal{H}_i \to \mathcal{G}_k$ be linear and bounded. It is assumed that $\min_{1\leq k\leq p}\sum_{i=1}^m \|L_{ki}\|^2 > 0$. The problem is to

$$\underset{x_1\in\mathcal{H}_1,\ldots,x_m\in\mathcal{H}_m}{\text{minimize}} \quad \sum_{i=1}^m f_i(x_i) + \sum_{k=1}^p \varphi_k\left(\sum_{i=1}^m L_{ki}x_i\right), \tag{4.1}$$

under the assumption that solutions exist.

Algorithm 4.2 Set

$$\beta = \frac{1}{p\max\limits_{1\leq k\leq p}\tau_k\displaystyle\sum_{i=1}^m\|L_{ki}\|^2}. \tag{4.2}$$

Fix $\varepsilon \in \left]0, \min\{1, \beta\}\right[$, $(\gamma_n)_{n\in\mathbb{N}}$ in $[\varepsilon, 2\beta-\varepsilon]$, $(\lambda_n)_{n\in\mathbb{N}}$ in $[0, 1-\varepsilon]$, and $(x_{i,0})_{1\le i\le m} \in \mathcal{H}_1 \times \cdots \times \mathcal{H}_m$. Set, for every $n \in \mathbb{N}$,

$$\begin{cases} x_{1,n+1} = \lambda_{1,n} x_{1,n} + \\ \qquad (1-\lambda_{1,n})\Bigg(\operatorname{prox}_{\gamma_n f_{1,n}}\bigg(x_{1,n} - \gamma_n\bigg(\sum_{k=1}^p L_{k1}^* \nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj} x_{j,n}\bigg) + b_{1,n}\bigg)\bigg) + a_{1,n}\Bigg), \\ \quad \vdots \\ x_{m,n+1} = \lambda_{m,n} x_{m,n} + \\ \qquad (1-\lambda_{m,n})\Bigg(\operatorname{prox}_{\gamma_n f_{m,n}}\bigg(x_{m,n} - \gamma_n\bigg(\sum_{k=1}^p L_{km}^* \nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj} x_{j,n}\bigg) + b_{m,n}\bigg)\bigg) + a_{m,n}\Bigg), \end{cases} \tag{4.3}$$

where, for every $i \in \{1, \ldots, m\}$, the following hold.

(i) $(f_{i,n})_{n\in\mathbb{N}}$ are functions in $\Gamma_0(\mathcal{H}_i)$ such that

$$(\forall \rho \in \left]0, +\infty\right[) \quad \sum_{n\in\mathbb{N}} \sup_{\|y\|\le\rho} \|\operatorname{prox}_{\gamma_n f_{i,n}} y - \operatorname{prox}_{\gamma_n f_i} y\| < +\infty. \tag{4.4}$$

(ii) $(a_{i,n})_{n\in\mathbb{N}}$ and $(b_{i,n})_{n\in\mathbb{N}}$ are sequences in $\mathcal{H}_i$ such that $\sum_{n\in\mathbb{N}} \|a_{i,n}\| < +\infty$ and $\sum_{n\in\mathbb{N}} \|b_{i,n}\| < +\infty$.

(iii) $(\lambda_{i,n})_{n\in\mathbb{N}}$ is a sequence in $\left[0, 1\right[$ such that $\sum_{n\in\mathbb{N}} |\lambda_{i,n} - \lambda_n| < +\infty$.

We now turn our attention to the asymptotic behavior of Algorithm 4.2 (strong convergence conditions can be derived from Theorem 2.10).

Theorem 4.3 *Let $((x_{i,n})_{n\in\mathbb{N}})_{1\le i\le m}$ be sequences generated by Algorithm 4.2. Then, for every $i \in \{1, \ldots, m\}$, $(x_{i,n})_{n\in\mathbb{N}}$ converges weakly to a point $x_i \in \mathcal{H}_i$, and $(x_i)_{1\le i\le m}$ is a solution to Problem 4.1.*

Proof. Problem 4.1 is a special case of Problem 1.1 where, for every $i \in \{1, \ldots, m\}$,

$$A_i = \partial f_i \quad \text{and} \quad B_i \colon (x_j)_{1\le j\le m} \mapsto \sum_{k=1}^p L_{ki}^* \nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj} x_j\bigg). \tag{4.5}$$

Indeed, define $\boldsymbol{\mathcal{H}}$ as in the proof of Theorem 2.9 and set

$$\boldsymbol{f} \colon \boldsymbol{\mathcal{H}} \to \left]-\infty, +\infty\right] : (x_i)_{1\le i\le m} \mapsto \sum_{i=1}^m f_i(x_i) \tag{4.6}$$

and

$$\boldsymbol{g} \colon \boldsymbol{\mathcal{H}} \to \mathbb{R} \colon (x_i)_{1\le i\le m} \mapsto \sum_{k=1}^p \varphi_k\bigg(\sum_{i=1}^m L_{ki} x_i\bigg). \tag{4.7}$$

Then $\boldsymbol{f}$ and $\boldsymbol{g}$ are in $\Gamma_0(\boldsymbol{\mathcal{H}})$ and it follows from Fermat's rule and elementary subdifferential calculus that, for every $(x_1,\ldots,x_m)\in\boldsymbol{\mathcal{H}}$,

$$\begin{aligned}(x_1,\ldots,x_m)\ \text{solves (4.1)} \quad &\Leftrightarrow \quad (0,\ldots,0)\in\partial(\boldsymbol{f}+\boldsymbol{g})(x_1,\ldots,x_m)\\ &\Leftrightarrow \quad (0,\ldots,0)\in\partial\boldsymbol{f}(x_1,\ldots,x_m)+\nabla\boldsymbol{g}(x_1,\ldots,x_m)\\ &\Leftrightarrow \quad (\forall i\in\{1,\ldots,m\})\ \ 0\in\partial f_i(x_i)+\sum_{k=1}^p L_{ki}^*\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}x_j\bigg)\\ &\Leftrightarrow \quad (\forall i\in\{1,\ldots,m\})\ \ 0\in A_ix_i+B_i(x_1,\ldots,x_m). \end{aligned}\tag{4.8}$$

Next, let us show that the family $(B_i)_{1\le i\le m}$ in (4.5) satisfies (1.6) with β as in (4.2). First, Lemma 2.6 asserts that, for every $k\in\{1,\ldots,p\}$, $\nabla\varphi_k$ is τ_k^{-1}-cocoercive. Hence, for every $(x_1,\ldots,x_m)\in\boldsymbol{\mathcal{H}}$ and every $(y_1,\ldots,y_m)\in\boldsymbol{\mathcal{H}}$, it follows from (4.5), (4.2), and the convexity of $\|\cdot\|^2$ that

$$\begin{aligned}&\sum_{i=1}^m\langle x_i-y_i\mid B_i(x_1,\ldots,x_m)-B_i(y_1,\ldots,y_m)\rangle\\
&\quad=\sum_{i=1}^m\sum_{k=1}^p\bigg\langle x_i-y_i\,\bigg|\,L_{ki}^*\bigg(\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}x_j\bigg)-\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}y_j\bigg)\bigg)\bigg\rangle\\
&\quad=\sum_{i=1}^m\sum_{k=1}^p\bigg\langle L_{ki}(x_i-y_i)\,\bigg|\,\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}x_j\bigg)-\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}y_j\bigg)\bigg\rangle\\
&\quad=\sum_{k=1}^p\bigg\langle\sum_{i=1}^m L_{ki}x_i-\sum_{i=1}^m L_{ki}y_i\,\bigg|\,\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}x_j\bigg)-\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}y_j\bigg)\bigg\rangle\\
&\quad\ge\sum_{k=1}^p\frac{1}{\tau_k}\bigg\|\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}x_j\bigg)-\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}y_j\bigg)\bigg\|^2\\
&\quad=\sum_{k=1}^p\frac{1}{\tau_k\sum_{i=1}^m\|L_{ki}\|^2}\sum_{i=1}^m\|L_{ki}\|^2\bigg\|\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}x_j\bigg)-\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}y_j\bigg)\bigg\|^2\\
&\quad\ge p\beta\sum_{k=1}^p\sum_{i=1}^m\|L_{ki}\|^2\bigg\|\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}x_j\bigg)-\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}y_j\bigg)\bigg\|^2\\
&\quad\ge\beta\sum_{i=1}^m p\sum_{k=1}^p\bigg\|L_{ki}^*\bigg(\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}x_j\bigg)-\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}y_j\bigg)\bigg)\bigg\|^2\\
&\quad\ge\beta\sum_{i=1}^m\bigg\|\sum_{k=1}^p L_{ki}^*\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}x_j\bigg)-\sum_{k=1}^p L_{ki}^*\nabla\varphi_k\bigg(\sum_{j=1}^m L_{kj}y_j\bigg)\bigg\|^2.\end{aligned}\tag{4.9}$$

This shows that (1.6) holds. Furthermore, upon setting

$$(\forall i\in\{1,\ldots,m\})(\forall n\in\mathbb{N})\quad A_{i,n}=\partial f_{i,n}\quad\text{and}\quad B_{i,n}=B_i,\tag{4.10}$$

we deduce from (4.4) that Algorithm 4.2 is a particular case of Algorithm 2.2. Altogether, Theorem 4.3 follows from Theorem 2.9. □

Here are a couple of applications of Problem 4.1.

Example 4.4 (network flows) Consider a network with M links indexed by $j \in \{1, \dots, M\}$ and N paths indexed by $l \in \{1, \dots, N\}$, linking a subset of Q origin-destination node pairs indexed by $k \in \{1, \dots, Q\}$. There are m types of users indexed by $i \in \{1, \dots, m\}$ transiting on the network. For every $i \in \{1, \dots, m\}$ and $l \in \{1, \dots, N\}$, let $\xi_{il} \in \mathbb{R}$ be the flux of user i on path l and let $x_i = (\xi_{il})_{1 \leq l \leq N}$ be the flow associated with user i. A standard problem in traffic theory is to find a Wardrop equilibrium [45] of the network, i.e., flows $(x_i)_{1 \leq i \leq m}$ such that the costs in all paths actually used are equal and less than those a single user would face on any unused path. Such an equilibrium can be obtained by solving the variational problem [13, 37, 41]

$$\underset{x_1 \in C_1, \dots, x_m \in C_m}{\text{minimize}} \quad \sum_{j=1}^{M} \int_0^{h_j(x_1, \dots, x_m)} \phi_j(h) dh, \tag{4.11}$$

where $\phi_j \colon \mathbb{R} \to [0, +\infty[$ is a strictly increasing τ-Lipschitz continuous function modeling the cost of transiting on link j and $h_j(x_1, \dots, x_m)$ is the total flow through link j, which can be expressed as $h_j(x_1, \dots, x_m) = \sum_{i=1}^{m} (L x_i)^\top e_j$, where e_j is the jth canonical basis vector of $\mathbb{R}^M$ and L is an $M \times N$ binary matrix with jlth entry equal to 1 or 0, according as link j belongs to path l or not. Furthermore, each closed and convex constraint set C_i in (4.11) is defined as $C_i = \big\{ (\eta_l)_{1 \leq l \leq N} \in [0, +\infty[^N \;\big|\; (\forall k \in \{1, \dots, Q\}) \; \sum_{l \in N_k} \eta_l = \delta_{ik} \big\}$, where $\varnothing \neq N_k \subset \{1, \dots, N\}$ is the set of paths linking the pair k and $\delta_{ik} \in [0, +\infty[$ is the flow of user i that must transit from the origin to the destination of pair k (for more details on network flows, see [40, 41]). Upon setting

$$\varphi_1 \colon \mathbb{R}^M \to \mathbb{R} \colon (\nu_j)_{1 \leq j \leq M} \mapsto \sum_{j=1}^{M} \int_0^{\nu_j} \phi_j(h) dh, \tag{4.12}$$

problem (4.11) can be written as

$$\underset{x_1 \in \mathbb{R}^N, \dots, x_m \in \mathbb{R}^N}{\text{minimize}} \quad \sum_{i=1}^{m} \iota_{C_i}(x_i) + \varphi_1\Bigg(\sum_{i=1}^{m} L x_i \Bigg). \tag{4.13}$$

Since φ_1 is strictly convex and differentiable with a τ-Lipschitz-continuous gradient, (4.13) is a particular instance of Problem 4.1 with $p = 1$, $\mathcal{G}_1 = \mathbb{R}^M$ and $(\forall i \in \{1, \dots, m\})$ $\mathcal{H}_i = \mathbb{R}^N$, $f_i = \iota_{C_i}$, and $L_{1i} = L$. Accordingly, Theorem 4.3 asserts that (4.13) can be solved by Algorithm 4.2 which, with the choice of parameters $\gamma_n \equiv \gamma \in \,]0, 2/\tau[$, $\lambda_{i,n} \equiv 0$, $\lambda_n \equiv 0$, $a_{i,n} \equiv 0$, and $b_{i,n} \equiv 0$, yields

$$(\forall i \in \{1, \dots, m\}) \quad x_{i,n+1} = P_{C_i}\Big(x_{i,n} - \gamma L^\top \big(\phi_1(\rho_{1,n}), \dots, \phi_M(\rho_{M,n}) \big) \Big), \tag{4.14}$$

where $(\rho_{1,n}, \dots, \rho_{M,n}) = \sum_{j=1}^{m} L x_{j,n}$. In the special case when $m = 1$ the algorithm described in (4.14) is proposed in [14]. Let us note that, as an alternative to (4.12), we can consider the function

$$\varphi_1 \colon \mathbb{R}^M \to \mathbb{R} \colon (\nu_j)_{1 \leq j \leq M} \mapsto \sum_{j=1}^{M} \nu_j \phi_j(\nu_j), \tag{4.15}$$

under suitable assumptions on $(\phi_j)_{1 \leq j \leq M}$. In this case, (4.13) reduces to the problem of finding the social optimum in the network [41], that is

$$\underset{x_1 \in C_1, \dots, x_m \in C_m}{\text{minimize}} \quad \sum_{j=1}^{M} h_j(x_1, \dots, x_m) \phi_j\big(h_j(x_1, \dots, x_m) \big), \tag{4.16}$$

which can also be solved with Algorithm 4.2.

Example 4.5 (best approximation) The convex feasibility problem is to find a point in the intersection of closed convex subsets $(C_i)_{1\leq i\leq m}$ of a real Hilbert space $\mathcal{H}$ [10, 21]. This problem arises in many applications in engineering and the physical sciences [17, 19]. In many instances, the intersection of the sets $(C_i)_{1\leq i\leq m}$ may turn out to be empty and a relaxation of this problem in the presence of a hard constraint represented by C_1 is to [23]

$$\underset{x_1\in C_1}{\text{minimize}} \;\; \frac{1}{2}\sum_{i=2}^{m}\omega_i d_{C_i}^2(x_1), \tag{4.17}$$

where $(\omega_i)_{2\leq i\leq m}$ are strictly positive weights such that $\max_{2\leq i\leq m}\omega_i = 1$. We assume that this problem admits at least one solution, as is the case when one of the sets in $(C_i)_{1\leq i\leq m}$ is bounded [23, Proposition 4]. Since, for every $i\in\{2,\ldots,m\}$ and every $x_1\in C_1$, $d_{C_i}^2(x_1)=\min_{x_i\in C_i}\|x_1-x_i\|^2$, (4.17) can be reformulated as

$$\underset{x_1\in C_1,\ldots,\,x_m\in C_m}{\text{minimize}} \;\; \frac{1}{2}\sum_{k=1}^{m-1}\omega_{k+1}\|x_1-x_{k+1}\|^2. \tag{4.18}$$

This is a special instance of Problem 4.1 with $p=m-1$ and, for every $i\in\{1,\ldots,m\}$, $f_i=\iota_{C_i}$ and

$$(\forall k\in\{1,\ldots,m-1\})\quad \varphi_k=\frac{\omega_{k+1}}{2}\|\cdot\|^2 \quad\text{and}\quad L_{ki}=\begin{cases}\mathrm{Id}\,, & \text{if } i=1;\\ -\mathrm{Id}\,, & \text{if } i=k+1;\\ 0, & \text{otherwise.}\end{cases} \tag{4.19}$$

We can derive from Algorithm 4.2 an algorithm which, by Theorem 4.3, generates orbits that are guaranteed to converge weakly to a solution to (4.18). Indeed, in this case, (4.2) yields $\beta = 1/(2(m-1))$. For example, upon setting $\gamma_n\equiv\gamma\in\,]0,1/(m-1)[$, $\lambda_n\equiv 0$, $\lambda_{i,n}\equiv 0$, $a_{i,n}\equiv 0$, $b_{i,n}\equiv 0$, and $f_{i,n}=\iota_{C_i}$ for simplicity, Algorithm 4.2 becomes

$$\begin{cases} x_{1,n+1}=P_{C_1}\big((1-\gamma\sum_{i=2}^m\omega_i)x_{1,n}+\gamma\sum_{i=2}^m\omega_i x_{i,n}\big)\\ (\forall i\in\{2,\ldots,m\})\quad x_{i,n+1}=P_{C_i}\big(\gamma\omega_i x_{1,n}+(1-\gamma\omega_i)x_{i,n}\big).\end{cases} \tag{4.20}$$

In the particular case when $m=2$ and $\gamma=1/2$, then $\omega_2=1$, (4.18) is equivalent to finding a best approximation pair relative to (C_1,C_2) [9, 11], and (4.20) reduces to

$$\begin{cases} x_{1,n+1}=P_{C_1}\big((x_{1,n}+x_{2,n})/2\big)\\ x_{2,n+1}=P_{C_2}\big((x_{1,n}+x_{2,n})/2\big).\end{cases} \tag{4.21}$$

References

[1] F. Acker and M. A. Prestel, Convergence d'un schéma de minimisation alternée, *Ann. Fac. Sci. Toulouse V. Sér. Math.*, vol. 2, pp. 1–9, 1980.

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

[3] 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.

[4] H. Attouch, P. Redont, and A. Soubeyran, A new class of alternating proximal minimization algorithms with costs-to-move, *SIAM J. Optim.,* vol. 18, pp. 1061–1081, 2007.

[5] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis,* Birkhäuser, Boston, MA, 1990.

[6] A. Auslender, *Problèmes de Minimax via l'Analyse Convexe et les Inégalités Variationnelles: Théorie et Algorithmes,* Lecture Notes in Econom. and Math. Systems, vol. 77, Springer-Verlag, New York, 1972.

[7] 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.

[8] H. H. Bauschke and J. M. Borwein, On the convergence of von Neumann's alternating projection algorithm for two sets, *Set-Valued Anal.,* vol. 1, pp. 185–212, 1993.

[9] H. H. Bauschke and J. M. Borwein, Dykstra's alternating projection algorithm for two sets, *J. Approx. Theory,* vol. 79, pp. 418–443, 1994.

[10] H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Rev.,* vol. 38, pp. 367–426, 1996.

[11] H. H. Bauschke, P. L. Combettes, and D. R. Luke, Finding best approximation pairs relative to two closed convex sets in Hilbert spaces, *J. Approx. Theory,* vol. 127, pp. 178–192, 2004.

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

[13] M. Beckmann, C. McGuire, and C. Winsten, *Studies in Economics of Transportation,* Yale University Press, New Haven, CT, 1956.

[14] D. P. Bertsekas and E. M. Gafni, Projection methods for variational inequalities with application to the traffic assignment problem, *Math. Program. Stud.*, vol. 17, pp. 139–159, 1982.

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

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

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

[18] W. Cheney and A. A. Goldstein, Proximity maps for convex sets, *Proc. Amer. Math. Soc.,* vol. 10, pp. 448–450, 1959.

[19] P. L. Combettes, The foundations of set theoretic estimation, *Proc. IEEE,* vol. 81, pp. 182–208, 1993.

[20] 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.

[21] P. L. Combettes, Hilbertian convex feasibility problem: Convergence of projection methods, *Appl. Math. Optim.,* vol. 35, pp. 311–330, 1997.

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

[23] P. L. Combettes and P. Bondon, Hard-constrained inconsistent signal feasibility problems, *IEEE Trans. Signal Process.,* vol. 47, pp. 2460–2468, 1999.

[24] P. L. Combettes and S. A. Hirstoaga, Approximating curves for nonexpansive and monotone operators, *J. Convex Anal.,* vol. 13, pp. 633–646, 2006.

[25] P. L. Combettes and T. Pennanen, Proximal methods for cohypomonotone operators, *SIAM J. Control Optim.,* vol. 43, pp. 731–742, 2004.

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

[27] J. C. Dunn, Convexity, monotonicity, and gradient processes in Hilbert space, *J. Math. Anal. Appl.,* vol. 53, pp. 145–158, 1976.

[28] M. Goldburg and R. J. Marks II, Signal synthesis in the presence of an inconsistent set of constraints, *IEEE Trans. Circuits Syst.,* vol. 32, pp. 647–663, 1985.

[29] L. G. Gubin, B. T. Polyak, and E. V. Raik, The method of projections for finding the common point of convex sets, *U.S.S.R. Comput. Math. Math. Phys.*, vol. 7, pp. 1–24, 1967.

[30] A. Haraux, *Nonlinear Evolution Equations: Global Behavior of Solutions,* Lecture Notes in Math., vol. 841, Springer-Verlag, New York, 1981.

[31] B. He, L.-Z. Liao, and S. Wang, Self-adaptive operator splitting methods for monotone variational inequalities, *Numer. Math.,* vol. 94, pp. 715–737, 2003.

[32] A. Kaplan and R. Tichatschke, Proximal point approach and approximation of variational inequalities, *SIAM J. Control Optim.,* vol. 39, pp. 1136–1159, 2000.

[33] N. Lehdili and B. Lemaire, The barycentric proximal method, *Comm. Appl. Nonlinear Anal.,* vol. 6, pp. 29–47, 1999.

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

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

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

[37] M. Patriksson, *The Traffic Assignment Problem: Models and Methods,* Utrecht, The Netherlands, 1994.

[38] W. V. Petryshyn, Construction of fixed points of demicompact mappings in Hilbert space, *J. Math. Anal. Appl.,* vol. 14, pp. 276–284, 1966.

[39] B. T. Polyak, *Introduction to Optimization,* Optimization Software Inc., New York, 1987.

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

[41] Y. Sheffi, *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods,* Prentice-Hall, Englewood Cliffs, NJ, 1985.

[42] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations*, Mathematical Surveys and Monographs, vol. 49, Amer. Math. Soc., Providence, RI, 1997.

[43] M. Sibony, Méthodes itératives pour les équations et inéquations aux dérivées partielles non linéaires de type monotone, *Calcolo,* vol. 7, pp. 65–183, 1970.

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

[45] J. G. Wardrop, Some theoretical aspects of road traffic research, *Proc. Inst. Civil Eng. II,* vol. 1, pp. 325–378, 1952.

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

[47] E. Zeidler, *Nonlinear Functional Analysis and Its Applications I*, Springer-Verlag, New York, 1993.

[48] E. Zeidler, *Nonlinear Functional Analysis and Its Applications II*, Springer-Verlag, New York, 1990.

[49] D. L. Zhu and P. Marcotte, Co-coercivity and its role in the convergence of iterative schemes for solving variational inequalities, *SIAM J. Optim.,* vol. 6, pp. 714–726, 1996.