Forward-Backward and Tseng's Type Penalty Schemes for Monotone Inclusion Problems

Radu Ioan Boţ · Ernö Robert Csetnek

Abstract We deal with monotone inclusion problems of the form $0 \in Ax + Dx + N_C(x)$ in real Hilbert spaces, where A is a maximally monotone operator, D a cocoercive operator and C the nonempty set of zeros of another cocoercive operator. We propose a forward-backward penalty algorithm for solving this problem which extends the one proposed by Attouch et al. (SIAM J. Optim. 21(4): 1251–1274, 2011). The condition which guarantees the weak ergodic convergence of the sequence of iterates generated by the proposed scheme is formulated by means of the Fitzpatrick function associated to the maximally monotone operator that describes the set C. In the second part we introduce a forward-backward-forward algorithm for monotone inclusion problems having the same structure, but this time by replacing the cocoercivity hypotheses with Lipschitz continuity conditions. The latter penalty type algorithm opens the gate to handle monotone inclusion problems with more complicated structures, for instance, involving compositions of maximally monotone operators with linear continuous ones.

R. I. Boţ (✉) · E. R. Csetnek
Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
e-mail: radu.bot@univie.ac.at

E. R. Csetnek
e-mail: ernoe.robert.csetnek@univie.ac.at

1 Introduction and Preliminaries

1.1 Motivation and Problem Formulation

In the last years one can observe in the literature an increasing interest in solving variational inequalities expressed as monotone inclusion problems of the form

$$0 \in Ax + N_C(x), \tag{1}$$

where $\mathcal{H}$ is a real Hilbert space, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator, $C = \text{argmin}\,\Psi$ is the set of global minima of the proper, convex and lower semicontinuous function $\Psi : \mathcal{H} \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ fulfilling $\min \Psi = 0$ and $N_C : \mathcal{H} \rightrightarrows \mathcal{H}$ is the normal cone of the set $C \subseteq \mathcal{H}$ (see [1–3, 18, 19]). Specifically, one can find in the literature forward-backward algorithms for solving (1) (see [2, 3, 18, 19]), which perform in each iteration a proximal step with respect to A and a subgradient step with respect to the penalization function Ψ.

In case $\Psi : \mathcal{H} \to \mathbb{R}$ is differentiable with Lipschitz continuous gradient, for the algorithm, that reads as

$$\texttt{Choose } x_1 \in \mathcal{H}. \texttt{ For } n \in \mathbb{N} \texttt{ set } x_{n+1} = (\text{Id} + \lambda_n A)^{-1}(x_n - \lambda_n \beta_n \nabla\Psi(x_n)),$$

with $(\lambda_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ sequences of positive real numbers, ergodic convergence results are usually obtained in the following hypotheses

$$(H) \begin{cases} (i) \ A + N_C \text{ is maximally monotone and } \{x \in \mathcal{H} : 0 \in Ax + N_C(x)\} \neq \emptyset; \\ (ii) \ \text{For every } p \in \text{ran } N_C, \sum_{n\in\mathbb{N}} \lambda_n\beta_n \left[\Psi^*\left(\frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] < +\infty; \\ (iii) \ (\lambda_n)_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1. \end{cases}$$

Here, $\Psi^* : \mathcal{H} \to \overline{\mathbb{R}}$ denotes the Fenchel conjugate function of Ψ and ran N_C the range of the normal cone operator $N_C : \mathcal{H} \rightrightarrows \mathcal{H}$. Let us mention that hypothesis (ii), which is the discretized counterpart of a condition introduced in [1] in the context of continuous-time nonautonomous differential inclusions, is satisfied, if $\sum_{n\in\mathbb{N}} \frac{\lambda_n}{\beta_n} < +\infty$ and Ψ is bounded below by a multiple of the square of the distance to C (see [2]). This is for instance the case when $C = \text{zer } L = \{x \in \mathcal{H} : Lx = 0\}$, $L : \mathcal{H} \to \mathcal{H}$ is a linear continuous operator with closed range and $\Psi : \mathcal{H} \to \mathbb{R}, \Psi(x) = \|Lx\|^2$ (see [2, 3]). For further situations for which condition (ii) is fulfilled we refer to [3, Section 4.1].

It is worth mentioning that when A is the convex subdifferential of a proper, convex and lower semicontinuous function $\Phi : \mathcal{H} \to \overline{\mathbb{R}}$ the above algorithm provides an iterative scheme for solving convex optimization problems which can be formulated as

$$\min_{x\in\mathcal{H}}\{\Phi(x) : x \in \text{argmin } \Psi\}. \tag{2}$$

Motivated by these considerations, we deal in this paper with monotone inclusion problems of the form

$$0 \in Ax + Dx + N_C(x), \tag{3}$$

where $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator, $D : \mathcal{H} \to \mathcal{H}$ is a (single-valued) cocoercive operator and $C \subseteq \mathcal{H}$ is the (nonempty) set of zeros of another cocoercive operator $B : \mathcal{H} \to \mathcal{H}$. Following [3] we propose a forward-backward penalty algorithm for solving (3) and prove weak ergodic convergence for the generated sequence of iterates under hypotheses which generalize the ones in (H). To this end we specially generalize (ii) to a condition which involves the Fitzpatrick function associated to the maximally monotone operator B. Added to that, we prove strong convergence for the sequence of iterates whenever A is strongly monotone.

The investigations made in this manuscript are completed in Section 3 with the treatment of the monotone inclusion problem (3), this time by relaxing the cocoercivity of D and B to monotonicity and Lipschitz continuity. We formulate in this more general setting a forward-backward-forward penalty type algorithm for solving (3) and study its convergence properties. The interest in having a suitable algorithmic scheme in this context is given by the fact that it allows via some primal-dual techniques to deal with monotone inclusion problems having more complicated structures, for instance, involving compositions of maximally monotone operators with linear continuous ones.

1.2 Notations and Preliminary Results

For the reader's convenience we present first some notations which are used throughout the paper (see [4, 7, 8, 16, 22, 25]). By $\mathbb{N} = \{1, 2, ...\}$ we denote the set of *positive integer numbers* and let $\mathcal{H}$ be a real Hilbert space with *inner product* $\langle \cdot, \cdot \rangle$ and associated *norm* $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$. The symbols $\rightharpoonup$ and $\to$ denote weak and strong convergence, respectively. When $\mathcal{G}$ is another Hilbert space and $K : \mathcal{H} \to \mathcal{G}$ a linear continuous operator, then the *norm* of K is defined as $\|K\| = \sup\{\|Kx\| : x \in \mathcal{H}, \|x\| \leq 1\}$, while $K^* : \mathcal{G} \to \mathcal{H}$, defined by $\langle K^*y, x \rangle = \langle y, Kx \rangle$ for all $(x, y) \in \mathcal{H} \times \mathcal{G}$, denotes the *adjoint operator* of K.

For a function $f : \mathcal{H} \to \overline{\mathbb{R}}$ we denote by $\operatorname{dom} f = \{x \in \mathcal{H} : f(x) < +\infty\}$ its *effective domain* and say that f is *proper*, if $\operatorname{dom} f \neq \emptyset$ and $f(x) \neq -\infty$ for all $x \in \mathcal{H}$. Concerning calculus rules where $\pm\infty$ are involved we make the following conventions (cf. [25]): $(+\infty) + (-\infty) = +\infty$, $0(+\infty) = +\infty$ and $0(-\infty) = 0$. Let $f^* : \mathcal{H} \to \overline{\mathbb{R}}$, $f^*(u) = \sup_{x \in \mathcal{H}}\{\langle x, u \rangle - f(x)\}$ for all $u \in \mathcal{H}$, be the *conjugate function* of f. The *subdifferential* of f at $x \in \mathcal{H}$, with $f(x) \in \mathbb{R}$, is the set $\partial f(x) := \{v \in \mathcal{H} : f(y) \geq f(x) + \langle y - x, v \rangle \, \forall y \in \mathcal{H}\}$. We take by convention $\partial f(x) := \emptyset$, if $f(x) \in \{\pm\infty\}$. We also denote by $\min f := \inf_{x \in \mathcal{H}} f(x)$ and by $\operatorname{argmin} f := \{x \in \mathcal{H} : f(x) = \min f\}$.

Let $S \subseteq \mathcal{H}$ be a nonempty set. The *indicator function* of S, $\delta_S : \mathcal{H} \to \overline{\mathbb{R}}$, is the function which takes the value 0 on S and $+\infty$ otherwise. The subdifferential of the indicator function is the *normal cone* of S, that is $N_S(x) = \{u \in \mathcal{H} : \langle y - x, u \rangle \leq 0 \, \forall y \in S\}$, if $x \in S$ and $N_S(x) = \emptyset$ for $x \notin S$. Notice that for $x \in S$, $u \in N_S(x)$ if and only if $\sigma_S(u) = \langle x, u \rangle$, where σ_S is the support function of S, defined by $\sigma_S(u) = \sup_{y \in S}\langle y, u \rangle$.

For an arbitrary set-valued operator $M : \mathcal{H} \rightrightarrows \mathcal{H}$ we denote by $\operatorname{Gr} M = \{(x, u) \in \mathcal{H} \times \mathcal{H} : u \in Mx\}$ its *graph*, by $\operatorname{dom} M = \{x \in \mathcal{H} : Mx \neq \emptyset\}$ its *domain*, by $\operatorname{ran} M = \{u \in \mathcal{H} : \exists x \in \mathcal{H} \text{ s.t. } u \in Mx\}$ its *range* and by $M^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ its *inverse operator*, defined by $(u, x) \in \operatorname{Gr} M^{-1}$ if and only if $(x, u) \in \operatorname{Gr} M$. We use also the notation $\operatorname{zer} M = \{x \in \mathcal{H} : 0 \in Mx\}$ for the *set of zeros* of the operator M. We say that M is *monotone* if $\langle x - y, u - v \rangle \geq 0$ for all $(x, u), (y, v) \in \operatorname{Gr} M$. A monotone operator M is said to be *maximally monotone*, if there exists no proper monotone extension of the graph of M on $\mathcal{H} \times \mathcal{H}$. Let us mention that in case M is maximally monotone, $\operatorname{zer} M$ is a convex and closed set [4, Proposition 23.39]. We refer to [4, Section 23.4] for conditions ensuring that $\operatorname{zer} M$ is nonempty. If M is maximally monotone, then one has the following characterization for the set of its zeros

$$z \in \operatorname{zer} M \text{ if and only if } \langle u - z, w \rangle \geq 0 \text{ for all } (u, w) \in \operatorname{Gr} M. \tag{4}$$

The operator M is said to be γ-*strongly monotone* with $\gamma > 0$, if $\langle x - y, u - v \rangle \geq \gamma \|x - y\|^2$ for all $(x, u), (y, v) \in \operatorname{Gr} M$. Notice that if M is maximally monotone and strongly monotone, then $\operatorname{zer} M$ is a singleton, thus nonempty (see [4, Corollary 23.37]).

The *resolvent* of M, $J_M : \mathcal{H} \rightrightarrows \mathcal{H}$, is defined by $J_M = (\operatorname{Id} + M)^{-1}$, where $\operatorname{Id} : \mathcal{H} \to \mathcal{H}$, $\operatorname{Id}(x) = x$ for all $x \in \mathcal{H}$, is the *identity operator* on $\mathcal{H}$. Moreover, if M

is maximally monotone, then $J_M : \mathcal{H} \to \mathcal{H}$ is single-valued and maximally monotone (cf. [4, Proposition 23.7 and Corollary 23.10]). For an arbitrary $\gamma > 0$ we have (see [4, Proposition 23.18])

$$J_{\gamma M} + \gamma J_{\gamma^{-1} M^{-1}} \circ \gamma^{-1} \, \mathrm{Id} = \mathrm{Id} \,. \tag{5}$$

The *Fitzpatrick function* associated to a monotone operator M, defined as

$$\varphi_M : \mathcal{H} \times \mathcal{H} \to \overline{\mathbb{R}}, \ \varphi_M(x, u) = \sup_{(y,v) \in \mathrm{Gr}\, M} \{\langle x, v \rangle + \langle y, u \rangle - \langle y, v \rangle\},$$

is a convex and lower semicontinuous function and it will play an important role throughout the paper. Introduced by Fitzpatrick in [17], this notion opened the gate towards the employment of convex analysis specific tools when investigating the maximality of monotone operators (see [4–9, 13, 22] and the references therein). In case M is maximally monotone, φ_M is proper and it fulfills

$$\varphi_M(x, u) \geq \langle x, u \rangle \ \forall (x, u) \in \mathcal{H} \times \mathcal{H},$$

with equality if and only if $(x, u) \in \mathrm{Gr}\, M$. Notice that if $f : \mathcal{H} \to \overline{\mathbb{R}}$, is a proper, convex and lower semi-continuous function, then ∂f is a maximally monotone operator (cf. [20]) and it holds $(\partial f)^{-1} = \partial f^*$. Furthermore, the following inequality is true (see [5])

$$\varphi_{\partial f}(x, u) \leq f(x) + f^*(u) \ \forall (x, u) \in \mathcal{H} \times \mathcal{H}. \tag{6}$$

We refer the reader to [5], for formulae of the corresponding Fitzpatrick functions computed for particular classes of monotone operators.

Let $\gamma > 0$ be arbitrary. A single-valued operator $M : \mathcal{H} \to \mathcal{H}$ is said to be γ-*cocoercive*, if $\langle x - y, Mx - My \rangle \geq \gamma \| Mx - My \|^2$ for all $(x, y) \in \mathcal{H} \times \mathcal{H}$, and γ-*Lipschitz continuous*, if $\| Mx - My \| \leq \gamma \| x - y \|$ for all $(x, y) \in \mathcal{H} \times \mathcal{H}$. A single-valued linear operator $M : \mathcal{H} \to \mathcal{H}$ is said to be *skew*, if $\langle x, Mx \rangle = 0$ for all $x \in \mathcal{H}$.

We close the section by presenting some convergence results that will be used several times in the paper. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence in $\mathcal{H}$ and $(\lambda_k)_{k \in \mathbb{N}}$ a sequence of positive numbers such that $\sum_{k \in \mathbb{N}} \lambda_k = +\infty$. Let $(z_n)_{n \in \mathbb{N}}$ be the sequence of weighted averages defined as (see [3])

$$z_n = \frac{1}{\tau_n} \sum_{k=1}^{n} \lambda_k x_k, \ \text{where} \ \tau_n = \sum_{k=1}^{n} \lambda_k \ \forall n \in \mathbb{N}. \tag{7}$$

Lemma 1 *(Opial-Passty) Let F be a nonempty subset of $\mathcal{H}$ and assume that the limes* $\lim_{n \to \infty} \| x_n - x \|$ *exists for every $x \in F$. If every weak cluster point of $(x_n)_{n \in \mathbb{N}}$ (respectively $(z_n)_{n \in \mathbb{N}}$) lies in F, then $(x_n)_{n \in \mathbb{N}}$ (respectively $(z_n)_{n \in \mathbb{N}}$) converges weakly to an element in F as $n \to +\infty$.*

The following result is taken from [3].

Lemma 2 *Let $(a_n)_{n \in \mathbb{N}}$, $(b_n)_{n \in \mathbb{N}}$ and $(\varepsilon_n)_{n \in \mathbb{N}}$ be real sequences. Assume that $(a_n)_{n \in \mathbb{N}}$ is bounded from below, $(b_n)_{n \in \mathbb{N}}$ is nonnegative, $(\varepsilon_n)_{n \in \mathbb{N}} \in \ell^1$ and $a_{n+1} - a_n + b_n \leq \varepsilon_n$ for any $n \in \mathbb{N}$. Then $(a_n)_{n \in \mathbb{N}}$ is convergent and $(b_n)_{n \in \mathbb{N}} \in \ell^1$.*

2 Forward-Backward Penalty Schemes

The problem we deal with at the beginning of this section has the following formulation.

Problem 3 Let $\mathcal{H}$ be a real Hilbert space, $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ maximally monotone operators, $D : \mathcal{H} \to \mathcal{H}$ an η-cocoercive operator with $\eta > 0$ and suppose that $C = \text{zer } B \neq \emptyset$. The monotone inclusion problem to solve is

$$0 \in Ax + Dx + N_C(x).$$

The following iterative scheme for solving Problem 3 is inspired by [3].

Algorithm 1

> Initialization: Choose $x_1 \in \mathcal{H}$
> For $n \in \mathbb{N}$:   Choose $w_n \in Bx_n$
> Set $x_{n+1} = J_{\lambda_n A}(x_n - \lambda_n Dx_n - \lambda_n \beta_n w_n)$,

---

where $(\lambda_n)_{n \in \mathbb{N}}$ and $(\beta_n)_{n \in \mathbb{N}}$ are sequences of positive real numbers. Notice that Algorithm 4 is well-defined, if dom $B = \mathcal{H}$, which will be the case in Subsection 2.2, when B is assumed to be cocoercive. For the convergence statement the following hypotheses are needed

$$(H_{fitz}) \begin{cases} (i) \ A + N_C \text{ is maximally monotone and zer } (A + D + N_C) \neq \emptyset; \\ (ii) \ \text{For every } p \in \text{ran } N_C, \sum_{n \in \mathbb{N}} \lambda_n \beta_n \left[\sup_{u \in C} \varphi_B \left(u, \frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] < +\infty; \\ (iii) \ (\lambda_n)_{n \in \mathbb{N}} \in \ell^2 \setminus \ell^1. \end{cases}$$

Since A is maximally monotone and C is a nonempty convex and closed set, $A + N_C$ is maximally monotone if a so-called regularity condition is fulfilled. We refer the reader to [4, 6–9, 22, 25] for conditions guaranteeing the maximal monotonicity of the sum of two maximally monotone operators.

Further, as D is maximally monotone (see [4, Example 20.28]) and dom $D = \mathcal{H}$, the hypothesis (i) above guarantees that $A + D + N_C$ is maximally monotone, too (see [4, Corollary 24.4]). Moreover, for each $p \in \text{ran } N_C$ we have

$$\sup_{u \in C} \varphi_B \left(u, \frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \geq 0 \ \forall n \in \mathbb{N}.$$

Indeed, if $p \in \text{ran } N_C$, then there exists $\overline{u} \in C$ such that $p \in N_C(\overline{u})$. This implies that

$$\sup_{u \in C} \varphi_B \left(u, \frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \geq \left\langle \overline{u}, \frac{p}{\beta_n} \right\rangle - \sigma_C \left(\frac{p}{\beta_n} \right) = 0 \ \forall n \in \mathbb{N}.$$

Remark 4 Let us mention that, if $Dx = 0$ for all $x \in \mathcal{H}$ and $B = \partial \Psi$, where $\Psi : \mathcal{H} \to \overline{\mathbb{R}}$ is a proper, convex and lower semicontinuous function with $\min \Psi = 0$, then the monotone inclusion problem in Problem 3 becomes (1), since in this case $C = \text{argmin } \Psi$. Moreover, as $\Psi(x) = 0$ for all $x \in C$, by (6) it follows that condition (ii) in (H) implies condition (ii) in (H_{fitz}), hence the hypothesis formulated by means of the Fitzpatrick function extends the one given [3] to the more general setting considered in Problem 3. It remains an open

question to find examples of proper, convex and lower semicontinuous functions $\Psi : \mathcal{H} \to \overline{\mathbb{R}}$ with $\min \Psi = 0$ for which (ii) in (H) is not fulfilled, while for $B = \partial \Psi$ condition (ii) in (H_{fitz}) holds.

The results provided throughout this section and the techniques and ideas used for proving them have been inspired by [3], however, they are here adapted and extended to the general context of the monotone inclusion problem stated in Problem 3.

2.1 The General Case

In this subsection we will prove an abstract convergence result for Algorithm (4), which will be subsequently refined in the case when B is a cocoercive operator.

Lemma 5 *Let $(x_n)_{n \in \mathbb{N}}$ and $(w_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 1 and take $(u, w) \in \mathrm{Gr}(A + D + N_C)$ such that $w = v + p + Du$, where $v \in Au$ and $p \in N_C(u)$. Then the following inequality holds for any $n \in \mathbb{N}$*

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \lambda_n(2\eta - 3\lambda_n)\|Dx_n - Du\|^2 \leq$$
$$2\lambda_n\beta_n\left[\sup_{u \in C} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] + 3\lambda_n^2\beta_n^2\|w_n\|^2 + 3\lambda_n^2\|Du + v\|^2 + 2\lambda_n\langle u - x_n, w\rangle. \tag{8}$$

Proof From the definition of the resolvent of A we have $\frac{x_n - x_{n+1}}{\lambda_n} - \beta_n w_n - Dx_n \in Ax_{n+1}$ and since $v \in Au$, the monotonicity of A guarantees

$$\langle x_{n+1} - u, x_n - x_{n+1} - \lambda_n(\beta_n w_n + Dx_n + v)\rangle \geq 0 \ \forall n \in \mathbb{N}, \tag{9}$$

thus

$$\langle u - x_{n+1}, x_n - x_{n+1}\rangle \leq \lambda_n\langle u - x_{n+1}, \beta_n w_n + Dx_n + v\rangle \ \forall n \in \mathbb{N}.$$

Further, since

$$\langle u - x_{n+1}, x_n - x_{n+1}\rangle = \frac{1}{2}\|x_{n+1} - u\|^2 - \frac{1}{2}\|x_n - u\|^2 + \frac{1}{2}\|x_{n+1} - x_n\|^2,$$

we get for any $n \in \mathbb{N}$

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 \leq 2\lambda_n\langle u - x_{n+1}, \beta_n w_n + Dx_n + v\rangle - \|x_{n+1} - x_n\|^2$$
$$= 2\lambda_n\langle u - x_n, \beta_n w_n + Dx_n + v\rangle + 2\lambda_n\langle x_n - x_{n+1}, \beta_n w_n + Dx_n + v\rangle - \|x_{n+1} - x_n\|^2$$
$$\leq 2\lambda_n\langle u - x_n, \beta_n w_n + Dx_n + v\rangle + \lambda_n^2\|\beta_n w_n + Dx_n + v\|^2$$
$$\leq 2\lambda_n\langle u - x_n, \beta_n w_n + Dx_n + v\rangle + 3\lambda_n^2\beta_n^2\|w_n\|^2 + 3\lambda_n^2\|Du + v\|^2 + 3\lambda_n^2\|Dx_n - Du\|^2.$$

Next we evaluate the first term on the right hand-side of the last of the above inequalities. By using the cocoercivity of D and the definition of the Fitzpatrick function and that $w_n \in Bx_n$ and $\sigma_C \left(\frac{p}{\beta_n} \right) = \langle u, \frac{p}{\beta_n} \rangle$ for every $n \in \mathbb{N}$, we obtain

$$2\lambda_n \langle u - x_n, \beta_n w_n + Dx_n + v \rangle$$
$$= 2\lambda_n \langle u - x_n, \beta_n w_n + Dx_n + w - p - Du \rangle$$
$$= 2\lambda_n \langle u - x_n, Dx_n - Du \rangle + 2\lambda_n \langle u - x_n, \beta_n w_n - p \rangle + 2\lambda_n \langle u - x_n, w \rangle$$
$$= 2\lambda_n \langle u - x_n, Dx_n - Du \rangle + 2\lambda_n \beta_n \left(\langle u, w_n \rangle + \langle x_n, \frac{p}{\beta_n} \rangle - \langle x_n, w_n \rangle - \langle u, \frac{p}{\beta_n} \rangle \right)$$
$$+ 2\lambda_n \langle u - x_n, w \rangle$$
$$\leq -2\eta\lambda_n \| Dx_n - Du \|^2 + 2\lambda_n \beta_n \left[\sup_{u \in C} \varphi_B \left(u, \frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] + 2\lambda_n \langle u - x_n, w \rangle.$$

This provides the desired conclusion. $\qquad\square$

Theorem 6 *Let $(x_n)_{n \in \mathbb{N}}$ and $(w_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 1 and $(z_n)_{n \in \mathbb{N}}$ the sequence defined in (7). If (H_{fitz}) is fulfilled and $(\lambda_n \beta_n \| w_n \|)_{n \in \mathbb{N}} \in \ell^2$, then $(z_n)_{n \in \mathbb{N}}$ converges weakly to an element in $\mathrm{zer}(A + D + N_C)$ as $n \to +\infty$.*

Proof As $\lim_{n \to +\infty} \lambda_n = 0$, there exists $n_0 \in \mathbb{N}$ such that $2\eta - 3\lambda_n \geq 0$ for all $n \geq n_0$. Thus, for $(u, w) \in \mathrm{Gr}(A + D + N_C)$, such that $w = v + p + Du$, where $v \in Au$ and $p \in N_C(u)$, by (8) it holds for any $n \geq n_0$

$$\| x_{n+1} - u \|^2 - \| x_n - u \|^2 \leq$$
$$2\lambda_n \beta_n \left[\sup_{u \in C} \varphi_B \left(u, \frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] + 3\lambda_n^2 \beta_n^2 \| w_n \|^2 + 3\lambda_n^2 \| Du + v \|^2 + 2\lambda_n \langle u - x_n, w \rangle.$$
$$(10)$$

By Lemma 1, it is sufficient to prove that the following two statements hold:

(a) for every $u \in \mathrm{zer}(A + D + N_C)$ the sequence $(\| x_n - u \|)_{n \in \mathbb{N}}$ is convergent;
(b) every weak cluster point of $(z_n)_{n \in \mathbb{N}}$ lies in $\mathrm{zer}(A + D + N_C)$.

(a) For every $u \in \mathrm{zer}(A + D + N_C)$ one can take $w = 0$ in (10) and the conclusion follows from Lemma 2.

(b) Let z be a weak cluster point of $(z_n)_{n \in \mathbb{N}}$. As we already noticed that $A + D + N_C$ is maximally monotone, in order to show that $z \in \mathrm{zer}(A + D + N_C)$ we will use the characterization given in (4). Take $(u, w) \in \mathrm{Gr}(A + D + N_C)$ such that $w = v + p + Du$, where $v \in Au$ and $p \in N_C(u)$. Let be $N \in \mathbb{N}$ with $N \geq n_0 + 2$. Summing up for $n = n_0 + 1, ..., N$ the inequalities in (10), we get

$$\| x_{N+1} - u \|^2 - \| x_{n_0+1} - u \|^2 \leq L + 2 \left\langle \sum_{n=1}^{N} \lambda_n u - \sum_{n=1}^{N} \lambda_n x_n - \sum_{n=1}^{n_0} \lambda_n u + \sum_{n=1}^{n_0} \lambda_n x_n, w \right\rangle,$$

where

$$L = 2 \sum_{n \geq n_0+1} \lambda_n \beta_n \left[\sup_{u \in C} \varphi_B \left(u, \frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right]$$

$$+ 3 \sum_{n \geq n_0+1} \lambda_n^2 \beta_n^2 \|w_n\|^2 + 3 \sum_{n \geq n_0+1} \lambda_n^2 \|Du + v\|^2 \in \mathbb{R}.$$

Discarding the nonnegative term $\|x_{N+1} - u\|^2$ and dividing by $2\tau_N = 2 \sum_{k=1}^{N} \lambda_k$ we obtain

$$-\frac{\|x_{n_0+1} - u\|^2}{2\tau_N} \leq \frac{\widetilde{L}}{2\tau_N} + \langle u - z_N, w \rangle,$$

where $\widetilde{L} := L + 2\langle -\sum_{n=1}^{n_0} \lambda_n u + \sum_{n=1}^{n_0} \lambda_n x_n, w \rangle \in \mathbb{R}$. By passing to the limit as $N \to +\infty$ and using that $\lim_{N \to +\infty} \tau_N = +\infty$, we get

$$\liminf_{N \to +\infty} \langle u - z_N, w \rangle \geq 0.$$

Since z is a weak cluster point of $(z_n)_{n \in \mathbb{N}}$, we obtain that $\langle u - z, w \rangle \geq 0$. Finally, as this inequality holds for arbitrary $(u, w) \in \mathrm{Gr}(A + D + N_C)$, the desired conclusion follows. $\quad\square$

In the following we show that strong monotonicity of the operator A ensures strong convergence of the sequence $(x_n)_{n \in \mathbb{N}}$.

Theorem 7 *Let $(x_n)_{n \in \mathbb{N}}$ and $(w_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 1. If (H_{fitz}) is fulfilled, $(\lambda_n \beta_n \|w_n\|)_{n \in \mathbb{N}} \in \ell^2$ and the operator A is γ-strongly monotone with $\gamma > 0$, then $(x_n)_{n \in \mathbb{N}}$ converges strongly to the unique element in $\mathrm{zer}(A + D + N_C)$ as $n \to +\infty$.*

Proof Let be $u \in \mathrm{zer}(A + D + N_C)$ and $w = 0 = v + p + Du$, where $v \in Au$ and $p \in N_C(u)$. Since A is γ-strongly monotone, inequality (9) becomes

$$\langle x_{n+1} - u, x_n - x_{n+1} - \lambda_n(\beta_n w_n + Dx_n + v) \rangle \geq \lambda_n \gamma \|x_{n+1} - u\|^2 \; \forall n \in \mathbb{N}. \quad (11)$$

Following the lines of the proof of Lemma 5 for $w = 0$ we obtain for any $n \in \mathbb{N}$

$$2\gamma \lambda_n \|x_{n+1} - u\|^2 + \|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \lambda_n(2\eta - 3\lambda_n)\|Dx_n - Du\|^2 \leq$$

$$2\lambda_n \beta_n \left[\sup_{u \in C} \varphi_B \left(u, \frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] + 3\lambda_n^2 \beta_n^2 \|w_n\|^2 + 3\lambda_n^2 \|Du + v\|^2.$$

Thus, as $\lim_{n \to +\infty} \lambda_n = 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$

$$2\gamma \lambda_n \|x_{n+1} - u\|^2 + \|x_{n+1} - u\|^2 - \|x_n - u\|^2$$

$$\leq 2\lambda_n \beta_n \left[\sup_{u \in C} \varphi_B \left(u, \frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] + 3\lambda_n^2 \beta_n^2 \|w_n\|^2 + 3\lambda_n^2 \|Du + v\|^2$$

and, so,

$$2\gamma \sum_{n \geq n_0} \lambda_n \|x_{n+1} - u\|^2 \leq \|x_{n_0} - u\|^2 + 2 \sum_{n \geq n_0} \lambda_n \beta_n \left[\sup_{u \in C} \varphi_B \left(u, \frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right]$$

$$+ 3 \sum_{n \geq n_0} \lambda_n^2 \beta_n^2 \|w_n\|^2 + 3\|Du + v\|^2 \sum_{n \geq n_0} \lambda_n^2 < +\infty.$$

Since $\sum_{n \in \mathbb{N}} \lambda_n = +\infty$ and $(\|x_n - u\|)_{n \in \mathbb{N}}$ is convergent (see the proof of Theorem 6 (a)), it follows $\lim_{n \to +\infty} \|x_n - u\| = 0$. $\quad\square$

2.2 The Case B is Cocoercive

In this subsection we deal with the situation when B is a (single-valued) cocoercive operator. Our aim is to show that in this setting the assumption $(\lambda_n \beta_n \|w_n\|)_{n\in\mathbb{N}} \in \ell^2$ in Theorem 22 and Theorem 7 can be replaced by a milder condition involving only the sequences $(\lambda_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$. The problem under consideration has the following formulation.

Problem 8 Let $\mathcal{H}$ be a real Hilbert space, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator, $D : \mathcal{H} \to \mathcal{H}$ an η-cocoercive operator with $\eta > 0$, $B : \mathcal{H} \to \mathcal{H}$ a μ-cocoercive operator with $\mu > 0$ and suppose that $C = \operatorname{zer} B \neq \emptyset$. The monotone inclusion problem to solve is

$$0 \in Ax + Dx + N_C(x).$$

Algorithm 1 has in this particular setting the following formulation.

Algorithm 2

  Initialization: Choose $x_1 \in \mathcal{H}$
      For $n \in \mathbb{N}$ set: $x_{n+1} = J_{\lambda_n A}(x_n - \lambda_n D x_n - \lambda_n \beta_n B x_n)$.

Remark 9

(a) If $Dx = 0$ for every $x \in \mathcal{H}$ and $B = \partial \Psi$, where $\Psi : \mathcal{H} \to \mathbb{R}$ is a convex and differentiable function with μ^{-1}-Lipschitz gradient for $\mu > 0$ fulfilling $\min \Psi = 0$, then we rediscover the setting considered in [3, Section 3], while Algorithm 2 becomes the iterative method investigated in that paper.

(b) In case $Bx = 0$ for all $x \in \mathcal{H}$ Algorithm 2 turns out to be classical forward-backward scheme (see [4, 14, 24]), since in this case $C = \mathcal{H}$, hence $N_C(x) = \{0\}$ for all $x \in \mathcal{H}$.

Before stating the convergence result for Algorithm 2 some technical results are in order.

Lemma 10 *Let be $u \in C \cap \operatorname{dom} A$ and $v \in Au$. Then for every $\varepsilon \geq 0$ and any $n \in \mathbb{N}$ we have*

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \frac{\varepsilon}{1+\varepsilon}\|x_{n+1} - x_n\|^2 + \frac{2\varepsilon}{1+\varepsilon}\lambda_n\beta_n\langle x_n - u, Bx_n\rangle \leq$$

$$\lambda_n\beta_n\left((1+\varepsilon)\lambda_n\beta_n - \frac{2\mu}{1+\varepsilon}\right)\|Bx_n\|^2 + 2\lambda_n\langle u - x_{n+1}, Dx_n + v\rangle. \tag{12}$$

Proof As in the proof of Lemma 5 we obtain for any $n \in \mathbb{N}$ that

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \|x_{n+1} - x_n\|^2 \leq 2\lambda_n\langle u - x_{n+1}, \beta_n Bx_n + Dx_n + v\rangle =$$
$$2\lambda_n\beta_n\langle u - x_n, Bx_n\rangle + 2\lambda_n\beta_n\langle x_n - x_{n+1}, Bx_n\rangle + 2\lambda_n\langle u - x_{n+1}, Dx_n + v\rangle.$$

Since B is μ-cocoercive and $Bu = 0$ we have that

$$\langle u - x_n, Bx_n\rangle \leq -\mu\|Bx_n\|^2 \; \forall n \in \mathbb{N},$$

hence

$$2\lambda_n\beta_n\langle u - x_n, Bx_n\rangle \le -\frac{2\mu}{1+\varepsilon}\lambda_n\beta_n\|Bx_n\|^2 + \frac{2\varepsilon}{1+\varepsilon}\lambda_n\beta_n\langle u - x_n, Bx_n\rangle \; \forall n \in \mathbb{N} \; \forall \varepsilon \ge 0.$$

Inequality (12) follows by taking into consideration also that

$$2\lambda_n\beta_n\langle x_n - x_{n+1}, Bx_n\rangle \le \frac{1}{1+\varepsilon}\|x_{n+1} - x_n\|^2 + (1+\varepsilon)\lambda_n^2\beta_n^2\|Bx_n\|^2 \; \forall n \in \mathbb{N} \; \forall \varepsilon \ge 0.$$

$\square$

Lemma 11 *Assume that* $\limsup_{n\to+\infty} \lambda_n\beta_n < 2\mu$ *and let be* $u \in C \cap \operatorname{dom} A$ *and* $v \in Au$. *Then there exist* $a, b > 0$ *and* $n_0 \in \mathbb{N}$ *such that for any* $n \ge n_0$ *it holds*

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + a\left(\|x_{n+1} - x_n\|^2 + \lambda_n\beta_n\langle x_n - u, Bx_n\rangle + \lambda_n\beta_n\|Bx_n\|^2\right) \le$$
$$\left(b\lambda_n^2 - 2\eta\lambda_n\right)\|Dx_n - Du\|^2 + 2\lambda_n\langle u - x_n, v + Du\rangle + b\lambda_n^2\|Du + v\|^2. \tag{13}$$

Proof We start by noticing that, by making use of the cocoercivity of D, for every $\varepsilon > 0$ and any $n \in \mathbb{N}$ it holds

$$2\lambda_n\langle u - x_{n+1}, Dx_n + v\rangle$$
$$= 2\lambda_n\langle x_n - x_{n+1}, Dx_n + v\rangle + 2\lambda_n\langle u - x_n, Dx_n + v\rangle$$
$$\le \frac{\varepsilon}{2(1+\varepsilon)}\|x_{n+1} - x_n\|^2 + \frac{2(1+\varepsilon)}{\varepsilon}\lambda_n^2\|Dx_n + v\|^2 + 2\lambda_n\langle u - x_n, Dx_n + v\rangle$$
$$\le \frac{\varepsilon}{2(1+\varepsilon)}\|x_{n+1} - x_n\|^2 + \frac{4(1+\varepsilon)}{\varepsilon}\lambda_n^2\|Dx_n - Du\|^2 + \frac{4(1+\varepsilon)}{\varepsilon}\lambda_n^2\|Du + v\|^2 +$$
$$2\lambda_n\langle u - x_n, Dx_n - Du\rangle + 2\lambda_n\langle u - x_n, v + Du\rangle$$
$$\le \frac{\varepsilon}{2(1+\varepsilon)}\|x_{n+1} - x_n\|^2 + \frac{4(1+\varepsilon)}{\varepsilon}\lambda_n^2\|Dx_n - Du\|^2 + \frac{4(1+\varepsilon)}{\varepsilon}\lambda_n^2\|Du + v\|^2 -$$
$$2\lambda_n\eta\|Dx_n - Du\|^2 + 2\lambda_n\langle u - x_n, v + Du\rangle.$$

In combination with (12) it yields for every $\varepsilon > 0$ and any $n \in \mathbb{N}$

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \frac{\varepsilon}{2(1+\varepsilon)}\|x_{n+1} - x_n\|^2 + \frac{2\varepsilon}{1+\varepsilon}\lambda_n\beta_n\langle x_n - u, Bx_n\rangle$$
$$+ \frac{\varepsilon}{1+\varepsilon}\lambda_n\beta_n\|Bx_n\|^2$$
$$\le \lambda_n\beta_n\left((1+\varepsilon)\lambda_n\beta_n - \frac{2\mu}{1+\varepsilon} + \frac{\varepsilon}{1+\varepsilon}\right)\|Bx_n\|^2 + \left(\frac{4(1+\varepsilon)}{\varepsilon}\lambda_n^2 - 2\eta\lambda_n\right)\|Dx_n - Du\|^2$$
$$+ 2\lambda_n\langle u - x_n, v + Du\rangle + \frac{4(1+\varepsilon)}{\varepsilon}\lambda_n^2\|Du + v\|^2.$$

Since $\limsup_{n\to+\infty} \lambda_n\beta_n < 2\mu$, there exists $\alpha > 0$ and $n_0 \in \mathbb{N}$ such that $\lambda_n\beta_n < \alpha < 2\mu$ for any $n \ge n_0$. Hence, for any $n \ge n_0$ and every $\varepsilon > 0$ it holds

$$\lambda_n\beta_n\left((1+\varepsilon)\lambda_n\beta_n - \frac{2\mu}{1+\varepsilon} + \frac{\varepsilon}{1+\varepsilon}\right) < \alpha\left((1+\varepsilon)\alpha - \frac{2\mu}{1+\varepsilon} + \frac{\varepsilon}{1+\varepsilon}\right)$$

and one can take $\varepsilon_0 > 0$ small enough such that $(1 + \varepsilon_0)\alpha - \frac{2\mu}{1+\varepsilon_0} + \frac{\varepsilon_0}{1+\varepsilon_0} < 0$. Taking $a = \frac{\varepsilon_0}{2(1+\varepsilon_0)}$ and $b = \frac{4(1+\varepsilon_0)}{\varepsilon_0}$ the desired conclusion follows. $\square$

Lemma 12 *Assume that* $\limsup_{n\to+\infty} \lambda_n \beta_n < 2\mu$ *and* $\lim_{n\to+\infty} \lambda_n = 0$ *and let be* (u, w) $\in \mathrm{Gr}(A + D + N_C)$ *such that* $w = v + p + Du$, *where* $v \in Au$ *and* $p \in N_C(u)$. *Then there exist* $a, b > 0$ *and* $n_1 \in \mathbb{N}$ *such that for all* $n \geq n_1$ *it holds*

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + a \left(\|x_{n+1} - x_n\|^2 + \frac{\lambda_n \beta_n}{2} \langle x_n - u, Bx_n \rangle + \lambda_n \beta_n \|Bx_n\|^2 \right)$$

$$\leq \frac{a\lambda_n \beta_n}{2} \left[\sup_{u \in C} \varphi_B \left(u, \frac{4p}{a\beta_n} \right) - \sigma_C \left(\frac{4p}{a\beta_n} \right) \right] + 2\lambda_n \langle u - x_n, w \rangle + b\lambda_n^2 \|Du + v\|^2. \quad (14)$$

Proof According to Lemma 11, there exist $a, b > 0$ and $n_0 \in \mathbb{N}$ such that for any $n \geq n_0$ inequality (13) holds. Since $\lim_{n\to\infty} \lambda_n = 0$, there exists $n_1 \in \mathbb{N}, n_1 \geq n_0$ such that $b\lambda_n^2 - 2\eta\lambda_n \leq 0$ for all $n \geq n_1$, hence,

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + a \left(\|x_{n+1} - x_n\|^2 + \lambda_n \beta_n \langle x_n - u, Bx_n \rangle + \lambda_n \beta_n \|Bx_n\|^2 \right)$$

$$\leq 2\lambda_n \langle u - x_n, v + Du \rangle + b\lambda_n^2 \|Du + v\|^2 \ \forall n \geq n_1.$$

The conclusion follows by combining this inequality with the subsequent estimation that holds for any $n \in \mathbb{N}$:

$$2\lambda_n \langle u - x_n, v + Du \rangle + \frac{a\lambda_n \beta_n}{2} \langle u - x_n, Bx_n \rangle$$

$$= 2\lambda_n \langle u - x_n, -p \rangle + \frac{a\lambda_n \beta_n}{2} \langle u - x_n, Bx_n \rangle + 2\lambda_n \langle u - x_n, w \rangle$$

$$= \frac{a\lambda_n \beta_n}{2} \left(\langle u, Bx_n \rangle + \left\langle x_n, \frac{4p}{a\beta_n} \right\rangle - \langle x_n, Bx_n \rangle - \left\langle u, \frac{4p}{a\beta_n} \right\rangle \right) + 2\lambda_n \langle u - x_n, w \rangle$$

$$\leq \frac{a\lambda_n \beta_n}{2} \left[\sup_{u \in C} \varphi_B \left(u, \frac{4p}{a\beta_n} \right) - \sigma_C \left(\frac{4p}{a\beta_n} \right) \right] + 2\lambda_n \langle u - x_n, w \rangle.$$

$\square$

Theorem 13 *Let* $(x_n)_{n\in\mathbb{N}}$ *and* $(w_n)_{n\in\mathbb{N}}$ *be the sequences generated by Algorithm 2 and* $(z_n)_{n\in\mathbb{N}}$ *the sequence defined in* (7). *If* (H_{fitz}) *is fulfilled and* $\limsup_{n\to+\infty} \lambda_n \beta_n < 2\mu$, *then the following statements are true:*

(i) *for every* $u \in \mathrm{zer}(A + D + N_C)$ *the sequence* $(\|x_n - u\|)_{n\in\mathbb{N}}$ *is convergent and the series* $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2$, $\sum_{n\in\mathbb{N}} \lambda_n \beta_n \langle Bx_n, x_n - u \rangle$ *and* $\sum_{n\in\mathbb{N}} \lambda_n \beta_n \|Bx_n\|^2$ *are convergent as well. In particular* $\lim_{n\to+\infty} \|x_{n+1} - x_n\| = 0$. *If, moreover,* $\liminf_{n\to+\infty} \lambda_n \beta_n > 0$, *then* $\lim_{n\to+\infty} \langle Bx_n, x_n - u \rangle = \lim_{n\to+\infty} \|Bx_n\| = 0$ *and every weak cluster point of* $(x_n)_{n\in\mathbb{N}}$ *lies in* C.

(ii) $(z_n)_{n\in\mathbb{N}}$ *converges weakly to an element in* $\mathrm{zer}(A + D + N_C)$ *as* $n \to +\infty$.

(iii) *if, additionally, A is strongly monotone, then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly to the unique element in* $\mathrm{zer}(A + D + N_C)$ *as* $n \to +\infty$.

Proof For every $u \in \mathrm{zer}(A + D + N_C)$, according to Lemma 12, there exist $a, b > 0$ and $n_1 \in \mathbb{N}$ such that for all $n \geq n_1$ inequality (14) is true for $w = 0$. This gives rise via Lemma 2 to the statements in (i). As the sequence $(\lambda_n \beta_n)_{n\in\mathbb{N}}$ is bounded above, it automatically follows that $(\lambda_n \beta_n \|Bx_n\|)_{n\in\mathbb{N}} \in \ell^2$. Hence, (ii) and (iii) follow as consequences of Theorem 6 and Theorem 7, respectively. $\square$

Remark 14 Let us emphasize the fact that the results obtained in this subsection by assuming that B is a cocoercive operator enables us to treat the more general case where C is the set of zeros of an arbitrary maximally monotone operator.

Indeed, we consider in Problem 3 that $C = \operatorname{zer} M \neq \emptyset$, where $M : \mathcal{H} \rightrightarrows \mathcal{H}$ is a (possibly set-valued) maximally monotone operator. The idea is to apply the results in Subsection 2.2 to the operator $B := J_{M^{-1}} : \mathcal{H} \to \mathcal{H}$, which according to [4, Proposition 20.22, Corollary 23.10 and Proposition 4.2] is μ-cocoercive with $\mu = 1$. By noticing that $\operatorname{zer} J_{M^{-1}} = \operatorname{zer} M$, we can address the monotone inclusion problem to be solved as a problem formulated in the framework of Problem 8. Obviously, in the iterative scheme given in Algorithm 2 the operator M will be evaluated by a backward step.

Further we will show that one can provide sufficient conditions for (ii) in (H_{fitz}) written in terms of the Fitzpatrick function of the operator M. To this end we use the following estimation for the Fitzpatrick function of $J_{M^{-1}}$, obtained by applying [5, Proposition 4.2], a result that gives an upper bound for the Fitzpatrick function of the sum of two maximally monotone operators in terms of the Fitzpatrick functions of the operators involved. Take an arbitrary $p \in \operatorname{ran} N_C$ and $u \in C$. We have for every $n \in \mathbb{N}$

$$\varphi_{J_{M^{-1}}}\left(u, \frac{p}{\beta_n}\right) = \varphi_{\operatorname{Id}+M^{-1}}\left(\frac{p}{\beta_n}, u\right) \le \varphi_{\operatorname{Id}}\left(\frac{p}{\beta_n}, 0\right) + \varphi_{M^{-1}}\left(\frac{p}{\beta_n}, u\right) =$$

$$\frac{1}{4}\left\|\frac{p}{\beta_n}\right\|^2 + \varphi_M\left(u, \frac{p}{\beta_n}\right),$$

where we used the fact that $\varphi_{\operatorname{Id}}(x, v) = \frac{1}{4}\|x + v\|^2$ for all $(x, v) \in \mathcal{H} \times \mathcal{H}$.

This means that the condition (ii) in (H_{fitz}) applied to the reformulation of Problem 3 described above is fulfilled, if we assume that $\sum_{n \in \mathbb{N}} \frac{\lambda_n}{\beta_n} < +\infty$ and that for every $p \in \operatorname{ran} N_C$, $\sum_{n \in \mathbb{N}} \lambda_n \beta_n \left[\sup_{u \in C} \varphi_M\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] < +\infty$.

3 Tseng's Type Penalty Schemes

In this section we deal first with the monotone inclusion problem stated in Problem 8 by relaxing the cocoercivity of B and D to monotonicity and Lipschitz continuity. The iterative method we propose in this setting is a forward-backward-forward penalty scheme and relies on a method introduced by Tseng in [23] (see [4, 12, 15] for further details and motivations). By making use of primal-dual techniques we will be able then to employ the proposed approach when solving monotone inclusion problems involving compositions of maximally monotone operators with linear continuous ones.

3.1 Relaxing cocoercivity to monotonicity and Lipschitz continuity

We deal first with the following problem.

Problem 15 Let $\mathcal{H}$ be a real Hilbert space, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator, $D : \mathcal{H} \to \mathcal{H}$ a monotone and η^{-1}-Lipschitz continuous operator with $\eta > 0$, $B : \mathcal{H} \to \mathcal{H}$ a monotone and μ^{-1}-Lipschitz continuous operator with $\mu > 0$ and suppose that $C = \operatorname{zer} B \neq \emptyset$. The monotone inclusion problem to solve is

$$0 \in Ax + Dx + N_C(x).$$

One can notice that we have relaxed the assumptions imposed on B and D in Problem 8, however, they are both maximally monotone, see [4, Corollary 20.25]. It is obvious that an η-cocoercive operator with $\eta > 0$ is monotone and η^{-1}-Lipschitz continuous, while the opposite is not the case. It is well-known that, due to the celebrated Baillon-Haddad Theorem (see, for instance, [4, Corollary 8.16]), the gradient of a convex and differentiable function does not provide a counterexample in this sense, however, nonzero linear, skew and Lipschitz continuous operators are monotone, but not cocoercive. For example, when $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces and $L : \mathcal{H} \to \mathcal{G}$ is nonzero linear continuous, then $(x, v) \mapsto (L^*v, -Lx)$ is an example in this sense. This operator appears in a natural way when employing primal-dual approaches in the context of monotone inclusion problems as done in [12] (see also [10, 11, 15, 24]).

Algorithm 3

> **Initialization:** Choose $x_1 \in \mathcal{H}$
>
> **For** $n \in \mathbb{N}$ **set:** $p_n = J_{\lambda_n A}(x_n - \lambda_n D x_n - \lambda_n \beta_n B x_n)$
>
> $$x_{n+1} = \lambda_n \beta_n (B x_n - B p_n) + \lambda_n (D x_n - D p_n) + p_n,$$

---

where $(\lambda_n)_{n \in \mathbb{N}}$ and $(\beta_n)_{n \in \mathbb{N}}$ are sequences of positive real numbers.

Remark 16 If $B x = 0$ for every $x \in \mathcal{H}$ (which corresponds to the situation $N_C(x) = \{0\}$ for all $x \in \mathcal{H}$), then Algorithm 3 turns out to be the error-free forward-backward-forward scheme from [12, Theorem 2.5] (see also [23]).

We start with the following technical statement.

Lemma 17 *Let $(x_n)_{n \in \mathbb{N}}$ and $(p_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 3 and let be $(u, w) \in \mathrm{Gr}(A + D + N_C)$ such that $w = v + p + Du$, where $v \in Au$ and $p \in N_C(u)$. Then the following inequality holds for all $n \in \mathbb{N}$:*

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \left(1 - \left(\frac{\lambda_n \beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)^2\right) \|x_n - p_n\|^2$$

$$\leq 2\lambda_n \beta_n \left[\sup_{u \in C} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] + 2\lambda_n \langle u - p_n, w \rangle. \tag{15}$$

Proof From the definition of the resolvent we have $\frac{x_n - p_n}{\lambda_n} - \beta_n B x_n - D x_n \in A p_n$ for every $n \in \mathbb{N}$ and since $v \in Au$, the monotonicity of A guarantees

$$\langle p_n - u, x_n - p_n - \lambda_n(\beta_n B x_n + D x_n + v) \rangle \geq 0 \ \forall n \in \mathbb{N},$$

thus

$$\langle u - p_n, x_n - p_n \rangle \leq \langle u - p_n, \lambda_n \beta_n B x_n + \lambda_n D x_n + \lambda_n v \rangle \ \forall n \in \mathbb{N}.$$

By using the definition of x_{n+1} given in the algorithm we obtain

$$\langle u - p_n, x_n - p_n \rangle$$
$$\leq \langle u - p_n, x_{n+1} - p_n + \lambda_n \beta_n B p_n + \lambda_n D p_n + \lambda_n v \rangle$$
$$= \langle u - p_n, x_{n+1} - p_n \rangle + \lambda_n \beta_n \langle u - p_n, B p_n \rangle + \lambda_n \langle u - p_n, D p_n \rangle + \lambda_n \langle u - p_n, v \rangle \ \forall n \in \mathbb{N}.$$

From here it follows

$$\frac{1}{2} \|u - p_n\|^2 - \frac{1}{2} \|x_n - u\|^2 + \frac{1}{2} \|x_n - p_n\|^2$$
$$\leq \frac{1}{2} \|u - p_n\|^2 - \frac{1}{2} \|x_{n+1} - u\|^2 + \frac{1}{2} \|x_{n+1} - p_n\|^2$$
$$+ \lambda_n \beta_n \langle u - p_n, B p_n \rangle + \lambda_n \langle u - p_n, D p_n \rangle + \lambda_n \langle u - p_n, v \rangle \ \forall n \in \mathbb{N}.$$

Since $v = w - p - Du$ and due to the fact that D is monotone, we obtain for every $n \in \mathbb{N}$

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2$$
$$\leq \|x_{n+1} - p_n\|^2 - \|x_n - p_n\|^2 + 2\lambda_n \beta_n \left(\langle u, B p_n \rangle + \left\langle p_n, \frac{p}{\beta_n} \right\rangle - \langle p_n, B p_n \rangle - \left\langle u, \frac{p}{\beta_n} \right\rangle \right)$$
$$+ 2\lambda_n \langle u - p_n, D p_n - D u \rangle + 2\lambda_n \langle u - p_n, w \rangle$$
$$\leq \|x_{n+1} - p_n\|^2 - \|x_n - p_n\|^2 + 2\lambda_n \beta_n \left[\sup_{u \in C} \varphi_B \left(u, \frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right] + 2\lambda_n \langle u - p_n, w \rangle.$$

The conclusion follows, by noticing that the Lipschitz continuity of B and D yields

$$\|x_{n+1} - p_n\| \leq \frac{\lambda_n \beta_n}{\mu} \|x_n - p_n\| + \frac{\lambda_n}{\eta} \|x_n - p_n\| = \left(\frac{\lambda_n \beta_n}{\mu} + \frac{\lambda_n}{\eta} \right) \|x_n - p_n\| \ \forall n \in \mathbb{N}.$$

$\square$

The convergence of Algorithm 3 is stated below.

Theorem 18 *Let $(x_n)_{n \in \mathbb{N}}$ and $(p_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 3 and $(z_n)_{n \in \mathbb{N}}$ the sequence defined in (7). If (H_{fitz}) is fulfilled and $\limsup_{n \to +\infty} \left(\frac{\lambda_n \beta_n}{\mu} + \frac{\lambda_n}{\eta} \right)$ < 1, then $(z_n)_{n \in \mathbb{N}}$ converges weakly to an element in $\mathrm{zer}(A + D + N_C)$ as $n \to +\infty$.*

Proof The proof of the theorem relies on the following three statements:

(a)   for every $u \in \mathrm{zer}(A + D + N_C)$ the sequence $(\|x_n - u\|)_{n \in \mathbb{N}}$ is convergent;
(b)   every weak cluster point of $(z_n')_{n \in \mathbb{N}}$, where

$$z_n' = \frac{1}{\tau_n} \sum_{k=1}^{n} \lambda_k p_k \text{ and } \tau_n = \sum_{k=1}^{n} \lambda_k \ \forall n \in \mathbb{N},$$

lies in $\mathrm{zer}(A + D + N_C)$;
(c)   every weak cluster point of $(z_n)_{n \in \mathbb{N}}$ lies in $\mathrm{zer}(A + D + N_C)$.

In order to show (a) and (b) one has only to slightly adapt the proof of Theorem 6 and this is why we omit to give further details. For (c) it is enough to prove that $\lim_{n \to +\infty} \|z_n - z_n'\| = 0$ and the statement of the theorem will be a consequence of Lemma 1.

Taking $u \in \mathrm{zer}(A + D + N_C)$ and $w = 0 = v + p + Du$, where $v \in Au$ and $p \in N_C(u)$, from (15) we have

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \left(1 - \left(\frac{\lambda_n \beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)^2\right) \|x_n - p_n\|^2$$

$$\leq 2\lambda_n \beta_n \left[\sup_{u \in C} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right].$$

As $\limsup_{n \to +\infty} \left(\frac{\lambda_n \beta_n}{\mu} + \frac{\lambda_n}{\eta}\right) < 1$, we obtain by Lemma 2 that $\sum_{n \in \mathbb{N}} \|x_n - p_n\|^2 < +\infty$.
Moreover, for any $n \in \mathbb{N}$ it holds

$$\|z_n - z_n'\|^2 = \frac{1}{\tau_n^2} \left\|\sum_{k=1}^n \lambda_k (x_k - p_k)\right\|^2 \leq \frac{1}{\tau_n^2} \left(\sum_{k=1}^n \lambda_k \|x_k - p_k\|\right)^2$$

$$\leq \frac{1}{\tau_n^2} \left(\sum_{k=1}^n \lambda_k^2\right) \left(\sum_{k=1}^n \|x_k - p_k\|^2\right).$$

Since $(\lambda_n)_{n \in \mathbb{N}} \in \ell^2 \setminus \ell^1$, taking into consideration that $\tau_n = \sum_{k=1}^n \lambda_k \to +\infty$ $(n \to +\infty)$, we obtain $\|z_n - z_n'\| \to 0$ $(n \to +\infty)$. □

As it happens for the forward-backward penalty scheme, strong monotonicity of the operator A ensures strong convergence of the sequence $(x_n)_{n \in \mathbb{N}}$.

Theorem 19 *Let $(x_n)_{n \in \mathbb{N}}$ and $(p_n)_{n \in \mathbb{N}}$ be the sequences generated by Algorithm 3. If (H_{fitz}) is fulfilled, $\limsup_{n \to +\infty} \left(\frac{\lambda_n \beta_n}{\mu} + \frac{\lambda_n}{\eta}\right) < 1$ and the operator A is γ-strongly monotone with $\gamma > 0$, then $(x_n)_{n \in \mathbb{N}}$ converges strongly to the unique element in $\mathrm{zer}(A + D + N_C)$ as $n \to +\infty$.*

Proof Let be $u \in \mathrm{zer}(A + D + N_C)$ and $w = 0 = v + p + Du$, where $v \in Au$ and $p \in N_C(u)$. Following the lines of the proof of Lemma 17 one can easily show that

$$2\gamma \lambda_n \|p_n - u\|^2 + \|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \left(1 - \left(\frac{\lambda_n \beta_n}{\mu} + \frac{\lambda_n}{\eta}\right)^2\right) \|x_n - p_n\|^2$$

$$\leq 2\lambda_n \beta_n \left[\sup_{u \in C} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] \ \forall n \in \mathbb{N}.$$

The hypotheses imply the existence of $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$

$$2\gamma \lambda_n \|p_n - u\|^2 + \|x_{n+1} - u\|^2 - \|x_n - u\|^2 \leq 2\lambda_n \beta_n \left[\sup_{u \in C} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right].$$

As in the proof of Theorem 7, from here it follows that

$$\sum_{n \in \mathbb{N}} \lambda_n \|p_n - u\|^2 < \infty.$$

Since $(\lambda_n)_{n \in \mathbb{N}}$ is bounded above and $\sum_{n \in N} \|x_n - p_n\|^2 < +\infty$ (see the proof of Theorem 18), it yields

$$\sum_{n=1}^\infty \lambda_n \|x_n - u\|^2 \leq 2 \sum_{n=1}^\infty \lambda_n \|x_n - p_n\|^2 + 2 \sum_{n=1}^\infty \lambda_n \|p_n - u\|^2 < +\infty.$$

As $\sum_{n\in\mathbb{N}} \lambda_n = +\infty$ and $(\|x_n - u\|)_{n\in\mathbb{N}}$ is convergent, it follows $\lim_{n\to+\infty} \|x_n - u\|$ $= 0$. $\square$

3.2 Monotone inclusion problems involving compositions with linear continuous operators

In this subsection we will show that the Tseng's type penalty scheme investigated in the previous section allows the solving of monotone inclusion problems with a more intricate formulation. The problem under consideration is the following.

Problem 20 Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, $A_1 : \mathcal{H} \rightrightarrows \mathcal{H}$ and $A_2 : \mathcal{G} \rightrightarrows \mathcal{G}$ maximally monotone operators, $K : \mathcal{H} \to \mathcal{G}$ linear continuous operator, $D : \mathcal{H} \to \mathcal{H}$ a monotone and η^{-1}-Lipschitz continuous operator with $\eta > 0$, $B : \mathcal{H} \to \mathcal{H}$ a monotone and μ^{-1}-Lipschitz continuous operator with $\mu > 0$ and suppose that $C = \operatorname{zer} B \neq \emptyset$. The monotone inclusion problem to solve is

$$0 \in A_1 x + K^* A_2 K x + Dx + N_C(x).$$

We use the product space approach (see [4, 10–12, 15]) in order to show that the above problem can be reformulated as the monotone inclusion problem treated in Subsection 3.1 in an appropriate product space. To this end we consider the real Hilbert space $\mathcal{H} \times \mathcal{G}$ endowed with the inner product

$$\langle (x, v), (x', v') \rangle_{\mathcal{H}\times\mathcal{G}} = \langle x, x' \rangle_{\mathcal{H}} + \langle v, v' \rangle_{\mathcal{G}} \ \forall (x, v), (x', v') \in \mathcal{H} \times \mathcal{G}$$

and corresponding norm. We define the following operators on $\mathcal{H} \times \mathcal{G}$. For $(x, v) \in \mathcal{H} \times \mathcal{G}$ we set

$$\widetilde{A}(x, v) = A_1 x \times A_2^{-1} v, \quad \widetilde{D}(x, v) = (Dx + K^* v, -Kx), \quad \widetilde{B}(x, v) = (Bx, 0)$$

and, for $\widetilde{C} = C \times \mathcal{G} = \operatorname{zer} \widetilde{B}$,

$$N_{\widetilde{C}}(x, v) = N_C(x) \times \{0\}.$$

One can easily show that if $(x, v) \in \operatorname{zer}(\widetilde{A}+\widetilde{D}+N_{\widetilde{C}})$, then $x \in \operatorname{zer}(A_1+K^* A_2 K+D+N_C)$. Conversely, when $x \in \operatorname{zer}(A_1 + K^* A_2 K + D + N_C)$, then exists $v \in A_2 K x$ such that $(x, v) \in \operatorname{zer}(\widetilde{A} + \widetilde{D} + N_{\widetilde{C}})$. Thus, determining the zeros of the operator $\widetilde{A} + \widetilde{D} + N_{\widetilde{C}}$ will provide a solution for the monotone inclusion problem in Problem 20.

One has that $\widetilde{A}$ is maximally monotone (see [4, Proposition 20.23]), $\widetilde{D}$ is monotone and $\widetilde{\eta}$-Lipschitz continuous, where $\widetilde{\eta} = \sqrt{2\left(\frac{1}{\eta^2} + \|K\|^2\right)}$ and $\widetilde{B}$ is monotone and μ^{-1}-Lipschitz continuous. All these considerations show that we are in the context of Problem 15, thus, in order to determine the zeros of $\widetilde{A} + \widetilde{D} + N_{\widetilde{C}}$, we can use Algorithm 3, the iterations of which read for any $n \in \mathbb{N}$ as follows:

$$\begin{cases} (p_n, q_n) = J_{\lambda_n \widetilde{A}} \left[(x_n, v_n) - \lambda_n \widetilde{D}(x_n, v_n) - \lambda_n \beta_n \widetilde{B}(x_n, v_n) \right] \\ (x_{n+1}, v_{n+1}) = \lambda_n \beta_n \left[\widetilde{B}(x_n, v_n) - \widetilde{B}(p_n, q_n) \right] + \lambda_n \left[\widetilde{D}(x_n, v_n) - \widetilde{D}(p_n, q_n) \right] + (p_n, q_n). \end{cases}$$

Since $J_{\lambda \widetilde{A}}(x, v) = (J_{\lambda A_1}(x), J_{\lambda A_2^{-1}}(v))$ for every $(x, v) \in \mathcal{H} \times \mathcal{G}$ and every $\lambda > 0$ (see [4, Proposition 23.16]), this gives rise to the following iterative scheme.

Algorithm 4

> Initialization: Choose $(x_1, v_1) \in \mathcal{H} \times \mathcal{G}$
>
> For $n \in \mathbb{N}$ set: $\quad p_n = J_{\lambda_n A_1}\Big(x_n - \lambda_n(Dx_n + K^*v_n) - \lambda_n\beta_n Bx_n\Big)$
>
> $\qquad\qquad\qquad q_n = J_{\lambda_n A_2^{-1}}(v_n + \lambda_n Kx_n)$
>
> $\qquad\qquad\qquad x_{n+1} = \lambda_n\beta_n(Bx_n - Bp_n) + \lambda_n(Dx_n - Dp_n) + \lambda_n K^*(v_n - q_n) + p_n$
>
> $\qquad\qquad\qquad v_{n+1} = \lambda_n K(p_n - x_n) + q_n,$

where $(\lambda_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ are sequences of positive real numbers. For the convergence of this iterative scheme the following hypotheses are needed:

$$\left(H_{fitz}^{comp}\right) \begin{cases} (i) \ A_1 + N_C \text{ is maximally monotone and } \mathrm{zer}(A_1 + K^*A_2K + D + N_C) \neq \emptyset; \\ (ii) \ \text{For each } p \in \mathrm{ran}(N_C), \ \sum_{n\in\mathbb{N}} \lambda_n\beta_n \left[\sup_{u\in C} \varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right] < +\infty; \\ (iii) \ (\lambda_n)_{n\in\mathbb{N}} \in \ell^2 \setminus \ell^1. \end{cases}$$

One will see that $\left(H_{fitz}^{comp}\right)$ implies the hypotheses (H_{fitz}) formulated in the context of the monotone inclusion problem of finding the zeros of $\widetilde{A} + \widetilde{D} + N_{\widetilde{C}}$. Indeed, hypothesis (i) in $\left(H_{fitz}^{comp}\right)$ guarantees that $\widetilde{A} + N_{\widetilde{C}}$ is maximally monotone and, as already seen, that $\mathrm{zer}(\widetilde{A} + \widetilde{D} + N_{\widetilde{C}}) \neq \emptyset$. Further, we have for all $(x, v), (x', v') \in \mathcal{H} \times \mathcal{G}$ that

$$\varphi_{\widetilde{B}}\big((x, v), (x', v')\big) = \begin{cases} \varphi_B(x, x'), & \text{if } v' = 0, \\ +\infty, & \text{otherwise} \end{cases}$$

and

$$\sigma_{\widetilde{C}}(x, v) = \begin{cases} \sigma_C(x), & \text{if } v = 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Moreover, $\mathrm{ran}\, N_{\widetilde{C}} = \mathrm{ran}\, N_C \times \{0\}$. Hence, condition (ii) in $\left(H_{fitz}^{comp}\right)$ is nothing else than

for each $(p, p') \in \mathrm{ran}(N_{\widetilde{C}})$, $\displaystyle\sum_{n\in\mathbb{N}} \lambda_n\beta_n \left[\sup_{(u,u')\in\widetilde{C}} \varphi_{\widetilde{B}}\left((u, u'), \frac{(p, p')}{\beta_n}\right) - \sigma_{\widetilde{C}}\left(\frac{(p, p')}{\beta_n}\right)\right] < +\infty.$

The following convergence statement is a direct consequence of Theorem 18 and Theorem 19.

Theorem 22 *Let $(x_n)_{n\in\mathbb{N}}$, $(v_n)_{n\in\mathbb{N}}$, $(p_n)_{n\in\mathbb{N}}$ and $(q_n)_{n\in\mathbb{N}}$ be the sequences generated by Algorithm 4 and $(z_n)_{n\in\mathbb{N}}$ the sequence defined in (7). If $\left(H_{fitz}^{comp}\right)$ is fulfilled and*

$$\limsup_{n\to+\infty} \left(\frac{\lambda_n\beta_n}{\mu} + \lambda_n\sqrt{2\left(\frac{1}{\eta^2} + \|K\|^2\right)}\right) < 1,$$

then $(z_n)_{n\in\mathbb{N}}$ converges weakly to an element in $\mathrm{zer}(A_1 + K^*A_2K + D + N_C)$ as $n \to +\infty$. If, additionally, A_1 and A_2^{-1} are strongly monotone, then $(x_n)_{n\in\mathbb{N}}$ converges strongly to the unique element in $\mathrm{zer}(A_1 + K^*A_2K + D + N_C)$ as $n \to +\infty$.

Remark 23 We applied the forward-backward-forward penalty scheme in the product space in order to solve monotone inclusion problems where also compositions with linear continuous operators are involved. Let us underline the fact that, even in the situation when B is

cocoercive and, hence, $\widetilde{B}$ is cocoercive, the forward-backward penalty scheme in Algorithm 2 cannot be applied in this context, because the operator $\widetilde{D}$ is definitely not cocoercive. This is due to the presence of the skew operator $(x, v) \mapsto (K^*v, -Kx)$ in its definition. This fact provides a good motivation for formulating, along the forward-backward penalty scheme, a forward-backward-forward penalty scheme for the monotone inclusion problem investigated in this paper.

Remark 24 In the particular case $Dx = 0$ for every $x \in \mathcal{H}$ and $Bx = 0$ for every $x \in \mathcal{H}$ (which corresponds to the situation when $N_C(x) = \{0\}$ for every $x \in \mathcal{H}$) Algorithm 4 turns out to be the error-free case of the forward-backward-forward scheme proposed and investigated from the point of view of its convergence in [12, Theorem 3.1].

References

1. Attouch, H., Czarnecki, M.-O.: Asymptotic behavior of coupled dynamical systems with multiscale aspects. J. Differ. Equ. **248**(6), 1315–1344 (2010)
2. Attouch, H., Czarnecki, M.-O., Peypouquet, J.: Prox-penalization and splitting methods for constrained variational problems. SIAM J. Optim. **21**(1), 149–173 (2011)
3. Attouch, H., Czarnecki, M.-O., Peypouquet, J.: Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities. SIAM J. Optim. **21**(4), 1251–1274 (2011)
4. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, CMS Books in Mathematics. Springer, New York (2011)
5. Bauschke, H.H., McLaren, D.A., Sendov, H.S.: Fitzpatrick functions: inequalities, examples and remarks on a problem by S. Fitzpatrick. J. Convex Anal. **13**(3–4), 499–523 (2006)
6. Borwein, J.M.: Maximal monotonicity via convex analysis. J. Convex Anal. **13**(3–4), 561–586 (2006)
7. Borwein, J.M., Vanderwerff, J.D.: Convex Functions: Constructions, Characterizations and Counter-examples. Cambridge University Press, Cambridge (2010)
8. Boţ, R.I.: Conjugate Duality in Convex Optimization: Lecture Notes in Economics and Mathematical Systems, vol. 637. Springer, Berlin Heidelberg (2010)
9. Boţ, R.I., Csetnek, E.R.: An application of the bivariate inf-convolution formula to enlargements of monotone operators. Set-Valued Anal. **16**(7–8), 983–997 (2008)
10. Boţ, R.I., Csetnek, E.R., Heinrich, A.: A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators. SIAM J. Optim. **23**(4), 2011–2036 (2013)
11. Boţ, R.I., Hendrich, C.: Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization. J. Math. Imaging Vision. (2014) doi:10.1007/s10851-013-0486-8
12. Briceño-Arias, L.M., Combettes, P.L.: A monotone + skew splitting model for composite monotone inclusions in duality. SIAM J. Optim. **21**(4), 1230-1250 (2011)
13. Burachik, R.S., Svaiter, B.F.: Maximal monotone operators, convex functions and a special family of enlargements. Set-Valued Anal. **10**(4), 297–316 (2002)
14. Combettes, P.L.: Solving monotone inclusions via compositions of nonexpansive averaged operators. Optim. **53**(5–6), 475–504 (2004)
15. Combettes, P.L., Pesquet, J.-C.: Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. Set-Valued Var. Anal. **20**(2), 307–330 (2012)
16. Ekeland, I., Temam, R.: Convex Analysis and Variational Problems. North-Holland Publishing Company, Amsterdam (1976)
17. Fitzpatrick, S.: Representing monotone operators by convex functions. In: Workshop/Miniconference on Functional Analysis and Optimization (Canberra, 1988), Proceedings of the Centre for Mathematical Analysis 20, Australian National University, Canberra, pp. 59–65 (1988)
18. Noun, N., Peypouquet, J.: Forward-backward penalty scheme for constrained convex minimization without inf-compactness. J. Optim. Theory Appl. **158**(3), 787–795 (2013)

19. Peypouquet, J.: Coupling the gradient method with a general exterior penalization scheme for convex minimization. J. Optim. Theory Appl. **153**(1), 123–138 (2012)
20. Rockafellar, R.T.: On the maximal monotonicity of subdifferential mappings. Pac. J. Math. **33**(1), 209–216 (1970)
21. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J Control. Optim. **14**(5), 877–898 (1976)
22. Simons, S.: From Hahn-Banach to Monotonicity. Springer, Berlin (2008)
23. Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. SIAM J. Control. Optim. **38**(2), 431–446 (2000)
24. Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. Adv. Comput. Math. **38**(3), 667–681 (2013)
25. Zălinescu, C.: Convex Analysis in General Vector Spaces. World Scientific, Singapore (2002)