Noname manuscript No.
(will be inserted by the editor)

Backward Penalty Schemes for Monotone Inclusion Problems

Sebastian Banert · **Radu Ioan Boţ**

Received: date / Accepted: date

Abstract In this paper, we are concerned with solving monotone inclusion problems expressed by the sum of a set-valued maximally monotone operator with a single-valued maximally monotone one and the normal cone to the nonempty set of zeros of another set-valued maximally monotone operator. Depending on the nature of the single-valued operator, we propose two iterative penalty schemes, both addressing the set-valued operators via backward steps. The single-valued operator is evaluated via a single forward step if it is cocoercive, and via two forward steps if it is monotone and Lipschitz continuous. The latter situation represents the starting point for dealing with complexly structured monotone inclusion problems from algorithmic point of view.

Keywords Backward penalty algorithm · Monotone inclusion · Maximally monotone operator · Fitzpatrick function · Convex subdifferential

Mathematics Subject Classification (2010) 47H05 · 65K05 · 90C25

The research of R.I. Boţ was partially supported by the German Research Foundation (DFG), project BO 2516/4-1.

S. Banert
University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
E-mail: sebastian.banert@univie.ac.at

R.I. Boţ (Corresponding author)
University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria
E-mail: radu.bot@univie.ac.at

1 Introduction

In this article, we address the algorithmic solution of variational inequalities stated as monotone inclusions expressed by the sum of a set-valued maximally monotone operator with a single-valued maximally monotone one and the normal cone to the nonempty set of zeros of another set-valued maximally monotone operator. In the case that the operators involved are subdifferentials, this problem particularizes to a convex bilevel optimization problem, where the feasible set of the upper-level problem is the set of solutions of the lower-level problem, and both are convex.

Our algorithmic scheme is based on the multiscale dynamical system considered in [1], which led to several algorithms for treating problems of different generality [2,3,4,5,6,7]. All these approaches have in common that they use backward (proximal) and forward (gradient) steps to evaluate the involved operators depending on their regularity, and the stepsizes of the upper- and lower-level problems are different to force constraint satisfaction. Typically, one can show weak ergodic convergence of the iterates to a solution of the problem and norm convergene under stronger monotonicity assumptions.

The investigations in the present article complement the ones made in [6], where, for the first time, the constraint set of the variational inequality was allowed to be given by a general monotone operator instead of a subdifferential. For this, the authors gave a new hypothesis in terms of the Fitzpatrick function associated to the operator. However, the operator was assumed to be single-valued and was evaluated in the iterative scheme via a forward step, whereas we only assume maximal monotonicity and address it accordingly by means of its resolvent.

Depending on the nature of the single-valued operator in the variational inequality under investigation, we propose two numerical schemes for both of which we undertake a convergence analysis. By assuming cocoercivity for this operator, we show an approach with one forward evaluation per iteration, on the other hand, if it is (only) monotone and Lipschitz continuous, it is evaluated twice. The latter scheme and the convergence statements provided in this context constitute a starting point for solving complexly structured variational inequalities, involving mixtures of sums of maximally monotone operators.

We close the paper by discussing the fulfillment of the assumption expressed via the Fitzpatrick function associated with some particular instances of the operator defining the constraint set.

2 Notation and Preliminary Results

For the reader's convenience we present first some notations which are used throughout the paper (see [8,9,10,11,12,13]). By $\mathbb{N} = \{1, 2, \ldots\}$ we denote the set of *positive integer numbers*. 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 $\rightarrow$ denote weak and strong convergence, respectively. When $\mathcal{G}$ is another Hilbert space and $K : \mathcal{H} \rightarrow \mathcal{G}$ is a continuous linear operator, then we define by $\|K\| := \sup \{\|Kx\| \mid x \in \mathcal{H}, \|x\| \leq 1\}$ its *norm*, while $K^* : \mathcal{G} \rightarrow \mathcal{H}$, defined by $\langle K^*y, x\rangle = \langle y, Kx\rangle$ for all $(x, y) \in \mathcal{H} \times \mathcal{G}$, denotes its *adjoint operator*.

For a function $f : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ we denote by $\operatorname{dom} f := \{x \in \mathcal{H} \mid f(x) < +\infty\}$ its *effective domain* and say that f is *proper* iff $\operatorname{dom} f \neq \emptyset$ and $f(x) \neq -\infty$ for all $x \in \mathcal{H}$. The *conjugate function* of f will be denoted by $f^* : \mathcal{H} \rightarrow \overline{\mathbb{R}}$, $f^*(u) := \sup \{\langle u, x\rangle - f(x) \mid x \in \mathcal{H}\}$ for all $u \in \mathcal{H}$. The *subdifferential* of f at $x \in \mathcal{H}$, with $f(x) \in \mathbb{R}$, is the set $\partial f(x) := \{v \in \mathcal{H} \mid f(y) \geq f(x) + \langle v, y - x\rangle \text{ for all } y \in \mathcal{H}\}$. We take by convention $\partial f(x) := \emptyset$ if $f(x) \in \{\pm\infty\}$. We also denote by $\arg\min f$ the set of global minima of the function f and set $\min f := \inf\{f(x)|x \in \arg\min f\}$. The *infimal convolution* of two functions $f, g : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ is defined as

$$(f \,\square\, g)(x) := \inf \{f(y) + g(x - y) \mid y \in \mathcal{H}\},$$

and we have $(f \,\square\, g)^* = f^* + g^*$ (see, e.g., [8, Proposition 13.21]).

Let $S \subseteq \mathcal{H}$ be a nonempty set. The *indicator function* of S, $\delta_S : \mathcal{H} \rightarrow \overline{\mathbb{R}}$, is the function which takes the value 0 on S and $+\infty$ elsewhere. The subdifferential of the indicator function is the *normal cone* of S, that is, $N_S(x) = \{u \in \mathcal{H} \mid \langle u, y - x\rangle \leq 0 \text{ for all } y \in S\}$ 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 $\sigma_S(u) = \langle u, x\rangle$, where σ_S is the *support function* of S, defined by $\sigma_S(u) := \sup\{\langle u, y\rangle \mid y \in S\}$.

For a set-valued operator $M : \mathcal{H} \rightrightarrows \mathcal{H}$ we denote by $\operatorname{Graph} M := \{(x,u) \in \mathcal{H} \times \mathcal{H} | u \in Mx\}$ its *graph*, by $\operatorname{Dom} M := \{x \in \mathcal{H} \mid Mx \neq \emptyset\}$ its *domain*, by $\operatorname{Ran} M := \bigcup\{Mx \mid x \in \mathcal{H}\}$ its *range* and by $M^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ its *inverse operator*, defined by $(u,x) \in \operatorname{Graph} M^{-1}$ if and only if $(x,u) \in \operatorname{Graph} M$. The *parallel sum* of two set-valued operators $M_1, M_2 : \mathcal{H} \rightrightarrows \mathcal{H}$ is denoted by $M_1 \,\square\, M_2 := \left(M_1^{-1} + M_2^{-1}\right)^{-1}$.

We also use the notation $\operatorname{zer} M := \{x \in \mathcal{H} \mid 0 \in Mx\}$ for the *set of zeros* of the operator M. We say that M is *monotone* iff $\langle x - y, u - v\rangle \geq 0$ for all $(x,u), (y,v) \in \operatorname{Graph} M$. A monotone operator M is said to be *maximally monotone* iff there exists no proper monotone extension of the graph of M on $\mathcal{H} \times \mathcal{H}$. Let us mention that in the case M is maximally monotone, one has the following characterization for the set of its zeros.

$$z \in \operatorname{zer} M \qquad \text{if and only if} \qquad \langle w, u - z\rangle \geq 0 \text{ for all } (u,w) \in \operatorname{Graph} M. \tag{1}$$

The operator M is said to be *strongly monotone with parameter* $\gamma > 0$ or γ*-strongly monotone* iff $\langle x - y, u - v\rangle \geq \gamma \|x - y\|^2$ for all $(x,u), (y,v) \in \operatorname{Graph} M$. Notice that if M is maximally monotone and strongly monotone (with a given parameter), then $\operatorname{zer} M$ is a singleton, thus nonempty (see [8, 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 $x \in \mathcal{H}$, denotes the *identity operator* on $\mathcal{H}$. If M is maximally monotone, then $J_M : \mathcal{H} \to \mathcal{H}$ is single-valued and maximally monotone (cf. [8, Proposition 23.7 and Corollary 23.10]). For an arbitrary $\gamma > 0$ we have (see [8, Proposition 23.18])

$$J_{\gamma M} + \gamma J_{\gamma^{-1} M^{-1}} \circ \gamma^{-1}\operatorname{Id} = \operatorname{Id}. \tag{2}$$

For the convergence statements that we provide in this paper we will assume that some hypotheses, one of them expressed in terms of the Fitzpatrick function associated to a certain maximally monotone operator, are fulfilled. In the following we will recall some properties of this function, which brought new and deep insights into the field of maximally monotone operators in the last decade (see [8,14,15,9,10,16,17,18,12] and the references therein). The *Fitzpatrick function* associated to a monotone operator M, defined as

$$\varphi_M : \mathcal{H} \times \mathcal{H} \to \overline{\mathbb{R}}, \qquad \varphi_M(x,u) := \sup\left\{\langle x,v\rangle + \langle y,u\rangle - \langle y,v\rangle \mid (y,v) \in \operatorname{Graph} M\right\},$$

is a convex and lower semicontinuous function. In case M is maximally monotone, φ_M is proper and it fulfills

$$\varphi_M(x,u) \geq \langle x,u\rangle \qquad \text{for all } (x,u) \in \mathcal{H} \times \mathcal{H},$$

with equality if and only if $(x,u) \in \operatorname{Graph} M$. Notice that if $f : \mathcal{H} \to \overline{\mathbb{R}}$ is a proper, convex and lower semicontinuous function, then ∂f is a maximally monotone operator (cf. [19]) and it holds $(\partial f)^{-1} = \partial f^*$. Furthermore, the following inequality is true (see [14])

$$\varphi_{\partial f}(x,u) \leq f(x) + f^*(u) \qquad \text{for all } (x,u) \in \mathcal{H} \times \mathcal{H}. \tag{3}$$

We refer the reader to [14] for formulas 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* iff $\langle x-y, Mx-My\rangle \geq \gamma \|Mx-My\|^2$ for all $(x,y) \in \mathcal{H} \times \mathcal{H}$, and γ-*Lipschitz continuous* iff $\|Mx-My\| \leq \gamma \|x-y\|$ 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* iff $\langle x, Mx\rangle = 0$ for all $x \in \mathcal{H}$.

We close the section by presenting some convergence results which will be used when carrying out a convergence analysis for the iterative schemes provided in the paper. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ and $(\lambda_k)_{k\in\mathbb{N}}$ be 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])

$$z_n := \frac{1}{\tau_n}\sum_{k=1}^{n} \lambda_k x_k, \qquad \text{where } \tau_n := \sum_{k=1}^{n} \lambda_k \qquad \text{for all } n \in \mathbb{N}. \tag{4}$$

Lemma 2.1 (Opial–Passty) *Let F be a nonempty subset of $\mathcal{H}$ and $(x_n)_{n\geq 0}$ be a sequence in $\mathcal{H}$ such that the limit $\lim_{n\to+\infty} \|x_n - x\|$ exists for every $x \in F$.*

(i) If every weak cluster point of $(x_n)_{n\geq 0}$ lies in F, then $(x_n)_{n\geq 0}$ converges weakly to an element in F as $n \to +\infty$.

(ii) If every weak cluster point of $(z_n)_{n\in\mathbb{N}}$ lies in F, then $(z_n)_{n\in\mathbb{N}}$ converges weakly to an element in F as $n \to +\infty$.

The following result is taken from [3, Lemma 2].

Lemma 2.2 *Let $(a_n)_{n\geq 0}$, $(b_n)_{n\geq 0}$ and $(\varepsilon_n)_{n\geq 0}$ be real sequences. Assume that $(a_n)_{n\geq 0}$ is bounded from below, $(b_n)_{n\geq 0}$ is nonnegative, $(\varepsilon_n)_{n\geq 0} \in \ell^1$ and $a_{n+1} - a_n + b_n \leq \varepsilon_n$ for any $n \geq 0$. Then $(a_n)_{n\geq 0}$ is convergent and $(b_n)_{n\geq 0} \in \ell^1$.*

3 A Backward Penalty Scheme with One Forward Step

The problem we deal with in this section has the following formulation.

Problem 3.1 Let $\mathcal{H}$ be a real Hilbert space, $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone operators, $D : \mathcal{H} \to \mathcal{H}$ be an η-cocoercive operator with $\eta > 0$ and suppose that $C := \operatorname{zer} B \neq \emptyset$. The monotone inclusion problem to solve is

$$0 \in Ax + Dx + N_C(x). \tag{5}$$

We propose for solving Problem 3.1 the following iteration scheme which has the particularity that it evaluates an appropriate penalization of the operator B via a backward step.

Algorithm 3.1 *Choose $x_0 \in \mathcal{H}$ and set for any $n \geq 1$:*

$$\begin{aligned} y_{n-1} &:= x_{n-1} - \lambda_n D x_{n-1}, \\ w_n &:= J_{\lambda_n A} y_{n-1}, \\ x_n &:= J_{\lambda_n \beta_n B} w_n, \end{aligned}$$

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

Remark 3.1 (a) The algorithmic treatment of Problem 3.1 is related to the solving of convex optimization problems of the form

$$\min_{x \in \arg\min \Psi} \Phi(x) + \Gamma(x),$$

where $\Phi, \Psi : \mathcal{H} \to \overline{\mathbb{R}}$ are proper, convex and lower semicontinuous functions with $\arg\min \Psi \neq \emptyset$, and $\Gamma : \mathcal{H} \to \mathbb{R}$ is a convex and differentiable function with Lipschitz continuous gradient. Different to the most common splitting algorithms, which rely on the calculation of the projection on the set $\arg\min \Psi$, the iterative scheme described in Algorithm 3.1 assumes an explicit evaluation of the function Ψ in terms of its proximal operator.

(b) 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 iterative scheme in Algorithm 3.1 reduces to the algorithm proposed and investigated in [2] for solving the monotone inclusion problem

$$0 \in Ax + N_{\arg\min \Psi}(x). \tag{6}$$

(c) Another penalty scheme for solving the monotone inclusion problem (5), in the case B is a cocoercive operator, which evaluates both B and D via forward steps and A via a backward step has been introduced and investigated from the point of view of its convergence properties in [6]. The mentioned algorithm is an extension of a numerical method proposed in [3] in the context of solving (6) when Ψ is, in addition, differentiable with Lipschitz continuous gradient.

The following lemma will be crucial for proving the convergence of Algorithm 3.1.

Lemma 3.1 *For $u \in C \cap \operatorname{dom} A$ take $w \in (A + D + N_C)(u)$ such that $w = v + Du + p$ for some $v \in Au$ and $p \in N_C(u)$. For each $n \in \mathbb{N}$, the following inequality holds:*

$$\begin{aligned}
&\|x_n - u\|^2 - \|x_{n-1} - u\|^2 + \lambda_n(2\eta - \lambda_n)\|Dx_{n-1} - Du\|^2 + \\
&\frac{1}{2}\|x_n - w_n\|^2 + \frac{1}{2}\|x_n - w_n - \lambda_n(Du + v)\|^2 + \|x_{n-1} - w_n - \lambda_n(Dx_{n-1} - Du)\|^2 \leq \\
&2\lambda_n\beta_n\left(\sup_{\tilde{u}\in C}\varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) + 2\lambda_n\langle w, u - x_n\rangle + 2\lambda_n^2\|Du + v\|^2.
\end{aligned}$$

Proof. Let be $n \geq 1$ fixed. We have $\lambda_n v \in \lambda_n Au$ and $y_{n-1} - w_n \in \lambda_n Aw_n$, so, by the monotonicity of A,

$$\langle y_{n-1} - w_n - \lambda_n v, w_n - u\rangle \geq 0, \tag{7}$$

which is equivalent to

$$2\lambda_n\langle v, w_n - u\rangle \leq 2\langle y_{n-1} - w_n, w_n - u\rangle = \|y_{n-1} - u\|^2 - \|y_{n-1} - w_n\|^2 - \|u - w_n\|^2. \tag{8}$$

Furthermore, we have $w_n - x_n \in \lambda_n \beta_n B x_n$, so, by definition of the Fitzpatrick function and noticing that $\sigma_C\left(u, \frac{p}{\beta_n}\right) = \left\langle \frac{p}{\beta_n}, u \right\rangle$,

$$\begin{aligned}
&2\lambda_n\beta_n\left(\sup_{\tilde{u}\in C}\varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) \geq 2\lambda_n\beta_n\left(\varphi_B\left(u, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right)\\
&\geq 2\left\langle u, w_n - x_n\right\rangle + 2\lambda_n\left\langle p, x_n\right\rangle - 2\left\langle x_n, w_n - x_n\right\rangle - 2\lambda_n\left\langle p, u\right\rangle\\
&= 2\left\langle u - x_n, w_n - x_n\right\rangle + 2\lambda_n\left\langle p, x_n - u\right\rangle\\
&= \left\|u - x_n\right\|^2 + \left\|x_n - w_n\right\|^2 - \left\|u - w_n\right\|^2 + 2\lambda_n\left\langle p, x_n - u\right\rangle. \qquad (9)
\end{aligned}$$

Adding (8) and (9), we obtain (recall that $w = v + Du + p$)

$$\begin{aligned}
&\left\|x_n - u\right\|^2 - \left\|x_{n-1} - u\right\|^2 - 2\lambda_n\beta_n\left(\sup_{\tilde{u}\in C}\varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) - 2\lambda_n\left\langle w, u - x_n\right\rangle\\
\leq{}& \left\|u - y_{n-1}\right\|^2 - \left\|x_{n-1} - u\right\|^2 - \left\|y_{n-1} - w_n\right\|^2 - \left\|x_n - w_n\right\|^2 + 2\lambda_n\left\langle v, x_n - w_n\right\rangle\\
&+ 2\lambda_n\left\langle Du, x_n - u\right\rangle\\
={}& \left\|u - x_{n-1} + \lambda_n Dx_{n-1}\right\|^2 - \left\|u - x_{n-1}\right\|^2 - \left\|x_{n-1} - w_n - \lambda_n Dx_{n-1}\right\|^2 - \left\|x_n - w_n\right\|^2\\
&+ 2\lambda_n\left\langle v, x_n - w_n\right\rangle + 2\lambda_n\left\langle Du, x_n - u\right\rangle\\
={}& 2\lambda_n\left\langle Dx_{n-1}, u - x_{n-1}\right\rangle - \left\|x_{n-1} - w_n\right\|^2 + 2\lambda_n\left\langle Dx_{n-1}, x_{n-1} - w_n\right\rangle - \left\|x_n - w_n\right\|^2\\
&+ 2\lambda_n\left\langle v, x_n - w_n\right\rangle + 2\lambda_n\left\langle Du, x_n - u\right\rangle\\
={}& 2\lambda_n\left\langle Dx_{n-1}, u - w_n\right\rangle - \left\|x_{n-1} - w_n\right\|^2 - \left\|x_n - w_n\right\|^2\\
&+ 2\lambda_n\left\langle v, x_n - w_n\right\rangle + 2\lambda_n\left\langle Du, x_n - u\right\rangle\\
={}& 2\lambda_n\left\langle Dx_{n-1} - Du, u - x_{n-1}\right\rangle + 2\lambda_n\left\langle Dx_{n-1}, x_{n-1} - w_n\right\rangle - \left\|x_{n-1} - w_n\right\|^2\\
&- \left\|x_n - w_n\right\|^2 + 2\lambda_n\left\langle v, x_n - w_n\right\rangle + 2\lambda_n\left\langle Du, x_n - x_{n-1}\right\rangle\\
\leq{}& -2\eta\lambda_n\left\|Dx_{n-1} - Du\right\|^2 + 2\lambda_n\left\langle Dx_{n-1} - Du, x_{n-1} - w_n\right\rangle - \left\|x_{n-1} - w_n\right\|^2\\
&- \left\|x_n - w_n\right\|^2 + 2\lambda_n\left\langle Du + v, x_n - w_n\right\rangle\\
={}& -\left\|x_{n-1} - w_n - \lambda_n(Dx_{n-1} - Du)\right\|^2 - \lambda_n(2\eta - \lambda_n)\left\|Dx_{n-1} - Du\right\|^2\\
&- \frac{1}{2}\left\|x_n - w_n - \lambda_n(Du + v)\right\|^2 - \frac{1}{2}\left\|x_n - w_n\right\|^2 + 2\lambda_n^2\left\|Du + v\right\|^2.
\end{aligned}$$

From here the conclusion is straightforward. ■

For the convergence statement of Algorithm 3.1, the following hypotheses are needed:

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

Remark 3.2 Some comments with respect to the hypotheses $(\mathrm{H_{fitz}})$ are in order.

(a) The hypotheses $(\mathrm{H_{fitz}})$ have already been used in [6] when showing the convergence of the iterative scheme proposed for solving (5) when B is a cocoercive operator. Still there it was pointed out that, since D is cocoercive and $\operatorname{dom}D=\mathcal{H}$, $A+D+N_C$ is maximally monotone, while, in the light of the properties of the Fitzpatrick function, for every $p\in\operatorname{Ran}N_C$ and any $n\in\mathbb{N}$ one has

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

(b) The convergence of the penalty iterative scheme proposed in [2] for solving the monotone inclusion problem (6), where $\Psi:\mathcal{H}\to\overline{\mathbb{R}}$ is a proper, convex and lower semicontinuous function with $\min\Psi=0$, have been shown under the following hypotheses:

$$\left.\begin{array}{rl} \text{(i)} & A+N_C \text{ is maximally monotone and } \operatorname{zer}(A+D+N_C)\neq\emptyset; \\ \text{(ii)} & \text{For every } p\in\operatorname{Ran}N_C,\ \displaystyle\sum_{n\in\mathbb{N}}\lambda_n\beta_n\left(\Psi^*\left(\frac{p}{\beta_n}\right)-\sigma_C\left(\frac{p}{\beta_n}\right)\right)<+\infty; \\ \text{(iii)} & (\lambda_n)_{n\in\mathbb{N}}\in\ell^2\setminus\ell^1. \end{array}\right\} \quad (\mathrm{H})$$

According to (3) it holds

$$\sum_{n\in\mathbb{N}}\lambda_n\beta_n\left(\sup_{\tilde u\in C}\varphi_{\partial\Psi}\left(\tilde u,\frac{p}{\beta_n}\right)-\sigma_C\left(\frac{p}{\beta_n}\right)\right)\leq\sum_{n\in\mathbb{N}}\lambda_n\beta_n\left(\Psi^*\left(\frac{p}{\beta_n}\right)-\sigma_C\left(\frac{p}{\beta_n}\right)\right),$$

thus condition (ii) in (H) implies condition (ii) in $(\mathrm{H_{fitz}})$ applied to $B=\partial\Psi$. This shows that the hypothesis formulated by means of the Fitzpatrick function extends the one from [2] to the more general setting considered in Problem 3.1. In the last section of this paper we will discuss the fulfillment of the hypotheses (H) and $(\mathrm{H_{fitz}})$ for specific instances of the operator B.

Theorem 3.1 *Let $(x_n)_{n\geq 0}$ and $(w_n)_{n\in\mathbb{N}}$ be the sequences generated by Algorithm 3.1 and $(z_n)_{n\in\mathbb{N}}$ be the sequence defined in* (4)*. If* $(\mathrm{H_{fitz}})$ *is fulfilled, then $(z_n)_{n\in\mathbb{N}}$ converges weakly to an element in* $\operatorname{zer}(A+D+N_C)$ *as $n\to+\infty$.*

Proof. As $\lim_{n\to+\infty}\lambda_n = 0$, there exists $n_0 \in \mathbb{N}$ such that $2\eta - \lambda_n \geq 0$ for all $n \geq n_0$. Thus, for $(u, w) \in \operatorname{Graph}(A + D + N_C)$ such that $w = v + p + Du$, where $v \in Au$ and $p \in N_C(u)$, by Lemma 3.1 it holds for any $n \geq n_0$

$$\begin{aligned}&\|x_n - u\|^2 - \|x_{n-1} - u\|^2 \leq \\ &2\lambda_n\beta_n\left(\sup_{\tilde{u}\in C}\varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) + 2\lambda_n\langle w, u - x_n\rangle + 2\lambda_n^2\|Du + v\|^2.\end{aligned} \tag{10}$$

By Lemma 2.1, it is sufficient to prove that the following two statements hold:

1. for every $u \in \operatorname{zer}(A + D + N_C)$ the sequence $\left(\|x_n - u\|_{n\geq 0}\right)$ is convergent;
2. every weak cluster point of $(z_n)_{n\in\mathbb{N}}$ lies in $\operatorname{zer}(A + D + N_C)$.

1. Take an arbitrary u in $\operatorname{zer}(A + D + N_C)$. By taking $w = 0$ in (10), we get

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

and the conclusion follows from Lemma 2.2.

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

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

where

$$\begin{aligned}L =& 2\sum_{n=n_0+1}^{\infty}\lambda_n\beta_n\left(\sup_{\tilde{u}\in C}\varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) + 2\sum_{n=n_0+1}^{\infty}\lambda_n^2\|Du + v\|^2 \\ &+ 2\sum_{n=1}^{n_0}\lambda_n\langle w, x_n - u\rangle\end{aligned}$$

is finite and independent of N. Discarding the nonnegative term $\|x_N - u\|^2$ and dividing by $2\tau_N = 2\sum_{n=1}^{N}\lambda_n$, we obtain

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

By passing to the limit $N \to +\infty$ and using that $\lim_{N\to+\infty} \tau_N = +\infty$, we get

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

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

In the following we show that strong monotonicity of the operator A ensures strong convergence of the sequence $(x_n)_{n\geq 0}$.

Theorem 3.2 *Let $(x_n)_{n\geq 0}$ and $(w_n)_{n\in\mathbb{N}}$ be the sequences generated by Algorithm 3.1. If* $(\mathrm{H_{fitz}})$ *is fulfilled and the operator A is γ-strongly monotone with $\gamma > 0$, then $(x_n)_{n\geq 0}$ converges strongly to the unique element in* $\operatorname{zer}(A + D + N_C)$ *as $n \to +\infty$.*

Proof. Let $u \in \operatorname{zer}(A + D + N_C)$ and $w = 0 = v + p + Du$, where $v \in Au$ and $p \in N_C(u)$. Since A is γ-strongly monotone, inequality (7) becomes for any $n \in \mathbb{N}$

$$\langle y_{n-1} - w_n - \lambda_n v, w_n - u \rangle \geq \lambda_n \gamma \|w_n - u\|^2. \tag{11}$$

Arguing as in the proof of Lemma 3.1 (for $w = 0$) we obtain for any $n \in \mathbb{N}$

$$\begin{aligned}
&\lambda_n \gamma \|w_n - u\|^2 + \|x_n - u\|^2 - \|x_{n-1} - u\|^2 + \lambda_n(2\eta - \lambda_n)\|Dx_{n-1} - Du\|^2 + \\
&\frac{1}{2}\|x_n - w_n\|^2 + \frac{1}{2}\|x_n - w_n - \lambda_n(Du + v)\|^2 + \|x_{n-1} - w_n - \lambda_n(Dx_{n-1} - Du)\|^2 \leq \\
&2\lambda_n\beta_n \left(\sup_{\tilde{u}\in C} \varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) + 2\lambda_n^2 \|Du + v\|^2.
\end{aligned}$$

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

$$\begin{aligned}
&\lambda_n \gamma \|w_n - u\|^2 + \frac{1}{2}\|x_n - w_n\|^2 + \|x_n - u\|^2 - \|x_{n-1} - u\|^2 \leq \\
&2\lambda_n\beta_n \left(\sup_{\tilde{u}\in C} \varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) + 2\lambda_n^2 \|Du + v\|^2
\end{aligned}$$

and, so,

$$\begin{aligned}
&\gamma \sum_{n\geq n_0} \lambda_n \|w_n - u\|^2 + \frac{1}{2}\sum_{n\geq n_0} \|x_n - w_n\|^2 \leq \\
&\|x_{n_0} - u\|^2 + 2\sum_{n\geq n_0} \lambda_n\beta_n \left(\sup_{\tilde{u}\in C} \varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) + 2\|Du + v\|^2 \sum_{n\geq n_0} \lambda_n^2 < +\infty.
\end{aligned}$$

From here it follows that $\sum_{n\geq n_0}\lambda_n(\|x_n-u\|-\|x_n-w_n\|)^2 \leq \sum_{n\geq n_0}\lambda_n\|w_n-u\|^2 < +\infty$ and $\sum_{n\geq n_0}\|x_n-w_n\|^2 < +\infty$. Since $(\|x_n-u\|-\|x_n-w_n\|)_{n\in\mathbb{N}}$ is convergent (see the proof of Theorem 3.1) and $\sum_{n\in\mathbb{N}}\lambda_n = +\infty$, it follows that $\lim_{n\to+\infty}(\|x_n-u\|-\|x_n-w_n\|) = 0$ and, so, $\lim_{n\to+\infty}\|x_n-u\| = 0$. ■

4 A Backward Penalty Scheme with Two Forward Steps

The problem we deal with in this section has the following formulation.

Problem 4.1 Let $\mathcal{H}$ be a real Hilbert space, $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ be maximally monotone operators, $D : \mathcal{H} \to \mathcal{H}$ be an η^{-1}-Lipschitz continuous and monotone operator with $\eta > 0$ and suppose that $C := \operatorname{zer} B \neq \emptyset$. The monotone inclusion problem to solve is

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

Problem 4.1 is a generalization of Problem 3.1, since every η-cocoercive operator is obviously monotone and η^{-1}-Lipschitz continuous. If $D = \nabla f$ for some convex and differentiable function $f : \mathcal{H} \to \mathbb{R}$ with η^{-1}-Lipschitzian gradient, then D is automatically η-cocoercive by the Baillon–Haddad theorem [20]. The investigations we make in Section 5 provide a strong motivation for treating monotone inclusion problems in the settings of Problem 4.1.

Algorithm 4.1 *Choose $x_1 \in \mathcal{H}$ and set for any $n \geq 1$*

$$\begin{aligned}
y_n &:= x_n - \lambda_n D x_n,\\
p_n &:= J_{\lambda_n A} y_n,\\
q_n &:= p_n - \lambda_n D p_n,\\
x_{n+1} &:= J_{\lambda_n \beta_n B}(x_n - y_n + q_n),
\end{aligned}$$

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

Lemma 4.1 *For $u \in C \cap \operatorname{dom} A$ take $w \in (A + D + N_C)(u)$ such that $w = v + Du + p$ for some $v \in Au$ and $p \in N_C(u)$. For each $n \in \mathbb{N}$, the following inequality holds:*

$$\begin{aligned}
&\|x_{n+1} - u\|^2 - \|x_n - u\|^2 + \left(1 - \frac{4\lambda_n^2}{\eta^2}\right) \|x_n - p_n\|^2 + \\
&\frac{1}{2}\|x_{n+1} - p_n\|^2 + \frac{1}{2}\|x_{n+1} - p_n + 2\lambda_n(Dp_n - Dx_n + p)\|^2 \leq \\
&2\lambda_n\beta_n\left(\sup_{\tilde{u}\in C} \varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) + 2\lambda_n \langle w, u - p_n\rangle + 4\lambda_n^2 \|p\|^2 .
\end{aligned}$$

Proof. Let be $n \geq 1$ fixed. We have $\lambda_n v \in \lambda_n Au$ and $y_n - p_n \in \lambda_n Ap_n$, so, by monotonicity of $\lambda_n A$,

$$\langle y_n - p_n - \lambda_n v, p_n - u\rangle \geq 0,$$

which is equivalent to

$$2\lambda_n \langle v, p_n - u\rangle \leq 2\langle y_n - p_n, p_n - u\rangle = \|y_n - u\|^2 - \|y_n - p_n\|^2 - \|p_n - u\|^2 . \tag{12}$$

Furthermore, we have $x_n - y_n + q_n - x_{n+1} \in \lambda_n\beta_n Bx_{n+1}$, so, by definition of the Fitzpatrick function,

$$\begin{aligned}
&2\lambda_n\beta_n\left(\sup_{\tilde{u}\in C} \varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) \\
&\geq 2\langle u, x_n - y_n + q_n - x_{n+1}\rangle + 2\lambda_n\langle p, x_{n+1}\rangle - 2\langle x_{n+1}, x_n - y_n + q_n - x_{n+1}\rangle \\
&\quad - 2\lambda_n\langle p, u\rangle \\
&= 2\lambda_n\langle p, x_{n+1} - u\rangle + 2\langle u - x_{n+1}, x_n - y_n + q_n - x_{n+1}\rangle \\
&= 2\lambda_n\langle p, x_{n+1} - u\rangle - \|u - x_n\|^2 + \|u - y_n\|^2 - \|u - q_n\|^2 + \|u - x_{n+1}\|^2 \\
&\quad + \|x_n - x_{n+1}\|^2 - \|x_{n+1} - y_n\|^2 + \|x_{n+1} - q_n\|^2 .
\end{aligned} \tag{13}$$

Adding (12) and (13), we obtain

$$\begin{aligned}
&\|x_{n+1} - u\|^2 - \|x_n - u\|^2 - 2\lambda_n\beta_n\left(\sup_{\tilde{u}\in C} \varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) - 2\lambda_n\langle w, u - p_n\rangle \\
&\leq 2\lambda_n\langle p, p_n - x_{n+1}\rangle + 2\lambda_n\langle Du, p_n - u\rangle - \|y_n - p_n\|^2 - \|u - p_n\|^2 + \|u - q_n\|^2 \\
&\quad - \|x_{n+1} - x_n\|^2 + \|x_{n+1} - y_n\|^2 - \|x_{n+1} - q_n\|^2 \\
&= 2\lambda_n\langle p, p_n - x_{n+1}\rangle + 2\lambda_n\langle Du, p_n - u\rangle - \|x_n - \lambda_n Dx_n - p_n\|^2 - \|u - p_n\|^2
\end{aligned}$$

$$
\begin{aligned}
&+\|u-p_n+\lambda_n Dp_n\|^2-\|x_{n+1}-x_n\|^2+\|x_{n+1}-x_n+\lambda_n Dx_n\|^2\\
&-\|x_{n+1}-p_n+\lambda_n Dp_n\|^2\\
=\;&2\lambda_n\langle p,p_n-x_{n+1}\rangle+2\lambda_n\langle Du,p_n-u\rangle-\|x_n-p_n\|^2+2\lambda_n\langle Dx_n,x_n-p_n\rangle\\
&-\lambda_n^2\|Dx_n\|^2+\lambda_n^2\|Dp_n\|^2+2\lambda_n\langle Dp_n,u-p_n\rangle+\lambda_n^2\|Dx_n\|^2+2\lambda_n\langle Dx_n,x_{n+1}-x_n\rangle\\
&-\|x_{n+1}-p_n\|^2-2\lambda_n\langle x_{n+1}-p_n,Dp_n\rangle-\lambda_n^2\|Dp_n\|^2\\
=\;&2\lambda_n\langle p,p_n-x_{n+1}\rangle+2\lambda_n\langle Du,p_n-u\rangle-\|x_n-p_n\|^2+2\lambda_n\langle Dp_n,u-x_{n+1}\rangle\\
&+2\lambda_n\langle Dx_n,x_{n+1}-p_n\rangle-\|x_{n+1}-p_n\|^2\\
=\;&2\lambda_n\langle Du-Dp_n,p_n-u\rangle+2\lambda_n\langle Dp_n-Dx_n+p,p_n-x_{n+1}\rangle\\
&-\|x_{n+1}-p_n\|^2-\|x_n-p_n\|^2\\
\leq\;&-\|x_{n+1}-p_n\|^2-\|x_n-p_n\|^2+2\lambda_n\langle Dp_n-Dx_n+p,p_n-x_{n+1}\rangle\\
=\;&-\|x_n-p_n\|^2-\frac{1}{2}\|x_{n+1}-p_n\|^2-\frac{1}{2}\|x_{n+1}-p_n+2\lambda_n(Dp_n-Dx_n+p)\|^2\\
&+2\lambda_n^2\|Dp_n-Dx_n+p\|^2\\
\leq\;&-\left(1-\frac{4\lambda_n^2}{\eta^2}\right)\|x_n-p_n\|^2-\frac{1}{2}\|x_{n+1}-p_n\|^2\\
&-\frac{1}{2}\|x_{n+1}-p_n+2\lambda_n(Dp_n-Dx_n+p)\|^2+4\lambda_n^2\|p\|^2,
\end{aligned}
$$

which leads to the desired conclusion. ■

For the convergence statement, the same additional hypotheses ($\mathrm{H}_{\mathrm{fitz}}$) are needed as for Algorithm 3.1.

Theorem 4.1 *Let $(x_n)_{n\in\mathbb{N}}$, $(p_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$ and $(q_n)_{n\in\mathbb{N}}$ be the sequences generated by Algorithm 4.1 and $(z_n)_{n\in\mathbb{N}}$ be the sequence defined in* (4). *If* ($\mathrm{H}_{\mathrm{fitz}}$) *is fulfilled, then $(z_n)_{n\in\mathbb{N}}$ converges weakly to an element in $\operatorname{zer}(A+D+N_C)$ as $n\to\infty$.*

Proof. As $\lim_{n\to+\infty}\lambda_n=0$, there exists $n_0\in\mathbb{N}$ such that $1-\frac{4\lambda_n^2}{\eta^2}\geq 0$ for all $n\geq n_0$. Thus, for $(u,w)\in\operatorname{Graph}(A+D+N_C)$ such that $w=v+p+Du$, where $v\in Au$ and $p\in N_C(u)$, by Lemma

4.1 it holds for any $n \geq n_0$

$$\|x_{n+1} - u\|^2 - \|x_n - u\|^2 \leq$$
$$2\lambda_n\beta_n\left(\sup_{\tilde{u}\in C}\varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) + 2\lambda_n\langle w, u - p_n\rangle + 4\lambda_n^2\|p\|^2. \tag{14}$$

Analogously to the proof of Theorem 3.1, one obtains from here that:

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

The conclusion follows by using again Lemma 2.1. ■

Arguing in the same way as in Theorem 3.2, one can show that the strong monotonicity of A guarantees strong convergence of the sequence $(x_n)_{n\in\mathbb{N}}$.

Theorem 4.2 *Let $(x_n)_{n\in\mathbb{N}}$, $(p_n)_{n\in\mathbb{N}}$, $(y_n)_{n\in\mathbb{N}}$ and $(q_n)_{n\in\mathbb{N}}$ be the sequences generated by Algorithm 4.1. If* $(\mathrm{H}_{\text{fitz}})$ *is fulfilled 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* $\operatorname{zer}(A + D + N_C)$ *as $n \to +\infty$.*

5 A Primal-Dual Algorithm Based on a Backward Penalty Scheme

In this section, we will derive a primal-dual algorithm for solving complexly structured monotone inclusion problems based on the backward penalty iterative scheme provided in Algorithm 4.1. The problem under investigation is the following one.

Problem 5.1 Let $\mathcal{H}$ be a real Hilbert space, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator and $D : \mathcal{H} \to \mathcal{H}$ a monotone and η^{-1}-Lipschitz continuous operator for some $\eta > 0$. Furthermore, let $m \geq 1$ and for any $i \in \{1, \ldots, m\}$, let $\mathcal{G}_i$ be real Hilbert spaces, $A_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ maximally monotone operators, $D_i : \mathcal{G}_i \rightrightarrows \mathcal{G}_i$ be maximally monotone operators such that D_i^{-1} are ν_i^{-1}-Lipschitz continuous for some $\nu_i > 0$ and $L_i : \mathcal{H} \to \mathcal{G}_i$ nonzero linear continuous operators. Consider also $B : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator and suppose that $C := \operatorname{zer} B \neq \emptyset$. The monotone inclusion problem to solve is to find $x \in \mathcal{H}$ with

$$0 \in Ax + \sum_{i=1}^{m} L_i^*(A_i \,\square\, D_i)(L_i x) + Dx + N_C(x), \tag{15}$$

together with its dual monotone inclusion problem in the sense of Attouch–Théra [21] of finding $v_i \in \mathcal{G}_i$, $i = 1, \ldots, m$, satisfying

$$\exists x \in \mathcal{H}: \qquad v_i \in (A_i \,\Box\, D_i)(L_i x) \quad \text{and} \quad 0 \in Ax + \sum_{i=1}^{m} L_i^* v_i + Dx + N_C(x). \tag{16}$$

We introduce the real Hilbert space $\boldsymbol{\mathcal{H}} := \mathcal{H} \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_m$, the operators

$$\begin{aligned}
&\boldsymbol{A}: \boldsymbol{\mathcal{H}} \rightrightarrows \boldsymbol{\mathcal{H}},\ \boldsymbol{A}(x, v_1, \ldots, v_m) = Ax \times A_1^{-1} v_1 \times \ldots \times A_m^{-1} v_m, \\
&\boldsymbol{D}: \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}},\ \boldsymbol{D}(x, v_1, \ldots, v_m) = \left(\sum_{i=1}^{m} L_i^* v_i + Dx, D_1^{-1} v_1 - L_1 x, \ldots, D_m^{-1} v_m - L_m x \right), \\
&\boldsymbol{B}: \boldsymbol{\mathcal{H}} \rightrightarrows \boldsymbol{\mathcal{H}},\ \boldsymbol{B}(x, v_1, \ldots, v_m) = Bx \times \{0\} \times \ldots \times \{0\}\,,
\end{aligned} \tag{17}$$

and the set

$$\boldsymbol{C} := \{\boldsymbol{x} \in \boldsymbol{\mathcal{H}} \,|\, \boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}\} = \operatorname{zer} B \times \mathcal{G}_1 \times \ldots \times \mathcal{G}_m.$$

In this setting, we have for $\boldsymbol{x} = (x, v_1, \ldots, v_m) \in \mathcal{H}$

$$\begin{aligned}
\boldsymbol{0} \in (\boldsymbol{A} + \boldsymbol{D} + N_{\boldsymbol{C}})\boldsymbol{x} \iff & \left\{ \begin{matrix} 0 \in Ax + \sum_{i=1}^{m} L_i^* v_i + Dx + N_C(x) \\ 0 \in A_1^{-1} v_1 + D_1^{-1} v_1 - L_1 x \\ \vdots \\ 0 \in A_m^{-1} v_m + D_m^{-1} v_m - L_m x \end{matrix} \right\} \\
\iff & \left\{ \begin{matrix} 0 \in Ax + \sum_{i=1}^{m} L_i^* v_i + Dx + N_C(x) \\ v_i \in (A_i \,\Box\, D_i)(L_i x), i = 1, \ldots, m \end{matrix} \right\} \\
\implies & x \text{ satisfies (15) and } (v_1, \ldots, v_m) \text{ satisfies (16).}
\end{aligned} \tag{18}$$

The resolvent of $\boldsymbol{B}$ is given by

$$J_{\gamma \boldsymbol{B}}(x, v_1, \ldots, v_m) = (J_{\gamma B} x, v_1, \ldots, v_m)$$

and its Fitzpatrick function $\varphi_{\boldsymbol{B}}: \boldsymbol{\mathcal{H}} \times \boldsymbol{\mathcal{H}} \to \overline{\mathbb{R}}$ by

$$\begin{aligned}
\varphi_{\boldsymbol{B}}(x, v_1, \ldots, v_m, x^*, v_1^*, \ldots, v_m^*) &= \sup_{\substack{y \in \mathcal{H} \\ y^* \in By \\ w_i \in \mathcal{G}_i, i = 1, \ldots, m}} \left\{ \langle x, y^* \rangle + \langle y, x^* \rangle + \sum_{i=1}^{m} \langle w_i, v_i^* \rangle - \langle y, y^* \rangle \right\} \\
&= \begin{cases} \varphi_B(x, x^*), & \text{if } v_i^* = 0,\ i = 1, \ldots, m, \\ +\infty, & \text{otherwise.} \end{cases}
\end{aligned}$$

Thus, in order to solve the primal-dual pair of monotone inclusion problems (15)–(16), one has to solve

$$\mathbf{0} \in \boldsymbol{A}\boldsymbol{x} + \boldsymbol{D}\boldsymbol{x} + N_{\boldsymbol{C}}(\boldsymbol{x}) \tag{19}$$

in the product space $\boldsymbol{\mathcal{H}}$. By doing this via Algorithm 4.1 one obtains the following iterative scheme:

Algorithm 5.1 *Choose $x_1 \in \mathcal{H}$ and $v_{i,1} \in \mathcal{G}_i$, $i = 1, \ldots, m$, and set for any $n \geq 1$*

$$\begin{aligned}
y_{1,n} &:= x_n - \lambda_n \left(\sum_{i=1}^m L_i^* v_{i,n} - D x_n \right), \\
y_{2,i,n} &:= v_{i,n} - \lambda_n \big(D_i^{-1} v_{i,n} - L_i x_n \big), i = 1, \ldots, m, \\
p_{1,n} &:= J_{\lambda_n A} y_{1,n}, \\
p_{2,i,n} &:= J_{\lambda_n A_i^{-1}} y_{2,i,n}, i = 1, \ldots, m, i = 1, \ldots, m, \\
q_n &:= p_{1,n} - \lambda_n \left(\sum_{i=1}^m L_i^* p_{2,i,n} - D p_{1,n} \right), \\
v_{i,n+1} &:= v_{i,n} - y_{2,i,n} + p_{2,i,n} - \lambda_n \big(D_i^{-1} p_{2,i,n} - L_i p_{1,n} \big), i = 1, \ldots, m, \\
x_{n+1} &:= J_{\lambda_n \beta_n B} (x_n - y_{1,n} + q_n).
\end{aligned}$$

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

For its convergence the following hypotheses are needed:

$$\left.\begin{array}{ll}
\text{(i)} & A + N_C \text{ is maximally monotone and } \operatorname{zer}\left(A + \sum\limits_{i=1}^m L_i^* \circ (A_i \,\Box\, D_i) \circ L_i + D + N_C \right) \neq \emptyset; \\
\text{(ii)} & \text{For every } p \in \operatorname{Ran} N_C, \ \sum\limits_{n \in \mathbb{N}} \lambda_n \beta_n \left(\sup\limits_{\tilde{u} \in C} \varphi_B\left(\tilde{u}, \dfrac{p}{\beta_n} \right) - \sigma_C\left(\dfrac{p}{\beta_n} \right) \right) < +\infty; \\
\text{(iii)} & (\lambda_n)_{n \geq 0} \in \ell^2 \setminus \ell^1.
\end{array}\right\} \tag{$\mathrm{H_{pd}}$}$$

Theorem 5.1 *Consider the sequences generated by Algorithm 5.1 and assume that* $(\mathrm{H}_{\mathrm{pd}})$ *is fulfilled. Then the sequence $(z_n)_{n \in \mathbb{N}}$ defined in* (4) *converges weakly to a solution of* (15) *and the sequence $\left(\frac{1}{\sum_{k=1}^n \lambda_k} \sum_{k=1}^n \lambda_k (v_{1,k}, .., v_{m,k}) \right)_{n \in \mathbb{N}}$ converges weakly to a solution of* (16) *as $n \to +\infty$. If, additionally, A and $A_i^{-1}, i = 1, ..., m$, are strongly monotone, then $(x_n)_{n \in \mathbb{N}}$ converges strongly to the unique solution of* (15) *and $(v_{1,n}, .., v_{m,n})_{n \in \mathbb{N}}$ converges strongly to the unique solution of* (16) *as $n \to +\infty$.*

Proof. Clearly, the iterations in Algorithm 5.1 can be for any $n \geq 1$ equivalently written as

$$\begin{aligned}
(y_{1,n}, y_{2,1,n}, \dots, y_{2,m,n}) &= (\mathbf{Id} - \lambda_n \boldsymbol{D})(x_n, v_{1,n}, \dots, v_{m,n}),\\
(p_{1,n}, p_{2,1,n}, \dots, p_{2,m,n}) &= J_{\boldsymbol{A}}(y_{1,n}, y_{2,1,n}, \dots, y_{2,m,n}),\\
(q_n, \tilde{q}_{1,n}, \dots, \tilde{q}_{m,n}) &= (\mathbf{Id} - \lambda_n \boldsymbol{D})(p_{1,n}, p_{2,1,n}, \dots, p_{2,m,n}),\\
(x_{n+1}, v_{1,n+1}, \dots, v_{m,n+1}) &= J_{\lambda_n \beta_n \boldsymbol{B}}\Big((x_n, v_{1,n}, \dots, v_{m,n}) - (y_{1,n}, y_{2,1,n}, \dots, y_{2,m,n})\\
&\qquad + (q_n, \tilde{q}_{1,n}, \dots, \tilde{q}_{m,n})\Big).
\end{aligned}$$

with the operators $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{D}$ defined in (17). The operators $\boldsymbol{A}$ and $\boldsymbol{B}$ are maximally monotone by [8, Proposition 20.23], and the operator $\boldsymbol{D}$ is monotone and Lipschitz continuous ([22]). If A and $A_i^{-1}, i = 1, \dots, m$, are strongly monotone, then $\boldsymbol{A}$ is strongly monotone, too. Thus the conclusion is a direct consequence of the Theorem 4.1 and Theorem 4.2 applied to the monotone inclusion problem (19), provided that the corresponding hypotheses ($\mathrm{H}_{\mathrm{fitz}}$) are fulfilled.

According to (H_{pd}), $A + N_C$ is maximally monotone, and so is $\boldsymbol{A} + N_{\boldsymbol{C}}$. Further, the assumption $\operatorname{zer}\left(A + \sum_{i=1}^{m} L_i^* \circ (A_i \,\Box\, D_i) \circ L_i + D + N_C\right) \neq \emptyset$ leads to $\operatorname{zer}(\boldsymbol{A} + \boldsymbol{D} + N_{\boldsymbol{C}}) \neq \emptyset$.

Furthermore, $\operatorname{Ran} N_{\boldsymbol{C}} = \operatorname{Ran} N_C \times \{0\} \times \ldots \times \{0\}$, so for all $(p, 0, \dots, 0) \in \operatorname{Ran} N_{\boldsymbol{C}}$

$$\sum_{n \in \mathbb{N}} \lambda_n \beta_n \sup_{(\tilde{u}, v_1, \dots, v_m) \in \boldsymbol{C}} \left(\varphi_{\boldsymbol{B}}\left(\tilde{u}, v_1, \dots, v_m, \frac{p}{\beta_n}, 0, \dots, 0\right) - \sigma_{\boldsymbol{C}}\left(\frac{p}{\beta_n}, 0, \dots, 0\right)\right) =$$
$$\sum_{n \in \mathbb{N}} \lambda_n \beta_n \sup_{\tilde{u} \in C} \left(\varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) < +\infty.$$

■

Remark 5.1 Even if the operators D and $D_i^{-1}, i = 1, ..., m$, are cocoercive, one cannot make use of Algorithm 3.1 and of the corresponding convergence theorem in order to solve the monotone inclusion problem (19). This is due to the fact that the operator

$$(x, v_1, ..., v_m) \mapsto \left(\sum_{i=1}^{m} L_i^* v_i, -L_1 x, \dots, -L_m x\right),$$

being skew, fails to be cocoercive, which means that $\boldsymbol{D}$ is not cocoercive as well. This shows the importance of having iterative schemes for monotone inclusion problems involving monotone and Lipschitz continuous operators, which are not necessarily cocoercive, as is Algorithm 4.1 (see, also [7]).

6 Examples

In this section, we discuss the fulfillment of condition (ii) in the hypotheses (H) and (H_{fitz}), for several particular instances of the operator B.

Example 6.1 For a convex and closed set $\emptyset \neq C \subseteq \mathcal{H}$, let $B := N_C$. Then $\operatorname{zer} B = C$ and (see [14, Example 3.1])

$$\varphi_B(x,u) = \varphi_{N_C}(x,u) = \delta_C(x) + \sigma_C(u),$$

and condition (ii) in (H_{fitz}) becomes

$$\sum_{n\in\mathbb{N}} \lambda_n\beta_n \left(\sup_{\tilde{u}\in C} \delta_C(\tilde{u}) + \sigma_C\left(\frac{p}{\beta_n}\right) - \sigma_C\left(\frac{p}{\beta_n}\right)\right) < +\infty,$$

which is satisfied for any choice of the sequences $(\lambda_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$. The same applies for condition (ii) in (H), where $\Psi(x) = \delta_C(x)$ and $\Psi^*(u) = \sigma_C(u)$.

Example 6.2 For a convex and closed set $\emptyset \neq C \subseteq \mathcal{H}$, let $\Psi : \mathcal{H} \to \mathbb{R}$, $\Psi(x) = \frac{1}{2} d_C(x)^2$, where $d_C(x) = \inf_{z\in C} \|x - z\|$ and $B := \partial\Psi$. Then $\operatorname{zer} B = C$ and (see [8, Corollary 12.30])

$$\nabla\Psi(x) = x - \operatorname{Proj}_C(x),$$

where $\operatorname{Proj}_C : \mathcal{H} \to C$ denotes the projection operator on C. We have $\Psi = \delta_C \,\Box \left(\frac{1}{2}\|\cdot\|^2\right)$, so $\Psi^* = \sigma_C + \frac{1}{2}\|\cdot\|^2$. If $C \neq \mathcal{H}$, condition (ii) in (H) is therefore equivalent to (see [2])

$$\sum_{n\in\mathbb{N}} \frac{\lambda_n}{\beta_n} < +\infty, \tag{20}$$

in which case condition (ii) in (H_{fitz}) is also fulfilled. Let also notice that the resolvent of B is given by

$$J_{\gamma B}(x) = \frac{x}{\gamma+1} + \frac{\gamma \operatorname{Proj}_C(x)}{\gamma+1} \ \forall x \in \mathcal{H}.$$

Notice that in this example the approach from [6] could also be applied to the single-valued operator B, resulting in slightly different algorithms.

Next, we present two examples, for which condition (ii) in (H_{fitz}) fails for any choice of the sequence of positive penalty parameters $(\beta_n)_{n\in\mathbb{N}}$.

Example 6.3 For a convex and closed set $\emptyset \neq C \subseteq \mathcal{H}$, let $\Psi : \mathcal{H} \to \mathbb{R}$, $\Psi(x) = d_C(x)$, and $B := \partial \Psi$. Then (see [8, Example 16.49])

$$Bx = \partial d_C(x) = \begin{cases} \{u \in N_C(x) \mid \|u\| \leq 1\}, & \text{if } x \in C, \\ \left\{ \frac{x - \operatorname{Proj}_C(x)}{\|x - \operatorname{Proj}_C(x)\|} \right\}, & \text{otherwise} \end{cases}$$

and $\operatorname{zer} B = C$. Since $\Psi^* = \sigma_C + \delta_{\mathbb{B}}$,

$$\sum_{n \in \mathbb{N}} \lambda_n \beta_n \left(\Psi^* \left(\frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \right) = \sum_{n \in \mathbb{N}} \lambda_n \beta_n \delta_{\mathbb{B}} \left(\frac{p}{\beta_n} \right). \tag{21}$$

For $C \neq \mathcal{H}$ and arbitrary $\beta_n > 0$, with $n \in \mathbb{N}$, there exists $p \in \operatorname{Ran} N_C$ with $\|p\| > \beta_n$, for which expression (21) is equal to $+\infty$. Thus, condition (ii) in (H) is not verified.

Condition (ii) in ($\mathrm{H_{fitz}}$) fails for similar reasons. Let be $\beta_n > 0$, with $n \in \mathbb{N}$, $y \in C$ and $p \in N_C(y)$ with $\|p\| > \beta_n$. Then

$$\left(y + tp, \frac{p}{\|p\|} \right) \in \operatorname{Graph} \partial d_C \ \forall t > 0,$$

which implies that

$$\begin{aligned} &\sup_{\tilde{u} \in C} \varphi_{\partial d_C} \left(\tilde{u}, \frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \geq \varphi_{\partial d_C} \left(y, \frac{p}{\beta_n} \right) - \sigma_C \left(\frac{p}{\beta_n} \right) \\ &\geq \sup_{t > 0} \left(\left\langle y, \frac{p}{\|p\|} \right\rangle + \left\langle y + tp, \frac{p}{\beta_n} \right\rangle - \left\langle y + tp, \frac{p}{\|p\|} \right\rangle - \left\langle y, \frac{p}{\beta_n} \right\rangle \right) \\ &= \sup_{t > 0} \left(t \|p\| \left(\frac{\|p\|}{\beta_n} - 1 \right) \right) = +\infty. \end{aligned}$$

Remark 6.1 One can notice that in the previous three examples, despite of the different choices of the operator B, the set of its zeros is the convex and closed set C. In what concerns condition (ii) in ($\mathrm{H_{fitz}}$) and (H), it is satisfied for any choice of $(\lambda_n)_{n \in \mathbb{N}}$ and $(\beta_n)_{n \in \mathbb{N}}$ when $B = N_C$ and for the two sequences fulfilling assumption (20) when $B = \partial \left(\frac{1}{2} d_C^2 \right)$, however, it fails for any choice of $(\lambda_n)_{n \in \mathbb{N}}$ and $(\beta_n)_{n \in \mathbb{N}}$ when $B = \partial d_C$. The applicability of ($\mathrm{H_{fitz}}$) and (H) therefore depends on the modeling of the variational inequality via the set-valued operator B.

Example 6.4 Let $B : \mathcal{H} \to \mathcal{H}$ be a nonzero skew linear continuous operator. So $\operatorname{zer} B = \ker B$ and by taking a nonzero element $p \in \operatorname{Ran} N_{\ker B} = (\ker B)^{\perp}$, it holds for any $n \in \mathbb{N}$

$$\begin{aligned}
&\sup_{\tilde{u} \in \ker B} \varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_{\ker B}\left(\frac{p}{\beta_n}\right) \\
&= \sup_{\tilde{u} \in \ker B} \sup_{y \in \mathcal{H}} \left(\langle \tilde{u}, By\rangle + \left\langle y, \frac{p}{\beta_n}\right\rangle - \langle y, By\rangle\right) \\
&\geq \sup_{y \in \mathcal{H}} \left\langle y, \frac{p}{\beta_n}\right\rangle = +\infty.
\end{aligned}$$

This shows that condition (ii) in ($\mathrm{H}_{\mathrm{fitz}}$) is not satisfied.

Example 6.5 On the other hand, let $S : \mathcal{H} \to \mathcal{H}$ be a nonzero skew linear continuous operator, and let $B := (\mathrm{Id} + S)^{-1}$ be its resolvent. This operator is not symmetric, therefore it cannot be the subdifferential of a convex function ([23, Proposition 2.51]). We have $\ker B = (\mathrm{Id} + S)(0) = \{0\}$. For every $p \in \operatorname{Ran} N_{\ker B} = (\ker B)^{\perp} = \mathcal{H}$, we have (recall that $\mathrm{Id} + S$ is surjective and $S^* = -S$)

$$\begin{aligned}
&\sum_{n\in\mathbb{N}} \lambda_n\beta_n \left(\sup_{\tilde{u} \in \ker B} \varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_{\ker B}\left(\frac{p}{\beta_n}\right)\right) \\
&= \sum_{n\in\mathbb{N}} \lambda_n\beta_n\varphi_B\left(0, \frac{p}{\beta_n}\right) = \sum_{n\in\mathbb{N}} \lambda_n\beta_n \sup_{y\in\mathcal{H}} \left\{\left\langle \frac{p}{\beta_n}, y\right\rangle - \langle y, By\rangle\right\} \\
&= \sum_{n\in\mathbb{N}} \lambda_n\beta_n \sup_{x\in\mathcal{H}} \left\{\left\langle \frac{p}{\beta_n}, x + Sx\right\rangle - \langle x + Sx, x\rangle\right\} = \sum_{n\in\mathbb{N}} \lambda_n\beta_n \sup_{x\in\mathcal{H}} \left\{\left\langle x, \frac{p - Sp}{\beta_n}\right\rangle - \|x\|^2\right\} \\
&= \sum_{n\in\mathbb{N}} \lambda_n\beta_n \sup_{x\in\mathcal{H}} \left\{-\left\|x - \frac{p - Sp}{2\beta_n}\right\|^2 + \frac{\|p\|^2 + \|Sp\|^2}{4\beta_n^2}\right\} = \left(\sum_{n\in\mathbb{N}} \frac{\lambda_n}{\beta_n}\right) \frac{\|p\|^2 + \|Sp\|^2}{4},
\end{aligned}$$

so condition (ii) in ($\mathrm{H}_{\mathrm{fitz}}$) is equivalent to (20).

Example 6.6 Let $B : \mathcal{H} \to \mathcal{H}$ be a self-adjoint linear continuous operator with closed range. We have for each $p \in \operatorname{Ran} N_{\ker B} = (\ker B)^{\perp} = \operatorname{Ran} B$

$$\begin{aligned}
&\sum_{n\in\mathbb{N}} \lambda_n\beta_n \left(\sup_{\tilde{u} \in \ker B} \varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_{\ker B}\left(\frac{p}{\beta_n}\right)\right) \\
&= \sum_{n\in\mathbb{N}} \lambda_n\beta_n \left(\sup_{\substack{\tilde{u} \in \ker B \\ y \in \mathcal{H}}} \left\{\langle \tilde{u}, By\rangle + \left\langle \frac{p}{\beta_n}, y\right\rangle - \langle y, By\rangle\right\}\right) \\
&= \sum_{n\in\mathbb{N}} \lambda_n\beta_n \left(\sup_{y\in\mathcal{H}} \left\{\left\langle \frac{p}{\beta_n}, y\right\rangle - \langle y, By\rangle\right\}\right).
\end{aligned}$$

For every $\beta_n > 0$ the function $y \mapsto \left\langle \frac{p}{\beta_n}, y\right\rangle - \langle y, By\rangle$ is concave and differentiable. Setting its derivative equal to zero yields for its maximizers the necessary and sufficient optimality condition

$By = \frac{p}{2\beta_n}$. Since $p \in \operatorname{Ran} B$, we find $x \in \mathcal{H}$ with $Bx = p$, so $y = \frac{x}{2\beta_n}$ furnishes the supremum and it holds

$$\begin{aligned}
&\sum_{n\in\mathbb{N}} \lambda_n \beta_n \left(\sup_{\tilde{u}\in\ker B} \varphi_B\left(\tilde{u}, \frac{p}{\beta_n}\right) - \sigma_{\ker B}\left(\frac{p}{\beta_n}\right)\right) \\
= &\sum_{n\in\mathbb{N}} \lambda_n \beta_n \left(\left\langle \frac{p}{\beta_n}, \frac{x}{2\beta_n} \right\rangle - \left\langle \frac{x}{2\beta_n}, \frac{p}{2\beta_n} \right\rangle \right) \\
= &\left(\sum_{n\in\mathbb{N}} \frac{\lambda_n}{\beta_n} \right) \frac{\langle p, x\rangle}{4}.
\end{aligned}$$

In other words, condition (ii) in ($\mathrm{H}_{\mathrm{fitz}}$) is equivalent to (20) (and to (H)).

7 Conclusions

We have proposed two iterative penalty schemes for solving variational inequalities on the set of zeros of a general maximally monotone operator depending on the structure of the problem. The operators involved were evaluated seperately according to their regularity properties. The theoretical results guarantee weak ergodic convergence in the general setting and norm convergence under strong monotonicity assumptions. We provided examples for the applicability of the schemes for several operators defining the constraint set.

Acknowledgements The authors thank the anonymous referees for several improvements of the paper.

References

1. Attouch, H., Czarnecki, M.-O.: Asymptotic behavior of coupled dynamical systems with multiscale aspects. J. Differential Equations **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 foward-backward with penalty schemes and parallel splitting for constrained variational inequalities. SIAM J. Optim. **21**(4), 1251–1274 (2011)
4. Peypouquet, J.: Coupling the gradient method with a general exterior penalization scheme for convex minimization. J. Optim. Theory Appl. **153**(1), 123–138 (2012)
5. Noun, N., Peypouquet, J.: Forward–backward penalty scheme for constrained convex minimization without inf-compactness. J. Optim. Theory Appl. **158**(3), 787–795 (2013)

6. Boţ, R. I., Csetnek, E. R.: Forward-backward and Tseng's type penalty schemes for monotone inclusion problems. Set-Valued Var. Anal. **22**(2), 313-331 (2013)
7. Boţ, R. I., Csetnek, E. R.: A Tseng's type penalty scheme for solving inclusion problems involving linearly composed and parallel-sum type monotone operators. Vietnam J. Math. **42**(4), 451–465 (2014)
8. Bauschke, H. H., Combettes, P. L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. CMS Books in Mathematics, Springer (2011)
9. Borwein, J. M., Vanderwerff, J. D.: Convex Functions: Constructions, Characterizations and Counterexamples. Cambridge University Press (2010)
10. Boţ, R. I.: Conjugate Duality in Convex Optimization. Lecture Notes in Economics and Mathematical Systems, vol. 637. Springer (2010)
11. Ekeland, I., Témam, R.: Convex Analysis and Variational Problems. Classics in Applied Mathematics, vol. 28. Society for Industrial and Applied Mathematics (1999)
12. Simons, S.: From Hahn–Banach to Monotonicity. Lecture Notes in Mathematics, vol. 1693. Springer (2008)
13. Zălinescu, C.: Convex Analysis in General Vector Spaces. World Scientific (2002)
14. 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)
15. Borwein, J. M.: Maximal monotonicity via convex analysis. J. Convex Anal. **13**(3–4), 561–586 (2006)
16. 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)
17. 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)
18. Fitzpatrick, S.: Representing monotone operators by convex functions. Workshop/Miniconference on Functional Analysis and Optimization (Canberra, 1988), Proceedings of the Centre for Mathematical Analysis, vol. 20, pp. 59–65 (1988)
19. Rockafellar, R. T.: On the maximal monotonicity of subdifferential mappings. Pac. J. Math. **33**(1), 209–216 (1970)
20. Baillon, J.-B., Haddad, G.: Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones. Isr. J. Math. **26**(2), 137–150 (1977)
21. Attouch, H., Théra, M.: A general duality principle for the sum of two operators. J. Convex Anal. **3**(1), 1–24 (1996)
22. 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)
23. Barbu, V., Precupanu, T.: Convexity and Optimization in Banach Spaces. Springer Monographs in Mathematics, Springer (2012)