FORWARD-BACKWARD-HALF FORWARD ALGORITHM WITH NON SELF-ADJOINT LINEAR OPERATORS FOR SOLVING MONOTONE INCLUSIONS*

LUIS M. BRICEÑO-ARIAS† AND DAMEK DAVIS‡

Abstract. Tseng's algorithm finds a zero of the sum of a maximally monotone operator and a monotone-Lipschitz operator by evaluating the latter twice per iteration. In this paper, we modify Tseng's algorithm for finding a zero of the sum of three operators, where we add a cocoercive operator to the inclusion. Since the sum of a cocoercive and a monotone-Lipschitz operator is monotone and Lipschitz, we could use Tseng's method for solving this problem, but implementing both operators twice per iteration and without taking into advantage the cocoercivity property of one operator. Instead, in our approach, although the Lipschitz operator must still be evaluated twice, we exploit the cocoercivity of one operator by evaluating it only once per iteration. Moreover, when the cocoercive or monotone-Lipschitz operators are zero it reduces to Tseng's or forward-backward splittings, respectively, unifying in this way both algorithms. In addition, we provide a variable metric version of the proposed method but including non self-adjoint linear operators in the computation of resolvents and the single-valued operators involved. This approach allows us to extend previous variable metric versions of Tseng's and forward-backward methods and simplify their conditions on the underlying metrics. We also exploit the case when non self-adjoint linear operators are triangular by blocks in the primal-dual product space for solving primal-dual composite monotone inclusions, obtaining Gauss-Seidel type algorithms which generalize several primal-dual methods available in the literature. Finally we explore two applications to the obstacle problem and Empirical Risk Minimization.

Key words. Convex optimization, forward-backward splitting, monotone operator theory, sequential algorithms, Tseng's splitting.

AMS subject classifications. 47H05, 65K05, 65K15, 90C25

1. Introduction.

This paper is devoted to the numerical resolution of following problem.

PROBLEM 1. *Let X be a nonempty closed convex subset of a real Hilbert space $\mathcal{H}$, let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $B_1 : \mathcal{H} \to \mathcal{H}$ be β-cocoercive*[1] *and let $B_2 : \mathcal{H} \to \mathcal{H}$ be monotone and L-Lipschitz continuous on* $\mathrm{dom} A \cup X$ *for some constants $\beta > 0$ and $L > 0$. The problem is to*

$$\textit{find} \quad x \in X \quad \textit{such that} \quad 0 \in Ax + B_1x + B_2x, \tag{1.1}$$

under the assumption that the set of solutions to (1.1) *is nonempty.*

The wide variety of applications of Problem 1 involving optimization problems, variational inequalities, partial differential equations, image processing, saddle point problems, game theory, among others can be explored in [**?**, **?**] and the references therein. As an important application, consider the case of composite optimization

*This work is partially supported by NSF GRFP grant DGE-0707424, by CONICYT grant FONDECYT 11140360, and by "Programa de financiamiento basal" from CMM, Universidad de Chile.

† Department of Mathematics, Universidad Técnica Federico Santa María, Santiago, Chile, (luis.briceno@usm.cl)

‡ School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14850 (dsd95@cornell.edu)

[1] An operator $C : \mathcal{H} \to \mathcal{H}$ is β-cocoercive for some $\beta > 0$ provided that $\langle Cx - Cy, x - y\rangle \geq \beta\|Cx - Cy\|^2$.

problems of the form

$$\underset{\mathrm{x}\in\mathrm{H}}{\text{minimize}}\, \mathrm{f}(\mathrm{x})+\mathrm{g}(\mathrm{Lx})+\mathrm{h}(\mathrm{x}), \tag{1.2}$$

where H and G are real Hilbert spaces, $\mathrm{L}:\mathrm{H}\to\mathrm{G}$ is linear and bounded, $\mathrm{f}:\mathrm{H}\to(-\infty,\infty]$ and $\mathrm{g}:\mathrm{G}\to(-\infty,\infty]$ are lower semicontinuous, convex, and proper, and $\mathrm{h}:\mathrm{H}\to\mathbb{R}$ is convex differentiable with β^{-1}-Lipschitz gradient. Since g may be non smooth, primal algorithms in this context need to evaluate $\mathbf{prox}_{\mathrm{g}\circ L}$ or invert L which can be costly numerically. In order to overcome this difficulty, fully split primal-dual algorithms are proposed, e.g., in [**?**, **?**], in which only $\mathbf{prox}_{\mathrm{g}}$, L, and L^* are computed. These algorithms follow from the first order optimality conditions of (1.2), which, under qualification conditions, can be written as Problem 1 with

$$X=\mathcal{H}=\mathrm{H}\times\mathrm{G},\quad A=\partial\mathrm{f}\times\partial\mathrm{g}^*,\quad B_1=\nabla\mathrm{h}\times\{0\},\quad B_2=\begin{bmatrix}0 & \mathrm{L}^*\\ -\mathrm{L} & 0\end{bmatrix}. \tag{1.3}$$

We have that, for any solution $x=(\mathrm{x}_1^*,\mathrm{x}_2^*)\in\mathrm{zer}(A+B_1+B_2)$, x_1^* solves (1.2), where we denote $\mathrm{zer}\,T=\big\{x\in\mathcal{H}\;\big|\;0\in Tx\big\}$ for any set valued operator $T\colon\mathcal{H}\to 2^{\mathcal{H}}$. A method proposed in [**?**] solves Problem 1 in a more general context by using forward-backward splitting (FB) in the product space with variable metric $\langle\cdot\mid\cdot\rangle_V=\langle V\cdot\mid\cdot\rangle$ for the operators $V^{-1}(A+B_2)$ and $V^{-1}B_1$ with a specific choice of self-adjoint strongly monotone linear operator V. We recall that the forward-backward splitting [**?**, **?**, **?**, **?**] finds a zero of the sum of a maximally monotone and a cocoercive operator, which is a particular case of Problem 1 when $X=\mathcal{H}$ and $B_2=0$. This method provides a sequence obtained from the fixed point iteration of the nonexpansive operator (for some $\gamma\in]0,2\beta[$)

$$T_{\mathrm{FB}}:=J_{\gamma A}\circ(I-\gamma B_1),$$

which converges weakly to a zero of $A+B_1$. Here I stands for the identity map in $\mathcal{H}$ and $J_{\gamma A}=(I+\gamma A)^{-1}$ is the resolvent of γA, which is single valued and nonexpansive. In the context of (1.3), the operators $V^{-1}(A+B_2)$ and $V^{-1}B_1$ are maximally monotone and β-cocoercive in the metric $\langle\cdot\mid\cdot\rangle_V=\langle V\cdot\mid\cdot\rangle$, respectively, which ensures the convergence of forward-backward splitting. The choice of V permits the explicit computation of $J_{V^{-1}(A+B_2)}$, which leads to a sequential method that generalizes the algorithm proposed in [**?**]. A variant for solving (1.2) in the case when $h=0$ is proposed in [**?**]. However, previous methods need the skew linear structure of B_2 in order to obtain an implementable method.

In the general case, since $B:=B_1+B_2$ is monotone and $(\beta^{-1}+L)$–Lipschitz continuous, the forward-backward-forward splitting (FBF) proposed by Tseng in [**?**] solves Problem 1. This method generates a sequence from the fixed point iteration of the operator

$$T_{\mathrm{FBF}}:=P_X\circ\left[(I-\gamma B)\circ J_{\gamma A}\circ(I-\gamma B)+\gamma B\right],$$

which converges weakly to a zero of $A+B$, provided that $\gamma\in]0,(\beta^{-1}+L)^{-1}[$. However, this approach has two drawbacks:

1. FBF needs to evaluate $B=B_1+B_2$ twice per iteration, without taking into advantage the cocoercivity property of B_1. In the particular case when $B_2=0$, this method computes B_1 twice at each iteration, while the forward-backward splitting needs only one computation of B_1 for finding a zero of

$A + B_1$. Even if we cannot ensure that FB is more efficient than FBF in this context, the cost of each iteration of FB is lower than that of FBF, especially when the computation cost of B_1 is high. This is usually the case, for instance, when A, B_1, and B_2 are as in (1.3) and we aim at solving (1.2) representing a variational formulation of some partial differential equation (PDE). In this case, the computation of ∇h frequently amounts to solving a PDE, which is computationally costly.

2. The step size γ in FBF is bounded above by $(\beta^{-1} + L)^{-1}$, which in the case when the influence of B_2 in the problem is low ($B_2 \approx 0$) leads to a method whose step size cannot go too far beyond β. In the case $B_2 = 0$, the step size γ in FB is bounded by 2β. This can affect the performance of the method, since very small stepsizes can lead to slow algorithms.

In this paper we propose a splitting algorithm for solving Problem 1 which overcomes previous drawbacks. The method is derived from the fixed point iteration of the operator $T_\gamma : \mathcal{H} \to \mathcal{H}$, defined by

$$T_\gamma := P_X \circ [(I - \gamma B_2) \circ J_{\gamma A} \circ (I - \gamma(B_1 + B_2)) + \gamma B_2], \tag{1.4}$$

for some $\gamma \in]0, \chi(\beta, L)[$, where $\chi(\beta, L) \leq \min\{2\beta, L^{-1}\}$. The algorithm thus obtained implements B_1 only once by iteration and it reduces to FB or FBF when $X = \mathcal{H}$ and $B_2 = 0$, or $B_1 = 0$, respectively, and in these cases we have $\chi(\beta, 0) = 2\beta$ and $\lim_{\beta\to+\infty} \chi(\beta, L) = L^{-1}$. These results can be found in Theorem 2.2 in Section 2. Moreover, a generalization of FB for finding a point in $X \cap \mathrm{zer}(A+B_1)$ can be derived when $B_2 = 0$. This can be useful when the solution is known to belong to a closed convex set X, which is the case, for example, in convex constrained minimization. The additional projection onto X can improve the performance of the method (see, e.g., [**?**]).

Another contribution of this paper is to include in our method non self-adjoint linear operators in the computation of resolvents and other operators involved. More precisely, in Theorem 3.1 in Section 3, for an invertible linear operator P (not necesarily self-adjoint) we justify the computation of $P^{-1}(B_1 + B_2)$ and $J_{P^{-1}A}$, respectively. In the case when P is self-adjoint and strongly monotone, the properties that A, B_1 and B_2 have with the standard metric are preserved by $P^{-1}A$, $P^{-1}B_1$, and $P^{-1}B_2$ in the metric $\langle \cdot \mid \cdot \rangle_P = \langle P \cdot \mid \cdot \rangle$. In this context, variable metric versions of FB and FBF have been developed in [**?**, **?**]. Of course, a similar generalization can be done for our algorithm, but we go beyond this self-adjoint case and we implement $P^{-1}(B_1 + B_2)$ and $J_{P^{-1}A}$, where the linear operator P is strongly monotone but non necesarily self-adjoint. The key for this implementation is the decomposition $P = S + U$, where U is self-adjoint and strongly monotone and S is skew linear. Our implementation follows after coupling S with the monotone and Lipschitz component B_2 and using some resolvent identities valid for the metric $\langle \cdot \mid \cdot \rangle_U$. One of the important implications of this issue is the justification of the convergence of some Gauss-Seidel type methods in product spaces, which are deduced from our setting for block triangular linear operators P.

Additionally, we provide a modification of the previous method in which linear operators P may vary among iterations in Theorem 4.2 in Section 4. In the case when, for every iteration $k \in \mathbb{N}$, P_k is self-adjoint, this feature has also been implemented for FB and FBF in [**?**, **?**] but with a strong dependence between P_{k+1} and P_k coming from the variable metric approach. Instead, in the general case, we modify our method for avoiding variable metrics, which allows us to ensure convergence under weaker conditions. For instance, in the case when $B_2 = 0$ and P_k is self-adjoint and ρ_k-strongly monotone for some $\rho_k > 0$, our condition on our FB variable metric version reduces to $(2\beta-\varepsilon)\rho_k > 1$ for every $k \in \mathbb{N}$. In the case when $P_k = I/\gamma_k$ this condition reduces to $\gamma_k < 2\beta-\varepsilon$ which is a standard assumption for FB with variable stepsizes. Hence, our condition on operators $(P_k)_{k\in\mathbb{N}}$ can be interpreted as "step-size" bounds.

Moreover, in Section 5 we use our methods in composite primal-dual inclusions, obtaining generalizations and new versions of several primal-dual methods [**?**, **?**, **?**, **?**]. We provide comparisons among methods and new bounds on stepsizes which improve several bounds in the literature. Finally, for illustrating the flexibility of the proposed methods, in Section 6 we apply them to the obstacle problem in PDE's and to Empirical Risk Minimization. In the first example, we take advantage to dropping the extra forward step on B_1, which amounts to reduce the computation of a PDE by iteration. In the second example, we use non self-adjoint linear operators in order to obtain a Gauss-Seidel structure which can be preferable to parallel architectures when the dimension is high.

2. Convergence theory. This section is devoted to study the conditions ensuring the convergence of the method $z^{k+1} = T_{\gamma_k} z^k$ for any starting point $z^0 \in \mathcal{H}$, where, for every $\gamma > 0$, T_γ is defined in (1.4). We first prove that T_γ is quasi-nonexpansive for a suitable choice of γ and satisfies $\mathrm{Fix}(T_\gamma) = \mathrm{zer}(A + B_1 + B_2) \cap X$. Using these results we prove the weak convergence of iterates $\{z^k\}_{k\in\mathbb{N}}$ to a solution to Problem 1.

PROPOSITION 2.1 (Properties of T_γ). *Let $\gamma > 0$ and assume that hypotheses of Problem 1 hold. Then,*

1. *If $\gamma < L^{-1}$ we have* $\mathrm{Fix}(T_\gamma) = \mathrm{zer}(A + B_1 + B_2) \cap X$.
2. *For all $z^* \in \mathrm{Fix}(T_\gamma)$ and $z \in \mathcal{H}$, we have*

$$\begin{aligned}\|T_\gamma z - z^*\|^2 \le \|z - z^*\|^2 &- L^2(\chi^2 - \gamma^2)\|z - J_{\gamma A}(z - \gamma B_1 z - \gamma B_2 z)\|^2\\ &- \frac{2\beta\gamma}{\chi}(\chi-\gamma)\|B_1 z - B_1 z^*\|^2\\ &- \frac{\chi}{2\beta}\left\|z - J_{\gamma A}(z - \gamma B_1 z - \gamma B_2 z) - \frac{2\beta\gamma}{\chi}(B_1 z - B_1 z^*)\right\|^2,\end{aligned} \tag{2.1}$$

where

$$\chi := \frac{4\beta}{1 + \sqrt{1 + 16\beta^2 L^2}} \le \min\{2\beta, L^{-1}\}. \tag{2.2}$$

Proof. Part 1: Let $z^* \in \mathcal{H}$. We have

$$\begin{aligned} z^* \in \mathrm{zer}(A + B_1 + B_2) \cap X \quad &\Leftrightarrow \quad z^* \in X \quad \text{and} \quad 0 \in Az^* + B_1 z^* + B_2 z^*\\ &\Leftrightarrow \quad z^* \in X \quad \text{and} \quad -\gamma(B_1 z^* + B_2 z^*) \in \gamma A z^*\\ &\Leftrightarrow \quad z^* \in X \quad \text{and} \quad z^* = J_{\gamma A}\left(z^* - \gamma(B_1 z^* + B_2 z^*)\right). \end{aligned} \tag{2.3}$$

Then, since B_2 is single-valued in $\mathrm{dom} A$, if $z^* \in \mathrm{zer}(A + B_1 + B_2) \cap X$ we have $B_2 z^* = B_2 J_{\gamma A}(z^* - \gamma(B_1 z^* + B_2 z^*))$ and, hence, $T_\gamma z^* = P_X(z^*) = z^*$ which yields $\mathrm{zer}(A + B_1 + B_2) \cap X \subset \mathrm{Fix}\, T_\gamma$. For the converse, if $z^* \in \mathrm{Fix}\, T_\gamma$ it is easy to see that $z^* \in X$ and

$$z^* - J_{\gamma A}(z^* - \gamma(B_1 + B_2)z^*) = \gamma\left(B_2 z^* - B_2 J_{\gamma A}(z^* - \gamma(B_1 + B_2)z^*)\right),$$

which, from the Lipschitz continuity of B_2 in $\mathrm{dom}A \cup X$ yields

$$\begin{aligned}\|z^* - J_{\gamma A}(z^* - \gamma(B_1+B_2)z^*)\| &= \gamma\|B_2 z^* - B_2 J_{\gamma A}(z^* - \gamma(B_1+B_2)z^*)\| \\ &\leq \gamma L\|z^* - J_{\gamma A}(z^* - \gamma(B_1+B_2)z^*)\|.\end{aligned}$$

Therefore, if $\gamma < L^{-1}$ we deduce $z^* = J_{\gamma A}(z^* - \gamma(B_1+B_2)z^*)$ and the result follows from (2.3).

Part 2: Let $z^* \in \mathrm{Fix}\, T_\gamma$ and define $B := B_1 + B_2$, $y := z - \gamma Bz$, $x := J_{\gamma A} y$, and $z^+ = T_\gamma z$. Note that $(x, y - x) \in \mathrm{gra}(\gamma A)$ and, from Part 1, $(z^*, -\gamma B z^*) \in \mathrm{gra}(\gamma A)$. Hence, by the monotonicity of A and B_2, we have $\langle x - z^*, x - y - \gamma B z^*\rangle \leq 0$ and $\langle x - z^*, \gamma B_2 z^* - \gamma B_2 x\rangle \leq 0$. Thus,

$$\begin{aligned}\langle x - z^*, x - y - \gamma B_2 x\rangle &= \langle x - z^*, \gamma B_1 z^*\rangle + \langle x - z^*, x - y - \gamma B z^*\rangle \\ &\quad + \langle x - z^*, \gamma B_2 z^* - \gamma B_2 x\rangle \\ &\leq \langle x - z^*, \gamma B_1 z^*\rangle.\end{aligned}$$

Therefore, we have

$$\begin{aligned}2\gamma\langle x - z^*, B_2 z - B_2 x\rangle &= 2\langle x - z^*, \gamma B_2 z + y - x\rangle + 2\langle x - z^*, x - y - \gamma B_2 x\rangle \\ &\leq 2\langle x - z^*, \gamma B z + y - x\rangle + 2\langle x - z^*, \gamma B_1 z^* - \gamma B_1 z\rangle \\ &= 2\langle x - z^*, z - x\rangle + 2\langle x - z^*, \gamma B_1 z^* - \gamma B_1 z\rangle \\ &= \|z - z^*\|^2 - \|x - z^*\|^2 - \|z - x\|^2 + 2\langle x - z^*, \gamma B_1 z^* - \gamma B_1 z\rangle.\end{aligned} \tag{2.4}$$

In addition, by cocoercivity of B_1, for all $\varepsilon > 0$, we have

$$\begin{aligned}2\langle x - z^*, \gamma B_1 z^* - \gamma B_1 z\rangle &= 2\langle z - z^*, \gamma B_1 z^* - \gamma B_1 z\rangle + 2\langle x - z, \gamma B_1 z^* - \gamma B_1 z\rangle \\ &\leq -2\gamma\beta\|B_1 z - B_1 z^*\|^2 + 2\langle x - z, \gamma B_1 z^* - \gamma B_1 z\rangle \\ &= -2\gamma\beta\|B_1 z - B_1 z^*\|^2 + \varepsilon\|z - x\|^2 + \frac{\gamma^2}{\varepsilon}\|B_1 z - B_1 z^*\|^2 \\ &\qquad - \varepsilon\left\|z - x - \frac{\gamma}{\varepsilon}(B_1 z - B_1 z^*)\right\|^2 \\ &= \varepsilon\|z - x\|^2 - \gamma\left(2\beta - \frac{\gamma}{\varepsilon}\right)\|B_1 z - B_1 z^*\|^2 \\ &\qquad - \varepsilon\left\|z - x - \frac{\gamma}{\varepsilon}(B_1 z - B_1 z^*)\right\|^2.\end{aligned} \tag{2.5}$$

Hence, combining (2.4) and (2.5), it follows from $z^* \in X$, the nonexpansivity of P_X, and the Lipschitz property of B_2 in $X \cup \mathrm{dom}A$ that

$$\begin{aligned}\|z^+ - z^*\|^2 &\leq \|x - z^* + \gamma B_2 z - \gamma B_2 x\|^2 \\ &= \|x - z^*\|^2 + 2\gamma\langle x - z^*, B_2 z - B_2 x\rangle + \gamma^2\|B_2 z - B_2 x\|^2 \\ &\leq \|x - z^*\|^2 + \|z - z^*\|^2 - \|x - z^*\|^2 - \|z - x\|^2 + \gamma^2\|B_2 z - B_2 x\|^2 \\ &\quad + \varepsilon\|z - x\|^2 - \gamma\left(2\beta - \frac{\gamma}{\varepsilon}\right)\|B_1 z - B_1 z^*\|^2 - \varepsilon\left\|z - x - \frac{\gamma}{\varepsilon}(B_1 z - B_1 z^*)\right\|^2 \\ &\leq \|z - z^*\|^2 - L^2\left(\frac{1-\varepsilon}{L^2} - \gamma^2\right)\|z - x\|^2 - \frac{\gamma}{\varepsilon}(2\beta\varepsilon - \gamma)\|B_1 z - B_1 z^*\|^2 \\ &\quad - \varepsilon\left\|z - x - \frac{\gamma}{\varepsilon}(B_1 z - B_1 z^*)\right\|^2.\end{aligned}$$

In order to obtain the largest interval for γ ensuring that the two rightmost terms in the above equation are negative, we choose the value ε so that $\sqrt{1-\varepsilon}/L = 2\beta\varepsilon$, which yields $\varepsilon = (-1+\sqrt{1+16\beta^2L^2})(8\beta^2L^2)^{-1}$. For this choice of ε we obtain $\chi = \sqrt{1-\varepsilon}/L = 2\beta\varepsilon$. □

THEOREM 2.2 (Forward-backward-half forward algorithm). *In Problem 1, suppose that $X \subset \mathrm{dom}B_2$ and that $A+B_2$ is maximally monotone. Let $z^0 \in \mathcal{H}$, let $\varepsilon \in]0,\chi/2[$, let $\{\gamma_k\}_{k\in\mathbb{N}}$ be a sequence of stepsizes in $[\varepsilon,\chi-\varepsilon]$, where χ is defined in* (2.2). *Then the sequence recursively defined by $z^{k+1} := T_{\gamma_k} z^k$ converges weakly to a solution to Problem 1 and satisfies the following recursion:*

$$(\forall k\in\mathbb{N}) \quad \left\lfloor \begin{array}{l} x^k = J_{\gamma_k A}(z^k - \gamma_k(B_1+B_2)z^k) \\ z^{k+1} = P_X\big(x^k + \gamma_k B_2 z^k - \gamma_k B_2 x^k\big). \end{array} \right. \tag{2.6}$$

Proof. It follows from Proposition 2.1 that the sequence $\{z^k\}_{k\in\mathbb{N}}$ is Fejér monotone with respect to $\mathrm{zer}(A+B_1+B_2)\cap X$. Thus, to show that $\{z^k\}_{k\in\mathbb{N}}$ converges weakly to a solution to Problem 1, we just need to show that all of its weak subsequential limits lie in $\mathrm{zer}(A+B_1+B_2)\cap X$ [**?**, Theorem 5.5]. Indeed, it follows from Proposition 2.1 and our hypotheses on the stepsizes that, for every $z^* \in \mathrm{Fix}\, T_\gamma$,

$$\begin{aligned} \|z^k - z^*\|^2 - \|z^{k+1}-z^*\|^2 \geq L^2\varepsilon^2\|z^k - x^k\|^2 + \frac{2\beta\varepsilon^2}{\chi}\|B_1 z^k - B_1 z^*\|^2 \\ + \frac{\chi}{2\beta}\left\| z^k - x^k - \frac{2\beta\gamma_k}{\chi}(B_1 z^k - B_1 z^*)\right\|^2. \end{aligned} \tag{2.7}$$

Therefore, we deduce from [**?**, Lemma 3.1] that

$$z^k - x^k \to 0 \quad \text{and} \quad B_1 z^k - B_1 z^* \to 0 \tag{2.8}$$

when $L > 0$ and $0 < \beta < \infty$[2]. Now let $z \in \mathcal{H}$ be the weak limit point of some subsequence of $\{z^k\}_{k\in\mathbb{N}}$. Since $z^k \in X$ for every $k \geq 1$ and X is weakly sequentially closed [**?**, Theorem 3.32] we deduce $z \in X$. Moreover, it follows from $x^k = J_{\gamma_k A}(z^k - \gamma_k B_1 z^k - \gamma_k B_2 z^k)$ that $u^k := \gamma_k^{-1}(z^k - x^k) - B_1 z^k + B_2 x^k - B_2 z^k \in (A+B_2)x^k$ and (2.8) yields $u^k \to -B_1 z^*$. Now, since B_1 and $A+B_2$ are maximally monotone, their graphs are closed in the weak-strong topology in $\mathcal{H}\times\mathcal{H}$, which yields $B_1 z^* = B_1 z$ and $-B_1 z^* = -B_1 z \in Az + B_2 z$ and the result follows. □

REMARK 1. *The maximal monotonicity assumption on $A+B_2$ is satisfied, for instance, if* $\mathrm{cone}(\mathrm{dom}A - \mathrm{dom}B_2) = \overline{\mathrm{span}}(\mathrm{dom}A - \mathrm{dom}B_2)$*, where, for any set $D \subset \mathcal{H}$,* $\mathrm{cone}(D) = \{\lambda d \mid \lambda \in \mathbb{R}_+, d \in D\}$ *and $\overline{\mathrm{span}}(D)$ is the smallest closed linear subspace of $\mathcal{H}$ containing D [**?**, Theorem 3.11.11]. Since* $\mathrm{dom}A \subset \mathrm{dom}B_2$*, a more tractable sufficient condition is* $\mathrm{span}(\mathrm{dom}A - \mathrm{dom}A)$ *is closed (see [**?**, Example 6.10]).*

REMARK 2. *The stepsize upper bound $\chi = \chi(\beta, L)$ defined in* (2.2) *depends on the cocoercivity parameter β of B_1 and the Lipschitz parameter L of B_2. In order to fully recover Tseng's splitting algorithm or the forward-backward algorithm in the cases when B_1 or B_2 are zero, respectively, we study the asymptotic behaviour of $\chi(\beta,L)$ when $L \to 0$ and $\beta \to +\infty$. It is easy to verify that*

$$\lim_{L\to 0}\chi(\beta,L) = 2\beta \quad \textit{and} \quad \lim_{\beta\to+\infty}\chi(\beta,L) = \frac{1}{L},$$

[2] The case $B_1 = 0$ ($\beta = +\infty$) has been studied by Tseng in [**?**]. In the case when $B_2 = 0$ we can also obtain convergence from Proposition 2.1, since $L = 0$ implies $\chi = 2\beta$ and even since the first term in the right hand side of (2.7) vanishes, the other two terms yield $z^k - x^k \to 0$.

which are exactly the bounds on the stepsizes of forward-backward and Tseng's splittings.

3. Forward-backward-half forward splitting with non self-adjoint linear operators. In this section, we introduce modified resolvents $J_{P^{-1}A} = (I + P^{-1}A)^{-1}$, which depend on an invertible linear mapping P. In some cases, it is preferable to compute the resolvent $J_{P^{-1}A}$ instead of the resolvent $J_A = (I + A)^{-1}$ because the former may be easier to compute than the latter or, when P is triangular by blocks in a product space, the former may *order the component computation* of the resolvent, replacing a parallel computation with a Gauss-Seidel style sequential computation. However, $P^{-1}A$ may not be maximally monotone. The following result allows us to use some non self-adjoint linear operators in the computation of the resolvent by using specific metrics.

THEOREM 3.1 (New Metrics and T_γ). *Let $P : \mathcal{H} \to \mathcal{H}$ be an bounded linear operator, let $U := (P + P^*)/2$ and $S := (P - P^*)/2$ be the the self-adjoint and skew symmetric components of P, respectively. Suppose that there exists $\rho > 0$ such that*

$$(\forall x \in \mathcal{H}) \qquad \rho\|x\|^2 \leq \langle Ux, x\rangle \quad \text{and} \quad K^2 < \rho\left(\rho - \frac{1}{2\beta}\right), \tag{3.1}$$

where $K \geq 0$ is the Lipschitz constant of $B_2 - S$. Let $z^0 \in \mathcal{H}$ and let $\{z^k\}_{k\in\mathbb{N}}$ be the sequence defined by the following iteration:

$$(\forall k \in \mathbb{N}) \qquad \left\lfloor \begin{array}{l} x^k = J_{P^{-1}A}(z^k - P^{-1}(B_1 + B_2)z^k) \\ z^{k+1} = P_X^U(x^k + U^{-1}(B_2 z^k - B_2 x^k - S(z^k - x^k))), \end{array} \right. \tag{3.2}$$

where P_X^U is the projection operator of X under the inner product $\langle\cdot,\cdot\rangle_U$. Then $\{z^k\}_{k\in\mathbb{N}}$ converges weakly to a solution to Problem 1.

Proof. Note that, since U is invertible from (3.1), by adding and subtracting the skew term S, Problem 1 is equivalent to

$$\text{find } x \in X \text{ such that } \quad 0 \in U^{-1}(A + S)x + U^{-1}B_1 + U^{-1}(B_2 - S)x. \tag{3.3}$$

Because S and $-S$ are both monotone and Lipschitz, $\mathcal{A} := U^{-1}(A + S)$ is monotone; $\mathcal{B}_1 := U^{-1}B_1$ is $\rho\beta$-cocoercive [**?**, Proposition 1.5]; and $\mathcal{B}_2 := U^{-1}(B_2 - S)$ is monotone and $\rho^{-1}K$-Lipschitz under the inner product $\langle\cdot,\cdot\rangle_U = \langle U\cdot \mid \cdot\rangle$, where K is the Lipschitz constant of $C := B_2 - S$[3]. For the last assertion note that, for every $x, y \in \mathcal{H}$,

$$\|\mathcal{B}_2 x - \mathcal{B}_2 y\|_U^2 = \langle U^{-1}(Cx - Cy), Cx - Cy\rangle \leq \rho^{-1}K^2\|x - y\|^2 \leq \rho^{-2}K^2\|x - y\|_U^2.$$

Moreover, the stepsize condition reduces to

$$\gamma = 1 < \frac{4\beta\rho}{1 + \sqrt{1 + 16\beta^2 K^2}} = \frac{-\rho + \sqrt{\rho^2 + 16\beta^2\rho^2 K^2}}{4\beta K^2} \tag{3.4}$$

or, equivalently,

$$(4\beta K^2 + \rho)^2 < \rho^2 + 16\beta^2\rho^2 K^2 \quad \Leftrightarrow \quad 2\beta K^2 + \rho < 2\beta\rho^2, \tag{3.5}$$

[3]Note that $K \leq L + \|S\|$, but this constant is not precise when, for instance, $B_2 = S$.

which yields the second condition in (3.1). Therefore, since $\mathcal{A}+\mathcal{B}_2 = U^{-1}(A+B_2)$ is maximally monotone in $(\mathcal{H}, \|\cdot\|_U)$, the inclusion (3.3) meets the conditions of Theorem 2.2 under this metric, and by iterating the quasi-nonexpansive operator

$$T_1 = P_X^U \circ \left[(I-\mathcal{B}_2)\circ J_{\mathcal{A}} \circ \big(I-(\mathcal{B}_1+\mathcal{B}_2)\big) + \mathcal{B}_2\right], \tag{3.6}$$

we obtain a sequence that weakly converges to a fixed point of T_1, and hence, to a solution of $\mathrm{zer}(A+B_1+B_2)\cap X$. Only the simplified form (3.2) remains to be proved. For every $z\in\mathcal{H}$, we have

$$\begin{aligned} x = J_{U^{-1}(A+S)}(z-U^{-1}(B_1+B_2-S)z) \\ \Leftrightarrow\quad & (z-U^{-1}(B_1+B_2-S)z)-x \in U^{-1}(A+S)x \\ \Leftrightarrow\quad & (U+S)z-(B_1+B_2)z-(U+S)x\in Ax \\ \Leftrightarrow\quad & x = J_{P^{-1}A}(z-P^{-1}(B_1+B_2)z), \end{aligned}$$

which yields

$$T_1 = P_X^U \circ \left[(I_{\mathcal{H}}-U^{-1}(B_2-S))\circ J_{P^{-1}A}(z-P^{-1}(B_1+B_2)z)+U^{-1}(B_2-S)\right]$$

and completes the proof.
□

Remark 3.

1. *Note that, in the particular case when* $P=\mathrm{Id}\,/\gamma$, *the algorithm* (3.2) *reduces to the constant case in* (2.6). *Moreover,* $U=P$, $S=0$, $K=L$, $\rho=1/\gamma$ *and the second condition in* (3.1) *reduces to* $\gamma<\chi$ *with* χ *defined in* (2.2). *Hence, this assumption can be seen as a kind of "step size" condition on* P.
2. *As in Remark 2, note that the second condition in* (3.1) *depends on the cocoercivity parameter* β *and the Lipschitz constant* L. *In the case when* B_1 *is zero, we can take* $\beta\to+\infty$ *and this condition reduces to* $K<\rho$. *On the other hand, if* B_2 *is zero we can take* $L=0$, *then* $K=\|S\|$ *and, hence, the condition reduces to* $\|S\|^2<\rho(\rho-1/(2\beta))$. *In this way we obtain convergent versions of Tseng's splitting and forward-backward algorithm with non self-adjoint linear operators by setting* $B_1=0$ *or* $B_2=0$ *in* (3.2), *respectively.*
3. *When* $S=0$, *from Theorem 3.1 we also recover the variable metric versions of Tseng's forward-backward-forward splitting [?, Theorem 3.1] and forward-backward [?, Theorem 4.1] in the cases* $B_1=0$ *and* $B_2=0$, *respectively, when the step-sizes are constant. Of course, when* $S=0$, $U=\mathrm{Id}\,/\gamma$, *and* $\rho=1/\gamma$, *we recover the classical bound for step-sizes in the standard metric case for each method.*
4. *For a particular choice of operators and metric, the variable forward-backward method discused before has been used for solving primal-dual composite inclusions in [?]. This approach generalizes, e.g., the method in [?]. In Section 5 we compare the application of our method in the primal-dual context with [?] and other methods in the literature.*
5. *In the particular instance when* $B_1=B_2=0$, *we need* $\|S\|<\rho$ *and we obtain from* (3.2) *the following version of the proximal point algorithm (we consider* $X=\mathcal{H}$ *for simplicity)*

$$\begin{aligned} z^0\in\mathcal{H},\quad (\forall k\in\mathbb{N})\quad z^{k+1} &= J_{P^{-1}A}z^k + U^{-1}S(J_{P^{-1}A}z^k - z^k) \\ &= (\mathrm{Id}\,-U^{-1}P)z^k + U^{-1}PJ_{P^{-1}A}z^k. \end{aligned} \tag{3.7}$$

Moreover, in the case when $A = B_2 = 0$*, since* $U^{-1} \circ S \circ P^{-1} = U^{-1} - P^{-1}$*, we recover from* (3.2) *the gradient-type method:*

$$z^0 \in \mathcal{H}, \quad (\forall k \in \mathbb{N}) \quad z^{k+1} = z^k - U^{-1} B_1 z^k. \tag{3.8}$$

6. *In the particular case when* $X = \mathcal{H}$ *and* B_2 *is linear, in [**?**] a method involving* B_2^* *is proposed. In the case when,* B_2 *is skew linear, i.e.,* $B_2^* = -B_2$ (3.1) *reduces to this method in the case* $\alpha_n \equiv 1$ *and* $S = P$*. The methods are different in general.*

4. Allowing variable P **and avoiding inversion of** U**.** In Algorithm (3.2), the linear operator U must be inverted. In this section, for the special case $X = \mathcal{H}$, we show how to replace this sometimes costly inversion with a single multiplication by the map P, which, in addition, may vary at each iteration. This new feature is a consequence of Proposition 4.1 below, which allows us to obtain from an operator of the class $\mathfrak{T}$ in $(\mathcal{H}, \|\cdot\|_U)$, another operator of the same class in $(\mathcal{H}, \|\cdot\|)$ preserving the set of fixed points. This change to the standard metric allows us to use different linear operators at each iteration by avoiding classical restrictive additional assumptions of the type $U_{n+1} \preccurlyeq U_n(1+\eta_n)$ with $(\eta_n)_{n\in\mathbb{N}}$ in ℓ^1_+. We recall that an operator $\mathcal{S}\colon \mathcal{H} \to \mathcal{H}$ belongs to the class $\mathfrak{T}$ in $(\mathcal{H}, \|\cdot\|)$ if and only if $\operatorname{dom}\mathcal{S} = \mathcal{H}$ and $(\forall y \in \operatorname{Fix}\mathcal{S})(\forall x \in \mathcal{H}) \quad \|x - \mathcal{S}x\|^2 \leq \langle x - \mathcal{S}x \mid x - y\rangle$.

Proposition 4.1. *Let* $U\colon \mathcal{H} \to \mathcal{H}$ *a self-adjoint bounded linear operator such that, for every* $x \in \mathcal{H}$, $\langle Ux \mid x\rangle \geq \rho\|x\|^2$*, for some* $\rho > 0$*, let* $0 < \mu \leq \|U\|^{-1}$*, and let* $\mathcal{S}\colon \mathcal{H} \to \mathcal{H}$ *be an operator in the class* $\mathfrak{T}$ *in* $(\mathcal{H}, \|\cdot\|_U)$*. Then, the operator* $\mathcal{Q} = \operatorname{Id} - \mu U(\operatorname{Id} - \mathcal{S})$ *belongs to the class* $\mathfrak{T}$ *in* $(\mathcal{H}, \|\cdot\|)$ *and* $\operatorname{Fix}\mathcal{S} = \operatorname{Fix}\mathcal{Q}$*.*

Proof. First note that, under the assumptions on U it is invertible and, from [**?**, Lemma 2.1], we deduce

$$(\forall x \in \mathcal{H}) \quad \|x\|_U^2 = \langle Ux \mid x\rangle = \left\langle Ux \mid U^{-1}Ux\right\rangle \geq \|U\|^{-1}\|Ux\|^2, \tag{4.1}$$

and $\operatorname{Fix}\mathcal{S} = \operatorname{Fix}\mathcal{Q}$ thus follows from the definition of $\mathcal{Q}$. Now let $y \in \operatorname{Fix}\mathcal{S}$ and $x \in \mathcal{H}$. We have from (4.1) that

$$\begin{aligned}
\|x - \mathcal{S}x\|_U^2 \leq \langle x - \mathcal{S}x \mid x - y\rangle_U \;&\Leftrightarrow\; \|x - \mathcal{S}x\|_U^2 \leq \langle U(x - \mathcal{S}x) \mid x - y\rangle \\
&\Rightarrow\; \|U\|^{-1}\|U(x - \mathcal{S}x)\|^2 \leq \langle U(x - \mathcal{S}x) \mid x - y\rangle \\
&\Leftrightarrow\; \frac{\|U\|^{-1}}{\mu}\|\mu U(x - \mathcal{S}x)\|^2 \leq \langle \mu U(x - \mathcal{S}x) \mid x - y\rangle \\
&\Leftrightarrow\; \frac{\|U\|^{-1}}{\mu}\|x - \mathcal{Q}x\|^2 \leq \langle x - \mathcal{Q}x \mid x - y\rangle
\end{aligned} \tag{4.2}$$

and, hence, if $\mu \in]0, \|U\|^{-1}]$ we deduce the result. □

Theorem 4.2. *Let* $\{P_k\}_{k\in\mathbb{N}}$ *be a sequence of bounded, linear maps from* $\mathcal{H}$ *to* $\mathcal{H}$*. For each* $k \in \mathbb{N}$*, let* $U_k := (P_k + P_k^*)/2$ *and* $S_k := (P_k - P_k^*)/2$ *be the self-adjoint and skew symmetric components of* P_k*, respectively. Suppose that* $M := \sup_{k\in\mathbb{N}} \|U_k\| < \infty$ *and that there exist* $\varepsilon \in]0, (2M)^{-1}[$*,* $\rho > 0$*, and* $\{\rho_k\}_{k\in\mathbb{N}} \subseteq [\rho, \infty[$ *such that, for every* $k \in \mathbb{N}$*,*

$$(\forall x \in \mathcal{H}) \qquad \rho_k\|x\|^2 \leq \langle U_k x, x\rangle \qquad \textit{and} \qquad K_k^2 \leq \frac{\rho_k}{1+\varepsilon}\left(\frac{\rho_k}{1+\varepsilon} - \frac{1}{2\beta}\right), \tag{4.3}$$

where $K_k \geq 0$ is the Lipschitz constant of $B_2 - S_k$. Let $\{\lambda_k\}_{k\in\mathbb{N}}$ be a sequence in $[\varepsilon, \|U_k\|^{-1} - \varepsilon]$, let $z^0 \in \mathcal{H}$, and let $\{z^k\}_{k\in\mathbb{N}}$ be a sequence of points defined by the following iteration:

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} x^k = J_{P_k^{-1}A}(z^k - P_k^{-1}(B_1 + B_2)z^k) \\ z^{k+1} = z^k + \lambda_k \left(P_k(x^k - z^k) + B_2 z^k - B_2 x^k\right). \end{array} \right. \tag{4.4}$$

Then $\{z^k\}_{k\in\mathbb{N}}$ converges weakly to a solution to Problem 1.

Proof. For every invertible and bounded linear map $P : \mathcal{H} \to \mathcal{H}$, let us denote by $\mathcal{T}_P : \mathcal{H} \to \mathcal{H}$ the forward-backward-forward operator of Theorem 3.1 in the case $X = \mathcal{H}$, which associates, to every $z \in \mathcal{H}$,

$$\mathcal{T}_P z = x_z + U^{-1}(B_2 z - B_2 x_z - S(z - x_z)),$$

where $x_z = J_{P^{-1}A}(z - P^{-1}(B_1 + B_2)z)$. Recall that, from (2.1) and the proof of Theorem 3.1, $\mathcal{T}_P$ is a quasi-nonexpansive mapping in $\mathcal{H}$ endowed with the scalar product $\langle \cdot \mid \cdot \rangle_U$. Observe that multiplying $I - \mathcal{T}_P$ by U on the left yields a U^{-1}-free expression:

$$\begin{aligned} & (I - \mathcal{T}_P)(z) = (z - x_z) + U^{-1}S(z - x_z) - U^{-1}(B_2 z - B_2 x_z) \\ \Leftrightarrow \quad & U(I - \mathcal{T}_P)(z) = (U + S)(z - x_z) + B_2 x_z - B_2 z \\ & \qquad = P(z - x_z) + B_2 x_z - B_2 z. \end{aligned} \tag{4.5}$$

Note that, since $\mathcal{T}_P$ is quasi-nonexpansive in $(\mathcal{H}, \|\cdot\|_U)$, it follows from [**?**, Proposition 2.2] that $\mathcal{S} := (\mathrm{Id} + \mathcal{T}_P)/2$ belongs to the class $\mathfrak{T}$ in $(\mathcal{H}, \|\cdot\|_U)$ and, from Proposition 4.1 and (4.5) we obtain that the operator

$$\mathcal{Q}_P := I - \|U\|^{-1}U(I - \mathcal{S}) = I - \frac{\|U\|^{-1}}{2}U(I - \mathcal{T}_P) \tag{4.6}$$

belongs to the class $\mathfrak{T}$ in $(\mathcal{H}, \|\cdot\|)$ and $\operatorname{Fix}\mathcal{S} = \operatorname{Fix}\mathcal{Q}_P = \operatorname{zer}(U(I - \mathcal{T}_P)) = \operatorname{Fix}(\mathcal{T}_P) = \operatorname{zer}(A + B_1 + B_2)$. Hence, from (4.5) and (4.6), the algorithm (4.4) can be written equivalently as

$$\begin{aligned} z^{k+1} &= z^k - \lambda_k(P_k(z^k - x_{z^k}) + B_2 x_{z^k} - B_2 z^k) \\ &= z^k + 2\lambda_k\|U_k\|(\mathcal{Q}_{P_k} z^k - z^k). \end{aligned} \tag{4.7}$$

Hence, since $0 < \liminf \lambda_k\|U_k\| \leq \limsup \lambda_k\|U_k\| < 1$, it follows from [**?**, Proposition 4.2 and Theorem 4.3] that $(\|z^k - \mathcal{Q}_{P_k} z^k\|^2)_{k\in\mathbb{N}}$ is a summable sequence and $\{z^k\}_{k\in\mathbb{N}}$ converges weakly in $(\mathcal{H}, \langle \cdot \mid \cdot \rangle)$ to a solution to $\cap_{k\in\mathbb{N}} \operatorname{Fix}\mathcal{T}_{P_k} = \operatorname{zer}(A + B_1 + B_2)$ if and only if every weak limit of the sequence is a solution. Note that, since (4.3) yields $\|U_k^{-1}\| \leq \rho_k^{-1}$, we have

$$\begin{aligned} \|z^k - \mathcal{T}_{P_k} z^k\|_{U_k}^2 &= \left\langle U_k(z^k - \mathcal{T}_{P_k} z^k) \mid z^k - \mathcal{T}_{P_k} z^k \right\rangle \\ &\leq \|U_k(z^k - \mathcal{T}_{P_k} z^k)\| \, \|z^k - \mathcal{T}_{P_k} z^k\| \\ &= \|U_k^{-1}\| \|U_k(z^k - \mathcal{T}_{P_k} z^k)\|^2 \\ &\leq 4\|U_k\|^2 \rho_k^{-1} \, \|z^k - \mathcal{Q}_{P_k} z^k\|^2 \\ &\leq 4M^2\rho^{-1}\|z^k - \mathcal{Q}_{P_k} z^k\|^2 \to 0. \end{aligned} \tag{4.8}$$

Moreover, since $\mathcal{T}_{P_k}$ coincides with T_1 defined in (3.6) involving the operators $\mathcal{A}_k := U_k^{-1}(A+S_k)$, $\mathcal{B}_{1,k} = U_k^{-1}B_1$, and $\mathcal{B}_{2,k} = U_k^{-1}(B_2 - S_k)$ which are monotone, $\rho_k\beta$-cocoercive, and monotone and $\rho_k^{-1}K_k$-lipschitzian in $(\mathcal{H}, \|\cdot\|_{U_k})$, respectively, we deduce from (2.1) that, for every $z^* \in \mathrm{zer}(A+B_1+B_2) = \cap_{k\in\mathbb{N}} \mathrm{zer}(\mathcal{A}_k + \mathcal{B}_{1,k} + \mathcal{B}_{2,k})$ we have

$$\begin{aligned}
&\rho_k^{-2}K_k^2(\chi_k^2-1)\|z^k - J_{P_k^{-1}A}(z^k - P_k^{-1}(B_1+B_2)z^k)\|_{U_k}^2\\
&+\frac{2\beta\rho_k}{\chi_k}(\chi_k-1)\|U_k^{-1}(B_1z^k - B_1z^*)\|_{U_k}^2\\
&+\frac{\chi_k}{2\beta\rho_k}\left\|z^k - J_{P_k^{-1}A}(z^k - P_k^{-1}(B_1z^k+B_2z^k)) - \frac{2\beta\rho_k}{\chi_k}U_k^{-1}(B_1z^k - B_1z^*)\right\|_{U_k}^2\\
&\le \|z^k - z^*\|_{U_k}^2 - \|\mathcal{T}_{P_k}z^k - z^*\|_{U_k}^2\\
&= -\|\mathcal{T}_{P_k}z^k - z^k\|_{U_k}^2 - 2\langle \mathcal{T}_{P_k}z^k - z^k, z^* - z^k\rangle_{U_k}\\
&\le -\|\mathcal{T}_{P_k}z^k - z^k\|_{U_k}^2 + 2M\|\mathcal{T}_{P_k}z^k - z^k\|_{U_k}\|z^* - z^k\|,
\end{aligned} \tag{4.9}$$

where

$$\chi_k := \frac{4\beta\rho_k}{1+\sqrt{1+16\beta^2K_k^2}} \le \rho_k \min\{2\beta, K_k^{-1}\}. \tag{4.10}$$

By straightforward computations in the line of (3.4) and (3.5) we deduce that (4.3) implies, for all $k\in\mathbb{N}$, $\chi_k \ge 1+\varepsilon$, $K_k \le \rho_k \le \|U_k\| \le M$ and, hence, we deduce from (4.9) and (4.3) that

$$\begin{aligned}
&\frac{\varepsilon\rho K_k^2}{M^2}\|z^k - J_{P_k^{-1}A}(z^k - P_k^{-1}(B_1+B_2)z^k)\|^2 + \varepsilon\rho\|U_k^{-1}(B_1z^k - B_1z^*)\|^2\\
&\quad+\frac{\rho}{2\beta M}\left\|z^k - J_{P_k^{-1}A}(z^k - P_k^{-1}(B_1z^k+B_2z^k)) - \frac{2\beta\rho_k}{\chi_k}U_k^{-1}(B_1z^k - B_1z^*)\right\|^2\\
&\qquad\le -\|\mathcal{T}_{P_k}z^k - z^k\|_{U_k}^2 + 2M\|\mathcal{T}_{P_k}z^k - z^k\|_{U_k}\|z^* - z^k\|.
\end{aligned} \tag{4.11}$$

Now, let z be a weak limit of some subsequence of $(z^k)_{k\in\mathbb{N}}$ called similarly for simplicity. We have that $(\|z^* - z^k\|)_{k\in\mathbb{N}}$ is bounded and, since (4.8) implies $\|z^k - \mathcal{T}_{P_k}z^k\|_{U_k}^2 \to 0$ we deduce from (4.11) that $u^k := z^k - J_{P_k^{-1}A}(z^k - P_k^{-1}(B_1+B_2)z^k) \to 0$. Hence, since, for every $x\in\mathcal{H}$,

$$\|S_kx\| \le \|(S_k-B_2)x - (S_k-B_2)0\| + \|B_2x - B_20\| \le (K_k+L)\|x\| \le (M+L)\,, \tag{4.12}$$

we have $\|P_ku_k\| = \|(U_k+S_k)u_k\| \le \|U_ku_k\| + \|S_ku_k\| \le (2M+L)\|u_k\| \to 0$. Moreover, since B_1 and B_2 are continuous with the strong topology, we have that $v^k := B_1z^k - B_1(z^k-u^k) \to 0$ and $w^k := B_1z^k - B_1(z^k-u^k) \to 0$ which yields

$$\begin{aligned}
z^k - u^k = J_{P_k^{-1}A}(z^k - P_k^{-1}(B_1+B_2)z^k) &\Leftrightarrow\ u_k - P_k^{-1}(B_1+B_2)z^k \in P_k^{-1}A(z^k-u^k)\\
&\Leftrightarrow\ P_ku_k - B_1z^k - B_2z^k \in A(z^k-u^k)\\
&\Leftrightarrow\ P_ku_k - v^k - w^k \in (A+B_1+B_2)(z^k-u^k).
\end{aligned}$$

Therefore, since the left hand side of the last equation converges strongly to 0 and $z^k - u^k \rightharpoonup z$, we conclude from the weak-strong closedness of the maximally monotone operator $A+B_1+B_2$ that $z\in\mathrm{zer}(A+B_1+B_2)$ and the result follows. □

Remark 4.

1. *Note that, in the particular case when* $S_k \equiv 0$ *and* $P_k = U_k = \gamma_k^{-1} V_k^{-1}$, *we have from [?, Lemma 2.1] that* $\rho_k = \gamma_k^{-1}\|V_k^{-1}\|$, *the conditions on the constants involved in Theorem 4.2 reduce to*

$$\frac{\|V_k^{-1}\|}{M} \leq \gamma_k \leq \frac{\|V_k^{-1}\|}{\rho}, \quad L^2 \leq \frac{\gamma_k^{-1}\|V_k^{-1}\|}{1+\varepsilon}\left(\frac{\gamma_k^{-1}\|V_k^{-1}\|}{1+\varepsilon} - \frac{1}{2\beta}\right), \tag{4.13}$$

for some $0 < \rho < M$, *for every* $k \in \mathbb{N}$, *and* (4.4) *reduces to*

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} x^k = J_{\gamma_k V_k A}(z^k - \gamma_k V_k (B_1 + B_2) z^k) \\ z^{k+1} = z^k + \frac{\lambda_k}{\gamma_k}\left(V_k^{-1}(x^k - z^k) + \gamma_k B_2 z^k - \gamma_k B_2 x^k\right). \end{array} \right. \tag{4.14}$$

If in addition we assume that $B_2 = 0$ *and, hence* $L = 0$, (4.13) *reduces to* $\gamma_k \leq \|V_k^{-1}\| 2\beta/(1+\varepsilon)$ *which is more general than the condition in [?] and, moreover, we do not need any additional hypothesis on the sequence of metrics* $(V_k)_{k\in\mathbb{N}}$ *for achieving convergence. Similarly, if* $B_1 = 0$, *and hence, we can take* $\beta \to \infty$, (4.13) *reduces to* $\gamma_k \leq \|V_k^{-1}\| L^{-1}$ *which is more general than the condition in [?] and no additional assumption on* $(V_k)_{k\in\mathbb{N}}$ *is needed. However,* (4.14) *involves an additional computation of* V_k^{-1} *in the last step of each iteration* $k \in \mathbb{N}$.

2. *In the particular case when, for every* $k \in \mathbb{N}$, $P_k = U_k = \operatorname{Id}/\gamma_k$, *where* $(\gamma_k)_{k\in\mathbb{N}}$ *is a real sequence, we have* $S_k \equiv 0$, $K_k \equiv L$, $\|U_k\| = \rho_k = 1/\gamma_k$, *and conditions* $\sup_{k\in\mathbb{N}} \|U_k\| < \infty$ *and* (4.3) *reduce to*

$$0 < \inf_{k\in\mathbb{N}} \gamma_k \leq \sup_{k\in\mathbb{N}} \gamma_k < \chi, \tag{4.15}$$

where χ *is defined in* (2.2) *and* (4.4) *reduces to*

$$(\forall k \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} x^k = J_{\gamma_k A}(z^k - \gamma_k (B_1 + B_2) z^k) \\ z^{k+1} = z^k + \eta_k\left(x^k + \gamma_k B_2 z^k - \gamma_k B_2 x^k - z^k\right), \end{array} \right.$$

where $\eta_k \in [\varepsilon, 1-\varepsilon]$, *which is a relaxed version of Theorem 2.2.*

3. *As in Remark 2, by setting* $B_1 = 0$ *or* $B_2 = 0$, *we can derive from* (4.4) *versions of Tseng's splitting and forward-backward algorithm with non self-adjoint linear operators but without needing the inversion of* U. *In particular, the proximal point algorithm in* (3.7) *reduces to*

$$z^0 \in \mathcal{H}, \quad (\forall k \in \mathbb{N}) \quad z^{k+1} = z^k + \lambda P(J_{P^{-1}A} z^k - z^k) \tag{4.16}$$

for $\lambda < \|U\|^{-1}$ *and, in the case of* (3.8), *to avoid inversion is to come back to the gradient-type method with the standard metric.*

5. Primal-dual composite monotone inclusions with non self-adjoint linear operators. In this section, we apply of our algorithm to composite primal-dual monotone inclusions involving a cocoercive and a lipschitzian monotone operator.

Problem 2. *Let* H *be a real Hilbert space, let* $\mathrm{X} \subset \mathrm{H}$ *be closed and convex, let* $z \in \mathrm{H}$, *let* $\mathrm{A}\colon \mathrm{H} \to 2^{\mathrm{H}}$ *be maximally monotone, let* $\mathrm{C}_1\colon \mathrm{H} \to \mathrm{H}$ *be* μ*-cocoercive, for some* $\mu \in \left]0, +\infty\right[$, *and let* $\mathrm{C}_2\colon \mathrm{H} \to \mathrm{H}$ *be a monotone and* δ*-lipschitzian operator, for* *some* $\delta \in]0,+\infty[$. *Let* $m \geq 1$ *be an integer, and, for every* $i \in \{1,\dots,m\}$, *let* G_i *be a real Hilbert space, let* $r_i \in \mathrm{G}_i$, *let* $\mathrm{B}_i \colon \mathrm{G}_i \to 2^{\mathrm{G}_i}$ *be maximally monotone, let* $\mathrm{D}_i \colon \mathrm{G}_i \to 2^{\mathrm{G}_i}$ *be maximally monotone and* ν_i*-strongly monotone, for some* $\nu_i \in]0,+\infty[$, *and suppose that* $\mathrm{L}_i \colon \mathrm{H} \to \mathrm{G}_i$ *is a nonzero linear bounded operator. The problem is to solve the primal inclusion.*

$$\textit{find} \quad \mathrm{x} \in \mathrm{X} \quad \textit{such that} \quad \mathrm{z} \in \mathrm{Ax} + \sum_{i=1}^{m} \mathrm{L}_i^*(\mathrm{B}_i \,\square\, \mathrm{D}_i)(\mathrm{L}_i\mathrm{x} - \mathrm{r}_i) + \mathrm{C}_1\mathrm{x} + \mathrm{C}_2\mathrm{x} \tag{5.1}$$

together with the dual inclusion

$$\begin{aligned} &\textit{find} \quad \mathrm{v}_1 \in \mathrm{G}_1, \dots, \mathrm{v}_m \in \mathrm{G}_m \\ &\textit{such that} \quad (\exists \mathrm{x} \in \mathrm{X}) \begin{cases} \mathrm{z} - \sum_{i=1}^{m} \mathrm{L}_i^* \mathrm{v}_i \in \mathrm{Ax} + \mathrm{C}_1\mathrm{x} + \mathrm{C}_2\mathrm{x} \\ (\forall i \in \{1,\dots,m\}) \ \mathrm{v}_i \in (\mathrm{B}_i \,\square\, \mathrm{D}_i)(\mathrm{L}_i\mathrm{x} - \mathrm{r}_i) \end{cases} \end{aligned} \tag{5.2}$$

under the assumption that a solution exists.

In the case when $\mathrm{X} = \mathrm{H}$ and $\mathrm{C}_2 = 0$, Problem 2 is studied in [**?**][4] and models a large class of problems including optimization problems, variational inequalities, equilibrium problems, among others (see [**?**, **?**, **?**] and the references therein). In [**?**] the author rewrite (5.1) and (5.2) in the case $\mathrm{X} = \mathrm{H}$ as

$$\text{find} \quad z \in \mathcal{H} \quad \text{such that} \quad 0 \in Mz + Sz + Qz, \tag{5.3}$$

where $\mathcal{H} = \mathrm{H} \times \mathrm{G}_1 \times \cdots \times \mathrm{G}_m$, $M \colon \mathcal{H} \to 2^{\mathcal{H}} \colon (\mathrm{x}, \mathrm{v}_1, \dots, \mathrm{v}_m) \mapsto (\mathrm{Ax} - \mathrm{z}) \times (\mathrm{B}_1^{-1}\mathrm{v}_1 + \mathrm{r}_1) \times \cdots \times (\mathrm{B}_m^{-1}\mathrm{v}_m + \mathrm{r}_m)$ is maximally monotone, $S \colon \mathcal{H} \to \mathcal{H} \colon (\mathrm{x}, \mathrm{v}_1, \dots, \mathrm{v}_m) \mapsto (\sum_{i=1}^{m} \mathrm{L}_i^*\mathrm{v}_i, -\mathrm{L}_1\mathrm{x}, \dots, -\mathrm{L}_m\mathrm{x})$ is skew linear, and $Q \colon \mathcal{H} \to \mathcal{H} \colon (\mathrm{x}, \mathrm{v}_1, \dots, \mathrm{v}_m) \mapsto (\mathrm{C}_1\mathrm{x}, \mathrm{D}_1^{-1}\mathrm{v}_1, \dots, \mathrm{D}_m^{-1}\mathrm{v}_m)$ is cocoercive. If $(\mathrm{x}, \mathrm{v}_1, \dots, \mathrm{v}_m)$ is a solution in the primal-dual space $\mathcal{H}$ to (5.3), then x is a solution to (5.1) and $(\mathrm{v}_1, \dots, \mathrm{v}_m)$ is a solution to (5.2). The author provide an algorithm for solving (5.1)–(5.2) in this particular instance, which is an application of the *forward-backward* splitting (FBS) applied to the inclusion

$$\text{find} \quad z \in \mathcal{H} \quad \text{such that} \quad 0 \in V^{-1}(M+S)z + V^{-1}Qz, \tag{5.4}$$

where V is a specific symmetric strongly monotone operator. Under the metric $\langle V\cdot \mid \cdot \rangle$, $V^{-1}(M+S)$ is maximally monotone and $V^{-1}Q$ is cocoercive and, therefore, the FBS converges weakly to a primal-dual solution.

In order to tackle the case $\mathrm{C}_2 \neq 0$, we propose to use the method in Theorem 4.2 for solving $0 \in Ax + B_1x + B_2x$ where $A = M$, $B_1 = Q$, $B_2 = S + C_2$, and $C_2 \colon (\mathrm{x}, \mathrm{v}_1, \dots, \mathrm{v}_m) \mapsto (\mathrm{C}_2\mathrm{x}, 0, \dots, 0)$ allowing, in that way, non self-adjoint linear operators which may vary among iterations. The following result provides the method thus obtained, where the dependence of the non self-adjoint linear operators with respect to iterations has been avoided for simplicity.

Theorem 5.1. *In Problem 2, set* $\mathrm{X} = \mathrm{H}$, *set* $\mathrm{G}_0 = \mathrm{H}$, *for every* $i \in \{0,1,\dots,m\}$ *and* $j \in \{0,\dots,i\}$, *let* $\mathrm{P}_{ij} \colon \mathrm{G}_j \to \mathrm{G}_i$ *be a linear operator satisfying*

$$(\forall \mathrm{x}_i \in \mathrm{G}_i) \quad \langle \mathrm{P}_{ii}\mathrm{x}_i \mid \mathrm{x}_i \rangle \geq \varrho_i \|\mathrm{x}_i\|^2 \tag{5.5}$$

[4]Note that in [?], weights $(\omega_i)_{1\leq i\leq m}$ multiplying operators $(\mathrm{B}_i \,\square\, \mathrm{D}_i)_{1\leq i\leq m}$ are considered. They can be retrieved in (5.1) by considering $(\omega_i\mathrm{B}_i)_{1\leq i\leq m}$ and $(\omega_i\mathrm{D}_i)_{1\leq i\leq m}$ instead of $(\mathrm{B}_i)_{1\leq i\leq m}$ and $(\mathrm{D}_i)_{1\leq i\leq m}$. Then both formulations are equivalent.

for some $\varrho_i > 0$*. Define the* $(m+1)\times(m+1)$ *symmetric real matrices* Υ*,* Σ*, and* Δ *by*

$$(\forall i \in \{0,\ldots,m\})(\forall j < i) \quad \Upsilon_{ij} = \begin{cases} 0, & \text{if } i = j; \\ \|\mathrm{P}_{ij}\|/2, & \text{if } i > j, \end{cases}$$

$$\Sigma_{ij} = \begin{cases} \|\mathrm{P}_{ii} - \mathrm{P}_{ii}^*\|/2, & \text{if } i = j; \\ \|\mathrm{L}_i + \mathrm{P}_{i0}/2\|, & \text{if } i \geq 1; j = 0; \\ \|\mathrm{P}_{ij}\|/2, & \text{if } i > j > 0, \end{cases} \tag{5.6}$$

and $\Delta = \mathrm{Diag}(\varrho_0,\ldots,\varrho_m)$*. Assume that* $\Delta - \Upsilon$ *is positive definite with smallest eigenvalue* $\rho > 0$ *and that*

$$(\|\Sigma\|_2 + \delta)^2 < \rho\left(\rho - \frac{1}{2\beta}\right), \tag{5.7}$$

where $\beta = \min\{\mu,\nu_1,\ldots,\nu_m\}$*. Let* $M = \max_{i=0,\ldots,m}\|\mathrm{P}_{ii}\| + \|\Upsilon\|_2$*, let* $\lambda \in]0, M^{-1}[$*, let* $(\mathrm{x}^0,\mathrm{u}_1^0,\ldots,\mathrm{u}_m^0) \in \mathrm{H}\times\mathrm{G}_1\times\cdots\times\mathrm{G}_m$*, and let* $\{\mathrm{x}^k\}_{k\in\mathbb{N}}$ *and* $\{\mathrm{u}_i^k\}_{k\in\mathbb{N},1\leq i\leq m}$ *the sequences generated by the following routine: for every* $k \in \mathbb{N}$

$$\left\lfloor\begin{array}{l}
\mathrm{y}^k = J_{\mathrm{P}_{00}^{-1}\mathrm{A}}\left(\mathrm{x}^k - \mathrm{P}_{00}^{-1}\left(\mathrm{C}_1\mathrm{x}^k + \mathrm{C}_2\mathrm{x}^k + \sum_{i=1}^m \mathrm{L}_i^*\mathrm{u}_i^k\right)\right)\\
\mathrm{v}_1^k = J_{\mathrm{P}_{11}^{-1}\mathrm{B}_1^{-1}}\left(\mathrm{u}_1^k - \mathrm{P}_{11}^{-1}\left(\mathrm{D}_1^{-1}\mathrm{u}_1^k - \mathrm{L}_1\mathrm{x}^k - \mathrm{P}_{10}(\mathrm{x}^k - \mathrm{y}^k)\right)\right)\\
\mathrm{v}_2^k = J_{\mathrm{P}_{22}^{-1}\mathrm{B}_2^{-1}}\left(\mathrm{u}_2^k - \mathrm{P}_{22}^{-1}\left(\mathrm{D}_2^{-1}\mathrm{u}_2^k - \mathrm{L}_2\mathrm{x}^k - \mathrm{P}_{20}(\mathrm{x}^k - \mathrm{y}^k) - \mathrm{P}_{21}(\mathrm{u}_1^k - \mathrm{v}_1^k)\right)\right)\\
\vdots\\
\mathrm{v}_m^k = J_{\mathrm{P}_{mm}^{-1}\mathrm{B}_m^{-1}}\left(\mathrm{u}_m^k - \mathrm{P}_{mm}^{-1}\left(\mathrm{D}_m^{-1}\mathrm{u}_m^k - \mathrm{L}_m\mathrm{x}^k - \mathrm{P}_{m0}(\mathrm{x}^k - \mathrm{y}^k) - \sum_{j=1}^{m-1}\mathrm{P}_{mj}(\mathrm{u}_j^k - \mathrm{v}_j^k)\right)\right)\\
\mathrm{x}^{k+1} = \mathrm{x}^k + \lambda\left(\mathrm{P}_{00}(\mathrm{y}^k - \mathrm{x}^k) + \left(\mathrm{C}_2\mathrm{x}^k - \mathrm{C}_2\mathrm{y}^k + \sum_{i=1}^m \mathrm{L}_i^*(\mathrm{u}_i^k - \mathrm{v}_i^k)\right)\right)\\
\mathrm{u}_1^{k+1} = \mathrm{u}_1^k + \lambda\left(\mathrm{P}_{10}(\mathrm{y}^k - \mathrm{x}^k) + \mathrm{P}_{11}(\mathrm{v}_1^k - \mathrm{u}_1^k) - \mathrm{L}_1(\mathrm{x}^k - \mathrm{y}^k)\right)\\
\vdots\\
\mathrm{u}_m^{k+1} = \mathrm{u}_m^k + \lambda\left(\mathrm{P}_{m0}(\mathrm{y}^k - \mathrm{x}^k) + \sum_{j=1}^m \mathrm{P}_{mj}(\mathrm{v}_j^k - \mathrm{u}_j^k) - \mathrm{L}_m(\mathrm{x}^k - \mathrm{y}^k)\right).
\end{array}\right. \tag{5.8}$$

Then there exists a primal-dual solution $(\mathrm{x}^*,\mathrm{u}_1^*,\ldots,\mathrm{u}_m^*) \in \mathrm{H}\times\mathrm{G}_1\times\cdots\times\mathrm{G}_m$ *to Problem 2 such that* $\mathrm{x}^k \rightharpoonup \mathrm{x}^*$ *and, for every* $i \in \{1,\ldots,m\}$*,* $\mathrm{u}_i^k \rightharpoonup \mathrm{u}_i^*$*.*

Proof. Consider the real Hilbert space $\mathcal{H} = \mathrm{H}\oplus\mathrm{G}_1\oplus\cdots\oplus\mathrm{G}_m$, where its scalar product and norm are denoted by $\langle\langle\cdot\mid\cdot\rangle\rangle$ and $|||\cdot|||$, respectively, and $x = (\mathrm{x}_0,\mathrm{x}_1,\ldots,\mathrm{x}_m)$ and $y = (\mathrm{y}_0,\mathrm{y}_1,\ldots,\mathrm{y}_m)$ denote generic elements of $\mathcal{H}$. Similarly as in [?], note that the set of primal-dual solutions $x^* = (\mathrm{x}^*,\mathrm{u}_1^*,\ldots,\mathrm{u}_m^*) \in \mathcal{H}$ to Problem 2 in the case $\mathrm{X} = \mathrm{H}$ coincides with the set of solutions to the monotone inclusion

$$\text{find} \quad x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + B_1x + B_2x, \tag{5.9}$$

where the operators $A\colon \mathcal{H} \to 2^{\mathcal{H}}$, $B_1\colon \mathcal{H}\to\mathcal{H}$, and $B_2\colon\mathcal{H}\to\mathcal{H}$ defined by

$$\begin{cases} A & : (\mathrm{x},\mathrm{v}_1,\ldots,\mathrm{v}_m) \mapsto (\mathrm{A}\mathrm{x} - \mathrm{z}) \times (\mathrm{B}_1^{-1}\mathrm{v}_1 + \mathrm{r}_1) \times\cdots\times (\mathrm{B}_m^{-1}\mathrm{v}_m + \mathrm{r}_m) \\ B_1 & : (\mathrm{x},\mathrm{v}_1,\ldots,\mathrm{v}_m) \mapsto (\mathrm{C}_1\mathrm{x}, \mathrm{D}_1^{-1}\mathrm{v}_1,\ldots,\mathrm{D}_m^{-1}\mathrm{v}_m) \\ B_2 & : (\mathrm{x},\mathrm{v}_1,\ldots,\mathrm{v}_m) \mapsto (\mathrm{C}_2\mathrm{x} + \sum_{i=1}^m \mathrm{L}_i^*\mathrm{v}_i, -\mathrm{L}_1\mathrm{x},\ldots,-\mathrm{L}_m\mathrm{x}), \end{cases} \tag{5.10}$$

are maximally monotone, β-cocoercive, and monotone-Lipschitz, respectively (see [**?**, Proposition 20.22 and 20.23] and [**?**, Eq. (3.12)]).

Now let $P\colon \mathcal{H} \to \mathcal{H}$ defined by

$$P\colon x \mapsto \left(\mathrm{P}_{00}\mathrm{x}_0, \mathrm{P}_{10}\mathrm{x}_0 + \mathrm{P}_{11}\mathrm{x}_1, \ldots, \sum_{j=0}^{m} \mathrm{P}_{mj}\mathrm{x}_j\right) = \left(\sum_{j=0}^{i} \mathrm{P}_{ij}\mathrm{x}_j\right)_{i=0}^{m}. \tag{5.11}$$

Then $P^*\colon x \mapsto (\sum_{j=i}^{m} \mathrm{P}_{ji}^*\mathrm{x}_j)_{i=0}^{m}$ and $U\colon \mathcal{H} \to \mathcal{H}$ and $S\colon \mathcal{H} \to \mathcal{H}$ defined by

$$U\colon x \mapsto \left(\frac{1}{2}\sum_{j=0}^{i-1} \mathrm{P}_{ij}\mathrm{x}_j + \left(\frac{\mathrm{P}_{ii} + \mathrm{P}_{ii}^*}{2}\right)\mathrm{x}_i + \frac{1}{2}\sum_{j=i+1}^{m} \mathrm{P}_{ji}^*\mathrm{x}_j\right)_{i=0}^{m} \tag{5.12}$$

$$S\colon x \mapsto \left(\frac{1}{2}\sum_{j=0}^{i-1} \mathrm{P}_{ij}\mathrm{x}_j + \left(\frac{\mathrm{P}_{ii} - \mathrm{P}_{ii}^*}{2}\right)\mathrm{x}_i - \frac{1}{2}\sum_{j=i+1}^{m} \mathrm{P}_{ji}^*\mathrm{x}_j\right)_{i=0}^{m} \tag{5.13}$$

are the self-adjoint and skew components of P, respectively, satisfying $P = U + S$. Moreover, for every $x = (\mathrm{x}_0, \mathrm{x}_1, \ldots, \mathrm{x}_m)$ in $\mathcal{H}$, we have

$$\begin{aligned}
\langle\langle Ux \mid x\rangle\rangle &= \sum_{i=0}^{m} \frac{1}{2}\sum_{j=0}^{i-1} \langle \mathrm{P}_{ij}\mathrm{x}_j \mid \mathrm{x}_i\rangle + \langle \mathrm{P}_{ii}\mathrm{x}_i \mid \mathrm{x}_i\rangle + \frac{1}{2}\sum_{j=i+1}^{m} \left\langle \mathrm{P}_{ji}^*\mathrm{x}_j \mid \mathrm{x}_i\right\rangle \\
&= \sum_{i=0}^{m} \langle \mathrm{P}_{ii}\mathrm{x}_i \mid \mathrm{x}_i\rangle + \sum_{i=1}^{m}\sum_{j=0}^{i-1} \langle \mathrm{P}_{ij}\mathrm{x}_j \mid \mathrm{x}_i\rangle \\
&\geq \sum_{i=0}^{m} \varrho_i\|\mathrm{x}_i\|^2 - \sum_{i=1}^{m}\sum_{j=0}^{i-1} \|\mathrm{P}_{ij}\|\,\|\mathrm{x}_i\|\,\|\mathrm{x}_j\| \\
&= \xi\cdot(\Delta - \Upsilon)\xi \geq \rho|\xi|^2 = \rho\,|||x|||^2,
\end{aligned} \tag{5.14}$$

where $\xi := (\|\mathrm{x}_i\|)_{i=0}^{m} \in \mathbb{R}^{m+1}$, Υ is defined in (5.6), and ρ is the smallest (strictly positive) eigenvalue of $\Delta - \Upsilon$. In addition, we can write $B_2 - S = C_2 + R$, where $C_2\colon x \mapsto (\mathrm{C}_2\mathrm{x}, 0, \ldots, 0)$ is monotone and δ-lipschitzian, and R is a skew linear operator satisfying, for every $x = (\mathrm{x}_0, \mathrm{x}_1, \ldots, \mathrm{x}_m) \in \mathcal{H}$, $Rx = (\sum_{j=0}^{m} R_{i,j}\mathrm{x}_j)_{0\leq i\leq m}$, where the operators $R_{i,j}\colon \mathrm{G}_j \to \mathrm{G}_i$ are defined by $R_{i,j} = -\mathrm{P}_{ij}/2$ if $i > j > 0$, $R_{i,j} = -(\mathrm{L}_i + \mathrm{P}_{i0})/2$ if $i > j = 0$, $R_{i,i} = (\mathrm{P}_{ii}^* - \mathrm{P}_{ii})/2$ and the other components follow from the skew property of R. Therefore,

$$|||Rx|||^2 = \sum_{i=0}^{m}\left\|\sum_{j=0}^{m} R_{i,j}\mathrm{x}_j\right\|^2 \leq \sum_{i=0}^{m}\left(\sum_{j=0}^{m} \|R_{i,j}\|\,\|\mathrm{x}_j\|\right)^2 = |\Sigma\xi|^2 \leq \|\Sigma\|_2^2|\xi|^2 = \|\Sigma\|_2^2|||x|||^2, \tag{5.15}$$

from which we obtain that $B_2 - S$ is $(\delta + \|\Sigma\|_2)$-lipschitzian. Altogether, by noting that, for every $x \in \mathcal{H}$, $\|Ux\| \leq M$, all the hypotheses of Theorem 4.2 hold in this instance and by developing (4.4) for this specific choices of A, B_1, B_2, P, γ, and setting, for every $k \in \mathbb{N}$, $z^k = (\mathrm{x}^k, \mathrm{u}_1^k, \ldots, \mathrm{u}_m^k)$ and $x^k = (\mathrm{y}^k, \mathrm{v}_1^k, \ldots, \mathrm{v}_m^k)$, we obtain (5.8) after straighforward computations and using

$$x^k = J_{\gamma P^{-1}A}(z^k - \gamma P^{-1}(B_1 z^k + B_2 z^k)) \quad\Leftrightarrow\quad P(z^k - x^k) - \gamma(B_1 z^k + B_2 z^k) \in \gamma A x^k. \tag{5.16}$$

The result follows, hence, as a consequence of Theorem 4.2. ☐

REMARK 5.

1. *As in Theorem 4.2, the algorithm in Theorem 5.1 allows for linear operators* $(\mathrm{P}_{ij})_{0\le i,j\le m}$ *depending on the iteration, whenever* (4.3) *holds for the corresponding operators defined in* (5.11)–(5.13)*. We omit this generalization in Theorem 5.1 for the sake of simplicity.*
2. *In the particular case when, for every* $i\in\{1,\dots,m\}$*,* $B_i=\widetilde{B}_i\,\square\,M_i$*, where* M_i *is such that* M_i^{-1} *is monotone and* σ_i*-Lipschitz, for some* $\sigma_i>0$*, Problem* (2) *can be solved in a similar way if, instead of* B_2 *and* δ*, we consider* $\widetilde{B}_2\colon(\mathrm{x},\mathrm{v}_1,\dots,\mathrm{v}_m)\mapsto(\mathrm{C}_2\mathrm{x}+\sum_{i=1}^m\mathrm{L}_i^*\mathrm{v}_i,\mathrm{M}_1^{-1}\mathrm{v}_1-\mathrm{L}_1\mathrm{x},\dots,\mathrm{M}_m^{-1}\mathrm{v}_m-\mathrm{L}_m\mathrm{x})$ *and* $\widetilde{\delta}=\max\{\delta,\sigma_1,\dots,\sigma_m\}$*. Again, for the sake of simplicity, this extension has not been considered in Problem 2.*
3. *If the inversion of the matrix* U *is not difficult or no variable metric is used and the projection onto* $\mathrm{X}\subset\mathrm{H}$ *is computable, we can also use Theorem 3.1 for solving Problem 2 in the general case* $\mathrm{X}\subset\mathrm{H}$.

COROLLARY 5.2. *In Problem 2, let* $\theta\in[-1,1]$*, let* $\sigma_0,\dots,\sigma_m$ *be strictly positive real numbers and let* Ω *the* $(m+1)\times(m+1)$ *symmetric real matrix given by*

$$(\forall i,j\in\{0,\dots,m\})\quad \Omega_{ij}=\begin{cases}\frac{1}{\sigma_i}, & \text{if } i=j;\\ -(\frac{1+\theta}{2})\|\mathrm{L}_i\|, & \text{if } 0=j<i;\\ 0, & \text{if } 0<j<i.\end{cases}\tag{5.17}$$

Assume that Ω *is positive definite with* $\rho>0$ *its smallest eigenvalue and that*

$$\left(\delta+\left(\frac{1-\theta}{2}\right)\sqrt{\sum_{i=1}^m\|\mathrm{L}_i\|^2}\right)^2<\rho\left(\rho-\frac{1}{2\beta}\right),\tag{5.18}$$

where $\beta=\min\{\mu,\nu_1,\dots,\nu_m\}$*. Let* $M=(\min\{\sigma_0,\dots,\sigma_m\})^{-1}+(\frac{1+\theta}{2})\sqrt{\sum_{i=1}^m\|\mathrm{L}_i\|^2}$*, let* $\lambda\in]0,M^{-1}[$*, let* $(\mathrm{x}^0,\mathrm{u}_1^0,\dots,\mathrm{u}_m^0)\in\mathrm{H}\times\mathrm{G}_1\times\cdots\times\mathrm{G}_m$*, and let* $\{\mathrm{x}^k\}_{k\in\mathbb{N}}$ *and* $\{\mathrm{u}_i^k\}_{k\in\mathbb{N},1\le i\le m}$ *the sequences generated by the following routine:*

$$(\forall k\in\mathbb{N})\quad\left\lfloor\begin{array}{l}\mathrm{y}^k=J_{\sigma_0\mathrm{A}}\left(\mathrm{x}^k-\sigma_0\left(\mathrm{C}_1\mathrm{x}^k+\mathrm{C}_2\mathrm{x}^k+\sum_{i=1}^m\mathrm{L}_i^*\mathrm{u}_i^k\right)\right)\\ \text{For every } i=1,\dots,m\\ \left\lfloor\mathrm{v}_i^k=J_{\sigma_i\mathrm{B}_i^{-1}}\left(\mathrm{u}_i^k-\sigma_i\left(\mathrm{D}_i^{-1}\mathrm{u}_i^k-\mathrm{L}_i(\mathrm{y}^k+\theta(\mathrm{y}^k-\mathrm{x}^k))\right)\right)\right.\\ \mathrm{x}^{k+1}=\mathrm{x}^k+\frac{\lambda}{\sigma_0}\left(\mathrm{y}^k-\mathrm{x}^k+\sigma_0\left(\mathrm{C}_2\mathrm{x}^k-\mathrm{C}_2\mathrm{y}^k+\sum_{i=1}^m\mathrm{L}_i^*(\mathrm{u}_i^k-\mathrm{v}_i^k)\right)\right)\\ \text{For every } i=1,\dots,m\\ \left\lfloor\mathrm{u}_i^{k+1}=\mathrm{u}_i^k+\frac{\lambda}{\sigma_i}\left(\mathrm{v}_i^k-\mathrm{u}_i^k-\sigma_i\theta\mathrm{L}_i(\mathrm{y}^k-\mathrm{x}^k)\right),\right.\end{array}\right.\tag{5.19}$$

Then there exists a primal-dual solution $(\mathrm{x}^*,\mathrm{u}_1^*,\dots,\mathrm{u}_m^*)\in\mathrm{H}\times\mathrm{G}_1\times\cdots\times\mathrm{G}_m$ *to Problem 2 such that* $\mathrm{x}^k\rightharpoonup\mathrm{x}^*$ *and, for every* $i\in\{1,\dots,m\}$*,* $\mathrm{u}_i^k\rightharpoonup\mathrm{u}_i^*$.

Proof. This result is a consequence of Theorem 5.1 when, for every $i\in\{0,\dots,m\}$, $\mathrm{P}_{ii}=\mathrm{Id}/\sigma_i$, $\mathrm{P}_{i0}=-(1+\theta)\mathrm{L}_i$, and, for every $0<j<i$, $\mathrm{P}_{ij}=0$. Indeed, we have from (5.5) that $\varrho_i=1/\sigma_i$, and from (5.6) we deduce that, for every $x=(\xi_i)_{0\le i\le m}\in\mathbb{R}^{m+1}$,

$$\|\Sigma x\|^2=\left(\frac{1-\theta}{2}\right)^2\left[\left(\sum_{i=0}^m\|\mathrm{L}_i\|\xi_i\right)^2+\xi_0^2\sum_{i=1}^m\|\mathrm{L}_i\|^2\right]\le\left(\frac{1-\theta}{2}\right)^2\left(\sum_{i=0}^m\|\mathrm{L}_i\|^2\right)\|x\|^2,\tag{5.20}$$

from which we obtain $\|\Sigma\|_2 \le (\frac{1-\theta}{2})\sqrt{\sum_{i=1}^m \|\mathrm{L}_i\|^2}$. Actually, we have the equality by choosing $\bar{x} = (\bar{\xi}_i)_{0\le i\le m}$ defined by $\bar{\xi}_i = \|\mathrm{L}_i\|/\sqrt{\sum_{j=1}^m \|\mathrm{L}_j\|^2}$ for every $i \in \{1,\dots,m\}$ and $\bar{\xi}_0 = 0$, which satisfies $\|\bar{x}\| = 1$ and $\|\Sigma\bar{x}\| = (\frac{1-\theta}{2})\sqrt{\sum_{i=1}^m \|\mathrm{L}_i\|^2}$. Therefore, condition (5.7) reduces to (5.18). On the other hand, from (5.6) we deduce that $\Omega = \Delta - \Upsilon$ and $\Upsilon = (\frac{1+\theta}{1-\theta})\Sigma$, which yields $\|\Upsilon\|_2 = (\frac{1+\theta}{2})\sqrt{\sum_{i=1}^m \|\mathrm{L}_i\|^2}$ and $\max_{i=0,\dots,m}\|\mathrm{P}_{ii}\| = (\min\{\sigma_0,\dots,\sigma_m\})^{-1}$. Altogether, since (5.19) is exactly (5.8) for this choice of matrices $(\mathrm{P}_{i,j})_{0\le i,j,\le m}$, the result is a consequence of Theorem 5.1. ☐

Remark 6.

1. *Note that, the condition $\rho > 0$ where ρ is the smallest eigenvalue of Ω defined in* (5.17)*, is guaranteed if $\sigma_0(\frac{1+\theta}{2})^2\sum_{i=1}^m \sigma_i\|\mathrm{L}_i\|^2 < 1$. Indeed, by repeating the procedure in [?, (3.20)] in finite dimension we obtain, for every $x = (\xi_i)_{0\le i\le m} \in \mathbb{R}^{m+1}$,*

$$\begin{aligned}
x\cdot\Omega x &= \sum_{i=0}^m \frac{\xi_i^2}{\sigma_i} - \sum_{i=1}^m 2\left(\frac{1+\theta}{2}\right)\xi_0\|\mathrm{L}_i\|\xi_i \\
&= \sum_{i=0}^m \frac{\xi_i^2}{\sigma_i} - \left(\frac{1+\theta}{2}\right)\sum_{i=1}^m 2\frac{\sqrt{\sigma_i}\|\mathrm{L}_i\|\xi_0}{(\sigma_0\sum_{j=1}^m \sigma_j\|\mathrm{L}_j\|^2)^{1/4}}\frac{(\sigma_0\sum_{j=1}^m \sigma_j\|\mathrm{L}_j\|^2)^{1/4}\xi_i}{\sqrt{\sigma_i}} \qquad (5.21)\\
&\ge \sum_{i=0}^m \frac{\xi_i^2}{\sigma_i} - \left(\frac{1+\theta}{2}\right)\left(\frac{\xi_0^2}{\sqrt{\sigma_0}}\sqrt{\sum_{j=1}^m \sigma_j\|\mathrm{L}_j\|^2} + \sqrt{\sigma_0\sum_{j=1}^m \sigma_j\|\mathrm{L}_j\|^2}\sum_{j=1}^m \frac{\xi_j^2}{\sigma_j}\right)\\
&= \left(1 - \left(\frac{1+\theta}{2}\right)\sqrt{\sigma_0\sum_{j=1}^m \sigma_j\|\mathrm{L}_j\|^2}\right)\sum_{i=0}^m \frac{\xi_i^2}{\sigma_i}\\
&\ge \rho_v\|x\|^2 \qquad (5.22)
\end{aligned}$$

with

$$\rho_v = \max\{\sigma_0,\dots,\sigma_m\}^{-1}\left(1 - \left(\frac{1+\theta}{2}\right)\sqrt{\sigma_0\sum_{j=1}^m \sigma_j\|\mathrm{L}_j\|^2}\right). \qquad (5.23)$$

Note that ρ_v coincides with the constant obtained in [?] in the case $\theta = 1$ and we have $\rho \ge \rho_v$. Moreover, $\sigma_0(\frac{1+\theta}{2})^2\sum_{i=1}^m \sigma_i\|\mathrm{L}_i\|^2 < 1$ is also necessary for obtaining $\rho > 0$, since in (5.21) *we can choose a particular vector x for obtaining the equality. Of course, this choice does not guarantee to also have equality in the last inequality in* (5.22) *and, hence, $\rho \ge \rho_v$ in general.*

2. *If we set $\theta = 1$ and $\mathrm{C}_2 = 0$ and, hence, $\delta = 0$,* (5.18) *reduces to $2\beta\rho > 1$ and we obtain from* (5.19) *a variant of [?, Theorem 3.1] including an extra forward step involving only the operators $(\mathrm{L}_i)_{1\le i\le m}$. However, our condition is less restrictive, since $\rho \ge \rho_v$, where ρ_v is defined in* (5.23) *and it is obtained in [?] as we have seen in the last remark. Actually, in the particular case when $m = 1$, $\mathrm{L}_1 = \alpha\,\mathrm{Id}$, $\sigma_0 = \eta^2\sigma_1 =: \eta\sigma$ for some $0 < \eta < 1$, constants ρ_v and ρ reduce to*

$$\rho_v(\eta) = \frac{1-\eta\sigma\alpha}{\sigma} \quad \textit{and} \quad \rho(\eta) = \frac{1}{2\sigma}\left(\frac{\eta^2+1}{\eta^2} - \sqrt{\left(\frac{\eta^2-1}{\eta^2}\right)^2 + 4\alpha^2\sigma^2}\right),$$

respectively. By straightforward computations we deduce that $\rho(\eta) > \rho_v(\eta)$ *for every* $0 < \eta < (\alpha\sigma)^{-1}$*, and hence our constant can strictly improve the condition* $2\beta\rho > 1$*, needed in both approaches. Moreover, since Theorem 5.1 allows for non self-adjoint linear operators varying among iterations, we can permit variable stepsizes* $\sigma_0^k, \dots, \sigma_m^k$ *in Theorem 5.1, which could not be used in [?] because of the variable metric framework.*

3. *In the particular case when* $\mathrm{C}_1 = 0$ *and* $\mathrm{C}_2 = 0$ *we can take* $\beta \to +\infty$ *and, hence, condition* (5.18) *reduces to*

$$\left(\frac{1-\theta}{2}\right)\sqrt{\sum_{i=1}^{m}\|\mathrm{L}_i\|^2} < \rho, \tag{5.24}$$

which is stronger than the condition in [?] for the case $m = 1$*, in which it is only needed that* $\rho > 0$ *for achieving convergence. Indeed, in the case* $m = 1$*,* (5.24) *reduces to* $2-2\theta\sigma_0\sigma_1\|\mathrm{L}_1\|^2 > (1-\theta)(\sigma_0+\sigma_1)\|\mathrm{L}_1\|$*, which coincides with the condition in [?] in the case* $\theta = 1$*, but they differ if* $\theta \neq 1$ *because of the extra forward step coming from the Tseng's splitting framework. Actually, in the case* $\theta = 0$ *it reduces to* $\sigma_0+\sigma_1 < 2/\|\mathrm{L}_1\|$ *and in the case* $\theta = -1$ *we obtain the stronger condition* $\max\{\sigma_0, \sigma_1\} < 1/\|\mathrm{L}_1\|$*. Anyway, in our context we can use constants* $\sigma_0^k, \dots, \sigma_m^k$ *varying among iterations and we have a variant of the method in [?] and, in the case when* $\theta = 1$*, of Chambolle-Pock's splitting [?].*

4. *Since* ρ_v *defined in* (5.23) *satisfies* $\rho_v \leq \rho$ *in the case when* $\mathrm{C}_1 = \mathrm{C}_2 = 0$*, a sufficient condition for guaranteeing* (5.24) *is* $(1-\theta)\sqrt{\sum_{i=1}^m \|\mathrm{L}_i\|^2}/2 < \rho_v$*, which implied by the condition*

$$\max\{\sigma_0, \dots, \sigma_m\}\sqrt{\sum_{i=1}^{m}\|\mathrm{L}_i\|^2} < 1. \tag{5.25}$$

5. *Consider the case of composite optimization problems, i.e., when* $\mathrm{A} = \partial \mathrm{f}$*,* $\mathrm{C}_1 = \nabla \mathrm{h}$ *for every* $i = 1, \dots, m$*,* $\mathrm{B}_i = \partial \mathrm{g}_i$ *and* $\mathrm{D}_i = \partial \ell_i$*, where, for every* $i = 1, \dots, m$*,* $\mathrm{f}\colon \mathrm{H} \to \left]-\infty, +\infty\right]$ *and* $\mathrm{g}_i\colon \mathrm{G}_i \to \left]-\infty, +\infty\right]$ *are proper lower semicontinuous and convex functions and* $\mathrm{h}\colon \mathrm{H} \to \mathbb{R}$ *is differentiable, convex, with* β^{-1}*-Lipschitz gradient. In this case, any solution to Problem 2 when* $\mathrm{C}_2 = 0$ *is a solution to the primal-dual optimization problems*

$$\min_{\mathrm{x}\in\mathrm{H}} \mathrm{f}(\mathrm{x}) + \mathrm{h}(\mathrm{x}) + \sum_{i=1}^{m}(\mathrm{g}_i \,\square\, \ell_i)(\mathrm{L}_i\mathrm{x}) \tag{5.26}$$

and

$$\min_{\mathrm{u}_1\in\mathrm{G}_1,\dots,\mathrm{u}_m\in\mathrm{G}_m} (\mathrm{f}^* \,\square\, \mathrm{h}^*)\left(-\sum_{i=1}^{m}\mathrm{L}_i^*\mathrm{u}_i\right) + \sum_{i=1}^{m}\mathrm{g}_i^*(\mathrm{u}_i) + \ell_i^*(\mathrm{u}_i), \tag{5.27}$$

and the equivalence holds under some qualification condition. In this partic-

ular case, (5.19) *reduces to*

$$\left\lfloor\begin{array}{l} \mathrm{y}^k = \mathbf{prox}_{\sigma_0 \mathrm{f}}\left(\mathrm{x}^k - \sigma_0\left(\nabla \mathrm{h}(\mathrm{x}^k) + \sum_{i=1}^m \mathrm{L}_i^* \mathrm{u}_i^k\right)\right) \\ \textit{For every } i = 1, \ldots, m \\ \left\lfloor \mathrm{v}_i^k = \mathbf{prox}_{\sigma_i \mathrm{g}_i^*}\left(\mathrm{u}_i^k - \sigma_i\left(\nabla \ell_i^*(\mathrm{u}_i^k) - \mathrm{L}_i(\mathrm{y}^k + \theta(\mathrm{y}^k - \mathrm{x}^k))\right)\right)\right. \\ \mathrm{x}^{k+1} = \mathrm{x}^k + \frac{\lambda}{\sigma_0}\left(\mathrm{y}^k - \mathrm{x}^k + \sigma_0 \sum_{i=1}^m \mathrm{L}_i^*(\mathrm{u}_i^k - \mathrm{v}_i^k)\right) \\ \textit{For every } i = 1, \ldots, m \\ \left\lfloor \mathrm{u}_i^{k+1} = \mathrm{u}_i^k + \frac{\lambda}{\sigma_i}\left(\mathrm{v}_i^k - \mathrm{u}_i^k - \sigma_i \theta \mathrm{L}_i(\mathrm{y}^k - \mathrm{x}^k)\right),\right. \end{array}\right. \tag{5.28}$$

*which, in the case $m = 1$, is very similar to the method proposed in [**?**, Algorithm 3] (by taking $\mu = (1-\theta)^{-1}$ for $\theta \in [-1, 0]$), with a slightly different choice of the parameters involved in the last two lines in* (5.28). *An advantage of our method, even in the case $m = 1$, is that the stepsizes σ_0 and σ_1 may vary among iterations.*

6. Applications. In this section we explore two applications for illustrating the advantages and flexibility of the methods proposed in the previous sections. First we provide an application of Theorem 2.2 to the obstacle problem in PDE's in which dropping the extra forward step decreases the computational cost by iteration because the computation of an extra gradient step is numerically expensive. In the second application, devoted to empirical risk minimization (ERM), we illustrate the flexibility of using non self-adjoint linear operators. We derive different sequential algorithms depending on the nature of the linear operator involved.

6.1. Obstacle problem. The obstacle problem is to find the equilibrium position of an elastic membrane on a domain Ω, whose boundary is fixed and is restricted to remain above the some obstacle, given by the function $\varphi\colon \Omega \to \mathbb{R}$. This problem can be applied to fluid filtration in porous media, elasto-plasticity, optimal control among other disciplines (see, e.g., [**?**] and the references therein). Let $u\colon \Omega \to \mathbb{R}$ be a function representing the vertical displacement of the membrane and let $\psi\colon \Gamma \to \mathbb{R}$ be the function representing the fixed boundary, where Γ is the smooth boundary of Ω. Assume that $\psi \in H^{1/2}(\Gamma)$ and $\varphi \in C^{1,1}(\Omega)$ satisfy $\mathrm{T}\varphi \leq \psi$, and consider the problem

$$\min_{\mathrm{u} \in H^1(\Omega)} \frac{1}{2} \int_\Omega |\nabla \mathrm{u}|^2 \mathrm{dx} \tag{6.1}$$

$$\text{s.t. } \mathrm{Tu} = \psi, \quad \text{a.e. on } \Gamma; \tag{6.2}$$

$$\mathrm{u} \geq \varphi, \quad \text{a.e. in } \Omega, \tag{6.3}$$

where $\mathrm{T}\colon H^1(\Omega) \to H^{1/2}(\Gamma)$ is the (linear) trace operator and $H^1(\Omega)$ is endowed with the scalar product $\langle \cdot \mid \cdot \rangle\colon (\mathrm{u}, \mathrm{v}) \mapsto \int_\Omega \mathrm{uv}\, dx + \int_\Omega \nabla \mathrm{u} \cdot \nabla \mathrm{v}\, dx$. There is a unique solution to this obstacle problem [**?**].

In order to set this problem in our context, let us define the operator

$$Q\colon H^{-1}(\Omega) \times H^{-1/2}(\Gamma) \to H^1(\Omega) \tag{6.4}$$

which associates to each $(\mathrm{q}, \mathrm{w}) \in H^{-1}(\Omega) \times H^{-1/2}(\Gamma)$ the unique weak solution (in the sense of distributions) to [**?**, Section 25]

$$\begin{cases} -\Delta \mathrm{u} + \mathrm{u} = \mathrm{q}, & \text{in } \Omega; \\ \frac{\partial \mathrm{u}}{\partial \nu} = \mathrm{w}, & \text{on } \Gamma, \end{cases} \tag{6.5}$$

where ν is outer unit vector normal to Γ. Hence, Q satisfies

$$(\forall \mathrm{v} \in \mathrm{H}) \qquad \langle Q(\mathrm{q},\mathrm{w}) \mid \mathrm{v}\rangle = \langle \mathrm{w} \mid \mathrm{Tv}\rangle_{-1/2,1/2} + \langle \mathrm{q} \mid \mathrm{v}\rangle_{-1,1}, \tag{6.6}$$

where $\langle\cdot \mid \cdot\rangle_{-1/2,1/2}$ and $\langle\cdot \mid \cdot\rangle_{-1,1}$ stand for the dual pairs $H^{-1/2}(\Gamma) - H^{1/2}(\Gamma)$ and $H^{-1}(\Omega) - H^{1}(\Omega)$, respectively. Then, by defining $\mathrm{H} = H^1(\Omega)$, $\mathrm{G} = H^{1/2}(\Gamma)$, $\mathrm{f}\colon \mathrm{u} \mapsto \frac{1}{2}\int_\Omega |\nabla \mathrm{u}|^2 \mathrm{dx}$, $\mathrm{g} = \iota_{\mathrm{C}}$, where $\mathrm{C} = \big\{\mathrm{u} \in \mathrm{H} \;\big|\; \mathrm{u} \geq \varphi \text{ a.e. in } \Omega\big\}$, let $\mathrm{D} = \{\psi\}$, and let $\mathrm{L} = \mathrm{T}$, (6.7) can be written equivalently as

$$\min_{\mathrm{Lu} \in \mathrm{D}} \mathrm{f}(\mathrm{u}) + \mathrm{g}(\mathrm{u}). \tag{6.7}$$

Moreover, it is easy to verify that f is convex and, by using integration by parts and (6.6), for every $\mathrm{h} \in \mathrm{H}$ we have

$$\begin{aligned}
\mathrm{f}(\mathrm{u}+\mathrm{h}) - \mathrm{f}(\mathrm{u}) - \left\langle Q\left(-\Delta \mathrm{u}, \frac{\partial \mathrm{u}}{\partial \nu}\right) \middle| \mathrm{h}\right\rangle &= \frac{1}{2}\int_\Omega |\nabla \mathrm{h}|^2 dx + \int_\Omega \nabla \mathrm{u} \cdot \nabla \mathrm{h}\, dx + \langle \Delta \mathrm{u} \mid \mathrm{h}\rangle_{-1,1} \\
&\qquad - \left\langle \frac{\partial \mathrm{u}}{\partial \nu} \mid \mathrm{Th}\right\rangle_{-1/2,1/2} \\
&= \frac{1}{2}\int_\Omega |\nabla \mathrm{h}|^2 dx,
\end{aligned} \tag{6.8}$$

which yields

$$\lim_{\|\mathrm{h}\| \to 0} \frac{\left|\mathrm{f}(\mathrm{u}+\mathrm{h}) - \mathrm{f}(\mathrm{u}) - \left\langle Q\left(-\Delta \mathrm{u}, \frac{\partial \mathrm{u}}{\partial \nu}\right) \;\middle|\; \mathrm{h}\right\rangle\right|}{\|\mathrm{h}\|} = \frac{1}{2}\lim_{\|\mathrm{h}\| \to 0} \frac{\|\nabla \mathrm{h}\|_{L^2}^2}{\|\mathrm{h}\|} = 0. \tag{6.9}$$

Hence, f is Fréchet differentiable with a linear gradient given by $\nabla \mathrm{f}\colon \mathrm{u} \mapsto Q\left(-\Delta \mathrm{u}, \frac{\partial \mathrm{u}}{\partial \nu}\right)$. Moreover, from integration by parts we have

$$\left\langle Q\left(-\Delta \mathrm{u}, \frac{\partial \mathrm{u}}{\partial \nu}\right) \;\middle|\; \mathrm{h}\right\rangle = \left\langle \frac{\partial \mathrm{u}}{\partial \nu} \;\middle|\; \mathrm{Th}\right\rangle_{-1/2,1/2} - \langle \Delta \mathrm{u} \mid \mathrm{h}\rangle_{-1,1} = \int_\Omega \nabla \mathrm{u} \cdot \nabla \mathrm{h}\, dx \leq \|\mathrm{u}\|\|\mathrm{h}\|, \tag{6.10}$$

which yields $\|\nabla \mathrm{f}(\mathrm{u})\| \leq \|\mathrm{u}\|$ and, hence, it is 1-cocoercive [**?**]. In addition, the trace operator is linear and bounded [**?**] and we have from (6.6) that

$$(\forall \mathrm{v} \in \mathrm{H})(\forall \mathrm{w} \in H^{1/2}(\Gamma)) \qquad \langle Q(0,\mathrm{w}) \mid \mathrm{v}\rangle = \langle \mathrm{w} \mid \mathrm{Tv}\rangle_{-1/2,1/2}, \tag{6.11}$$

which yields $\mathrm{L}^*\colon \mathrm{w} \mapsto \mathrm{Q}(0,\mathrm{w})$ and since C is non-empty closed convex, g is convex, proper, lower semicontinuous and $\mathbf{prox}_{\gamma \mathrm{g}} = P_{\mathrm{C}}$, for any $\gamma > 0$.

Since first order conditions of (6.7) reduce to find $(\mathrm{u},\mathrm{w}) \in \mathrm{H} \times \mathrm{G}$ such that $0 \in N_{\mathrm{C}}(\mathrm{u}) + \nabla \mathrm{f}(\mathrm{u}) + \mathrm{T}^* N_{\mathrm{D}}(\mathrm{Tu})$, which is a particular case of Problem 2 and from Corollary 5.2 when $\theta = 1$ the method

$$\left\lfloor \begin{array}{l}
\mathrm{v}^k = P_{\mathrm{C}}\left(\mathrm{u}^k - \sigma_0 Q\left(-\Delta \mathrm{u}^k, \frac{\partial \mathrm{u}^k}{\partial \nu} + \mathrm{w}^k\right)\right) \\
\mathrm{t}^k = \mathrm{w}^k + \sigma_1\left(\mathrm{T}(2\mathrm{y}^k - \mathrm{x}^k) - \psi\right) \\
\mathrm{u}^{k+1} = \mathrm{u}^k + \frac{\lambda}{\sigma_0}\left(\mathrm{v}^k - \mathrm{u}^k + \sigma_0 Q(0, \mathrm{w}^k - \mathrm{t}^k)\right) \\
\mathrm{w}^{k+1} = \mathrm{w}^k + \frac{\lambda}{\sigma_1}\left(\mathrm{t}^k - \mathrm{w}^k - \sigma_1 \mathrm{T}(\mathrm{v}^k - \mathrm{u}^k)\right)
\end{array}\right. \tag{6.12}$$

generates a weakly convergent sequence $(\mathrm{u}^k)_{k\in\mathbb{N}}$ to the unique solution to the obstacle problem provided, for instance (see Remark 6.1), that $\max\{\sigma_0,\sigma_1\}+2\sqrt{\sigma_0\sigma_1}\|\mathrm{T}\|<2$. Note that $\nabla \mathrm{f}$ must be computed only once at each iteration, improving the performance with respect to primal-dual methods following Tseng's approach, in which $\nabla \mathrm{f}$ must be computed twice by iteration (see, e.g., [**?**, **?**]). The method proposed in [**?**] can also solve this problem but with stronger conditions on constants σ_0 and σ_1 as studied in Remark 6. Moreover, our approach may include variable stepsizes together with different assymetric linear operators which may improve the performance of the method.

On the other hand, the general version of our method in Theorem 3.1 allows for an additional projection onto a closed convex set. In this case this can be useful to impose some of the constraints of the problem in order to guarantee that iterates at each iteration satisfy such constraints. An additional projection step may accelerate the method as it has been studied in [**?**]. Numerical comparisons between these methods are part of further research.

6.2. An Incremental Algorithm for Nonsmooth Empirical Risk Minimization. In machine learning [**?**], the the Empirical Risk Minimization (ERM) problem seeks to minimize a finite sample approximation of an expected loss, under conditions on the feasible set and the loss function. If the solution to the sample approximation converges to a minimizer of the expected loss when the size of the sample increases, we say that the problem is learnable. Suppose that we have a sample of size m, and, for every $i\in\{1,\ldots,m\}$, the loss function associated to the sample z_i is given by $l(\cdot;\mathrm{z}_i)\colon \mathrm{x}\mapsto \mathrm{f}_i(\mathrm{a}_i^\top x)$, where each $\mathrm{a}_i\in\mathbb{R}^d\backslash\{0\}$ and each $\mathrm{f}_i:\mathbb{R}\to(-\infty,\infty]$ is closed, proper, and convex. Then the ERM problem is to

$$\underset{\mathrm{x}\in\mathbb{R}^d}{\text{minimize}}\ \frac{1}{m}\sum_{i=1}^m \mathrm{f}_i(\mathrm{a}_i^\top \mathrm{x}). \tag{6.13}$$

This form features in support vector machines, logistic regression, linear regression, least-absolute deviations, and many other common models in machine learning.

The parameter m indicates the size of the training set and is typically large. Parallelizing a (sub)gradient computation of (6.13) is straightforward, but in general, because training sets are large, we may not have enough processors to do so. Thus, when only a few processors are available, incremental iterative algorithms, in which one or a few training samples are used per iteration to update our solution estimate, are a natural choice.

Several incremental algorithms are available for solving (6.13), including incremental (sub)gradient descent and incremental aggregated gradient methods [**?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**, **?**]. The former class requires diminishing stepsizes (e.g., of size $O(k^{-1/2})$) and, hence, their convergence may be very slow, while the latter class of algorithms is usually restricted to the cases in which either f_i is smooth or the dual problem of (6.13) is smooth (in which case (6.13) is strongly convex). In contrast, we now develop an incremental proximal algorithm, which imposes no smoothness or strong convexity assumptions. It has a Gauss-Seidel structure and is obtained by an application of Theorem 5.1. The involved stepsizes may vary among iterations but they are set to be constants for simplicity.

The method follows from the following first-order optimality conditions obtained assuming some qualification condition:

$$\mathrm{x} \quad \text{solves (6.13)} \quad \Leftrightarrow \quad 0 \in \sum_{i=1}^{m} \mathrm{a}_i \partial \mathrm{f}_i(\mathrm{a}_i^\top \mathrm{x}), \tag{6.14}$$

which is a particular case of Problem 2 when $\mathrm{H} = \mathbb{R}^d$, $\mathrm{A} \equiv \{0\}$, $\mathrm{C}_1 = \mathrm{C}_2 \equiv 0$ and, for every $i \in \{1, \ldots, m\}$, $\mathrm{G}_i = \mathbb{R}$, $\mathrm{D}_i^{-1} = 0$, $\mathrm{L}_i = \mathrm{a}_i^\top$, and $\mathrm{B}_i = \partial \mathrm{f}_i$. By using Theorem 5.1 in this case for matrices $(\mathrm{P}_{ij})_{0 \leq i < j \leq m}$ given by

$$(\forall 0 \leq j < i \leq m) \quad \mathrm{P}_{ij} = \begin{cases} \frac{\mathrm{Id}}{\sigma_0}, & \text{if } i = j = 0; \\ \frac{1}{\sigma_i}, & \text{if } i = j > 0; \\ -\mathrm{a}_i^\top, & \text{if } j = 0; \\ \sigma_0 \mathrm{a}_i^\top \mathrm{a}_j, & \text{if } 0 < j < i, \end{cases} \tag{6.15}$$

we obtain

$$\left\lfloor \begin{array}{l} \mathrm{v}_1^k = \mathbf{prox}_{\sigma_1 \mathrm{f}_1^*}\left(\mathrm{u}_1^k + \sigma_1\left(\mathrm{a}_1^\top \mathrm{x}^k - \sigma_0 \sum_{i=1}^m \mathrm{a}_1^\top \mathrm{a}_i \mathrm{u}_i^k\right)\right) \\ \mathrm{v}_2^k = \mathbf{prox}_{\sigma_2 \mathrm{f}_2^*}\left(\mathrm{u}_2^k + \sigma_2\left(\mathrm{a}_2^\top \mathrm{x}^k - \sigma_0\left(\mathrm{a}_2^\top \mathrm{a}_1 \mathrm{v}_1^k + \sum_{i=2}^m \mathrm{a}_2^\top \mathrm{a}_i \mathrm{u}_i^k\right)\right)\right) \\ \quad \vdots \\ \mathrm{v}_m^k = \mathbf{prox}_{\sigma_m \mathrm{f}_m^*}\left(\mathrm{u}_m^k + \sigma_m\left(\mathrm{a}_m^\top \mathrm{x}^k - \sigma_0\left(\sum_{i=1}^{m-1} \mathrm{a}_m^\top \mathrm{a}_i \mathrm{v}_i^k + \|\mathrm{a}_m\|^2 \mathrm{u}_m^k\right)\right)\right) \\ \mathrm{x}^{k+1} = \mathrm{x}^k - \lambda \sum_{i=1}^m \mathrm{a}_i \mathrm{v}_i^k \\ \mathrm{u}_1^{k+1} = \mathrm{u}_1^k + \frac{\lambda}{\sigma_1}\left(\mathrm{v}_1^k - \mathrm{u}_1^k\right) \\ \qquad \vdots \\ \mathrm{u}_m^{k+1} = \mathrm{u}_m^k + \frac{\lambda}{\sigma_m}\left(\mathrm{v}_m^k - \mathrm{u}_m^k\right) + \sigma_0 \sum_{j=1}^{m-1} \mathrm{a}_m^\top \mathrm{a}_j (\mathrm{v}_j^k - \mathrm{u}_j^k). \end{array} \right. \tag{6.16}$$

Since conditions (5.5)-(5.7) hold if

$$\sqrt{\sum_{i=1}^m \|\mathrm{a}_i\|^2} + \sigma_0 \sum_{i=1}^m \|\mathrm{a}_i\|^2 + \frac{\sigma_0}{2}\left(\max_{i=1,\ldots,m} \|\mathrm{a}_i\|^2 - \min_{i=1,\ldots,m} \|\mathrm{a}_i\|^2\right) < \frac{1}{\max\limits_{i=0,\ldots,m} \sigma_i}, \tag{6.17}$$

by choosing $(\sigma_i)_{0 \leq i \leq m}$ satisfying (6.17) the sequence $(\mathrm{x}^k)_{k \in \mathbb{N}}$ generated by (6.16) converges to a solution provided that $\lambda < M^{-1}$ where

$$M = \left(\min_{i=0,\ldots,m} \sigma_i\right)^{-1} + \frac{1}{2}\sqrt{\sum_{i=1}^m \|\mathrm{a}_i\|^2} + \frac{\sigma_0}{2}\left(\sum_{i=1}^m \|\mathrm{a}_i\|^2 + \max_{i=1,\ldots,m} \|\mathrm{a}_i\|^2\right).$$

Note that, without loss of generality, we can assume, for every $i \in \{1, \ldots, m\}$, $\|\mathrm{a}_i\| = 1$, since $\mathrm{f}_i(\mathrm{a}_i^\top \mathrm{x}) = \mathrm{g}_i((\mathrm{a}_i/\|\mathrm{a}_i\|)^\top \mathrm{x})$ with $\mathrm{g}_i \colon \mathrm{x} \mapsto \mathrm{f}_i(\|\mathrm{a}_i\| \mathrm{x})$ and $\mathbf{prox}_{\mathrm{g}_i} \colon \mathrm{x} \mapsto \mathbf{prox}_{\|\mathrm{a}_i\|^2 \mathrm{f}_i}(\|\mathrm{a}_i\| \mathrm{x})/\|\mathrm{a}_i\|$. Therefore, condition (6.17) can be reduced to $\sqrt{m} + m\sigma_0 < (\max_{i=0,\ldots,m} \sigma_i)^{-1}$, which, in the case $\sigma_0 = \cdots = \sigma_m$ reduces to $\sigma_0 < (\sqrt{5} - 1)/(2\sqrt{m})$.

7. Conclusion. In this paper, we systematically investigated a new extension of Tseng's forward-backward-forward and forward-backward methods. The three primary contributions of this investigation are (1) a lower per-iteration complexity variant of Tseng's method which activates the cocoercive operator only once; (2) the ability to incorporate variable metrics in operator-splitting schemes, which, unlike typical variable metric methods, we do not enforce compatibility conditions between metrics employed at successive time steps; and (3) the ability to incorporate modified resolvents $J_{P^{-1}A}$ in iterative fixed-point algorithms, which, unlike typical preconditioned fixed point iterations, can be formed from non self-adjoint linear operators P, which lead to new Gauss-Seidel style operator-splitting schemes. Future work on this topic should investigate whether and when such modifications lead to better practical performance.