Chapter 1
Variable metric ADMM for solving variational inequalities with monotone operators over affine sets

Radu Ioan Boţ, Ernö Robert Csetnek and Dennis Meier

Abstract We propose an iterative scheme for solving variational inequalities with monotone operators over affine sets in an infinite dimensional Hilbert space setting. We show that several primal-dual algorithms in the literature as well as the classical ADMM algorithm for convex optimization problems, together with some of its variants, are encompassed by the proposed numerical scheme. Furthermore, we carry out a convergence analysis of the generated iterates and provide convergence rates by using suitable dynamical step sizes together with variable metric techniques.

Key words: ADMM algorithm, primal-dual algorithm, monotone operators, convex optimization
AMS 2010 Subject Classification: 47H05, 65K05, 90C25

1.1 Introduction

Many problems in fields like signal and image processing, portfolio optimization, cluster analysis, location theory, network communication and machine learning as well as inverse problems can be formulated as a convex optimization problem of the form

Radu Ioan Boţ
University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, e-mail: radu.bot@univie.ac.at

Ernö Robert Csetnek
University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, e-mail: ernoe.robert.csetnek@univie.ac.at

Dennis Meier
University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, e-mail: meierd61@univie.ac.at

$$\inf_{x\in\mathscr{H},z\in\mathscr{G}}\{f(x)+h(x)+g(z)\}, \qquad (1.1)$$
$$\text{s.t. } L_1x+L_2z=d$$

where $\mathscr{H},\mathscr{G}$ and $\mathscr{Z}$ are real Hilbert spaces, $f:\mathscr{H}\to\overline{\mathbb{R}}:=\mathbb{R}\cup\{\pm\infty\}$ and $g:\mathscr{G}\to\overline{\mathbb{R}}$ are proper, convex and lower semicontinuous functions, $h:\mathscr{H}\to\mathbb{R}$ is a convex and Fréchet differentiable function with Lipschitz continuous gradient, $L_1:\mathscr{H}\to\mathscr{Z},L_2:\mathscr{G}\to\mathscr{Z}$ are linear continuous operators and $d\in\mathscr{Z}$.

One of the most prominent numerical algorithms one can find in the literature for solving optimization problems of the form (1.1) is the *alternating direction method of multipliers (ADMM)*. In the case $h=0$, which represents the standard setting in the literature addressing ADMM methods, the augmented Lagrangian associated with problem (1.1) is given for a fixed real number $c>0$ as

$$L_c:\mathscr{H}\times\mathscr{G}\times\mathscr{Z}\to\overline{\mathbb{R}},$$
$$L_c(x,z,y)=f(x)+g(z)+\langle y,L_1x+L_2z-d\rangle+\frac{c}{2}\|L_1x+L_2z-d\|^2.$$

The ADMM algorithm generates a sequence $(x^k,z^k,y^k)_{k\geq 0}\in\mathscr{H}\times\mathscr{G}\times\mathscr{Z}$ by iterating for every $k\geq 0$

$$x^{k+1}\in\arg\min_{x\in\mathscr{H}}L_c(x,z^k,y^k)=\arg\min_{x\in\mathscr{H}}\left\{f(x)+\frac{c}{2}\|L_1x+L_2z^k-d+c^{-1}y^k\|^2\right\}$$
$$z^{k+1}\in\arg\min_{z\in\mathscr{G}}L_c(x^{k+1},z,y^k)=\arg\min_{x\in\mathscr{G}}\left\{g(z)+\frac{c}{2}\|L_1x^{k+1}+L_2z-d+c^{-1}y^k\|^2\right\}$$
$$y^{k+1}=y^k+c(L_1x^{k+1}+L_2z^{k+1}-d).$$

Since the function f and the operator L_1 are not evaluated independently in the first line of the algorithm, the minimization with respect to the variable x does not lead to a proximal step (the same is true for the second minimization). This results in less attractiveness for implementations than for primal-dual splitting algorithms, which represent the second class of prominent iterative methods for solving (1.1). This drawback has been overcome in the literature by introducing a suitable regularizer equipped with a (semi-)metric, see for example [15] for a finite dimensional approach (in case $\mathscr{G}=\mathscr{Z}$, $L_2=-\text{Id}$ and $d=0$ see also [20], and also [3] for an extension of the ADMM algorithm by involving also smooth parts in the objective, by employing variable metrics and by working in an infinite dimensional Hilbert setting). This so-called alternating direction *proximal* method of multipliers (AD-PMM) reveals a bridge which connects the classical ADMM algorithm with primal-dual methods. This observation served as the starting point for the investigations made in [6], where a generalization of the AD-PMM algorithm to monotone inclusions was proposed and investigated from the point of view of its convergence properties.

In this paper we propose an iterative algorithm for solving variational inequalities with monotone operators of the type

$$\text{find } (x,z) \in \mathscr{H} \times \mathscr{G} \text{ such that } 0 \in (Ax + Cx) \times Bz + N_S(x,z),$$

where $A : \mathscr{H} \rightrightarrows \mathscr{H}$ and $B : \mathscr{G} \rightrightarrows \mathscr{G}$ are maximally monotone operators, $C : \mathscr{H} \to \mathscr{H}$ is an η-cocoercive operator, for $\eta \geq 0$, $S := \{(x,z) \in \mathscr{H} \times \mathscr{G} : L_1 x + L_2 z = d\}$, and N_S denotes the normal cone operator to the set S. This delivers a unifying framework for solving monotone inclusions in Hilbert spaces which encompasses in particular the ADMM algorithm in [6], several primal-dual iterative methods [7, 10, 14, 21] as well as the classical ADMM algorithm designed to solve problems of type 1.1 (and its variants from [15, 20], see also [16, 17]). After giving the necessary preliminaries, we formulate the ADMM iterative scheme for variational inequalities and carry out a convergence analysis. Furthermore, under additional strong monotonicity assumptions, we derive convergence rates for the primal iterates by using a dynamic step size strategy combined with variable metric techniques.

1.2 Notation and preliminaries

Throughout, $\mathscr{H}$, $\mathscr{G}$ and $\mathscr{Z}$ denote real Hilbert spaces with scalar products $\langle \cdot, \cdot \rangle$ and associated norms $\| \cdot \|$ (since there is no risk of confusion, they are denoted in the same way). Let $M : \mathscr{H} \rightrightarrows \mathscr{H}$ be an arbitrary set-valued operator. We denote by $\operatorname{gra} M := \{(x,u) \in \mathscr{H} \times \mathscr{H} : u \in Mx\}$ its graph, by $\operatorname{dom} M := \{x \in \mathscr{H} : Mx \neq \emptyset\}$ its domain and by $M^{-1} : \mathscr{H} \rightrightarrows \mathscr{H}$ its inverse operator, defined by $(u,x) \in \operatorname{gra} M^{-1}$ if and only if $(x,u) \in \operatorname{gra} M$. M is said to be monotone, if $\langle x - y, u - v \rangle \geq 0$ for all $(x,u),(y,v) \in \operatorname{gra} M$. A monotone operator M is said to be maximally monotone, if there exists no proper monotone extension of the graph of M on $\mathscr{H} \times \mathscr{H}$. For an arbitrary $\gamma > 0$, the operator M is called γ-strongly monotone, if $\langle x - y, u - v \rangle \geq \gamma \|x - y\|^2$ for all $(x,u),(y,v) \in \operatorname{gra} M$.

The resolvent of M is the mapping $J_M : \mathscr{H} \rightrightarrows \mathscr{H}$, $J_M := (\operatorname{Id} + M)^{-1}$, where Id denotes the identity operator on $\mathscr{H}$. If M is maximally monotone, then $J_M : \mathscr{H} \to \mathscr{H}$ is single-valued and maximally monotone (see [4, Proposition 23.7 and Corollary 23.10]). Furthermore, for an arbitrary $\gamma > 0$ we have (see [4, Proposition 23.18])

$$J_{\gamma M} + \gamma J_{\gamma^{-1} M^{-1}} \circ \gamma^{-1} \operatorname{Id} = \operatorname{Id}. \tag{1.2}$$

For a linear continuous operator $L : \mathscr{H} \to \mathscr{G}$, its adjoint operator $L^* : \mathscr{G} \to \mathscr{H}$ is defined by $\langle L^* y, x \rangle = \langle y, Lx \rangle$ for all $(x,y) \in \mathscr{H} \times \mathscr{G}$. The norm of L is defined by $\|L\| := \sup\{\|Lx\| : x \in \mathscr{H}, \|x\| \leq 1\}$. The linear operator L is said to be skew, if $\langle x, Lx \rangle = 0$ for all $x \in \mathscr{H}$. A single-valued operator $M : \mathscr{H} \to \mathscr{H}$ is said to be β-cocoercive, for $\beta \geq 0$, if $\beta \langle x - y, Mx - My \rangle \geq \|Mx - My\|^2$ for all $(x,y) \in \mathscr{H} \times \mathscr{H}$. Moreover, M is β-Lipschitz continuous, if $\|Mx - My\| \leq \beta \|x - y\|$ for all $(x,y) \in \mathscr{H} \times \mathscr{H}$.

We write

$$\mathscr{S}_+(\mathscr{H}) := \{L : \mathscr{H} \to \mathscr{H} : L \text{ is linear, bounded, positive semidefinite and } L = L^*\}.$$

The Loewner partial ordering on $\mathcal{S}_+(\mathcal{H})$ is defined by

$$(\forall U \in \mathcal{S}_+(\mathcal{H}))(\forall V \in \mathcal{S}_+(\mathcal{H})) \quad U \succeq V :\Leftrightarrow (\forall x \in \mathcal{H}) \ \langle x, Ux\rangle \geq \langle x, Vx\rangle.$$

Further, for every $U \in \mathcal{S}_+(\mathcal{H})$, we define a semi-scalar product and a semi-norm by

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x,y\rangle_U := \langle x, Uy\rangle \ \text{ and } \ \|x\|_U := \sqrt{\langle x, Ux\rangle},$$

respectively. For $\alpha > 0$ we set

$$\mathcal{P}_\alpha(\mathcal{H}) := \{U \in \mathcal{S}_+(\mathcal{H}) \mid U \succeq \alpha \operatorname{Id}\}.$$

Since we will also address convex optimization problems, we recall some elements of convex analysis. For a function $f : \mathcal{H} \to \overline{\mathbb{R}}$ we denote by $\operatorname{dom} f := \{x \in \mathcal{H} : f(x) < +\infty\}$ its effective domain and say that f is proper, if $\operatorname{dom} f \neq \emptyset$ and $f(x) \neq -\infty$ for all $x \in \mathcal{H}$. The (convex) conjugate function $f^* : \mathcal{H} \to \overline{\mathbb{R}}$ of f is defined by $f^*(u) := \sup_{x\in\mathcal{H}}\{\langle u,x\rangle - f(x)\}$ for all $u \in \mathcal{H}$. The (convex) subdifferential $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$ of f is given by

$$\partial f(x) := \{p \in \mathcal{H} : f(y) - f(x) \geq \langle p, y-x\rangle \ \forall y \in \mathcal{H}\},$$

for $x \in \mathcal{H}$ with $f(x) \in \mathbb{R}$ and $\partial f(x) = \emptyset$, otherwise. In case f is a proper, convex and lower semi-continuous function, $\partial f : \mathcal{H} \rightrightarrows \mathcal{H}$ is a maximally monotone operator [19].

For $f, g : \mathcal{H} \to \overline{\mathbb{R}}$ two proper functions, the infimal convolution $f \Box g : \mathcal{H} \to \overline{\mathbb{R}}$ is defined by $(f \Box g)(x) = \inf_{u\in\mathcal{H}}\{f(u) + g(x-u)\}$ for all $x \in \mathcal{H}$.

For a proper, convex and lower semi-continuous function $f : \mathcal{H} \to \overline{\mathbb{R}}$ and $\gamma > 0$, for every $x \in \mathcal{H}$ we denote by $\operatorname{prox}_{\gamma f}(x)$ the proximal point of parameter γ of f at x, which is defined by

$$\operatorname{prox}_{\gamma f}(x) := \operatorname{argmin}_{y\in\mathcal{H}} \left\{ f(y) + \frac{1}{2\gamma}\|y-x\|^2 \right\}.$$

Since $J_{\gamma\partial f} = (\operatorname{Id} + \gamma\partial f)^{-1} = \operatorname{prox}_{\gamma f}$ (see [4, Example 23.3]), this gives a single-valued operator $\operatorname{prox}_{\gamma f} : \mathcal{H} \to \mathcal{H}$ fulfilling the extended Moreau's decomposition formula

$$\operatorname{prox}_{\gamma f} + \gamma \operatorname{prox}_{\gamma^{-1} f^*} \circ \gamma^{-1} \operatorname{Id} = \operatorname{Id}.$$

Last, for a nonempty convex subset S of $\mathcal{H}$ and for $x \in \mathcal{H}$, the normal cone to S at x is

$$N_S(x) = \begin{cases} \{u \in \mathcal{H} \mid \sup\langle s - x, u\rangle \leq 0 \ \forall s \in S\}, & \text{if } x \in S \\ \emptyset, & \text{if } x \notin S \end{cases}.$$

1.3 A variable metric ADMM for monotone operators

In this section we present the variational inequality problem to solve, formulate the iterative numerical scheme and prove convergence for the sequence of generated iterates.

1.3.1 Problem formulation and algorithm

We start by describing the problem under investigation.

Problem 1. Let $\mathscr{H}$, $\mathscr{G}$ and $\mathscr{Z}$ be real Hilbert spaces, $A : \mathscr{H} \rightrightarrows \mathscr{H}$ and $B : \mathscr{G} \rightrightarrows \mathscr{G}$ be maximally monotone operators and $C : \mathscr{H} \to \mathscr{H}$ an η-cocoercive operator, for $\eta \geq 0$. Further, let $L_1 : \mathscr{H} \to \mathscr{Z}$ and $L_2 : \mathscr{G} \to \mathscr{Z}$ be linear continuous operators and $S := \{(x,z) \in \mathscr{H} \times \mathscr{G} : L_1x + L_2z = d\}$. The aim is to solve the variational inequality with monotone operators over the set S

$$\text{find } (x,z) \in \mathscr{H} \times \mathscr{G} \text{ such that } 0 \in (Ax+Cx) \times Bz + N_S(x,z), \tag{1.3}$$

which can be reformulated as

$$\begin{gathered}\text{find } (x,z) \in S \text{ such that } \exists (p,q) \in -(Ax+Cx) \times (-Bz) \text{ with the property} \\ \langle (p,q),(u,v)-(x,z)\rangle \leq 0 \ \forall (u,v) \in S.\end{gathered}$$

We will propose an algorithm for determining the KKT points associated to the variational inequality (1.3), namely, those $(x,z,y) \in \mathscr{H} \times \mathscr{G} \times \mathscr{Z}$ which fulfill

$$-L_1^*y \in Ax + Cx, \ \ -L_2^*y \in Bz \text{ and } L_1x + L_2z = d. \tag{1.4}$$

Remark 1. If $(x,z,y) \in \mathscr{H} \times \mathscr{G} \times \mathscr{Z}$ is a KKT point of (1.3), then, obviously,

$$(-L_1^*y, -L_2^*y) \in (Ax+Cx) \times Bz + N_S(x,z),$$

which means that (x,z) is a solution of (1.3).

On the other hand, if $(x,z) \in \mathscr{H} \times \mathscr{G}$ is a solution of (1.3), then there exists $(p,q) \in -(Ax+Cx) \times (-Bz)$ such that

$$L_1x + L_2z = d \text{ and } (x,z) \in \underset{(u,v)\in S}{\arg\min} \langle (-p,-q),(u,v)\rangle.$$

Using duality theory, we obtain under suitable constraint qualifications the existence of $y \in \mathscr{Z}$ such that

$$\begin{aligned}\langle (-p,-q),(x,z)\rangle &= \inf_{(u,v)\in \mathscr{H}\times\mathscr{G}} \{\langle (-p,-q),(u,v)\rangle + \langle y, L_1u + L_2v - d\rangle\} \\ &= \inf_{u\in\mathscr{H}} \langle u, -p + L_1^*y\rangle + \inf_{v\in\mathscr{G}} \langle z, -q + L_2^*y\rangle - \langle y, d\rangle.\end{aligned}$$

Since the term on the left-hand side is finite, this holds only when $p = L_1^* y$ and $q = L_2^* y$. In other words, (x,z,y) is a KKT point of (1.3).

Next, we relate Problem 1 to a particular convex optimization problem with affine constraints.

Problem 2. Let $\mathscr{H}$, $\mathscr{G}$ and $\mathscr{Z}$ be real Hilbert spaces, $f : \mathscr{H} \to \overline{\mathbb{R}}$, $g : \mathscr{G} \to \overline{\mathbb{R}}$ be proper, convex and lower semicontinuous functions, $h : \mathscr{H} \to \mathbb{R}$ a convex and Fréchet differentiable function with η-Lipschitz continuous gradient, for $\eta \geq 0$, $L_1 : \mathscr{H} \to \mathscr{Z}$, $L_2 : \mathscr{G} \to \mathscr{Z}$ linear continuous operators and $d \in \mathscr{Z}$. We consider the convex optimization problem

$$\inf_{x\in\mathscr{H}, z\in\mathscr{G}} \{f(x) + h(x) + g(z)\}. \quad \text{s.t. } L_1 x + L_2 z = d \tag{1.5}$$

The system of KKT optimality conditions associated to this optimization problem is given by

$$-L_1^* y \in \partial f(x) + \nabla h(x), \ -L_2^* y \in \partial g(z) \text{ and } L_1 x + L_2 z = d. \tag{1.6}$$

If (x,z,y) is a solution of (1.6), then (x,z) is an optimal solution of (1.5) and y is an optimal solution of its dual problem

$$\sup_{y\in\mathscr{G}} \{-(f^* \Box h^*)(-L_1^* y) - g^*(-L_2^* y) - \langle d, y\rangle\}, \tag{1.7}$$

For

$$A := \partial f, B := \partial g \text{ and } C := \nabla h, \tag{1.8}$$

the system of KKT optimality conditions (1.6) is nothing else than (1.4). Notice that ∂f and ∂g are maximally monotone operators, while, by the Baillon-Haddad Theorem (see [4, Corollary 18.16]), the gradient of h is η-cocoercive.

Remark 2. Consider the optimization problem

$$\inf_{y\in\mathscr{Z}} \{f(L_1 y) + g(L_2 y)\}, \tag{1.9}$$

where $f : \mathscr{H} \to \overline{\mathbb{R}}$ and $g : \mathscr{G} \to \overline{\mathbb{R}}$ are proper, convex and lower semicontinuous functions, and $L_1 : \mathscr{Z} \to \mathscr{H}$ and $L_2 : \mathscr{Z} \to \mathscr{G}$ are linear continuous operators. The associated dual problem can be written as

$$\inf_{(p,e)\in\mathscr{H}\times\mathscr{G}} \{f^*(p) + g^*(q)\}, \quad \text{s.t. } L_1^* p + L_2^* q = 0 \tag{1.10}$$

while (1.9) is on its turn the dual problem of (1.10). Finding a solution (p,q,y) of the system of KKT optimality conditions associated to (1.10)

$$-L_1y \in \partial f^*(p),\ -L_2y \in \partial g^*(y) \text{ and } L_1^*p + L_2^*q = 0$$

provides an optimal solution y of problem (1.9) and an optimal solution (p,q) of problem (1.10).

We propose the following iterative scheme for determining the KKT points of the variational inequality (1.3).

Algorithm 1 *Let $M_1^k \in \mathscr{S}_+(\mathscr{H})$, $M_2^k \in \mathscr{S}_+(\mathscr{G})$ and $c > 0$ be such that $cL_1^*L_1 + M_1^k \in \mathscr{P}_{\alpha_k}(\mathscr{H})$ and $cL_2^*L_2 + M_2^k \in \mathscr{P}_{\beta_k}(\mathscr{G})$, with $\alpha_k, \beta_k > 0$, for all $k \geq 0$. Choose $(x^0, z^0, y^0) \in \mathscr{H} \times \mathscr{G} \times \mathscr{Z}$. For all $k \geq 0$ generate the sequence $(x^k, z^k, y^k)_{k\geq 0}$ as follows:*

$$x^{k+1} := (cL_1^*L_1 + M_1^k + A)^{-1}\left[cL_1^*(-L_2z^k + d - c^{-1}y^k) + M_1^kx^k - Cx^k\right] \quad (1.11)$$

$$z^{k+1} := (cL_2^*L_2 + M_2^k + B)^{-1}\left[cL_2^*(-L_1x^{k+1} + d - c^{-1}y^k) + M_2^kz^k\right] \quad (1.12)$$

$$y^{k+1} := y^k + c(L_1x^{k+1} + L_2z^{k+1} - d). \quad (1.13)$$

Remark 3. For the choice $\mathscr{G} := \mathscr{Z}$, $L_2 := -\operatorname{Id}$ and $d := 0$, the variational inequality to solve simplifies to the following monotone inclusion problem

$$\text{find } x \in \mathscr{H} \text{ such that } 0 \in Ax + Cx + (L_1^* \circ B \circ L_1)(x),$$

while Algorithm 1 becomes the iterative scheme proposed in [6] for solving it.

We show in the following that the numerical scheme above encompasses several other algorithms from the literature. For all $k \geq 0$, the equations (1.11) and (1.12) are equivalent to

$$-cL_1^*(L_1x^{k+1} + L_2z^k - d + c^{-1}y^k) + M_1^k(x^k - x^{k+1}) - Cx^k \in Ax^{k+1}, \quad (1.14)$$

and, respectively,

$$-cL_2^*(L_1x^{k+1} + L_2z^{k+1} - d + c^{-1}y^k) + M_2^k(z^k - z^{k+1}) \in Bz^{k+1}. \quad (1.15)$$

In the variational setting of Problem 2, i.e., considering the particular choice (1.8), the inclusion (1.14) becomes

$$0 \in \partial f(x^{k+1}) + cL_1^*(L_1x^{k+1} + L_2z^k - d + c^{-1}y^k) + M_1^k(x^{k+1} - x^k) + \nabla h(x^k),$$

which is equivalent to

$$x^{k+1} = \operatorname*{arg\,min}_{x \in \mathscr{H}} \Big\{ f(x) + \langle x - x^k, \nabla h(x^k) \rangle + \frac{c}{2}\|L_1x + L_2z^k - d + c^{-1}y^k\|^2 + \frac{1}{2}\|x - x^k\|^2_{M_1^k} \Big\}.$$

On the other hand, (1.15) becomes

$$-cL_2^*(L_1x^{k+1}+L_2z^{k+1}-d+c^{-1}y^k)+M_2^k(z^k-z^{k+1})\in\partial g(z^{k+1}),$$

which is equivalent to

$$z^{k+1}=\underset{x\in\mathscr{G}}{\arg\min}\left\{g(z)+\frac{c}{2}\|L_1x^{k+1}+L_2z-d+c^{-1}y^k\|^2+\frac{1}{2}\|z-z^k\|^2_{M_2^k}\right\}.$$

In conclusion, the iterative scheme (1.11) - (1.13) applied to the variational setting of problem 2 reads

$$x^{k+1}=\underset{x\in\mathscr{H}}{\arg\min}\left\{f(x)+\langle x-x^k,\nabla h(x^k)\rangle+\frac{c}{2}\|L_1x+L_2z^k-d+c^{-1}y^k\|^2\right.$$
$$\left.+\frac{1}{2}\|x-x^k\|^2_{M_1^k}\right\} \quad (1.16)$$
$$z^{k+1}=\underset{x\in\mathscr{G}}{\arg\min}\left\{g(z)+\frac{c}{2}\|L_1x^{k+1}+L_2z-d+c^{-1}y^k\|^2+\frac{1}{2}\|z-z^k\|^2_{M_2^k}\right\} \quad (1.17)$$
$$y^{k+1}=y^k+c(L_1x^{k+1}+L_2z^{k+1}-d). \quad (1.18)$$

The situation when $h=0$ and the sequences $(M_1^k)_{k\geq 0},(M_2^k)_{k\geq 0}$ are constant has been considered for example in [15]. The case $\mathscr{G}=\mathscr{X},L_2=-\operatorname{Id}$ and $d=0$ delivers the algorithm formulated and investigated by Banert, Boţ and Csetnek in [3]. The latter is a generalization of the iterative scheme proposed by Shefi and Teboulle [20], which addresses the case when $h=0$ and the sequences $(M_1^k)_{k\geq 0},(M_2^k)_{k\geq 0}$ are constant in the setting of finite dimensional Hilbert spaces. Finally, when $h=0$ and $M_1^k=M_2^k=0$ for all $k\geq 0$, the iterative scheme (1.16) - (1.18) collapses into the classical version of the ADMM algorithm (see for example [16, 17]).

We refer the reader to [6, Remark 5], where it is shown that several primal-dual-type algorithms from the literature [7, 10, 14, 21] can be embedded in the algorithm designed for the situation where $\mathscr{G}=\mathscr{X},L_2=-\operatorname{Id}$ and $d=0$ and, consequently, in the general algorithm considered above.

1.3.2 Convergence analysis

An important ingredient for our convergence analysis will be the following version of the Opial Lemma (see [13, Theorem 3.3]).

Lemma 1. *Let C be a nonempty subset of $\mathscr{H}$ and $(x^k)_{k\geq 0}$ be a sequence in $\mathscr{H}$. Let $\alpha>0$ and $W^k\in\mathscr{P}_\alpha(\mathscr{H})$ be such that $W^k\succcurlyeq W^{k+1}$ for all $k\geq 0$. Assume that:*
(i) for all $z\in C$ and for all $k\geq 0$: $\|x^{k+1}-z\|_{W^{k+1}}\leq\|x^k-z\|_{W^k}$.
(ii) every weak sequential cluster point of $(x^k)_{k\geq 0}$ belongs to C.
Then $(x^k)_{k\geq 0}$ converges weakly to an element in C.

The following theorem is the main result of this section.

Theorem 1. *In the context of Problem 1, assume that the set of KKT points of the variational inequality with monotone operators* (1.3) *is nonempty and that* $M_1^k - \frac{\eta}{2}\operatorname{Id} \in \mathscr{S}_+(\mathscr{H})$, $M_1^k \succeq M_1^{k+1}$, $M_2^k \in \mathscr{S}_+(\mathscr{G})$, $M_2^k \succeq M_2^{k+1}$, *and* $M_2^k + cL_2^*L_2 \in \mathscr{S}_+(\mathscr{G})$ *for all* $k \geq 0$. *Let* $(x^k, z^k, y^k)_{k\geq 0}$ *be the sequence generated by Algorithm 1. Suppose that one of the following assumptions is fulfilled:*
(I) there exist $\alpha_1, \beta_1 > 0$ *such that* $M_1^k - \frac{\eta}{2}\operatorname{Id} \in \mathscr{P}_{\alpha_1}(\mathscr{H})$ *and* $M_2^k + cL_2^*L_2 \in \mathscr{P}_{\beta_1}(\mathscr{G})$ *for all* $k \geq 0$*;*
(II) there exist $\alpha_2, \beta_2 > 0$ *such that* $L_1^*L_1 \in \mathscr{P}_{\alpha_2}(\mathscr{H})$ *and* $M_2^k \in \mathscr{P}_{\beta_2}(\mathscr{G})$ *for all* $k \geq 0$*;*
(III) there exist $\alpha_3, \beta_3 > 0$ *such that* $M_1^k - \frac{\eta}{2}\operatorname{Id} + cL_1^*L_1 \in \mathscr{P}_{\alpha_3}(\mathscr{H})$, $L_2^*L_2 \in \mathscr{P}_{\beta_3}(\mathscr{G})$ *and* $2M_2^{k+1} \succeq M_2^k \succeq M_2^{k+1}$ *for all* $k \geq 0$.
Then $(x^k, z^k, y^k)_{k\geq 0}$ *converges weakly to a KKT point of the variational inequality* (1.3).

Proof. Let (x^*, z^*, y^*) be a KKT point of the variational inequality with monotone operators (1.3). Then

$$-L_1^*y^* - Cx^* \in Ax^*, \quad -L_2^*y^* \in Bz^* \ \text{ and } \ L_1x^* + L_2z^* = d.$$

Let $k \geq 0$ be fixed. By (1.14), (1.15) and the monotonicity of A and B, we obtain the inequalities

$$\langle -cL_1^*(L_1x^{k+1} + L_2z^k - d + c^{-1}y^k) + M_1^k(x^k - x^{k+1}) - Cx^k + L_1^*y^* + Cx^*, x^{k+1} - x^*\rangle \geq 0$$

and

$$\langle -cL_2^*(L_1x^{k+1} + L_2z^{k+1} - d + c^{-1}y^k) + M_2^k(z^k - z^{k+1}) + L_2^*y^*, z^{k+1} - z^*\rangle \geq 0.$$

Since C is η-cocoercive, we have

$$\eta\langle Cx^* - Cx^k, x^* - x^k\rangle \geq \|Cx^* - Cx^k\|^2.$$

We consider first the case when $\eta > 0$. Summing up the three inequalities from above we get

$$\begin{aligned}
c\langle -L_1x^{k+1} - L_2z^k + d, L_1x^{k+1} - L_1x^*\rangle + \langle y^* - y^k, L_1x^{k+1} - L_1x^*\rangle \\
+\langle Cx^* - Cx^k, x^{k+1} - x^*\rangle + \langle M_1^k(x^k - x^{k+1}), x^{k+1} - x^*\rangle \\
+c\langle L_2^*(-L_1x^{k+1} - L_2z^{k+1} + d), z^{k+1} - z^*\rangle + \langle -L_2^*y^k + L_2^*y^*, z^{k+1} - z^*\rangle \\
+\langle M_2^k(z^k - z^{k+1}), z^{k+1} - z^*\rangle + \langle Cx^* - Cx^k, x^* - x^k\rangle - \eta^{-1}\|Cx^* - Cx^k\|^2 \geq 0.
\end{aligned}$$

By taking into account (1.13) we also obtain

$$\begin{aligned}
&\langle y^* - y^k, L_1x^{k+1} - L_1x^* \rangle + \langle -L_2^* y^k + L_2^* y^*, z^{k+1} - z^* \rangle \\
&= \langle y^* - y^k, L_1x^{k+1} - L_1x^* \rangle + \langle y^* - y^k, L_2(z^{k+1} - z^*) \rangle \\
&= \langle y^* - y^k, L_1x^{k+1} + L_2z^{k+1} - \underbrace{(L_1x^* + L_2z^*)}_{=d} \rangle \\
&= c^{-1} \langle y^* - y^k, y^{k+1} - y^k \rangle.
\end{aligned}$$

Hence the above inequality reads as

$$\begin{aligned}
c\langle (d - L_2z^k) - L_1x^{k+1}, L_1x^{k+1} - L_1x^* \rangle + c^{-1} \langle y^* - y^k, y^{k+1} - y^k \rangle \\
+ \langle Cx^* - Cx^k, x^{k+1} - x^k \rangle + \langle M_1^k (x^k - x^{k+1}), x^{k+1} - x^* \rangle \\
+ c\langle (d - L_1x^{k+1}) - L_2z^{k+1}, L_2z^{k+1} - L_2z^* \rangle + \langle M_2^k (z^k - z^{k+1}), z^{k+1} - z^* \rangle \\
- \eta^{-1} \|Cx^* - Cx^k\|^2 \geq 0.
\end{aligned}$$

By expressing the inner products through norms the above inequality becomes

$$\begin{aligned}
\frac{c}{2} \left(\|(d - L_2z^k) - L_1x^*\|^2 - \|(d - L_2z^k) - L_1x^{k+1}\|^2 - \|L_1x^{k+1} - L_1x^*\|^2 \right) \\
+ \frac{c}{2} \left(\|(d - L_1x^{k+1}) - L_2z^*\|^2 - \|(d - L_1x^{k+1}) - L_2z^{k+1}\|^2 - \|L_2z^{k+1} - L_2z^*\|^2 \right) \\
+ \frac{1}{2c} \left(\|y^* - y^k\|^2 + \|y^{k+1} - y^k\|^2 - \|y^{k+1} - y^*\|^2 \right) \\
+ \frac{1}{2} \left(\|x^k - x^*\|^2_{M_1^k} - \|x^k - x^{k+1}\|^2_{M_1^k} - \|x^{k+1} - x^*\|^2_{M_1^k} \right) \\
+ \frac{1}{2} \left(\|z^k - z^*\|^2_{M_2^k} - \|z^k - z^{k+1}\|^2_{M_2^k} - \|z^{k+1} - z^*\|^2_{M_2^k} \right) \\
+ \langle Cx^* - Cx^k, x^{k+1} - x^k \rangle - \eta^{-1} \|Cx^* - Cx^k\|^2 \geq 0.
\end{aligned}$$

By using again that $y^{k+1} = y^k + c(L_1x^{k+1} + L_2z^{k+1} - d)$ and by taking into account that

$$\begin{aligned}
\langle Cx^* - Cx^k, x^{k+1} - x^k \rangle - \eta^{-1} \|Cx^* - Cx^k\|^2 = \\
-\eta \left\| \eta^{-1}(Cx^* - Cx^k) + \frac{1}{2}(x^k - x^{k+1}) \right\|^2 + \frac{\eta}{4} \|x^k - x^{k+1}\|^2,
\end{aligned}$$

we obtain

$$
\begin{aligned}
\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^{k+1}-z^*\|^2_{M_2^k}+\frac{1}{2}\|L_2z^{k+1}-L_2z^*\|^2_{c\,\mathrm{Id}}+\frac{1}{2c}\|y^{k+1}-y^*\|^2 \leq \\
\frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-z^*\|^2_{M_2^k}+\frac{1}{2}\|(d-L_2z^k)-L_1x^*\|^2_{c\,\mathrm{Id}}+\frac{1}{2c}\|y^*-y^k\|^2 \\
-\frac{c}{2}\|(d-L_2z^k)-L_1x^{k+1}\|^2-\frac{1}{2}\|x^k-x^{k+1}\|^2_{M_1^k-\frac{\eta}{2}\mathrm{Id}}-\frac{1}{2}\|z^k-z^{k+1}\|^2_{M_2^k} \\
-\eta\left\|\eta^{-1}(Cx^*-Cx^k)+\frac{1}{2}(x^k-x^{k+1})\right\|^2.
\end{aligned}
$$

Since $(d-L_2z^k)-L_1x^*=-L_2z^k+L_2z^*$ and by using the monotonicity assumptions on $(M_1^k)_{k\geq 0}$ and $(M_2^k)_{k\geq 0}$ it yields

$$
\begin{aligned}
\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-z^*\|^2_{M_2^{k+1}+cL_2^*L_2}+\frac{1}{2c}\|y^{k+1}-y^*\|^2 \leq \\
\frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-z^*\|^2_{M_2^k+cL_2^*L_2}+\frac{1}{2c}\|y^*-y^k\|^2 \\
-\frac{c}{2}\|L_1x^{k+1}+L_2z^k-d\|^2-\frac{1}{2}\|x^k-x^{k+1}\|^2_{M_1^k-\frac{\eta}{2}\mathrm{Id}}-\frac{1}{2}\|z^k-z^{k+1}\|^2_{M_2^k} \\
-\eta^{-1}\left\|\eta(Cx^*-Cx^k)+\frac{1}{2}(x^k-x^{k+1})\right\|^2. \qquad (1.19)
\end{aligned}
$$

In the case when $\eta=0$, by repeating the above calculations, we obtain

$$
\begin{aligned}
\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-z^*\|^2_{M_2^{k+1}+cL_2^*L_2}+\frac{1}{2c}\|y^{k+1}-y^*\|^2 \leq \\
\frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-z^*\|^2_{M_2^k+cL_2^*L_2}+\frac{1}{2c}\|y^*-y^k\|^2 \\
-\frac{c}{2}\|L_1x^{k+1}+L_2z^k-d\|^2-\frac{1}{2}\|x^k-x^{k+1}\|^2_{M_1^k}-\frac{1}{2}\|z^k-z^{k+1}\|^2_{M_2^k}. \qquad (1.20)
\end{aligned}
$$

By using arguments involving telescoping sums, each of the inequalities (1.19) and (1.20) yield

$$
\begin{aligned}
&\sum_{k\geq 0}\|L_1x^{k+1}+L_2z^k-d\|^2<+\infty,\ \sum_{k\geq 0}\|x^k-x^{k+1}\|^2_{M_1^k-\frac{\eta}{2}\mathrm{Id}}<+\infty, \\
&\sum_{k\geq 0}\|z^k-z^{k+1}\|^2_{M_2^k}<+\infty. \qquad (1.21)
\end{aligned}
$$

Assume that condition (I) holds. By neglecting the negative terms (notice that $M_1^k-\frac{\eta}{2}\mathrm{Id}\in\mathscr{S}_+(\mathscr{H})$ for all $k\geq 0$), from each of the inequalities (1.19) and (1.20) it follows that assumption (i) the Opial Lemma holds, when applied in the product space $\mathscr{H}\times\mathscr{G}\times\mathscr{Z}$, for the sequence $(x^k,z^k,y^k)_{k\geq 0}$, for $W^k:=(M_1^k,M_2^k+cL_2^*L_2,c^{-1}\mathrm{Id})$ for $k\geq 0$, and for $C\subseteq\mathscr{H}\times\mathscr{G}\times\mathscr{Z}$ the set of KKT points of the variational inequality (1.3).

Since $M_1^k-\frac{\eta}{2}\mathrm{Id}\in\mathscr{P}_{\alpha_1}(\mathscr{H})$ for all $k\geq 0$ with $\alpha_1>0$, we get from (1.21)

$$x^k - x^{k+1} \to 0 \ (k \to +\infty) \tag{1.22}$$

and

$$L_1 x^{k+1} + L_2 z^k - d \to 0 \ (k \to +\infty). \tag{1.23}$$

Therefore

$$\begin{aligned}\|z^{k+1} - z^k\|_{L_2^* L_2} =& \|L_2 z^{k+1} - L_2 z^k\| \\ \leq& \|L_1 x^{k+2} + L_2 z^{k+1} - d\| + \|L_1 x^{k+1} + L_2 z^k - d\| \\ &+ \|L_1 x^{k+1} - L_1 x^{k+2}\|,\end{aligned}$$

which means that

$$\|z^{k+1} - z^k\|_{L_2^* L_2} \to 0 \ (k \to +\infty).$$

Using the third condition in (1.21) and the fact that $M_2^k + cL_2^* L_2 \in \mathscr{P}_{\beta_1}(\mathscr{G})$ we conclude

$$z^k - z^{k+1} \to 0 \ (k \to +\infty). \tag{1.24}$$

From (1.13) we derive

$$\begin{aligned}\|y^k - y^{k+1}\| &= c\|L_1 x^{k+1} + L_2 z^{k+1} - d\| \\ &\leq c\left(\|L_1 x^{k+1} + L_2 z^k - d\| + \|L_2 z^{k+1} - L_2 z^k\|\right),\end{aligned}$$

hence, by (1.23) and (1.24)

$$y^k - y^{k+1} \to 0 \ (k \to +\infty). \tag{1.25}$$

Now we are able to verify the second assumption in the Opial Lemma for C taken as the set of KKT points of (1.3). Let $(\bar{x}, \bar{z}, \bar{y}) \in \mathscr{H} \times \mathscr{G} \times \mathscr{Z}$ be such that there exists $(k_n)_{n \geq 0}$, $k_n \to +\infty$ (as $n \to +\infty$), and $(x^{k_n}, z^{k_n}, y^{k_n})$ converges weakly to $(\bar{x}, \bar{z}, \bar{y})$ (as $n \to +\infty$). From (1.22) and the linearity of L_1 we obtain that $(L_1 x^{k_n+1} + L_2 z^{k_n})_{n \geq 0}$ converges weakly to $L_1 \bar{x} + L_2 \bar{z}$ (as $n \to +\infty$), which combined with (1.23) yields $L_1 \bar{x} + L_2 \bar{z} = d$. For $n \geq 0$, we use now the following notations

$$\begin{aligned}
a_n^* &:= cL_1^*(-L_1x^{k_n+1} - L_2z^{k_n} + d) + L_1^*(y^{k_n+1} - y^{k_n}) \\
&\quad + M_1^{k_n}(x^{k_n} - x^{k_n+1}) + Cx^{k_n+1} - Cx^{k_n} \\
a_n &:= x^{k_n+1} \\
b_n^* &:= M_2^{k_n}(z^{k_n} - z^{k_n+1}) \\
b_n &:= z^{k_n+1} \\
c_n^* &:= -L_1x^{k_n+1} - L_2z^{k_n+1} + d \\
c_n &:= y^{k_n+1}.
\end{aligned}$$

From (1.14) we have

$$a_n^* \in (A+C)a_n + L_1^*c_n \tag{1.26}$$

and by combining (1.15) with (1.13) we obtain

$$b_n^* \in Bb_n + L_2^*c_n \tag{1.27}$$

for all $n \geq 0$. From (1.22), (1.24) and (1.25) we have that

$$(a_n, b_n, c_n) \text{ converges weakly to } (\bar{x}, \bar{z}, \bar{y}) \ \text{ (as } n \to +\infty). \tag{1.28}$$

Moreover, by (1.22) - (1.25) and the Lipschitz continuity of C we obtain

$$(a_n^*, b_n^*, c_n^*) \text{ converges strongly to } (0,0,0) \ \text{ (as } n \to +\infty). \tag{1.29}$$

Next we define the maximally monotone operator

$$T : \mathscr{H} \times \mathscr{G} \times \mathscr{Z} \rightrightarrows \mathscr{H} \times \mathscr{G} \times \mathscr{Z}, T(x,z,y) := ((A+C)x, Bz, 0),$$

and the linear continuous operator

$$\tilde{K} : \mathscr{H} \times \mathscr{G} \times \mathscr{Z} \to \mathscr{H} \times \mathscr{G} \times \mathscr{Z}, \tilde{K}(x,z,y) := (L_1^*y, L_2^*y, -L_1x - L_2z).$$

For all $(x,z,y) \in \mathscr{H} \times \mathscr{G} \times \mathscr{Z}$ we have

$$\begin{aligned}
\langle \tilde{K}(x,z,y), (x,z,y) \rangle &= \langle L_1^*y, x \rangle + \langle L_2^*y, z \rangle + \langle -L_1x - L_2z, y \rangle \\
&= \langle y, L_1x \rangle + \langle y, L_2z \rangle - \langle L_1x, y \rangle - \langle L_2z, y \rangle = 0,
\end{aligned}$$

hence $\tilde{K}$ is maximally monotone and therefore the shifted operator

$$K : \mathscr{H} \times \mathscr{G} \times \mathscr{Z} \to \mathscr{H} \times \mathscr{G} \times \mathscr{Z}, K(x,y,z) := \tilde{K}(x,y,z) + (0,0,d),$$

is maximally monotone, as well. Since K has full domain we obtain that

$$T + K \text{ is a maximally monotone operator.} \tag{1.30}$$

On the other hand, from (1.26) and (1.27) we have that

$$((a_n,b_n,c_n),(a_n^*,b_n^*,c_n^*)) \in \text{gra}(T+K)\ \forall n \geq 0. \tag{1.31}$$

Since the graph of a maximally monotone operator is sequentially closed with respect to the weak×strong topology (see [4, Proposition 20.33]), from (1.28), (1.29), (1.30) and (1.31) we derive that

$$((\bar{x},\bar{z},\bar{y}),(0,0,0)) \in \text{gra}(T+K),$$

which is equivalent to

$$(0,0,0) \in ((A+C)\bar{x}+L_1^*\bar{y}, B\bar{z}+L_2^*\bar{y}, -L_1\bar{x}-L_2\bar{z}+d).$$

The latter means nothing else than saying that $(\bar{x},\bar{z},\bar{y})$ is a KKT point of (1.3), thus assumption (ii) in the Opial Lemma is verified, too. In conclusion, $(x^k,z^k,y^k)_{k\geq 0}$ converges weakly to a KKT point of (1.3).

Consider now the situation in assumption (II). From (1.19) and (1.20) it follows that (1.23) and (1.24) hold. From (1.13), (1.23) and (1.24) we obtain (1.25). Finally, by using that $L_1^*L_1 \in \mathscr{P}_{\alpha_2}(\mathscr{H})$ for $\alpha_2 > 0$, relation (1.22) holds, too.

On the other hand, (1.19) and (1.20) yield that

$$\exists \lim_{k\to+\infty}\left(\frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-z^*\|^2_{M_2^k+cL_2^*L_2}+\frac{1}{2c}\|y^k-y^*\|^2\right), \tag{1.32}$$

hence $(y^k)_{k\geq 0}$ and $(z^k)_{k\geq 0}$ are bounded. Combining this with

$$\begin{aligned}\alpha_2\|x^k-x^*\|^2 &\leq \|L_1x^k-L_1x^*\|^2\\ &\leq \frac{1}{3}\|L_1x^k+L_2z^k-d\|^2+\frac{1}{3}\|L_1x^*+L_2z^*-d\|^2+\frac{1}{3}\|L_2z^*-L_2z^k\|^2,\end{aligned}$$

which holds for all $k \geq 0$, and using (1.13), we derive that $(x^k)_{k\geq 0}$ is bounded, too. Hence there exists a weakly convergent subsequence of $(x^k,z^k,y^k)_{k\geq 0}$. By using the same arguments as in the second part of the proof of (I) it follows that every weak sequential cluster point of $(x^k,z^k,y^k)_{k\geq 0}$ is a KKT point of (1.3).

Now we show that the set of weak sequential cluster points of $(x^k,z^k,y^k)_{k\geq 0}$ is a singleton. Let (x_1,z_1,y_1), (x_2,z_2,y_2) be two such weak sequential cluster points. Then there exist $(k_p)_{p\geq 0}$, $(k_q)_{q\geq 0}$, $k_p \to +\infty$ (as $p \to +\infty$), $k_q \to +\infty$ (as $q \to +\infty$), a subsequence $(x^{k_p},z^{k_p},y^{k_p})_{p\geq 0}$ which converges weakly to (x_1,z_1,y_1) (as $p \to +\infty$), and a subsequence $(x^{k_q},z^{k_q},y^{k_q})_{q\geq 0}$ which converges weakly to (x_2,z_2,y_2) (as $q \to +\infty$). As seen above, (x_1,z_1,y_1) and (x_2,z_2,y_2) are KKT points of (1.3). , thus $L_1x_i + L_2z_i = d$ for $i \in \{1,2\}$. From (1.32), which is true for every KKT point of (1.3), we derive

$$\exists \lim_{k\to+\infty}\Big(E(x^k,z^k,y^k;x_1,z_1,y_1)-E(x^k,z^k,y^k;x_2,z_2,y_2)\Big), \tag{1.33}$$

where

$$E(x^k,z^k,y^k;x,z,y) := \frac{1}{2}\|x^k - x\|^2_{M_1^k} + \frac{1}{2}\|z^k - z\|^2_{M_2^k + cL_2^*L_2} + \frac{1}{2c}\|y^k - y\|^2.$$

We have for all $k \geq 0$

$$\frac{1}{2}\|x^k - x_1\|^2_{M_1^k} - \frac{1}{2}\|x^k - x_2\|^2_{M_1^k} = \frac{1}{2}\|x_2 - x_1\|^2_{M_1^k} + \langle x^k - x_2, M_1^k(x_2 - x_1)\rangle,$$

$$\begin{aligned} &\frac{1}{2}\|z^k - z_1\|^2_{M_2^k + cL_2^*L_2} - \frac{1}{2}\|z^k - z_2\|^2_{M_2^k + cL_2^*L_2} \\ =\ &\frac{1}{2}\|z_2 - z_1\|^2_{M_2^k + cL_2^*L_2} + \langle z^k - z_2, (M_2^k + cL_2^*L_2)(z_2 - z_1)\rangle \end{aligned}$$

and

$$\frac{1}{2c}\|y^k - y_1\|^2 - \frac{1}{2c}\|y^k - y_2\|^2 = \frac{1}{2c}\|y_2 - y_1\|^2 + \frac{1}{c}\langle y^k - y_2, y_2 - y_1\rangle.$$

According to [18, Théorème 104.1] there exists $M_1 \in \mathscr{S}_+(\mathscr{H})$ such that $(M_1^k)_{k\geq 0}$ converges pointwise to M_1 in the strong topology (as $k \to +\infty$). Similarly, the monotonicity condition imposed on $(M_2^k)_{k\geq 0}$ implies that $\sup_{k\geq 0}\|M_2^k + cL_2^*L_2\| < +\infty$. Thus, according to [13, Lemma 2.3], there exists $\alpha' > 0$ and $M_2 \in \mathscr{P}_{\alpha'}(\mathscr{G})$ such that $(M_2^k + cL_2^*L_2)_{k\geq 0}$ converges pointwise to M_2 in the strong topology (as $k \to +\infty$). Taking the limit in (1.33) along the subsequences $(k_p)_{p\geq 0}$ and $(k_q)_{q\geq 0}$ and using the last three identities above, we obtain

$$\begin{aligned} &\frac{1}{2}\|x_1 - x_2\|^2_{M_1} + \langle x_1 - x_2, M_1(x_2 - x_1)\rangle + \frac{1}{2}\|z_1 - z_2\|^2_{M_2} + \langle z_1 - z_2, M_2(z_2 - z_1)\rangle \\ &+ \frac{1}{2c}\|y_1 - y_2\|^2 + \frac{1}{c}\langle y_1 - y_2, y_2 - y_1\rangle \\ =&\frac{1}{2}\|x_1 - x_2\|^2_{M_1} + \frac{1}{2}\|z_1 - z_2\|^2_{M_2} + \frac{1}{2c}\|y_1 - y_2\|^2, \end{aligned}$$

hence

$$-\|x_1 - x_2\|^2_{M_1} - \|z_1 - z_2\|^2_{M_2} - \frac{1}{c}\|y_1 - y_2\|^2 = 0,$$

thus $z_1 = z_2$ and $y_1 = y_2$. Further, since $L_1x_i + L_2z_i = d$ for $i \in \{1,2\}$,

$$\begin{aligned} \alpha_2\|x_1 - x_2\|^2 &\leq \|L_1x_1 - L_1x_2\|^2 \\ &\leq \frac{1}{3}\|L_1x_1 + L_2z_1 - d\|^2 + \frac{1}{3}\|L_1x_2 + L_2z_2 - d\|^2 + \frac{1}{3}\|L_2z_1 - L_2z_2\|^2 \\ &= 0, \end{aligned}$$

thus $x_1 = x_2$. In conclusion, $(x^k, z^k, y^k)_{k\geq 0}$ converges weakly to a KKT point of (1.3).

Finally, we consider the situation when the hypotheses in assumption (III) hold. Let $k \geq 1$ be fixed. Combining (1.15) with (1.13) gives

$$-L_2^* y^{k+1} + M_2^k (z^k - z^{k+1}) \in B z^{k+1}.$$

Considering this monotone inclusion for consecutive iterates and by taking into account the monotonicity of B, we obtain

$$\langle z^{k+1} - z^k, -L_2^*(y^{k+1} - y^k) + M_2^k (z^k - z^{k+1}) - M_2^{k-1}(z^{k-1} - z^k) \rangle \geq 0,$$

hence

$$\begin{aligned} &\langle z^{k+1} - z^k, -L_2^*(y^{k+1} - y^k) \rangle \\ \geq\ & \|z^{k+1} - z^k\|^2_{M_2^k} + \langle z^{k+1} - z^k, M_2^{k-1}(z^{k-1} - z^k) \rangle \\ \geq\ & \|z^{k+1} - z^k\|^2_{M_2^k} - \frac{1}{2}\|z^{k+1} - z^k\|^2_{M_2^{k-1}} - \frac{1}{2}\|z^k - z^{k-1}\|^2_{M_2^{k-1}}. \end{aligned}$$

Using that $y^{k+1} - y^k = c(L_1 x^{k+1} + L_2 z^{k+1} - d)$, the last inequality yields

$$\begin{aligned} &\|z^{k+1} - z^k\|^2_{M_2^k} - \frac{1}{2}\|z^{k+1} - z^k\|^2_{M_2^{k-1}} - \frac{1}{2}\|z^k - z^{k-1}\|^2_{M_2^{k-1}} \\ \leq\ & c\langle L_2 z^{k+1} - L_2 z^k, -L_1 x^{k+1} - L_2 z^{k+1} + d \rangle \\ =\ & \frac{c}{2}\left(\|L_1 x^{k+1} + L_2 z^k - d\|^2 - \|L_2 z^{k+1} - L_2 z^k\|^2 - \|L_1 x^{k+1} + L_2 z^{k+1} - d\|^2 \right), \end{aligned}$$

which, after adding it with (1.19) and using (1.13), leads to

$$\begin{aligned} &\frac{1}{2}\|x^{k+1} - x^*\|^2_{M_1^{k+1}} + \frac{1}{2}\|z^{k+1} - z^*\|^2_{M_2^{k+1} + cL_2^* L_2} + \frac{1}{2c}\|y^{k+1} - y^*\|^2 + \\ &\frac{1}{2}\|z^{k+1} - z^k\|^2_{3M_2^k - M_2^{k-1}} \\ \leq\ & \frac{1}{2}\|x^k - x^*\|^2_{M_1^k} + \frac{1}{2}\|z^k - z^*\|^2_{M_2^k + cL_2^* L_2} + \frac{1}{2c}\|y^* - y^k\|^2 + \frac{1}{2}\|z^k - z^{k-1}\|^2_{M_2^{k-1}} - \\ &\frac{1}{2}\|x^k - x^{k+1}\|^2_{M_1^k - \frac{\eta}{2}\operatorname{Id}} - \frac{c}{2}\|L_2 z^{k+1} - L_2 z^k\|^2 - \frac{1}{2c}\|y^{k+1} - y^k\|^2 - \\ &\eta\|\eta^{-1}(Cx^* - Cx^k) + \frac{1}{2}(x^k - x^{k+1})\|^2. \end{aligned}$$

Taking into account that according to (III) we have $3M_2^k - M_2^{k-1} \succeq M_2^k$, we can conclude that for all $k \geq 1$ it holds

$$\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-z^*\|^2_{M_2^{k+1}+cL_2^*L_2}+\frac{1}{2c}\|y^{k+1}-y^*\|^2+\frac{1}{2}\|z^{k+1}-z^k\|^2_{M_2^k}$$
$$\leq \frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-z^*\|^2_{M_2^k+cL_2^*L_2}+\frac{1}{2c}\|y^*-y^k\|^2-\frac{1}{2}\|x^k-x^{k+1}\|^2_{M_1^k-\frac{\eta}{2}\operatorname{Id}}-$$
$$\frac{1}{2}\|z^{k+1}-z^k\|^2_{cL_2^*L_2}-\frac{1}{2c}\|y^{k+1}-y^k\|^2+\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}-$$
$$\eta^{-1}\|\eta(Cx^*-Cx^k)+\frac{1}{2}(x^k-x^{k+1})\|^2, \tag{1.34}$$

while, by using when $\eta=0$ (1.20) instead of (1.19), it yields

$$\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-z^*\|^2_{M_2^{k+1}+cL_2^*L_2}+\frac{1}{2c}\|y^{k+1}-y^*\|^2+\frac{1}{2}\|z^{k+1}-z^k\|^2_{M_2^k}$$
$$\leq \frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-z^*\|^2_{M_2^k+cL_2^*L_2}+\frac{1}{2c}\|y^*-y^k\|^2+\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}-$$
$$\frac{1}{2}\|x^k-x^{k+1}\|^2_{M_1^k}-\frac{1}{2}\|z^{k+1}-z^k\|^2_{cL_2^*L_2}-\frac{1}{2c}\|y^{k+1}-y^k\|^2. \tag{1.35}$$

Using telescoping sum arguments, we obtain that $\|x^k-x^{k+1}\|^2_{M_1^k-\frac{\eta}{2}\operatorname{Id}}\to 0$, $y^k-y^{k+1}\to 0$ and $z^k-z^{k+1}\to 0$ as $k\to+\infty$. Using (1.13), it follows that $L_1(x^k-x^{k+1})\to 0$ as $k\to+\infty$, which, combined with the hypotheses imposed on $M_1^k-\frac{\eta}{2}\operatorname{Id}+cL_1^*L_1$, implies that $x^k-x^{k+1}\to 0$ as $k\to+\infty$. Consequently, $L_1x^{k+1}+L_2z^k-d\to 0$ as $k\to+\infty$. Hence the relations (1.22) - (1.25) are fulfilled. On the other hand, from (1.34) and (1.35) if follows that the limit

$$\lim_{k\to+\infty}\left(\frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-z^*\|^2_{M_2^k+cL_2^*L_2}+\frac{1}{2c}\|y^k-y^*\|^2+\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}\right)$$

exists. By using that

$$\|z^k-z^{k-1}\|^2_{M_2^{k-1}}\leq\|z^k-z^{k-1}\|^2_{M_2^0}\leq\|M_2^0\|\|z^k-z^{k-1}\|^2 \quad \forall k\geq 1,$$

it follows that $\lim_{k\to+\infty}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}=0$, which further implies that (1.32) holds. From here the conclusion follows by arguing as in the second part of the proof provided in the setting of assumption (II). □

1.4 Convergence rates in the case when $A+C$ is strongly monotone

In this section we adress the following modification of the Problem 1.

Problem 3. In the context of Problem 1, we replace the cocoercivity of C by the assumptions that C is monotone and μ-Lipschitz continuous, for $\mu\geq 0$. Further, we assume that the sum $A+C$ is γ-strongly monotone for $\gamma>0$, and that $d=0$.

We have the following characterization for a KKT point of (1.3):

$$\exists (x,z,y)\in\mathscr{H}\times\mathscr{G}\times\mathscr{Z}: \begin{cases} -L_1^*y & \in Ax+Cx\\ -L_2^*y & \in Bz\\ L_1x & = -L_2z\end{cases}$$

$$\Leftrightarrow\ \exists (x,z,y)\in\mathscr{H}\times\mathscr{G}\times\mathscr{Z}: \begin{cases} -L_1^*y & \in Ax+Cx\\ z & \in B^{-1}\circ(-L_2^*)y\\ L_1x & = -L_2z\end{cases}$$

$$\Leftrightarrow\ \exists (x,y)\in\mathscr{H}\times\mathscr{Z}: \begin{cases} -L_1^*y & \in Ax+Cx\\ L_1x & \in (-L_2)\circ B^{-1}\circ(-L_2^*)y.\end{cases}$$

The latter means that (x,y) is a so-called primal-dual solution associated to the monotone inclusion problem

$$\text{find } x\in\mathscr{H} \text{ such that } 0\in Ax+Cx+(L_1^*\circ\bar{B}\circ L_1)(x),$$

and its Attouch-Thera dual inclusion problem, where $\bar{B}:=[(-L_2)\circ B^{-1}\circ(-L_2^*)]^{-1}$. Algorithm 14 in [6] which is designed to determine these primal-dual solutions in a setting which is similar to the one in Problem 3, gives rise to the following iterative scheme.

Algorithm 2 *For all $k\geq 0$, let $M_2^k:\mathscr{Z}\to\mathscr{Z}$ be a linear, continuous and self-adjoint operator such that $\tau_kL_1L_1^*+M_2^k\in\mathscr{P}_{\alpha_k}(\mathscr{Z})$, for $\alpha_k>0$. Choose $(x^0,z^0,y^0)\in\mathscr{H}\times\mathscr{G}\times\mathscr{Z}$. For all $k\geq 0$ generate the sequence $(x^k,z^k,y^k)_{k\geq 0}$ as follows:*

$$y^{k+1}=\left(\tau_kL_1L_1^*+M_2^k+(-L_2)\circ B^{-1}\circ(-L_2^*)\right)^{-1}[-\tau_kL_1(z^k-\tau_k^{-1}x^k)+M_2^ky^k] \tag{1.36}$$

$$z^{k+1}=\left(\frac{\theta_k}{\lambda}-1\right)L_1^*y^{k+1}+\frac{\theta_k}{\lambda}Cx^k + \frac{\theta_k}{\lambda}(\mathrm{Id}+\lambda\tau_{k+1}^{-1}A^{-1})^{-1}(-L_1^*y^{k+1}+\lambda\tau_{k+1}^{-1}x^k-Cx^k) \tag{1.37}$$

$$x^{k+1}=x^k+\frac{\tau_{k+1}}{\theta_k}(-L_1^*y^{k+1}-z^{k+1}), \tag{1.38}$$

where $\lambda,\tau_k,\theta_k>0$.

In case $\mathscr{G}=\mathscr{Z}$ and the linear continuous operator $L_2:\mathscr{G}\to\mathscr{G}$ is invertible, we obtain the following full splitting formulation for Algorithm 2.

Algorithm 3 *For all $k\geq 0$, let $M_2^k:\mathscr{G}\to\mathscr{G}$ be a linear, continuous and self-adjoint operator such that $\tau_kL_2^{-1}L_1(L_2^{-1}L_1)^*+L_2^{-1}M_2^k(L_2^*)^{-1}\in\mathscr{P}_{\alpha_k}(\mathscr{Z})$ for $\alpha_k>0$. Choose $(x^0,z^0,y^0)\in\mathscr{H}\times\mathscr{G}\times\mathscr{G}$. For all $k\geq 0$ generate the sequence $(x^k,z^k,y^k)_{k\geq 0}$ as follows:*

$$y^{k+1} = (-L_2^*)^{-1} \circ \left(\tau_k L_2^{-1} L_1 (L_2^{-1} L_1)^* + L_2^{-1} M_2^k (L_2^*)^{-1} + B^{-1} \right)^{-1} \circ \\ (-L_2)^{-1} [-\tau_k L_1 (z^k - \tau_k^{-1} x^k) + M_2^k y^k] \tag{1.39}$$

$$z^{k+1} = \left(\frac{\theta_k}{\lambda} - 1 \right) L_1^* y^{k+1} + \frac{\theta_k}{\lambda} C x^k + \\ \frac{\theta_k}{\lambda} (\mathrm{Id} + \lambda \tau_{k+1}^{-1} A^{-1})^{-1} (-L_1^* y^{k+1} + \lambda \tau_{k+1}^{-1} x^k - C x^k) \tag{1.40}$$

$$x^{k+1} = x^k + \frac{\tau_{k+1}}{\theta_k} (-L_1^* y^{k+1} - z^{k+1}), \tag{1.41}$$

where $\lambda, \tau_k, \theta_k > 0$.

Concerning the parameters involved in Algorithm 2, we assume that

$$\mu \tau_1 < 2\gamma, \ \lambda \geq \mu + 1, \tag{1.42}$$

that there exists $\sigma_0 > 0$ such that

$$\sigma_0 \tau_1 \|L_1\|^2 \leq 1, \tag{1.43}$$

and that for all $k \geq 0$

$$\theta_k = \frac{1}{\sqrt{1 + \tau_{k+1} \lambda^{-1} (2\gamma - \mu \tau_{k+1})}} \tag{1.44}$$

$$\tau_{k+2} = \theta_k \tau_{k+1} \tag{1.45}$$

$$\sigma_{k+1} = \theta_k^{-1} \sigma_k \tag{1.46}$$

$$\tau_k L_1 L_1^* + M_2^k \succeq \sigma_k^{-1} \mathrm{Id} \tag{1.47}$$

$$\frac{\tau_k}{\tau_{k+1}} L_1 L_1^* + \frac{1}{\tau_{k+1}} M_2^k \succeq \frac{\tau_{k+1}}{\tau_{k+2}} L_1 L_1^* + \frac{1}{\tau_{k+2}} M_2^{k+1}. \tag{1.48}$$

The following convergence rate result follows from [6, Theorem 19].

Theorem 2. *Consider the setting of Problem 3 in the hypothesis* $(-L_2) \circ B^{-1} \circ (-L_2^*)$ *is maximally monotone. Let* (x, z, y) *be a KKT point of the variational inequality* (1.3). *Let* $(x^k, z^k, y^k)_{k \geq 0}$ *be the sequence generated by Algorithm 2 and assume that the relations* (1.42) - (1.48) *are fulfilled. Then we have for all* $n \geq 2$

$$\frac{\lambda \|x^n - x\|^2}{\tau_{n+1}^2} + \frac{1 - \sigma_0 \tau_1 \|L_1\|^2}{\sigma_0 \tau_1} \|y^n - y\|^2 \leq \\ \frac{\lambda \|x^1 - x\|^2}{\tau_2^2} + \frac{\|y^1 - y\|^2_{\tau_1 L_1 L_1^* + M_2^1}}{\tau_2} + \frac{\|x^1 - x^0\|^2}{\tau_1^2} + \frac{2}{\tau_1} \langle L_1 (x^1 - x^0), y^1 - y \rangle.$$

Moreover, $\lim\limits_{n \to +\infty} n \tau_n = \frac{\lambda}{\gamma}$, *hence one obtains for* $(x^n)_{n \geq 0}$ *an order of convergence of* $\mathcal{O}(\frac{1}{n})$.

Remark 4. Conditions guaranteeing the maximal monotonicity of compositions of a maximally monotone operator with a linear continuous operator have been intensively studied in the Hilbert space setting; for more insights we refer the reader to [4] and [5] and to the references therein.

Acknowledgements RIB has been partially supported by FWF (Austrian Science Fund), project I 2419-N32. ERC has been supported by FWF (Austrian Science Fund), project P 29809-N32. DM has been partially supported by FWF (Austrian Science Fund), project I 2419-N32, and by the Doctoral Programme Vienna Graduate School on Computational Optimization (VGSCO), project W1260-N35.

References

1. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Alternating proximal algorithms for weakly coupled convex minimization problems. Applications to dynamical games and PDE's. J. Convex Anal. **15**, 485–506 (2008)
2. Attouch, H., Cabot, A., Frankel, P., Peypouquet, J.: Alternating proximal algorithms for linearly constrained variational inequalities: Applications to domain decomposition for PDE's. Nonlinear Anal. **74**, 7455–7473 (2011)
3. Banert, S., Boţ, R.I., Csetnek, E.R.: Fixing and extending some recent results on the ADMM algorithm (2017). Available via arXiv. https://arxiv.org/abs/1612.05057
4. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. CMS Books in Mathematics, Springer, New York (2011)
5. Boţ, R.I.: Conjugate Duality in Convex Optimization. Lecture Notes in Economics and Mathematical Systems **637**, Springer-Verlag Berlin Heidelberg (2010)
6. Boţ, R.I., Csetnek, E.R.: ADMM for monotone operators: convergence analysis and rates. Adv. Comput. Math., DOI: 10.1007/s10444-018-9619-3 (2017)
7. Boţ, R.I., Csetnek, E.R., Heinrich, A.: A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators. SIAM J. Optim. **23**, 2011–2036 (2013)
8. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning **3**, 1–12 (2010)
9. Briceno-Arias, L.M., Combettes, P.L.: A monotone + skew splitting model for composite monotone inclusions in duality. SIAM J. Optim. **21**, 1230–1250 (2011)
10. Chambolle, P.L., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vision **40**, 120–145 (2011)
11. 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**, 307–330 (2012)
12. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Model. Simul. **4**, 1168–1200 (2005)
13. Combettes, P.L., Vũ, B.C.: Variable metric quasi-Feyér monotonicity. Nonlinear Anal. **78**, 17–31 (2013)
14. Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J. Optim. Theory Appl. **158**, 460–479 (2013)
15. Fazel, M., Pong, T.K., Sun, D., Tseng, P.: Hankel matrix rank minimization with applications in system identification and realization. SIAM J. Matrix Anal. **34**, 946–977 (2013)
16. Fortin, M., Glowinski, R.: On decomposition-coordination methods using an augmented Lagrangian. In: Fortin, M. and Glowinski, R. (eds.), Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems, North-Holland, Amsterdam (1983)

17. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M. and Glowinski, R. (eds.), Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems, North-Holland, Amsterdam (1983)
18. Riesz, F., Nagy, B.Sz.: Leons d'Analyse Fonctionelle. Fifth ed., Gauthier-Villars, Paris (1968)
19. Rockafellar, R.T.: On the maximal monotonicity of subdifferential mappings. Pacific J. Math. **33**, 209–216 (1970)
20. Shefi, R., Teboulle, M.: Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. SIAM J. Optim. **24**, 269–297 (2014)
21. Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. Adv. Comput. Math. **38**, 667–681 (2013)