

ADMM for monotone operators: convergence analysis and rates

Radu Ioan Boţ^{*} Ernő Robert Csetnek[†]

May 4, 2017

Abstract. We propose in this paper a unifying scheme for several algorithms from the literature dedicated to the solving of monotone inclusion problems involving compositions with linear continuous operators in infinite dimensional Hilbert spaces. We show that a number of primal-dual algorithms for monotone inclusions and also the classical ADMM numerical scheme for convex optimization problems, along with some of its variants, can be embedded in this unifying scheme. While in the first part of the paper convergence results for the iterates are reported, the second part is devoted to the derivation of convergence rates obtained by combining variable metric techniques with strategies based on suitable choice of dynamical step sizes.

Key Words. monotone operators, primal-dual algorithm, ADMM algorithm, subdifferential, convex optimization, Fenchel duality

AMS subject classification. 47H05, 65K05, 90C25

1 Introduction and preliminaries

Consider the convex optimization problem

$$\inf_{x \in \mathcal{H}} \{f(x) + g(Lx) + h(x)\}, \quad (1)$$

where \mathcal{H} and \mathcal{G} are real Hilbert spaces, $f : \mathcal{H} \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ and $g : \mathcal{G} \rightarrow \overline{\mathbb{R}}$ are proper, convex and lower semicontinuous functions, $h : \mathcal{H} \rightarrow \mathbb{R}$ is a convex and Fréchet differentiable function with Lipschitz continuous gradient and $L : \mathcal{H} \rightarrow \mathcal{G}$ is a linear continuous operator.

Due to numerous applications in fields like signal and image processing, portfolio optimization, cluster analysis, location theory, network communication, machine learning, the design and investigation of numerical algorithms for solving convex optimization problems of type (1) attracted in the last couple of years huge interest from the applied mathematics community. The most prominent methods one can find in the literature for solving (1) are the *primal-dual proximal splitting algorithms* and the *ADMM algorithms*. We briefly describe the two classes of algorithms.

First proximal splitting algorithms for solving convex optimization problems involving compositions with linear continuous operators have been reported by Combettes and Wajs [16],

^{*}University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: radu.bot@univie.ac.at. Research partially supported by FWF (Austrian Science Fund), project I 2419-N32.

[†]University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: ernoe.robert.csetnek@univie.ac.at. Research supported by FWF (Austrian Science Fund), project P 29809-N32.

Esser, Zhang and Chan [22] and Chambolle and Pock [13]. Further investigations have been made in the more general framework of finding zeros of sums of linearly composed maximally monotone operators, and monotone and Lipschitz, respectively, cocoercive operators. The resulting numerical schemes have been employed in the solving of the inclusion problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in \partial f(x) + (L^* \circ \partial g \circ L)(x) + \nabla h(x), \quad (2)$$

which represents the system of optimality conditions of problem (1).

Briceño-Arias and Combettes pioneered this approach in [12], by reformulating the general monotone inclusion in an appropriate product space as the sum of a maximally monotone operator and a linear and skew one, and by solving the resulting inclusion problem via a forward-backward-forward type algorithm (see also [14]). Afterwards, by using the same product space approach, this time in a suitable renormed space, Vũ succeeded in [29] in formulating a primal-dual splitting algorithm of forward-backward type, in other words, by saving a forward step. Condat has presented in [17] in the variational case algorithms of the same nature with the one in [29]. Under strong monotonicity/convexity assumptions and the use of dynamic step size strategies convergence rates have been provided in [8] for the primal-dual algorithm in [29] (see also [13]), and in [9] for the primal-dual algorithm in [14].

We describe the ADMM algorithm for solving (1) in the case $h = 0$, which corresponds to the standard setting in the literature. By introducing an auxiliary variable one can rewrite (1) as

$$\inf_{\substack{(x,z) \in \mathcal{H} \times \mathcal{G} \\ Lx-z=0}} \{f(x) + g(z)\}. \quad (3)$$

For a fixed real number $c > 0$ we consider the augmented Lagrangian associated with problem (3), which is defined as

$$L_c : \mathcal{H} \times \mathcal{G} \times \mathcal{G} \rightarrow \overline{\mathbb{R}}, \quad L_c(x, z, y) = f(x) + g(z) + \langle y, Ax - z \rangle + \frac{c}{2} \|Ax - z\|^2.$$

The ADMM algorithm relies on the alternating minimization of the augmented Lagrangian with respect to the variables x and z (see [11, 20, 23–25] and Remark 4 for the exact formulation of the iterative scheme). Generally, the minimization with respect to the variable x does not lead to a proximal step. This drawback has been overcome by Shefi and Teboulle in [27] by introducing additional suitably chosen metrics, and also in [5] for an extension of the ADMM algorithm designed for problems which involve also smooth parts in the objective.

The aim of this paper is to provide a unifying algorithmic scheme for solving monotone inclusion problems which encompasses several primal-dual iterative methods [7, 13, 17, 29], and the ADMM algorithm (and its variants from [27]) in the particular case of convex optimization problems. A closer look at the structure of the new algorithmic scheme shows that it translates the paradigm behind ADMM methods for optimization problems to the solving of monotone inclusions. We carry out a convergence analysis for the proposed iterative scheme by making use of techniques relying on the Opial Lemma applied in a variable metric setting. Furthermore, we derive convergence rates for the iterates under supplementary strong monotonicity assumptions. To this aim we use a dynamic step strategy, based on which we can provide a unifying scheme for the algorithms in [8, 13]. Not least we also provide accelerated versions for the classical ADMM algorithm (and its variable metric variants).

In what follows we recall some elements of the theory of monotone operators in Hilbert spaces and refer for more details to [2, 3, 28].

Let \mathcal{H} be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and associated norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$. For an arbitrary set-valued operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ we denote by $\text{Gr } A = \{(x, u) \in \mathcal{H} \times \mathcal{H} : u \in Ax\}$ its graph, by $\text{dom } A = \{x \in \mathcal{H} : Ax \neq \emptyset\}$ its domain and by $A^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ its inverse operator, defined by $(u, x) \in \text{Gr } A^{-1}$ if and only if $(x, u) \in \text{Gr } A$. We say that A is monotone if $\langle x - y, u - v \rangle \geq 0$ for all $(x, u), (y, v) \in \text{Gr } A$. A monotone operator A is said to be maximal monotone, if there exists no proper monotone extension of the graph of A on $\mathcal{H} \times \mathcal{H}$.

The resolvent of A , $J_A : \mathcal{H} \rightrightarrows \mathcal{H}$, is defined by $J_A = (\text{Id} + A)^{-1}$, where $\text{Id} : \mathcal{H} \rightarrow \mathcal{H}$, $\text{Id}(x) = x$ for all $x \in \mathcal{H}$, is the identity operator on \mathcal{H} . If A is maximal monotone, then $J_A : \mathcal{H} \rightarrow \mathcal{H}$ is single-valued and maximal monotone (see [2, Proposition 23.7 and Corollary 23.10]). For an arbitrary $\gamma > 0$ we have (see [2, Proposition 23.2])

$$p \in J_{\gamma A} x \text{ if and only if } (p, \gamma^{-1}(x - p)) \in \text{Gr } A$$

and (see [2, Proposition 23.18])

$$J_{\gamma A} + \gamma J_{\gamma^{-1} A^{-1}} \circ \gamma^{-1} \text{Id} = \text{Id}. \quad (4)$$

When \mathcal{G} is another Hilbert space and $L : \mathcal{H} \rightarrow \mathcal{G}$ is a linear continuous operator, then $L^* : \mathcal{G} \rightarrow \mathcal{H}$, defined by $\langle L^* y, x \rangle = \langle y, Lx \rangle$ for all $(x, y) \in \mathcal{H} \times \mathcal{G}$, denotes the adjoint operator of L , while the norm of L is defined as $\|L\| = \sup\{\|Lx\| : x \in \mathcal{H}, \|x\| \leq 1\}$.

Let $\gamma > 0$ be arbitrary. We say that A is γ -strongly monotone, if $\langle x - y, u - v \rangle \geq \gamma \|x - y\|^2$ for all $(x, u), (y, v) \in \text{Gr } A$. A single-valued operator $A : \mathcal{H} \rightarrow \mathcal{H}$ is said to be γ -cocoercive, if $\langle x - y, Ax - Ay \rangle \geq \gamma \|Ax - Ay\|^2$ for all $(x, y) \in \mathcal{H} \times \mathcal{H}$. Moreover, A is γ -Lipschitz continuous, if $\|Ax - Ay\| \leq \gamma \|x - y\|$ for all $(x, y) \in \mathcal{H} \times \mathcal{H}$. A single-valued linear operator $A : \mathcal{H} \rightarrow \mathcal{H}$ is said to be skew, if $\langle x, Ax \rangle = 0$ for all $x \in \mathcal{H}$. The parallel sum of two operators $A, B : \mathcal{H} \rightrightarrows \mathcal{H}$ is defined by $A \square B : \mathcal{H} \rightrightarrows \mathcal{H}$, $A \square B = (A^{-1} + B^{-1})^{-1}$.

Since the variational case will be also in the focus of our investigations, we recall next some elements of convex analysis.

For a function $f : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ we denote by $\text{dom } f = \{x \in \mathcal{H} : f(x) < +\infty\}$ its effective domain and say that f is proper, if $\text{dom } f \neq \emptyset$ and $f(x) \neq -\infty$ for all $x \in \mathcal{H}$. We denote by $\Gamma(\mathcal{H})$ the family of proper convex and lower semi-continuous extended real-valued functions defined on \mathcal{H} . Let $f^* : \mathcal{H} \rightarrow \overline{\mathbb{R}}$, $f^*(u) = \sup_{x \in \mathcal{H}} \{\langle u, x \rangle - f(x)\}$ for all $u \in \mathcal{H}$, be the conjugate function of f . The subdifferential of f at $x \in \mathcal{H}$, with $f(x) \in \mathbb{R}$, is the set $\partial f(x) := \{v \in \mathcal{H} : f(y) \geq f(x) + \langle v, y - x \rangle \forall y \in \mathcal{H}\}$. We take by convention $\partial f(x) := \emptyset$, if $f(x) \in \{\pm\infty\}$. If $f \in \Gamma(\mathcal{H})$, then ∂f is a maximally monotone operator (cf. [26]) and it holds $(\partial f)^{-1} = \partial f^*$. For $f, g : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ two proper functions, we consider also their infimal convolution, which is the function $f \square g : \mathcal{H} \rightarrow \overline{\mathbb{R}}$, defined by $(f \square g)(x) = \inf_{y \in \mathcal{H}} \{f(y) + g(x - y)\}$, for all $x \in \mathcal{H}$.

When $f \in \Gamma(\mathcal{H})$ and $\gamma > 0$, for every $x \in \mathcal{H}$ we denote by $\text{prox}_{\gamma f}(x)$ the proximal point of parameter γ of f at x , which is the unique optimal solution of the optimization problem

$$\inf_{y \in \mathcal{H}} \left\{ f(y) + \frac{1}{2\gamma} \|y - x\|^2 \right\}. \quad (5)$$

Notice that $J_{\gamma \partial f} = (\text{Id} + \gamma \partial f)^{-1} = \text{prox}_{\gamma f}$, thus $\text{prox}_{\gamma f} : \mathcal{H} \rightarrow \mathcal{H}$ is a single-valued operator fulfilling the extended Moreau's decomposition formula

$$\text{prox}_{\gamma f} + \gamma \text{prox}_{(1/\gamma) f^*} \circ \gamma^{-1} \text{Id} = \text{Id}. \quad (6)$$

Finally, we say that the function $f : \mathcal{H} \rightarrow \overline{\mathbb{R}}$ is γ -strongly convex for $\gamma > 0$, if $f - \frac{\gamma}{2} \|\cdot\|^2$ is a convex function. This property implies that ∂f is γ -strongly monotone (see [2, Example 22.3]).

2 The ADMM paradigm employed to monotone inclusions

In this section we propose an algorithm for solving monotone inclusion problems involving compositions with linear continuous operators in infinite dimensional Hilbert spaces which is designed in the spirit of the ADMM paradigm.

2.1 Problem formulation, algorithm and particular cases

The following problem represents the central point of our investigations.

Problem 1 Let \mathcal{H} and \mathcal{G} be real Hilbert spaces, $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and $B : \mathcal{G} \rightrightarrows \mathcal{G}$ be maximally monotone operators and $C : \mathcal{H} \rightarrow \mathcal{H}$ an η -cocoercive operator for $\eta > 0$. Let $L : \mathcal{H} \rightarrow \mathcal{G}$ be a linear continuous operator. To solve is the primal monotone inclusion

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + (L^* \circ B \circ L)x + Cx, \quad (7)$$

together with its dual monotone inclusion

$$\text{find } v \in \mathcal{G} \text{ such that } \exists x \in \mathcal{H} : -L^*v \in Ax + Cx \text{ and } v \in B(Lx). \quad (8)$$

Simple algebraic manipulations yield that (8) is equivalent to the problem

$$\text{find } v \in \mathcal{G} \text{ such that } 0 \in B^{-1}v + \left((-L) \circ (A + C)^{-1} \circ (-L^*) \right)v,$$

which can be equivalently written as

$$\text{find } v \in \mathcal{G} \text{ such that } 0 \in B^{-1}v + \left((-L) \circ (A^{-1} \square C^{-1}) \circ (-L^*) \right)v. \quad (9)$$

We say that $(x, v) \in \mathcal{H} \times \mathcal{G}$ is a primal-dual solution to the primal-dual pair of monotone inclusions (7)-(8), if

$$-L^*v \in Ax + Cx \text{ and } v \in B(Lx). \quad (10)$$

If $x \in \mathcal{H}$ is a solution to (7), then there exists $v \in \mathcal{G}$ such that (x, v) is a primal-dual solution to (7)-(8). On the other hand, if $v \in \mathcal{G}$ is a solution to (8), then there exists $x \in \mathcal{H}$ such that (x, v) is a primal-dual solution to (7)-(8). Furthermore, if $(x, v) \in \mathcal{H} \times \mathcal{G}$ is a primal-dual solution to (7)-(8), then x is a solution to (7) and v is a solution to (8).

Next we relate this general setting to the solving of a primal-dual pair of convex optimization problems.

Problem 2 Let \mathcal{H} and \mathcal{G} be real Hilbert spaces, $f \in \Gamma(\mathcal{H})$, $g \in \Gamma(\mathcal{G})$, $h : \mathcal{H} \rightarrow \mathbb{R}$ a convex and Fréchet differentiable function with η^{-1} -Lipschitz continuous gradient, for $\eta > 0$, and $L : \mathcal{H} \rightarrow \mathcal{G}$ a linear continuous operator. Consider the primal convex optimization problem

$$\inf_{x \in \mathcal{H}} \{f(x) + h(x) + g(Lx)\} \quad (11)$$

and its Fenchel dual problem

$$\sup_{v \in \mathcal{G}} \{-(f^* \square h^*)(-L^*v) - g^*(v)\}. \quad (12)$$

The system of optimality conditions for the primal-dual pair of optimization problems (11)-(12) reads:

$$-L^*\bar{v} - \nabla h(\bar{x}) \in \partial f(\bar{x}) \text{ and } \bar{v} \in \partial g(L\bar{x}), \quad (13)$$

which is actually a particular formulation of (10) when

$$A := \partial f, \quad C := \nabla h, \quad B := \partial g. \quad (14)$$

Notice that, due to the Baillon-Haddad Theorem (see [2, Corollary 18.16]), ∇h is η -cocoercive.

If (11) has an optimal solution $x \in \mathcal{H}$ and a suitable qualification condition is fulfilled, then there exists $v \in \mathcal{G}$, an optimal solution to (12), such that (13) holds. If (12) has an optimal solution $v \in \mathcal{G}$ and a suitable qualification condition is fulfilled, then there exists $x \in \mathcal{H}$, an optimal solution to (11), such that (13) holds. Furthermore, if the pair $(x, v) \in \mathcal{H} \times \mathcal{G}$ satisfies relation (13), then x is an optimal solution to (11), v is an optimal solution to (12) and the optimal objective values of (11) and (12) coincide.

One of the most popular and useful qualification conditions guaranteeing the existence of a dual optimal solution is the one known under the name Attouch-Brézis and which requires that:

$$0 \in \text{sqli}(\text{dom } g - L(\text{dom } f)) \quad (15)$$

holds. Here, for $S \subseteq \mathcal{G}$ a convex set, we denote by

$$\text{sqli } S := \{x \in S : \cup_{\lambda > 0} \lambda(S - x) \text{ is a closed linear subspace of } \mathcal{G}\}$$

its strong quasi-relative interior. The topological interior is contained in the strong quasi-relative interior: $\text{int } S \subseteq \text{sqli } S$, however, in general this inclusion may be strict. If \mathcal{G} is finite-dimensional, then for a nonempty and convex set $S \subseteq \mathcal{G}$, one has $\text{sqli } S = \text{ri } S$, which denotes the topological interior of S relative to its affine hull. Considering again the infinite dimensional setting, we remark that condition (15) is fulfilled, if there exists $x' \in \text{dom } f$ such that $Lx' \in \text{dom } g$ and g is continuous at Lx' . For further considerations on convex duality we refer to [2–4, 21, 30].

Throughout the paper the following additional notations and facts will be used. We denote by $\mathcal{S}_+(\mathcal{H})$ the family of operators $U : \mathcal{H} \rightarrow \mathcal{H}$ which are linear, continuous, self-adjoint and positive semidefinite. For $U \in \mathcal{S}_+(\mathcal{H})$ we consider the semi-norm defined by

$$\|x\|_U^2 = \langle x, Ux \rangle \quad \forall x \in \mathcal{H}.$$

The Loewner partial ordering is defined for $U_1, U_2 \in \mathcal{S}_+(\mathcal{H})$ by

$$U_1 \succcurlyeq U_2 \Leftrightarrow \|x\|_{U_1}^2 \geq \|x\|_{U_2}^2 \quad \forall x \in \mathcal{H}.$$

Finally, for $\alpha > 0$, we set

$$\mathcal{P}_\alpha(\mathcal{H}) := \{U \in \mathcal{S}_+(\mathcal{H}) : U \succcurlyeq \alpha \text{Id}\}.$$

Let $\alpha > 0$, $U \in \mathcal{P}_\alpha(\mathcal{H})$ and $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator. Then the operator $(U + A)^{-1} : \mathcal{H} \rightarrow \mathcal{H}$ is single-valued with full domain; in other words

$$\text{for every } x \in \mathcal{H} \text{ there exists a unique } p \in \mathcal{H} \text{ such that } p = (U + A)^{-1}x.$$

Indeed, this is a consequence of the relation

$$(U + A)^{-1} = (\text{Id} + U^{-1}A)^{-1} \circ U^{-1}$$

and of the maximal monotonicity of the operator $U^{-1}A$ in the renormed Hilbert space $(\mathcal{H}, \langle \cdot, \cdot \rangle_U)$ (see for example [15, Lemma 3.7]), where

$$\langle x, y \rangle_U := \langle x, Uy \rangle \quad \forall x, y \in \mathcal{H}.$$

We are now in the position to formulate the algorithm relying on the ADMM paradigm for solving the primal-dual pair of monotone inclusions (7)-(8).

Algorithm 3 For all $k \geq 0$, let $M_1^k \in \mathcal{S}_+(\mathcal{H})$, $M_2^k \in \mathcal{S}_+(\mathcal{G})$ and $c > 0$ be such that $cL^*L + M_1^k \in \mathcal{P}_{\alpha_k}(\mathcal{H})$ for $\alpha_k > 0$. Choose $(x^0, z^0, y^0) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}$. For all $k \geq 0$ generate the sequence $(x^k, z^k, y^k)_{k \geq 0}$ as follows:

$$x^{k+1} = \left(cL^*L + M_1^k + A \right)^{-1} \left[cL^*(z^k - c^{-1}y^k) + M_1^k x^k - Cx^k \right] \quad (16)$$

$$z^{k+1} = \left(\text{Id} + c^{-1}M_2^k + c^{-1}B \right)^{-1} \left[Lx^{k+1} + c^{-1}y^k + c^{-1}M_2^k z^k \right] \quad (17)$$

$$y^{k+1} = y^k + c(Lx^{k+1} - z^{k+1}). \quad (18)$$

As shown below, several algorithms from the literature can be embedded in this numerical scheme.

Remark 4 For all $k \geq 0$, the equations (16) and (17) are equivalent to

$$cL^*(z^k - Lx^{k+1} - c^{-1}y^k) + M_1^k(x^k - x^{k+1}) - Cx^k \in Ax^{k+1}, \quad (19)$$

and, respectively,

$$c(Lx^{k+1} - z^{k+1} + c^{-1}y^k) + M_2^k(z^k - z^{k+1}) \in Bz^{k+1}. \quad (20)$$

Notice that the latter is equivalent to

$$y^{k+1} + M_2^k(z^k - z^{k+1}) \in Bz^{k+1}. \quad (21)$$

In the variational setting as described in Problem 2, namely, by choosing the operators as in (14), the inclusion (19) becomes

$$0 \in \partial f(x^{k+1}) + cL^*(Lx^{k+1} - z^k + c^{-1}y^k) + M_1^k(x^{k+1} - x^k) + \nabla h(x^k),$$

which is equivalent to

$$x^{k+1} = \operatorname{argmin}_{x \in \mathcal{H}} \left\{ f(x) + \langle x - x^k, \nabla h(x^k) \rangle + \frac{c}{2} \|Lx - z^k + c^{-1}y^k\|^2 + \frac{1}{2} \|x - x^k\|_{M_1^k}^2 \right\}. \quad (22)$$

On the other hand, (20) becomes

$$c(Lx^{k+1} - z^{k+1} + c^{-1}y^k) + M_2^k(z^k - z^{k+1}) \in \partial g(z^{k+1}),$$

which is equivalent to

$$z^{k+1} = \operatorname{argmin}_{z \in \mathcal{G}} \left\{ g(z) + \frac{c}{2} \|Lx^{k+1} - z + c^{-1}y^k\|^2 + \frac{1}{2} \|z - z^k\|_{M_2^k}^2 \right\}. \quad (23)$$

Consequently, the iterative scheme (16)-(18) reads

$$x^{k+1} = \operatorname{argmin}_{x \in \mathcal{H}} \left\{ f(x) + \langle x - x^k, \nabla h(x^k) \rangle + \frac{c}{2} \|Ax - z^k + c^{-1}y^k\|^2 + \frac{1}{2} \|x - x^k\|_{M_1^k}^2 \right\} \quad (24)$$

$$z^{k+1} = \operatorname{argmin}_{z \in \mathcal{G}} \left\{ g(z) + \frac{c}{2} \|Ax^{k+1} - z + c^{-1}y^k\|^2 + \frac{1}{2} \|z - z^k\|_{M_2^k}^2 \right\} \quad (25)$$

$$y^{k+1} = y^k + c(Ax^{k+1} - z^{k+1}), \quad (26)$$

which is the algorithm formulated and investigated by Banert, Boş and Csetnek in [5]. The case when $h = 0$ and M_1^k, M_2^k are constant for every $k \geq 0$ has been considered in the setting of finite dimensional Hilbert spaces by Shefi and Teboulle [27]. We want to emphasize that when $h = 0$ and $M_1^k = M_2^k = 0$ for all $k \geq 0$ the iterative scheme (24)-(26) collapses into the classical version of the ADMM algorithm.

Remark 5 For all $k \geq 0$, consider the particular choices $M_1^k := \frac{1}{\tau_k} \operatorname{Id} - cL^*L$ for $\tau_k > 0$, and $M_2^k := 0$.

(i) Let $k \geq 0$ be fixed. Relation (16) (written for x^{k+2}) reads

$$x^{k+2} = (\tau_{k+1}^{-1} \operatorname{Id} + A)^{-1} \left[cL^*(z^{k+1} - c^{-1}y^{k+1}) + \tau_{k+1}^{-1}x^{k+1} - cL^*Lx^{k+1} - Cx^{k+1} \right].$$

From (18) we have

$$cL^*(z^{k+1} - c^{-1}y^{k+1}) = -L^*(2y^{k+1} - y^k) + cL^*Lx^{k+1},$$

hence

$$\begin{aligned} x^{k+2} &= (\tau_{k+1}^{-1} \operatorname{Id} + A)^{-1} \left[\tau_{k+1}^{-1}x^{k+1} - L^*(2y^{k+1} - y^k) - Cx^{k+1} \right] \\ &= J_{\tau_{k+1}A} \left(x^{k+1} - \tau_{k+1}Cx^{k+1} - \tau_{k+1}L^*(2y^{k+1} - y^k) \right). \end{aligned} \quad (27)$$

On the other hand, by using (4), relation (17) reads

$$z^{k+1} = J_{c^{-1}B} \left(Lx^{k+1} + c^{-1}y^k \right) = Lx^{k+1} + c^{-1}y^k - c^{-1}J_{cB^{-1}} \left(cLx^{k+1} + y^k \right)$$

which is equivalent to

$$cLx^{k+1} + y^k - cz^{k+1} = J_{cB^{-1}} \left(cLx^{k+1} + y^k \right).$$

By using again (18), this can be reformulated as

$$y^{k+1} = J_{cB^{-1}} \left(y^k + cLx^{k+1} \right). \quad (28)$$

The iterative scheme in (27)- (28) generates for a given starting point $(x^1, y^0) \in \mathcal{H} \times \mathcal{G}$ and $c > 0$ a sequence $(x^k, y^k)_{k \geq 1}$ which is generated for all $k \geq 0$ as follows:

$$y^{k+1} = J_{cB^{-1}} \left(y^k + cLx^{k+1} \right) \quad (29)$$

$$x^{k+2} = J_{\tau_{k+1}A} \left(x^{k+1} - \tau_{k+1}Cx^{k+1} - \tau_{k+1}L^*(2y^{k+1} - y^k) \right). \quad (30)$$

When $\tau_k = \tau$ for all $k \geq 1$, the algorithm (29)-(30) recovers a numerical scheme for solving monotone inclusion problems proposed by Vũ in [29, Theorem 3.1]. More precisely, the error-free variant of the algorithm in [29, Theorem 3.1] formulated for a constant sequence $(\lambda_n)_{n \in \mathbb{N}}$ equal to 1 and employed to the solving of the primal-dual pair (9)-(7) (by reversing the order in Problem 1, that is, by treating (9) as the primal monotone inclusion and (7) as its dual monotone inclusion) is nothing else than the iterative scheme (29)-(30).

In case $C = 0$, (29)-(30) becomes for all $k \geq 0$

$$x^{k+1} = J_{\tau_k A} \left(x^k - \tau_k L^*(2y^k - y^{k-1}) \right) \quad (31)$$

$$y^{k+1} = J_{cB^{-1}} \left(y^k + cLx^{k+1} \right), \quad (32)$$

which, in case $\tau_k = \tau$ for all $k \geq 1$ and $c\tau\|L\|^2 < 1$, is nothing else than the algorithm introduced by Boţ, Csetnek and Heinrich in [7, Algorithm 1, Theorem 2] applied to the solving of the primal-dual pair (9)-(7) (by reversing the order in Problem 1).

(ii) Considering again the variational setting as described in Problem 2, the algorithm (29)-(30) reads for all $k \geq 0$

$$y^{k+1} = \text{prox}_{cg^*} \left(y^k + cLx^{k+1} \right) \quad (33)$$

$$x^{k+2} = \text{prox}_{\tau_{k+1}f} \left(x^{k+1} - \tau_{k+1}\nabla h(x^{k+1}) - \tau_{k+1}L^*(2y^{k+1} - y^k) \right). \quad (34)$$

When $\tau_k = \tau > 0$ for all $k \geq 1$, one recovers a primal-dual algorithm investigated under the assumption $\frac{1}{\tau} - c\|L\|^2 > \frac{1}{2\eta}$ by Condat in [17, Algorithm 3.2, Theorem 3.1].

Not least, (31)-(32) reads in the variational setting (which corresponds to the case $h = 0$) for all $k \geq 0$

$$x^{k+1} = \text{prox}_{\tau_k f} \left(x^k - \tau_k L^*(2y^k - y^{k-1}) \right) \quad (35)$$

$$y^{k+1} = \text{prox}_{cg^*} \left(y^k + cLx^{k+1} \right). \quad (36)$$

When $\tau_k = \tau > 0$ for all $k \geq 1$, this iterative schemes becomes the algorithm proposed by Chambolle and Pock in [13, Algorithm 1, Theorem 1] for solving the primal-dual pair of optimization problems (12)-(11) (in this order).

2.2 Convergence analysis

In this subsection we will address the convergence of the sequence of iterates generated by Algorithm 3. One of the tools we will use in the proof of the convergence statement is the following version of the Opial Lemma formulated in the setting of variable metrics (see [15, Theorem 3.3]).

Lemma 6 *Let S be a nonempty subset of \mathcal{H} and $(x^k)_{k \geq 0}$ be a sequence in \mathcal{H} . Let $\alpha > 0$ and $W^k \in \mathcal{P}_\alpha(\mathcal{H})$ be such that $W^k \succcurlyeq W^{k+1}$ for all $k \geq 0$. Assume that:*

- (i) *for all $z \in S$ 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 S .*

Then $(x^k)_{k \geq 0}$ converges weakly to an element in S .

We present the first main theorem of this manuscript.

Theorem 7 *Consider the setting of Problem 1 and assume that the set of primal-dual solutions to the primal-dual pair of monotone inclusions (7)-(8) is nonempty. Let $(x^k, z^k, y^k)_{k \geq 0}$ be the sequence generated by Algorithm 3 and assume that $M_1^k - \frac{1}{2\eta} \text{Id} \in \mathcal{S}_+(\mathcal{H})$, $M_1^k \succcurlyeq M_1^{k+1}$, $M_2^k \in \mathcal{S}_+(\mathcal{G})$, $M_2^k \succcurlyeq M_2^{k+1}$ for all $k \geq 0$. If one of the following assumptions:*

(I) *there exists $\alpha_1 > 0$ such that $M_1^k - \frac{1}{2\eta} \text{Id} \in \mathcal{P}_{\alpha_1}(\mathcal{H})$ for all $k \geq 0$;*

(II) *there exist $\alpha, \alpha_2 > 0$ such that $L^*L \in \mathcal{P}_\alpha(\mathcal{H})$ and $M_2^k \in \mathcal{P}_{\alpha_2}(\mathcal{G})$ for all $k \geq 0$;*

is fulfilled, then there exists (x, v) , a primal-dual solution to (7)-(8), such that $(x^k, z^k, y^k)_{k \geq 0}$ converges weakly to (x, Lx, v) .

Proof. Let $S \subseteq \mathcal{H} \times \mathcal{G} \times \mathcal{G}$ be defined by

$$S = \{(x, Lx, v) : (x, v) \text{ is a primal dual solution to (7)-(8)}\}. \quad (37)$$

Let $(x^*, Lx^*, y^*) \in S$ be fixed. Then it holds

$$-L^*y^* - Cx^* \in Ax^* \text{ and } y^* \in B(Lx^*).$$

Let $k \geq 0$ be fixed. From (19) and the monotonicity of A we have

$$\langle cL^*(z^k - Lx^{k+1} - c^{-1}y^k) + M_1^k(x^k - x^{k+1}) - Cx^k + L^*y^* + Cx^*, x^{k+1} - x^* \rangle \geq 0, \quad (38)$$

while from (20) and the monotonicity of B we have

$$\langle c(Lx^{k+1} - z^{k+1} + c^{-1}y^k) + M_2^k(z^k - z^{k+1}) - y^*, z^{k+1} - Lx^* \rangle \geq 0. \quad (39)$$

Since C is η -cocoercive, we have

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

Summing up the three inequalities obtained above we get

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

According to (18) we also have

$$\langle y^* - y^k, Lx^{k+1} - Lx^* \rangle + \langle y^k - y^*, z^{k+1} - Lx^* \rangle = \langle y^* - y^k, Lx^{k+1} - z^{k+1} \rangle = c^{-1} \langle y^* - y^k, y^{k+1} - y^k \rangle.$$

By expressing the inner products through norms we further derive

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

By expressing $Lx^{k+1} - z^{k+1}$ using again relation (18) and by taking into account that

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

we obtain

$$\begin{aligned} & \frac{1}{2} \|x^{k+1} - x^*\|_{M_1^k}^2 + \frac{1}{2} \|z^{k+1} - Lx^*\|_{M_2^k + c\text{Id}}^2 + \frac{1}{2c} \|y^{k+1} - y^*\|^2 \leq \\ & \frac{1}{2} \|x^k - x^*\|_{M_1^k}^2 + \frac{1}{2} \|z^k - Lx^*\|_{M_2^k + c\text{Id}}^2 + \frac{1}{2c} \|y^k - y^*\|^2 \\ & - \frac{c}{2} \|z^k - Lx^{k+1}\|^2 - \frac{1}{2} \|x^k - x^{k+1}\|_{M_1^k}^2 - \frac{1}{2} \|z^k - z^{k+1}\|_{M_2^k}^2 \\ & - \eta^{-1} \left\| \eta (Cx^* - Cx^k) + \frac{1}{2} (x^k - x^{k+1}) \right\|^2 + \frac{1}{4\eta} \|x^k - x^{k+1}\|^2. \end{aligned}$$

From here, using the monotonicity assumptions on $(M_1^k)_{k \geq 0}$ and $(M_2^k)_{k \geq 0}$, it yields

$$\begin{aligned} & \frac{1}{2} \|x^{k+1} - x^*\|_{M_1^{k+1}}^2 + \frac{1}{2} \|z^{k+1} - Lx^*\|_{M_2^{k+1} + c\text{Id}}^2 + \frac{1}{2c} \|y^{k+1} - y^*\|^2 \leq \\ & \frac{1}{2} \|x^k - x^*\|_{M_1^k}^2 + \frac{1}{2} \|z^k - Lx^*\|_{M_2^k + c\text{Id}}^2 + \frac{1}{2c} \|y^k - y^*\|^2 \\ & - \frac{c}{2} \|z^k - Lx^{k+1}\|^2 - \frac{1}{2} \|x^k - x^{k+1}\|_{M_1^k - \frac{1}{2\eta}\text{Id}}^2 - \frac{1}{2} \|z^k - z^{k+1}\|_{M_2^k}^2 \\ & - \eta^{-1} \left\| \eta (Cx^* - Cx^k) + \frac{1}{2} (x^k - x^{k+1}) \right\|^2. \end{aligned} \quad (40)$$

Discarding the negative terms on the right-hand side of the above inequality (notice that $M_1^k - \frac{1}{2\eta}\text{Id} \in \mathcal{S}_+(\mathcal{H})$ for all $k \geq 0$), it follows that statement (i) in Opial Lemma (Lemma 6) holds, when applied in the product space $\mathcal{H} \times \mathcal{G} \times \mathcal{G}$, for the sequence $(x^k, z^k, y^k)_{k \geq 0}$, for $W^k := (M_1^k, M_2^k + c\text{Id}, c^{-1}\text{Id})$ for $k \geq 0$, and for S defined as in (37).

Furthermore, summing up the inequalities in (40), we get

$$\sum_{k \geq 0} \|z^k - Lx^{k+1}\|^2 < +\infty, \quad \sum_{k \geq 0} \|x^k - x^{k+1}\|_{M_1^k - \frac{1}{2\eta}\text{Id}}^2 < +\infty, \quad \sum_{k \geq 0} \|z^k - z^{k+1}\|_{M_2^k}^2 < +\infty. \quad (41)$$

Consider first the hypotheses in assumption (I). Since $M_1^k - \frac{1}{2\eta}\text{Id} \in \mathcal{P}_{\alpha_1}(\mathcal{H})$ for all $k \geq 0$ with $\alpha_1 > 0$, we get

$$x^k - x^{k+1} \rightarrow 0 \quad (k \rightarrow +\infty) \quad (42)$$

and

$$z^k - Lx^{k+1} \rightarrow 0 \quad (k \rightarrow +\infty). \quad (43)$$

A direct consequence of (42) and (43) is

$$z^k - z^{k+1} \rightarrow 0 \quad (k \rightarrow +\infty). \quad (44)$$

From (18), (43) and (44) we derive

$$y^k - y^{k+1} \rightarrow 0 \quad (k \rightarrow +\infty). \quad (45)$$

Next we show that the relations (42)-(45) are fulfilled also under assumption (II). Indeed, in this situation we derive from (41) that (43) and (44) hold. From (18), (43) and (44) we obtain (45). Finally, the inequalities

$$\alpha \|x^{k+1} - x^k\|^2 \leq \|Lx^{k+1} - Lx^k\|^2 \leq 2\|Lx^{k+1} - z^k\|^2 + 2\|z^k - Lx^k\|^2 \quad \forall k \geq 0 \quad (46)$$

yield (42).

The relations (42)-(45) will play an essential role when verifying assumption (ii) in the Opial Lemma for S taken as in (37). Let $(\bar{x}, \bar{z}, \bar{y}) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}$ be such that there exists $(k_n)_{n \geq 0}$, $k_n \rightarrow +\infty$ (as $n \rightarrow +\infty$), and $(x^{k_n}, z^{k_n}, y^{k_n})$ converges weakly to $(\bar{x}, \bar{z}, \bar{y})$ (as $n \rightarrow +\infty$).

From (42) and the linearity and the continuity of L we obtain that $(Lx^{k_n+1})_{n \in \mathbb{N}}$ converges weakly to $L\bar{x}$ (as $n \rightarrow +\infty$), which combined with (43) yields $\bar{z} = L\bar{x}$. We use now the following notations for $n \geq 0$:

$$\begin{aligned} a_n^* &:= cL^*(z^{k_n} - Lx^{k_n+1} - c^{-1}y^{k_n}) + M_1^{k_n}(x^{k_n} - x^{k_n+1}) + Cx^{k_n+1} - Cx^{k_n} \\ a_n &:= x^{k_n+1} \\ b_n^* &:= y^{k_n+1} + M_2^{k_n}(z^{k_n} - z^{k_n+1}) \\ b_n &:= z^{k_n+1}. \end{aligned}$$

From (19) we have for all $n \geq 0$

$$a_n^* \in (A + C)(a_n). \quad (47)$$

Further, from (20) and (18) we have for all $n \geq 0$

$$b_n^* \in Bb_n. \quad (48)$$

Furthermore, from (42) we have

$$a_n \text{ converges weakly to } \bar{x} \text{ (as } n \rightarrow +\infty). \quad (49)$$

From (45) and (44) we obtain

$$b_n^* \text{ converges weakly to } \bar{y} \text{ (as } n \rightarrow +\infty). \quad (50)$$

Moreover, (18) and (45) yield

$$La_n - b_n \text{ converges strongly to } 0 \text{ (as } n \rightarrow +\infty). \quad (51)$$

Finally, we have

$$\begin{aligned} a_n^* + L^*b_n^* &= cL^*(z^{k_n} - Lx^{k_n+1}) + L^*(y^{k_n+1} - y^{k_n}) \\ &\quad + M_1^{k_n}(x^{k_n} - x^{k_n+1}) + L^*M_2^{k_n}(z^{k_n} - z^{k_n+1}) \\ &\quad + Cx^{k_n+1} - Cx^{k_n}. \end{aligned}$$

By using the fact that C is η^{-1} -Lipschitz continuous, from (42)-(45) we get

$$a_n^* + L^*b_n^* \text{ converges strongly to } 0 \text{ (as } n \rightarrow +\infty). \quad (52)$$

Let us define $T : \mathcal{H} \times \mathcal{G} \rightrightarrows \mathcal{H} \times \mathcal{G}$ by $T(x, y) = (A(x) + C(x)) \times B^{-1}(y)$ and $K : \mathcal{H} \times \mathcal{G} \rightarrow \mathcal{H} \times \mathcal{G}$ by $K(x, y) = (L^*y, -Lx)$ for all $(x, y) \in \mathcal{H} \times \mathcal{G}$. Since C is maximally monotone with full domain (see [2]), $A + C$ is maximally monotone, too (see [2]), thus T is maximally monotone. Since K is a skew operator, it is also maximally monotone (see [2]). Due to the fact that K has full domain, we conclude that

$$T + K \text{ is a maximally monotone operator.} \quad (53)$$

Moreover, from (47) and (48) we have

$$(a_n^* + L^*b_n^*, b_n - La_n) \in (T + K)(a_n, b_n^*) \quad \forall n \geq 0. \quad (54)$$

Since the graph of a maximally monotone operator is sequentially closed with respect to the weak \times strong topology (see [2, Proposition 20.33]), from (53), (54), (49), (50), (51) and (52) we derive that

$$(0, 0) \in (T + K)(\bar{x}, \bar{y}) = (A + C, B^{-1})(\bar{x}, \bar{y}) + (L^*\bar{y}, -L\bar{x}).$$

The latter is nothing else than saying that (\bar{x}, \bar{y}) is a primal dual-solution to (7)-(8), which combined with $\bar{z} = L\bar{x}$ implies that the second assumption of the Opial Lemma is verified, too. In conclusion, $(x^k, z^k, y^k)_{k \geq 0}$ converges weakly to (x, Lx, v) , where (x, v) a primal-dual solution to (7)-(8). \blacksquare

Remark 8 (i) Choosing as in Remark 5 $M_1^k := \frac{1}{\tau_k} \text{Id} - cL^*L$, with $\tau_k > 0$ and $\tau := \sup_{k \geq 0} \tau_k \in \mathbb{R}$, and $M_2^k := 0$ for all $k \geq 0$, we have

$$\left\langle x, \left(M_1^k - \frac{1}{2\eta} \text{Id} \right) x \right\rangle \geq \left(\frac{1}{\tau_k} - c\|L\|^2 - \frac{1}{2\eta} \right) \|x\|^2 \geq \left(\frac{1}{\tau} - c\|L\|^2 - \frac{1}{2\eta} \right) \|x\|^2 \quad \forall x \in \mathcal{H},$$

which means that under the assumption $\frac{1}{\tau} - c\|L\|^2 > \frac{1}{2\eta}$ (which recovers the one in Algorithm 3.2 and Theorem 3.1 in [17]) the operators $M_1^k - \frac{1}{2\eta} \text{Id}$ belong for all $k \geq 0$ to the class $\mathcal{P}_{\alpha_1}(\mathcal{H})$ with $\alpha_1 := \frac{1}{\tau} - c\|L\|^2 - \frac{1}{2\eta} > 0$.

(ii) Let us briefly discuss the condition considered in (II):

$$\exists \alpha > 0 \text{ such that } L^*L \in \mathcal{P}_\alpha(\mathcal{H}). \quad (55)$$

If L is injective and $\text{ran } L^*$ is closed, then (55) holds (see [2, Fact 2.19]). Moreover, (55) implies that L is injective. This means that if $\text{ran } L^*$ is closed, then (55) is equivalent to L is injective. Hence, in finite dimensional spaces, namely, if $\mathcal{H} = \mathbb{R}^n$ and $\mathcal{G} = \mathbb{R}^m$, with $m \geq n \geq 1$, (55) is nothing else than saying that L has full column rank, which is a widely used assumption in the proof of the convergence of the classical ADMM algorithm.

In the second convergence result of this section we consider the case when C is identically 0. We notice that this cannot be encompassed in the above theorem due to the assumptions which involve the cocercivity constant η in the denominator of some fractions and which do not allow us to take it equal to zero.

Theorem 9 *Consider the setting of Problem 1 in the situation when $C = 0$ and assume that the set of primal-dual solutions to the primal-dual pair of monotone inclusions (7)-(8) is nonempty. Let $(x^k, z^k, y^k)_{k \geq 0}$ be the sequence generated by Algorithm 3 and assume that $M_1^k \in \mathcal{S}_+(\mathcal{H})$, $M_1^k \succcurlyeq M_1^{k+1}$, $M_2^k \in \mathcal{S}_+(\mathcal{G})$, $M_2^k \succcurlyeq M_2^{k+1}$ for all $k \geq 0$. If one of the following assumptions:*

(I) there exists $\alpha_1 > 0$ such that $M_1^k \in \mathcal{P}_{\alpha_1}(\mathcal{H})$ for all $k \geq 0$;

(II) there exist $\alpha, \alpha_2 > 0$ such that $L^*L \in \mathcal{P}_\alpha(\mathcal{H})$, $M_1^k = 0$ and $M_2^k \in \mathcal{P}_{\alpha_2}(\mathcal{G})$ for all $k \geq 0$;

(III) there exists $\alpha > 0$ such that $L^*L \in \mathcal{P}_\alpha(\mathcal{H})$, $M_1^k = 0$ and $2M_2^{k+1} \succcurlyeq M_2^k \succcurlyeq M_2^{k+1}$ for all $k \geq 0$;

is fulfilled, then there exists (x, v) , a primal-dual solution to (7)-(8), such that $(x^k, z^k, y^k)_{k \geq 0}$ converges weakly to (x, Lx, v) .

Proof. We fix an element (x^*, Lx^*, y^*) with the property that (x^*, y^*) a primal-dual solution to (7)-(8).

Take an arbitrary $k \geq 0$. As in the proof of Theorem 7, we can derive the inequality

$$\begin{aligned} & \frac{1}{2} \|x^{k+1} - x^*\|_{M_1^{k+1}}^2 + \frac{1}{2} \|z^{k+1} - Lx^*\|_{M_2^{k+1} + c\text{Id}}^2 + \frac{1}{2c} \|y^{k+1} - y^*\|^2 \leq \\ & \quad \frac{1}{2} \|x^k - x^*\|_{M_1^k}^2 + \frac{1}{2} \|z^k - Lx^*\|_{M_2^k + c\text{Id}}^2 + \frac{1}{2c} \|y^k - y^*\|^2 \\ & \quad - \frac{c}{2} \|z^k - Lx^{k+1}\|^2 - \frac{1}{2} \|x^k - x^{k+1}\|_{M_1^k}^2 - \frac{1}{2} \|z^k - z^{k+1}\|_{M_2^k}^2. \end{aligned} \quad (56)$$

Under assumption (I) the conclusion follows as in the proof of Theorem 7 by making use of the Opial Lemma.

Consider the situation when the hypotheses in assumption (II) are fulfilled. In this case, (56) becomes

$$\begin{aligned} & \frac{1}{2} \|z^{k+1} - Lx^*\|_{M_2^{k+1} + c\text{Id}}^2 + \frac{1}{2c} \|y^{k+1} - y^*\|^2 \leq \\ & \frac{1}{2} \|z^k - Lx^*\|_{M_2^k + c\text{Id}}^2 + \frac{1}{2c} \|y^k - y^*\|^2 - \frac{c}{2} \|z^k - Lx^{k+1}\|^2 - \frac{1}{2} \|z^k - z^{k+1}\|_{M_2^k}^2. \end{aligned} \quad (57)$$

By using telescopic sum techniques, it follows that (43) and (44) hold. From (18), (43) and (44) we obtain (45). Finally, by using again the inequality (46), relation (42) holds, too.

On the other hand, (57) yields that

$$\exists \lim_{k \rightarrow +\infty} \left(\frac{1}{2} \|z^k - Lx^*\|_{M_2^k + c\text{Id}}^2 + \frac{1}{2c} \|y^k - y^*\|^2 \right), \quad (58)$$

hence $(y^k)_{k \geq 0}$ and $(z^k)_{k \geq 0}$ are bounded. Combining this with the condition imposed on L , we derive that $(x^k)_{k \geq 0}$ is bounded, too. Hence there exists a weak 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 Theorem 7, one can see that every sequential weak cluster point of $(x^k, z^k, y^k)_{k \geq 0}$ belongs to the set S defined in (37).

In the remaining of the proof we show that the set of sequential weak 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 sequential weak cluster points. Then there exist $(k_p)_{p \geq 0}, (k_q)_{q \geq 0}$, $k_p \rightarrow +\infty$ (as $p \rightarrow +\infty$), $k_q \rightarrow +\infty$ (as $q \rightarrow +\infty$), 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 $p \rightarrow +\infty$), 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 $q \rightarrow +\infty$). As shown

above, (x_1, z_1, y_1) and (x_2, z_2, y_2) belong to the set S (see (37)), thus $z_i = Lx_i$, $i \in \{1, 2\}$. From (58), which is true for every primal-dual solution to (7)-(8), we derive

$$\exists \lim_{k \rightarrow +\infty} \left(\frac{1}{2} \|z^k - Lx_1\|_{M_2^k + c\text{Id}}^2 - \frac{1}{2} \|z^k - Lx_2\|_{M_2^k + c\text{Id}}^2 + \frac{1}{2c} \|y^k - y_1\|^2 - \frac{1}{2c} \|y^k - y_2\|^2 \right). \quad (59)$$

Further, we have for all $k \geq 0$

$$\frac{1}{2} \|z^k - Lx_1\|_{M_2^k + c\text{Id}}^2 - \frac{1}{2} \|z^k - Lx_2\|_{M_2^k + c\text{Id}}^2 = \frac{1}{2} \|Lx_2 - Lx_1\|_{M_2^k + c\text{Id}}^2 + \langle z^k - Lx_2, (M_2^k + c\text{Id})(Lx_2 - Lx_1) \rangle$$

and

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

Now, the monotonicity condition imposed on $(M_2^k)_{k \geq 0}$ implies that $\sup_{k \geq 0} \|M_2^k + c\text{Id}\| < +\infty$. Thus, according to [15], there exists $\alpha' > 0$ and $M_2 \in \mathcal{P}_{\alpha'}(\mathcal{G})$ such that $(M_2^k + c\text{Id})_{k \geq 0}$ converges pointwise to M_2 (as $k \rightarrow +\infty$).

Taking the limit in (59) along the subsequences $(k_p)_{p \geq 0}$ and $(k_q)_{q \geq 0}$ and using the last two relations above we obtain

$$\begin{aligned} & \frac{1}{2} \|Lx_1 - Lx_2\|_{M_2}^2 + \langle Lx_1 - Lx_2, M_2(Lx_2 - Lx_1) \rangle + \frac{1}{2c} \|y_1 - y_2\|^2 + \frac{1}{c} \langle y_1 - y_2, y_2 - y_1 \rangle \\ &= \frac{1}{2} \|Lx_1 - Lx_2\|_{M_2}^2 + \frac{1}{2c} \|y_1 - y_2\|^2, \end{aligned}$$

hence

$$-\|Lx_1 - Lx_2\|_{M_2}^2 - \frac{1}{c} \|y_1 - y_2\|^2 = 0,$$

thus $Lx_1 = Lx_2$ and $y_1 = y_2$. Assumption (II) implies that $x_1 = x_2$. In conclusion, $(x^k, z^k, y^k)_{k \geq 0}$ converges weakly to an element in S (see(37)).

Finally, we consider the situation when the hypotheses in assumption (III) hold. As noticed above, since $M_1^k = 0$ for all $k \geq 0$, relation (56) becomes (57). Let $k \geq 1$ be fixed. By considering the relation (21) for consecutive iterates and by taking into account the monotonicity of B we derive

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

hence

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

Substituting $y^{k+1} - y^k = c(Lx^{k+1} - z^{k+1})$ in the last inequality it follows

$$\begin{aligned} & \|z^{k+1} - z^k\|_{M_2^k}^2 - \frac{1}{2} \|z^{k+1} - z^k\|_{M_2^{k-1}}^2 - \frac{1}{2} \|z^k - z^{k-1}\|_{M_2^{k-1}}^2 \leq \\ & \frac{c}{2} \left(\|z^k - Lx^{k+1}\|^2 - \|z^{k+1} - z^k\|^2 - \|Lx^{k+1} - z^{k+1}\|^2 \right), \end{aligned}$$

which, after adding it with (57), leads to

$$\begin{aligned} & \frac{1}{2} \|z^{k+1} - Lx^*\|_{M_2^{k+1} + c\text{Id}}^2 + \frac{1}{2c} \|y^{k+1} - y^*\|^2 + \frac{1}{2} \|z^{k+1} - z^k\|_{3M_2^k - M_2^{k-1}}^2 \leq \\ & \frac{1}{2} \|z^k - Lx^*\|_{M_2^k + c\text{Id}}^2 + \frac{1}{2c} \|y^k - y^*\|^2 + \frac{1}{2} \|z^k - z^{k-1}\|_{M_2^{k-1}}^2 - \frac{c}{2} \|z^{k+1} - z^k\|^2 - \frac{1}{2c} \|y^{k+1} - y^k\|^2. \end{aligned} \quad (60)$$

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

$$\begin{aligned} & \frac{1}{2} \|z^{k+1} - Lx^*\|_{M_2^{k+1} + c\text{Id}}^2 + \frac{1}{2c} \|y^{k+1} - y^*\|^2 + \frac{1}{2} \|z^{k+1} - z^k\|_{M_2^k}^2 \leq \\ & \frac{1}{2} \|z^k - Lx^*\|_{M_2^k + c\text{Id}}^2 + \frac{1}{2c} \|y^k - y^*\|^2 + \frac{1}{2} \|z^k - z^{k-1}\|_{M_2^{k-1}}^2 - \frac{c}{2} \|z^{k+1} - z^k\|^2 - \frac{1}{2c} \|y^{k+1} - y^k\|^2. \end{aligned} \quad (61)$$

Using telescoping sum arguments, we obtain that $y^k - y^{k+1} \rightarrow 0$ and $z^k - z^{k+1} \rightarrow 0$ as $k \rightarrow +\infty$. Using (18), it follows that $L(x^k - x^{k+1}) \rightarrow 0$ as $k \rightarrow +\infty$, which, combined with the hypotheses imposed to L , further implies that $x^k - x^{k+1} \rightarrow 0$ as $k \rightarrow +\infty$. Consequently, $z^k - Lx^{k+1} \rightarrow 0$ as $k \rightarrow +\infty$. Hence the relations (42)-(45) are fulfilled. On the other hand, from (61) we also derive that

$$\exists \lim_{k \rightarrow +\infty} \left(\frac{1}{2} \|z^k - Lx^*\|_{M_2^k + c\text{Id}}^2 + \frac{1}{2c} \|y^k - y^*\|^2 + \frac{1}{2} \|z^k - z^{k-1}\|_{M_2^{k-1}}^2 \right). \quad (62)$$

By using that

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

it follows that $\lim_{k \rightarrow +\infty} \|z^k - z^{k-1}\|_{M_2^{k-1}}^2 = 0$, which further implies that (58) holds. From here the conclusion follows by arguing as in the second part of the proof provide above under hypotheses (II). \blacksquare

Remark 10 In the variational case, the sequences generated by the classical ADMM algorithm, which corresponds to the iterative scheme (24)-(26) for $h = 0$ and $M_1^k = M_2^k = 0$ for all $k \geq 0$, are convergent, provided that L has full column rank. This situation is covered by the theorem above in the context of assumption (III).

3 Convergence rates under strong monotonicity and by means of dynamic step sizes

We state the problem on which we focus throughout this section.

Problem 11 In the setting of Problem 1 we replace the cocoercivity of C by the assumptions that C is monotone and μ -Lipschitz continuous for $\mu \geq 0$. Moreover, we assume that $A + C$ is γ -strongly monotone for $\gamma > 0$.

Remark 12 If C is a η -cocoercive operator for $\eta > 0$, then C is monotone and η^{-1} -Lipschitz continuous. Though, the converse statement may fail. The skew operator $(x, y) \mapsto (L^*y, -Lx)$ is for instance monotone and Lipschitz continuous, and not cocoercive. This operator appears in a natural way when considering formulating the system of optimality conditions for convex optimization problems involving compositions with linear continuous operators (see [12]). Notice that due to the celebrated Baillon-Haddad Theorem (see, for instance, [2, Corollary 8.16]), the gradient of a convex and Fréchet differentiable function is η -cocoercive if and only if it is η^{-1} -Lipschitz continuous.

Remark 13 In the setting of Problem 11 the operator $A + L^* \circ B \circ L + C$ is strongly monotone, thus the monotone inclusion problem (7) has at most one solution. Hence, if (x, v) is a primal-dual solution to the primal-dual pair (7)-(8), then x is the unique solution to (7). Notice that the problem (8) may not have an unique solution.

We propose the following algorithm for the formulation of which we use dynamic step sizes.

Algorithm 14 For all $k \geq 0$, let $M_2^k : \mathcal{G} \rightarrow \mathcal{G}$ be a linear, continuous and self-adjoint operator such that $\tau_k LL^* + M_2^k \in \mathcal{P}_{\alpha_k}(\mathcal{G})$ for $\alpha_k > 0$ for all $k \geq 0$. Choose $(x^0, z^0, y^0) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}$. For all $k \geq 0$ generate the sequence $(x^k, z^k, y^k)_{k \geq 0}$ as follows:

$$y^{k+1} = \left(\tau_k LL^* + M_2^k + B^{-1} \right)^{-1} \left[-\tau_k L(z^k - \tau_k^{-1} x^k) + M_2^k y^k \right] \quad (63)$$

$$z^{k+1} = \left(\frac{\theta_k}{\lambda} - 1 \right) L^* y^{k+1} + \frac{\theta_k}{\lambda} C x^k + \frac{\theta_k}{\lambda} \left(\text{Id} + \lambda \tau_{k+1}^{-1} A^{-1} \right)^{-1} \left[-L^* y^{k+1} + \lambda \tau_{k+1}^{-1} x^k - C x^k \right] \quad (64)$$

$$x^{k+1} = x^k + \frac{\tau_{k+1}}{\theta_k} \left(-L^* y^{k+1} - z^{k+1} \right), \quad (65)$$

where $\lambda, \tau_k, \theta_k > 0$ for all $k \geq 0$.

Remark 15 We would like to emphasize that when $C = 0$ Algorithm 14 has a similar structure to Algorithm 3. Indeed, in this setting, the monotone inclusion problems (7) and (9) become

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + (L^* \circ B \circ L)x \quad (66)$$

and, respectively,

$$\text{find } v \in \mathcal{G} \text{ such that } 0 \in B^{-1}v + \left((-L) \circ (A^{-1}) \circ (-L^*) \right) v. \quad (67)$$

The two problems (66) and (67) are dual to each other in the sense of the Attouch-Théra duality (see [1]). By taking in (63)-(65) $\lambda = 1$, $\theta_k = 1$ (which corresponds to the limit case $\mu = 0$ and $\gamma = 0$ in the equation (74) below) and $\tau_k = c > 0$ for all $k \geq 0$, then the resulting iterative scheme reads

$$y^{k+1} = \left(cLL^* + M_2^k + B^{-1} \right)^{-1} \left[-cL(z^k - c^{-1}x^k) + M_2^k y^k \right] \quad (68)$$

$$z^{k+1} = \left(\text{Id} + c^{-1}A^{-1} \right)^{-1} \left[-L^* y^{k+1} + c^{-1}x^k \right] \quad (69)$$

$$x^{k+1} = x^k + c \left(-L^* y^{k+1} - z^{k+1} \right). \quad (70)$$

This is nothing else than Algorithm 3 employed to the solving of the primal-dual system of monotone inclusions (67)-(66), that is, by treating (67) as the primal monotone inclusion and (66) as its dual monotone inclusion (notice that in this case we take in relation (17) of Algorithm 3 $M_2^k = 0$ for all $k \geq 0$).

Concerning the parameters involved in Algorithm 14 we assume that

$$\mu\tau_1 < 2\gamma, \quad (71)$$

$$\lambda \geq \mu + 1, \quad (72)$$

there exists $\sigma_0 > 0$ such that

$$\sigma_0\tau_1\|L\|^2 \leq 1, \quad (73)$$

and for all $k \geq 0$:

$$\theta_k = \frac{1}{\sqrt{1 + \tau_{k+1}\lambda^{-1}(2\gamma - \mu\tau_{k+1})}} \quad (74)$$

$$\tau_{k+2} = \theta_k\tau_{k+1} \quad (75)$$

$$\sigma_{k+1} = \theta_k^{-1}\sigma_k \quad (76)$$

$$\tau_k LL^* + M_2^k \succcurlyeq \sigma_k^{-1} \text{Id} \quad (77)$$

$$\frac{\tau_k}{\tau_{k+1}} LL^* + \frac{1}{\tau_{k+1}} M_2^k \succcurlyeq \frac{\tau_{k+1}}{\tau_{k+2}} LL^* + \frac{1}{\tau_{k+2}} M_2^{k+1}. \quad (78)$$

Remark 16 Fix an arbitrary $k \geq 1$. From (63) we have

$$-\tau_k L(z^k - \tau_k^{-1}x^k) + M_2^k y^k \in \widetilde{M}_2^k y^{k+1} + B^{-1}y^{k+1}, \quad (79)$$

where

$$\widetilde{M}_2^k := \tau_k LL^* + M_2^k. \quad (80)$$

Due to (65) we have

$$-\tau_k z^k = \tau_k L^* y^k + \theta_{k-1}(x^k - x^{k-1}),$$

which combined with (79) delivers

$$\widetilde{M}_2^k (y^k - y^{k+1}) + L \left[x^k + \theta_{k-1}(x^k - x^{k-1}) \right] \in B^{-1}y^{k+1}. \quad (81)$$

Fix now an arbitrary $k \geq 0$. From (4) and (64) we have

$$\begin{aligned} -L^* y^{k+1} + \frac{\lambda}{\theta_k} (z^{k+1} + L^* y^{k+1}) - Cx^k &= -L^* y^{k+1} + \frac{\lambda}{\tau_{k+1}} x^k - Cx^k \\ &\quad - \frac{\lambda}{\tau_{k+1}} J_{(\tau_{k+1}/\lambda)A} \left[x^k + \frac{\tau_{k+1}}{\lambda} (-L^* y^{k+1} - Cx^k) \right]. \end{aligned}$$

By using (65) we obtain

$$x^{k+1} = J_{(\tau_{k+1}/\lambda)A} \left[x^k + \frac{\tau_{k+1}}{\lambda} (-L^* y^{k+1} - Cx^k) \right]. \quad (82)$$

Finally, the definition of the resolvent yields the relation

$$\frac{\lambda}{\tau_{k+1}} (x^k - x^{k+1}) - L^* y^{k+1} + Cx^{k+1} - Cx^k \in (A + C)x^{k+1}. \quad (83)$$

Remark 17 The choice

$$\tau_k LL^* + M_2^k = \sigma_k^{-1} \text{Id} \quad \forall k \geq 0 \quad (84)$$

leads to so-called accelerated versions of primal-dual algorithms that have been intensively studied in the literature. Indeed, under these auspices (63) becomes (by taking into account also (65))

$$\begin{aligned} y^{k+1} &= (\sigma_k^{-1} \text{Id} + B^{-1})^{-1} \left[L \left(-\tau_k z^k + x^k - \tau_k L^* y^k \right) + \sigma_k^{-1} y^k \right] \\ &= (\text{Id} + \sigma_k B^{-1})^{-1} \left[y^k + \sigma_k L \left(x^k + \theta_k (x^k - x^{k-1}) \right) \right]. \end{aligned}$$

This together with (82) gives rise for all $k \geq 0$ to the following numerical scheme

$$x^{k+1} = J_{(\tau_{k+1}/\lambda)A} \left[x^k + \frac{\tau_{k+1}}{\lambda} \left(-L^* y^{k+1} - C x^k \right) \right] \quad (85)$$

$$y^{k+2} = J_{\sigma_{k+1} B^{-1}} \left[y^{k+1} + \sigma_{k+1} L \left(x^{k+1} + \theta_{k+1} (x^{k+1} - x^k) \right) \right], \quad (86)$$

which has been investigated by Bot, Csetnek, Heinrich and Hendrich in [8, Algorithm 5]. Not least, assuming that $C = 0$ and $\lambda = 1$, the variational case $A = \partial f$ and $B = \partial g$ leads for all $k \geq 0$ to the numerical scheme

$$y^{k+1} = \text{prox}_{\sigma_k g^*} \left[y^k + \sigma_k L \left(x^k + \theta_k (x^k - x^{k-1}) \right) \right] \quad (87)$$

$$x^{k+1} = \text{prox}_{\tau_{k+1} f} \left(x^k - \tau_{k+1} L^* y^{k+1} \right), \quad (88)$$

which has been considered by Chambolle and Pock in [13, Algorithm 2].

We also notice that condition (84) guarantees the fulfillment of both (77) and (78), due to the fact that the sequence $(\tau_{k+1} \sigma_k)_{k \geq 0}$ is constant (see (75) and (76)).

Remark 18 Assume again that $C = 0$ and consider the variational case as described in Problem 2. From (79) and (80) we derive for all $k \geq 1$ the relation

$$0 \in \partial g^*(y^{k+1}) + \tau_k L \left(L^* y^{k+1} + z^k - \tau_k^{-1} x^k \right) + M_2^k \left(y^{k+1} - y^k \right),$$

which in case $M_2^k \in \mathcal{S}_+(\mathcal{G})$ is equivalent to

$$y^{k+1} = \underset{y \in \mathcal{G}}{\text{argmin}} \left[g^*(y) + \frac{\tau_k}{2} \left\| L^* y + z^k - \tau_k^{-1} x^k \right\|^2 + \frac{1}{2} \|y - y^k\|_{M_2^k}^2 \right].$$

Algorithm 14 becomes in case $\lambda = 1$

$$\begin{aligned} y^{k+1} &= \underset{y \in \mathcal{G}}{\text{argmin}} \left[g^*(y) + \frac{\tau_k}{2} \left\| L^* y + z^k - \tau_k^{-1} x^k \right\|^2 + \frac{1}{2} \|y - y^k\|_{M_2^k}^2 \right] \\ z^{k+1} &= (\theta_k - 1) L^* y^{k+1} + \theta_k \underset{x \in \mathcal{H}}{\text{argmin}} \left[f^*(x) + \frac{\tau_{k+1}}{2} \left\| -L^* y^{k+1} - z + \tau_{k+1}^{-1} x^k \right\|^2 \right] \\ x^{k+1} &= x^k + \frac{\tau_{k+1}}{\theta_k} \left(-L^* y^{k+1} - z^{k+1} \right), \end{aligned}$$

which can be regarded as an accelerated version of the algorithm (24)-(26) in Remark 4.

We present the main theorem of this section.

Theorem 19 *Consider the setting of Problem 11 and let (x, v) be a primal-dual solution to the primal-dual system of monotone inclusions (7)-(8). Let $(x^k, z^k, y^k)_{k \geq 0}$ be the sequence generated by Algorithm 14 and assume that the relations (71)-(78) are fulfilled. Then we have for all $n \geq 2$*

$$\begin{aligned} & \frac{\lambda \|x^n - x\|^2}{\tau_{n+1}^2} + \frac{1 - \sigma_0 \tau_1 \|L\|^2}{\sigma_0 \tau_1} \|y^n - v\|^2 \leq \\ & \frac{\lambda \|x^1 - x\|^2}{\tau_2^2} + \frac{\|y^1 - v\|_{\tau_1 LL^* + M_2^1}^2}{\tau_2} + \frac{\|x^1 - x^0\|^2}{\tau_1^2} + \frac{2}{\tau_1} \langle L(x^1 - x^0), y^1 - v \rangle. \end{aligned}$$

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

Proof. Let $k \geq 1$ be fixed. From (81), the relation $Lx \in B^{-1}v$ (see (10)) and the monotonicity of B^{-1} we obtain

$$\left\langle y^{k+1} - v, \widetilde{M}_2^k (y^k - y^{k+1}) + L \left[x^k + \theta_{k-1} (x^k - x^{k-1}) \right] - Lx \right\rangle \geq 0$$

or, equivalently,

$$\frac{1}{2} \|y^k - v\|_{\widetilde{M}_2^k}^2 - \frac{1}{2} \|y^{k+1} - v\|_{\widetilde{M}_2^k}^2 - \frac{1}{2} \|y^k - y^{k+1}\|_{\widetilde{M}_2^k}^2 \geq \left\langle y^{k+1} - v, Lx - L \left[x^k + \theta_{k-1} (x^k - x^{k-1}) \right] \right\rangle. \quad (89)$$

Further, from (83), the relation $-L^*v \in (A + C)x$ (see (10)) and the γ -strong monotonicity of $A + C$ we obtain

$$\left\langle x^{k+1} - x, \frac{\lambda}{\tau_{k+1}} (x^k - x^{k+1}) - L^*y^{k+1} + Cx^{k+1} - Cx^k + L^*v \right\rangle \geq \gamma \|x^{k+1} - x\|^2$$

or, equivalently,

$$\begin{aligned} \frac{\lambda}{2\tau_{k+1}} \|x^k - x\|^2 - \frac{\lambda}{2\tau_{k+1}} \|x^{k+1} - x\|^2 - \frac{\lambda}{2\tau_{k+1}} \|x^k - x^{k+1}\|^2 & \geq \gamma \|x^{k+1} - x\|^2 \\ & + \langle x^{k+1} - x, Cx^k - Cx^{k+1} \rangle \\ & + \langle y^{k+1} - v, Lx^{k+1} - x \rangle. \quad (90) \end{aligned}$$

Since C is μ -Lipschitz continuous, we have that

$$\langle x^{k+1} - x, Cx^k - Cx^{k+1} \rangle \geq -\frac{\mu\tau_{k+1}}{2} \|x^{k+1} - x\|^2 - \frac{\mu}{2\tau_{k+1}} \|x^{k+1} - x^k\|^2,$$

which combined with (90) implies

$$\begin{aligned} \frac{\lambda}{2\tau_{k+1}} \|x^k - x\|^2 & \geq \left(\frac{\lambda}{2\tau_{k+1}} + \gamma - \frac{\mu\tau_{k+1}}{2} \right) \|x^{k+1} - x\|^2 + \frac{\lambda - \mu}{2\tau_{k+1}} \|x^{k+1} - x^k\|^2 \\ & + \langle y^{k+1} - v, Lx^{k+1} - Lx \rangle. \quad (91) \end{aligned}$$

By adding the inequalities (89) and (91), we obtain

$$\begin{aligned}
\frac{1}{2}\|y^k - v\|_{\widetilde{M}_2^k}^2 + \frac{\lambda}{2\tau_{k+1}}\|x^k - x\|^2 &\geq \frac{1}{2}\|y^{k+1} - v\|_{\widetilde{M}_2^k}^2 + \left(\frac{\lambda}{2\tau_{k+1}} + \gamma - \frac{\mu\tau_{k+1}}{2}\right)\|x^{k+1} - x\|^2 \\
&+ \frac{1}{2}\|y^k - y^{k+1}\|_{\widetilde{M}_2^k}^2 + \frac{\lambda - \mu}{2\tau_{k+1}}\|x^{k+1} - x^k\|^2 \\
&+ \left\langle y^{k+1} - \bar{v}, L \left[x^{k+1} - x^k - \theta_{k-1}(x^k - x^{k-1}) \right] \right\rangle. \quad (92)
\end{aligned}$$

Further, we have

$$\begin{aligned}
\left\langle L \left[x^{k+1} - x^k - \theta_{k-1}(x^k - x^{k-1}) \right], y^{k+1} - v \right\rangle &= \langle L(x^{k+1} - x^k), y^{k+1} - v \rangle \\
&- \theta_{k-1} \langle L(x^k - x^{k-1}), y^k - v \rangle \\
&+ \theta_{k-1} \langle L(x^k - x^{k-1}), y^k - y^{k+1} \rangle \\
&\geq \langle L(x^{k+1} - x^k), y^{k+1} - v \rangle \\
&- \theta_{k-1} \langle L(x^k - x^{k-1}), y^k - v \rangle \\
&- \frac{\theta_{k-1}^2 \|L\|^2 \sigma_k}{2} \|x^{k-1} - x^k\|^2 - \frac{\|y^k - y^{k+1}\|^2}{2\sigma_k}.
\end{aligned}$$

By combining this inequality with (92) we obtain (after dividing by τ_{k+1})

$$\begin{aligned}
\frac{\|y^k - v\|_{\widetilde{M}_2^k}^2}{2\tau_{k+1}} + \frac{\lambda}{2\tau_{k+1}^2}\|x^k - x\|^2 &\geq \frac{\|y^{k+1} - v\|_{\widetilde{M}_2^k}^2}{2\tau_{k+1}} + \left(\frac{\lambda}{2\tau_{k+1}^2} + \frac{\gamma}{\tau_{k+1}} - \frac{\mu}{2}\right)\|x^{k+1} - x\|^2 \\
&+ \frac{\|y^k - y^{k+1}\|_{\widetilde{M}_2^k}^2}{2\tau_{k+1}} - \frac{\|y^k - y^{k+1}\|^2}{2\tau_{k+1}\sigma_k} \quad (93) \\
&+ \frac{\lambda - \mu}{2\tau_{k+1}^2}\|x^{k+1} - x^k\|^2 - \frac{\theta_{k-1}^2 \|L\|^2 \sigma_k}{2\tau_{k+1}}\|x^k - x^{k-1}\|^2 \\
&+ \frac{1}{\tau_{k+1}} \langle L(x^{k+1} - x^k), y^{k+1} - v \rangle \\
&- \frac{\theta_{k-1}}{\tau_{k+1}} \langle L(x^k - x^{k-1}), y^k - v \rangle.
\end{aligned}$$

From (77) and (80) we have that the term in (93) is nonnegative. Further, noticing that (see (74), (75), (76) and (73))

$$\begin{aligned}
\frac{\theta_{k-1}}{\tau_{k+1}} &= \frac{1}{\tau_k} \\
\frac{\lambda}{2\tau_{k+1}^2} + \frac{\gamma}{\tau_{k+1}} - \frac{\mu}{2} &= \frac{\lambda}{2\tau_{k+2}^2}, \\
\tau_{k+1}\sigma_k &= \tau_k\sigma_{k-1} = \dots = \tau_1\sigma_0
\end{aligned}$$

and

$$\frac{\|L\|^2 \sigma_k \theta_{k-1}^2}{\tau_{k+1}} = \frac{\tau_{k+1} \|L\|^2 \sigma_k}{\tau_k^2} = \frac{\tau_1 \|L\|^2 \sigma_0}{\tau_k^2} \leq \frac{1}{\tau_k^2},$$

we obtain (see also (72) and (78))

$$\begin{aligned} \frac{\|y^k - v\|_{\widetilde{M}_2^k}^2}{2\tau_{k+1}} + \frac{\lambda}{2\tau_{k+1}^2} \|x^k - x\|^2 &\geq \frac{\|y^{k+1} - v\|_{\widetilde{M}_2^{k+1}}^2}{2\tau_{k+2}} + \frac{\lambda}{2\tau_{k+2}^2} \|x^{k+1} - x\|^2 \\ &+ \frac{1}{2\tau_{k+1}^2} \|x^{k+1} - x^k\|^2 - \frac{1}{2\tau_k^2} \|x^k - x^{k-1}\|^2 \\ &+ \frac{1}{\tau_{k+1}} \langle L(x^{k+1} - x^k), y^{k+1} - v \rangle - \frac{1}{\tau_k} \langle L(x^k - x^{k-1}), y^k - v \rangle. \end{aligned}$$

Let n be a natural number such that $n \geq 2$. Summing up the above inequality from $k = 1$ to $n - 1$, it follows

$$\begin{aligned} \frac{\|y^1 - v\|_{\widetilde{M}_2^1}^2}{2\tau_2} + \frac{\lambda}{2\tau_2^2} \|x^1 - x\|^2 &\geq \frac{\|y^n - v\|_{\widetilde{M}_2^n}^2}{2\tau_{n+1}} + \frac{\lambda}{2\tau_{n+1}^2} \|x^n - x\|^2 \\ &+ \frac{1}{2\tau_n^2} \|x^n - x^{n-1}\|^2 - \frac{1}{2\tau_1^2} \|x^1 - x^0\|^2 \\ &+ \frac{1}{\tau_n} \langle L(x^n - x^{n-1}), y^n - v \rangle - \frac{1}{\tau_1} \langle L(x^1 - x^0), y^1 - v \rangle. \end{aligned}$$

The inequality in the statement of the theorem follows by combining this relation with (see (77))

$$\begin{aligned} \frac{\|y^n - v\|_{\widetilde{M}_2^n}^2}{2\tau_{n+1}} &\geq \frac{\|y^n - v\|^2}{2\sigma_n\tau_{n+1}}, \\ \frac{1}{2\tau_n^2} \|x^n - x^{n-1}\|^2 + \frac{1}{\tau_n} \langle L(x^n - x^{n-1}), y^n - v \rangle &\geq -\frac{\|L\|^2}{2} \|y^n - v\|^2 \text{ and } \sigma_n\tau_{n+1} = \sigma_0\tau_1. \end{aligned}$$

Finally, we notice that for any $n \geq 0$ (see (74) and (75))

$$\tau_{n+2} = \frac{\tau_{n+1}}{\sqrt{1 + \frac{\tau_{n+1}}{\lambda}(2\gamma - \mu\tau_{n+1})}}. \quad (94)$$

From here it follows that $\tau_{n+1} < \tau_n$ for all $n \geq 1$ and $\lim_{n \rightarrow +\infty} n\tau_n = \lambda/\gamma$ (see [8, page 261]). The proof is complete. \blacksquare

Remark 20 In Remark 17 we provided an example of a family of linear, continuous and self-adjoint operators $(M_2^k)_{k \geq 0}$ for which the relations (77) and (78) are fulfilled. In the following we will furnish more examples in this sense.

To begin we notice that simple algebraic manipulations easily lead to the conclusion that if

$$\mu\tau_1 < \gamma, \quad (95)$$

then $(\theta_k)_{k \geq 0}$ is monotonically increasing. In the examples below we replace (71) with the stronger assumption (95).

(i) For all $k \geq 0$, take

$$M_2^k := \sigma_k^{-1} \text{Id}.$$

Then (77) trivially holds, while (78), which can be equivalently written as

$$\frac{1}{\theta_{k-1}}LL^* + \frac{1}{\tau_{k+1}}M_2^k \succcurlyeq \frac{1}{\theta_k}LL^* + \frac{1}{\tau_{k+2}}M_2^{k+1},$$

follows from the fact that $(\tau_{k+1}\sigma_k)_{k \geq 0}$ is constant (see (75) and (76)) and $(\theta_k)_{k \geq 0}$ is monotonically increasing.

(ii) For all $k \geq 0$, take

$$M_2^k := 0.$$

Relation (78) holds since $(\theta_k)_{k \geq 0}$ is monotonically increasing. Condition (77) becomes in this setting

$$\sigma_k \tau_k LL^* \succcurlyeq \text{Id} \quad \forall k \geq 0. \quad (96)$$

Since $\tau_k > \tau_{k+1}$ for all $k \geq 1$ and $(\tau_{k+1}\sigma_k)_{k \geq 0}$ is constant, (96) holds, if

$$LL^* \in \mathcal{P}_{\frac{1}{\sigma_0 \tau_1}}(\mathcal{G}). \quad (97)$$

In this case one has to take in (73)

$$\sigma_0 \tau_1 \|L\|^2 = 1.$$

In view of the above theorem, the iterative scheme obtained in this particular instance (see Remark 18) can be regarded as an accelerated version of the classical ADMM algorithm (see Remark 4 and Remark 8(ii)).

(iii) For all $k \geq 0$, take

$$M_2^k := \tau_k \text{Id}.$$

Relation (78) holds since $(\theta_k)_{k \geq 0}$ is monotonically increasing. On the other hand, condition (77) is equivalent to

$$\sigma_k \tau_k (LL^* + \text{Id}) \succcurlyeq \text{Id}. \quad (98)$$

Since $\tau_k > \tau_{k+1}$ for all $k \geq 1$ and $(\tau_{k+1}\sigma_k)_{k \geq 0}$ is constant, (98) holds, if

$$\sigma_0 \tau_1 LL^* \succcurlyeq (1 - \sigma_0 \tau_1) \text{Id}. \quad (99)$$

In case $\sigma_0 \tau_1 \geq 1$ (which is allowed according to (73) if $\|L\|^2 \leq 1$) this is obviously fulfilled. Otherwise, in order to guarantee (99), we have to impose that

$$LL^* \in \mathcal{P}_{\frac{1 - \sigma_0 \tau_1}{\sigma_0 \tau_1}}(\mathcal{G}). \quad (100)$$

References

- [1] H. Attouch, M. Théra, *A general duality principle for the sum of two operators*, Journal of Convex Analysis 3(1), 1–24, 1996
- [2] H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York, 2011
- [3] J.M. Borwein and J.D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge University Press, Cambridge, 2010

- [4] R.I. Boş, *Conjugate Duality in Convex Optimization*, Lecture Notes in Economics and Mathematical Systems, Vol. 637, Springer, Berlin Heidelberg, 2010
- [5] S. Banert, R.I. Boş, E.R. Csetnek, *Fixing and extending some recent results on the ADMM algorithm*, arXiv:1612.05057, 2016
- [6] R.I. Boş, E.R. Csetnek, *An inertial alternating direction method of multipliers*, Minimax Theory and its Applications 1(1), 29–49, 2016
- [7] R.I. Boş, E.R. Csetnek, A. Heinrich, *A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators*, SIAM Journal on Optimization 23(4), 2011–2036, 2013
- [8] R.I. Boş, E.R. Csetnek, A. Heinrich, C. Hendrich, *On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems*, Mathematical Programming 150(2), 251–279, 2015
- [9] R.I. Boş, C. Hendrich, *Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization*, Journal of Mathematical Imaging and Vision 49(3), 551–568, 2014
- [10] R.I. Boş, C. Hendrich, *A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators*, SIAM Journal on Optimization 23(4), 2541–2565, 2013
- [11] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning 3, 1–12, 2010
- [12] L.M. Briceño-Arias, P.L. Combettes, *A monotone + skew splitting model for composite monotone inclusions in duality*, SIAM Journal on Optimization 21(4), 1230–1250, 2011
- [13] A. Chambolle, T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision 40(1), 120–145, 2011
- [14] P.L. Combettes, J.-C. Pesquet, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators*, Set-Valued and Variational Analysis 20(2), 307–330, 2012
- [15] P.L. Combettes, B.C. Vũ, *Variable metric quasi-Fejér monotonicity*, Nonlinear Analysis 78, 17–31, 2013
- [16] P.L. Combettes, V.R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multi-scale Modeling and Simulation 4(4), 1168–1200, 2005
- [17] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, Journal of Optimization Theory and Applications 158(2), 460–479, 2013

- [18] J. Eckstein, *Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results*, Rutcor Research Report 32-2012, 2012
- [19] J. Eckstein, *Some saddle-function splitting methods for convex programming*, Optimization Methods and Software 4, 75–83, 1994
- [20] J. Eckstein, D.P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming 55, 293–318, 1992
- [21] I. Ekeland, R. Temam, *Convex Analysis and Variational Problems*, North-Holland Publishing Company, Amsterdam, 1976
- [22] E. Esser, *Applications of Lagrangian-based alternating direction methods and connections to split Bregman*, CAM Reports 09-31, UCLA, Center for Applied Mathematics, 2009
- [23] M. Fortin, R. Glowinski, *On decomposition-coordination methods using an augmented Lagrangian*, in: M. Fortin and R. Glowinski (eds.), *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, North-Holland, Amsterdam, 1983
- [24] D. Gabay, *Applications of the method of multipliers to variational inequalities*, in: M. Fortin and R. Glowinski (eds.), *Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems*, North-Holland, Amsterdam, 1983
- [25] D. Gabay, B. Mercier, *A dual algorithm for the solution of nonlinear variational problems via finite element approximations*, Computers and Mathematics with Applications 2, 17–40, 1976
- [26] R.T. Rockafellar, *On the maximal monotonicity of subdifferential mappings*, Pacific Journal of Mathematics 33(1), 209–216, 1970
- [27] R. Shefi, M. Teboulle, *Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization*, SIAM Journal on Optimization 24(1), 269–297, 2014
- [28] S. Simons, *From Hahn-Banach to Monotonicity*, Springer-Verlag, Berlin, 2008
- [29] B.C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Advances in Computational Mathematics 38(3), 667–681, 2013
- [30] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, Singapore, 2002