Adv Comput Math (2019) 45:327–359
https://doi.org/10.1007/s10444-018-9619-3

ADMM for monotone operators: convergence analysis and rates

Radu Ioan Boţ[1] · Ernö Robert Csetnek[1]

Received: 6 November 2017 / Accepted: 1 June 2018 /
Published online: 25 June 2018

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. The numerical performances, which can be obtained for different dynamical step size strategies, are compared in the context of solving an image denoising problem.

Keywords Monotone operators · Primal-dual algorithm · ADMM algorithm · Subdifferential · Convex optimization · Fenchel duality

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

Communicated by: Russell Luke

Research of the first author partially supported by FWF (Austrian Science Fund), project I 2419-N32.
Research of the second author supported by FWF (Austrian Science Fund), project P 29809-N32.

✉ Radu Ioan Boţ
radu.bot@univie.ac.at

Ernö Robert Csetnek
ernoe.robert.csetnek@univie.ac.at

1 Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria

1 Introduction and preliminaries

Consider the convex optimization problem

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

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

Due to numerous applications in fields like signal and image processing, portfolio optimization, cluster analysis, location theory, network communication, and 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.

Primal-dual algorithms have their origins in the works of Arrow, Hurwicz and Uzawa [1], and Korpelevich [33]. Tseng's algorithm [40], which stands at heart of primal-dual algorithms of forward-backward-forward type, is a modification of the iterative methods in these two fundamental works. Proximal splitting algorithms for solving convex optimization problems involving compositions with linear continuous operators have been proposed by Combettes and Ways [19], Esser et al. [26], Chambolle and Pock [14], and He and Yuan [32]. 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 to the solving of the inclusion problem as follows:

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

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

Briceño-Arias and Combettes pioneered this approach in [13], 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 [16]). Afterwards, by using the same product space approach, this time in a suitable renormed space, Vũ succeeded in [41] in formulating a primal-dual splitting algorithm of forward-backward type, in other words, by saving a forward step. Condat has presented in [20], in the variational case, algorithms of the same nature with the one in [41]. A primal-dual algorithm of Douglas-Rachford type has been proposed in [11]. Under strong monotonicity/convexity assumptions and the use of dynamic step size strategies, convergence rates have been provided in [9], for the primal-dual algorithm in [41] (see also [14, 15]), and in [10] and for the primal-dual algorithm in [16]. Among the recent developments in this field count, the primal-dual algorithm with linesearch introduced in [34], which avoids the exact calculation of the norm of the linear operator and the three-operator splitting algorithm for monotone inclusions introduced in [21].

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 follows:

$$\inf_{\substack{(x,z)\in\mathcal{H}\times\mathcal{G}\\ Lx-z=0}} \{f(x)+g(z)\}. \tag{2}$$

For a fixed real number $c > 0$, we consider the augmented Lagrangian associated with problem (2), which is defined as follows:

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

The ADMM algorithm relies on the alternating minimization of the augmented Lagrangian with respect to the variables x and z (see [12, 22–24, 28–30] 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 [38] by introducing additional suitably chosen metrics, and also in [3] 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 [8, 14, 20, 41] and the ADMM algorithm (and its variants from [38]) 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 [9, 14]. Not least, we also provide accelerated versions for the classical ADMM algorithm (and its variable metric variants). In the last section, we compare the performances of the accelerated algorithm under different dynamical step size strategies in the context of solving an image processing problem.

In what follows, we recall some elements of the theory of monotone operators in Hilbert spaces and refer for more details to [4, 6, 39].

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 $\operatorname{Gr} A = \{(x,u) \in \mathcal{H}\times\mathcal{H} : u \in Ax\}$ its graph, by $\operatorname{dom} A = \{x \in \mathcal{H} : Ax \neq \emptyset\}$ its domain and by $A^{-1} : \mathcal{H} \rightrightarrows \mathcal{H}$ its inverse operator, defined by $(u,x) \in \operatorname{Gr} A^{-1}$ if and only if $(x,u) \in \operatorname{Gr} A$. We say that A is monotone if $\langle x-y, u-v\rangle \geq 0$ for all $(x,u), (y,v) \in \operatorname{Gr} A$. 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 $J_A = (\mathrm{Id}+A)^{-1}$, where $\mathrm{Id} : \mathcal{H} \to \mathcal{H}$, $\mathrm{Id}(x) = x$ for all $x \in \mathcal{H}$, is the identity operator on $\mathcal{H}$. If A is maximal monotone, then $J_A : \mathcal{H} \to \mathcal{H}$ is single-valued and maximal monotone (see [4, Proposition 23.7 and Corollary 23.10]). For an arbitrary $\gamma > 0$, we have (see [4, Proposition 23.2])

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

and (see [4, Proposition 23.18])

$$J_{\gamma A} + \gamma J_{\gamma^{-1}A^{-1}} \circ \gamma^{-1} \operatorname{Id} = \operatorname{Id}. \tag{3}$$

When $\mathcal{G}$ is another Hilbert space and $L : \mathcal{H} \to \mathcal{G}$ is a linear continuous operator, then $L^* : \mathcal{G} \to \mathcal{H}$, 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 $\|L\| = \sup\{\|\text{Lx}\| : x \in \mathcal{H}, \|x\| \leq 1\}$.

Let $\gamma \geq 0$ be arbitrary. We say that A is γ-strongly monotone, if $\langle x - y, u - v\rangle \geq \gamma\|x - y\|^2$ for all $(x, u), (y, v) \in \operatorname{Gr} A$. A single-valued operator $A : \mathcal{H} \to \mathcal{H}$ is said to be γ-cocoercive, if $\gamma\langle x-y, \text{Ax}-\text{Ay}\rangle \geq \|\text{Ax}-\text{Ay}\|^2$ for all $(x, y) \in \mathcal{H}\times\mathcal{H}$. Notice that we slightly modify the classical definition of a coercive operator, without altering its sense, in order to cover also the situation when A is constant (in particular, when $A = 0$) and $\gamma = 0$. A is called γ-Lipschitz continuous, if $\|\text{Ax} - \text{Ay}\| \leq \gamma\|x - y\|$ for all $(x, y) \in \mathcal{H} \times \mathcal{H}$. A single-valued linear operator $A : \mathcal{H} \to \mathcal{H}$ 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 $A\Box B : \mathcal{H} \rightrightarrows \mathcal{H}$, $A\Box B = (A^{-1} + B^{-1})^{-1}$.

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

For a function $f : \mathcal{H} \to \overline{\mathbb{R}}$ 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 $x \in \mathcal{H}$. We denote by $\Gamma(\mathcal{H})$ the family of proper convex and lower semi-continuous extended real-valued functions defined on $\mathcal{H}$. Let $f^* : \mathcal{H} \to \overline{\mathbb{R}}$, $f^*(u) = \sup_{x\in\mathcal{H}}\{\langle u, x\rangle - f(x)\}$ 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. [37]) and it holds $(\partial f)^{-1} = \partial f^*$. For $f, g : \mathcal{H} \to \overline{\mathbb{R}}$ two proper functions, we consider also their infimal convolution, which is the function $f\Box g : \mathcal{H} \to \overline{\mathbb{R}}$, defined by $(f\Box g)(x) = \inf_{y\in\mathcal{H}}\{f(y) + g(x - y)\}$, for all $x \in \mathcal{H}$.

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

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

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

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

Finally, we say that the function $f : \mathcal{H} \to \overline{\mathbb{R}}$ 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 [4, 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 $C : \mathcal{H} \to \mathcal{H}$ an η-cocoercive operator for $\eta \geq 0$. Let $L : \mathcal{H} \to \mathcal{G}$ be a linear continuous operator. The aim is to solve the primal monotone inclusion

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in \text{A x} + (L^* \circ B \circ L)x + Cx \tag{4}$$

together with its dual monotone inclusion

$$\text{find } v \in \mathcal{G} \text{ such that } \exists x \in \mathcal{H} : -L^*v \in \text{Ax} + \text{Cx and } v \in B(\text{Lx}). \tag{5}$$

Simple algebraic manipulations yield that (5) is equivalent to the problem

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

which can be equivalently written as follows:

$$\text{find } v \in \mathcal{G} \text{ such that } \ 0 \in B^{-1}v + \Big((-L) \circ (A^{-1} \Box C^{-1}) \circ (-L^*)\Big)v. \tag{6}$$

We say that $(x, v) \in \mathcal{H} \times \mathcal{G}$ is a primal-dual solution to the primal-dual pair of monotone inclusions (4)–(5), if

$$-L^*v \in \text{Ax} + \text{Cx and } v \in B(\text{Lx}). \tag{7}$$

If $x \in \mathcal{H}$ is a solution to (4), then there exists $v \in \mathcal{G}$ such that (x, v) is a primal-dual solution to (4)–(5). On the other hand, if $v \in \mathcal{G}$ is a solution to (5), then there exists $x \in \mathcal{H}$ such that (x, v) is a primal-dual solution to (4)–(5). Furthermore, if $(x, v) \in \mathcal{H} \times \mathcal{G}$ is a primal-dual solution to (4)–(5), then x is a solution to (4) and v is a solution to (5).

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, $f \in \Gamma(\mathcal{H})$, $g \in \Gamma(\mathcal{G})$, $h : \mathcal{H} \to \mathbb{R}$ a convex and Fréchet differentiable function with η-Lipschitz continuous gradient, for $\eta \geq 0$, and $L : \mathcal{H} \to \mathcal{G}$ a linear continuous operator. Consider the primal convex optimization problem

$$\inf_{x \in \mathcal{H}} \{f(x) + h(x) + g(\text{Lx})\} \tag{8}$$

and its Fenchel dual problem

$$\sup_{v\in\mathcal{G}}\{-(f^*\,\square\, h^*)(-L^*v)-g^*(v)\}. \tag{9}$$

The system of optimality conditions for the primal-dual pair of optimization problems (8)–(9) reads as follows:

$$-L^*\overline{v}-\nabla h(\overline{x})\in\partial f(\overline{x}) \text{ and } \overline{v}\in\partial g(L\overline{x}), \tag{10}$$

which is actually a particular formulation of (7) when

$$A:=\partial f,\ C:=\nabla h,\ B:=\partial g. \tag{11}$$

Notice that, due to the Baillon-Haddad Theorem (see [4, Corollary 18.16]), ∇h is η-cocoercive.

If (8) 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 (9), such that (10) holds. If (9) 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 (8), such that (10) holds. Furthermore, if the pair $(x,v)\in\mathcal{H}\times\mathcal{G}$ satisfies relation (10), then x is an optimal solution to (8), v is an optimal solution to (9) and the optimal objective values of (8) and (9) 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:

$$0\in\operatorname{sqri}(\operatorname{dom} g-L(\operatorname{dom} f)) \tag{12}$$

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

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

its strong quasi-relative interior. The topological interior is contained in the strong quasi-relative interior: $\operatorname{int} S\subseteq\operatorname{sqri} S$, 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 $\operatorname{sqri} S=\operatorname{ri} S$, which denotes the topological interior of S relative to its affine hull. Considering again the infinite dimensional setting, we remark that condition (12) is fulfilled, if there exists $x'\in\operatorname{dom} f$ such that $\mathrm{Lx}'\in\operatorname{dom} g$ and g is continuous at Lx'. For further considerations on convex duality, we refer to [4, 6, 7, 25, 42].

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

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

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

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

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

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

Let $\alpha > 0$, $U \in \mathcal{P}_\alpha(\mathcal{H})$ and $A : \mathcal{H} \rightrightarrows \mathcal{H}$ a maximally monotone operator. Then, the operator $(U + A)^{-1} : \mathcal{H} \to \mathcal{H}$ 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

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

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 [18, Lemma 3.7]), where

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

We are now in the position to formulate the algorithm relying on the ADMM paradigm for solving the primal-dual pair of monotone inclusions (4)–(5).

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 $k \geq 0$, the generated the sequence $(x^k, z^k, y^k)_{k\geq 0}$ is as follows:*

$$x^{k+1} = \left(cL^*L + M_1^k + A\right)^{-1}\left[cL^*(z^k - c^{-1}y^k) + M_1^k x^k - Cx^k\right] \tag{13}$$

$$z^{k+1} = \left(\mathrm{Id} + c^{-1}M_2^k + c^{-1}B\right)^{-1}\left[Lx^{k+1} + c^{-1}y^k + c^{-1}M_2^k z^k\right] \tag{14}$$

$$y^{k+1} = y^k + c(Lx^{k+1} - z^{k+1}). \tag{15}$$

The choice of variable metrics is mainly motivated by the fact this allows make use of variable step sizes, as we will show in Section 3 in the context of primal-dual algorithms. In [31], variable metrics have been used in the context of an ADMM iterative scheme. We refer the reader to [35], where the positive impact of variable metrics on the performances of numerical optimization algorithms is emphasized.

We show below that several algorithms from the literature can be embedded in the iterative scheme of Algorithm 3.

Remark 4 For all $k \geq 0$, the Eqs. 13 and (14) are equivalent to

$$\mathrm{cL}^*(z^k - \mathrm{Lx}^{k+1} - c^{-1}y^k) + M_1^k(x^k - x^{k+1}) - \mathrm{Cx}^k \in \mathrm{Ax}^{k+1}, \tag{16}$$

and, respectively,

$$c(\mathrm{Lx}^{k+1} - z^{k+1} + c^{-1}y^k) + M_2^k(z^k - z^{k+1}) \in \mathrm{Bz}^{k+1}. \tag{17}$$

Notice that the latter is equivalent to

$$y^{k+1} + M_2^k(z^k - z^{k+1}) \in \mathrm{Bz}^{k+1}. \tag{18}$$

In the variational setting as described in Problem 2, namely, by choosing the operators as in (11), the inclusion (16) becomes

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

which is equivalent to

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

On the other hand, (17) becomes

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

which is equivalent to

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

Consequently, the iterative scheme (13)–(15) reads as follows:

$$x^{k+1} = \underset{x\in\mathcal{H}}{\operatorname{argmin}} \left\{ f(x) + \langle x - x^k, \nabla h(x^k)\rangle + \frac{c}{2}\|Lx - z^k + c^{-1}y^k\|^2 + \frac{1}{2}\|x - x^k\|^2_{M_1^k} \right\} \quad (19)$$

$$z^{k+1} = \underset{z\in\mathcal{G}}{\operatorname{argmin}} \left\{ g(z) + \frac{c}{2}\|\mathrm{L}x^{k+1} - z + c^{-1}y^k\|^2 + \frac{1}{2}\|z - z^k\|^2_{M_2^k} \right\} \quad (20)$$

$$y^{k+1} = y^k + c(\mathrm{L}x^{k+1} - z^{k+1}), \quad (21)$$

which is the algorithm formulated and investigated by Banert et al. in [3]. 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 [38] (see also [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 (19)–(21) collapses into the classical version of the ADMM algorithm.

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

(i) Let $k \geq 0$ be fixed. Relation (13) (written for x^{k+2}) reads

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

From (15), we have

$$c\mathrm{L}^*(z^{k+1} - c^{-1}y^{k+1}) = -L^*(2y^{k+1} - y^k) + c\mathrm{L}^*\mathrm{L}x^{k+1};$$

hence,

$$\begin{aligned} x^{k+2} &= \left(\tau_{k+1}^{-1}\mathrm{Id} + A\right)^{-1}\left[\tau_{k+1}^{-1}x^{k+1} - L^*(2y^{k+1} - y^k) - \mathrm{C}x^{k+1}\right] \\ &= J_{\tau_{k+1}A}\left(x^{k+1} - \tau_{k+1}\mathrm{C}x^{k+1} - \tau_{k+1}L^*(2y^{k+1} - y^k)\right). \end{aligned} \quad (22)$$

On the other hand, by using (3), relation (14) reads as folloes:

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

which is equivalent to

$$c\mathrm{L}x^{k+1} + y^k - cz^{k+1} = J_{cB^{-1}}\left(c\mathrm{L}x^{k+1} + y^k\right).$$

By using again (15), this can be reformulated as follows:

$$y^{k+1} = J_{\mathrm{cB}^{-1}}\left(y^k + \mathrm{cLx}^{k+1}\right). \tag{23}$$

The iterative scheme in (22)–(23) generates, for a given starting point $(x^1, y^0) \in \mathcal{H} \times \mathcal{G}$ and a fixed $c > 0$, a sequence $(x^k, y^k)_{k\geq 1}$ as follows:

$$y^{k+1} = J_{\mathrm{cB}^{-1}}\left(y^k + \mathrm{cLx}^{k+1}\right) \tag{24}$$

$$x^{k+2} = J_{\tau_{k+1}A}\left(x^{k+1} - \tau_{k+1}\mathrm{Cx}^{k+1} - \tau_{k+1}L^*(2y^{k+1} - y^k)\right). \tag{25}$$

When $\tau_k = \tau$ for all $k \geq 1$, the algorithm (24)–(25) recovers a numerical scheme for solving monotone inclusion problems proposed by Vũ in [41, Theorem 3.1]. More precisely, the error-free variant of the algorithm in [41, 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 (6)–(4) (by reversing the order in Problem 1, that is, by treating (6) as the primal monotone inclusion and (4) as its dual monotone inclusion) is nothing else than the iterative scheme (24)–(25).

In case $C = 0$, (24)–(25) becomes for all $k \geq 0$

$$x^{k+1} = J_{\tau_k A}\left(x^k - \tau_k L^*(2y^k - y^{k-1})\right) \tag{26}$$

$$y^{k+1} = J_{cB^{-1}}\left(y^k + \mathrm{cLx}^{k+1}\right), \tag{27}$$

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 [8, Algorithm 1, Theorem 2] applied to the solving of the primal-dual pair (6)–(4) (by reversing the order in Problem 1).

(ii) Considering again the variational setting as described in Problem 2, the algorithm (24)–(25) reads for all $k \geq 0$

$$y^{k+1} = \operatorname{prox}_{cg^*}\left(y^k + \mathrm{cLx}^{k+1}\right)$$

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

When $\tau_k = \tau > 0$ for all $k \geq 1$, one recovers a primal-dual algorithm investigated under the assumption $\frac{1}{\tau} - c\|L\|^2 > \frac{\eta}{2}$ by Condat in [20, Algorithm 3.2, Theorem 3.1].

Not least, (26)–(27) reads in the variational setting (which corresponds to the case $h = 0$) for all $k \geq 0$

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

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

When $\tau_k = \tau > 0$ for all $k \geq 1$, this iterative schemes becomes the algorithm proposed by Chambolle and Pock in [14, Algorithm 1, Theorem 1] for solving in case $h = 0$ the primal-dual pair of optimization problems (9)–(8) (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 [17, 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 (4)–(5) is nonempty. Let $(x^k, z^k, y^k)_{k\geq 0}$ be the sequence generated by Algorithm 3 and assume that $M_1^k - \frac{\eta}{2}\operatorname{Id} \in \mathcal{S}_+(\mathcal{H})$, $M_1^k \succcurlyeq M_1^{k+1}$, $M_2^k \in \mathcal{S}_+(\mathcal{G})$, $M_2^k \succcurlyeq M_2^{k+1}$ for all $k \geq 0$. If one of the following assumptions:*

(I) there exists $\alpha_1 > 0$ such that $M_1^k - \frac{\eta}{2}\operatorname{Id} \in \mathcal{P}_{\alpha_1}(\mathcal{H})$ for all $k \geq 0$;
(II) there exist $\alpha, \alpha_2 > 0$ such that $M_1^k - \frac{\eta}{2}\operatorname{Id} + L^*L \in \mathcal{P}_\alpha(\mathcal{H})$ and $M_2^k \in \mathcal{P}_{\alpha_2}(\mathcal{G})$ for all $k \geq 0$;
(III) there exists $\alpha > 0$ such that $M_1^k - \frac{\eta}{2}\operatorname{Id} + L^*L \in \mathcal{P}_\alpha(\mathcal{H})$ 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 (4)–(5), 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

$$S = \{(x, Lx, v) : (x, v) \text{ is a primal-dual solution to (4)–(5)}\}. \tag{28}$$

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

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

Let $k \geq 0$ be fixed. From (16) and the monotonicity of A, we have

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

while from (17) and the monotonicity of B, we have

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

Since C is η-cocoercive, we have

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

We consider first the case when $\eta > 0$.

Summing up the three inequalities obtained above, we get

$$
\begin{aligned}
&c\langle z^k - \mathrm{L}x^{k+1}, \mathrm{L}x^{k+1} - \mathrm{L}x^*\rangle + \langle y^* - y^k, \mathrm{L}x^{k+1} - \mathrm{L}x^*\rangle + \langle \mathrm{C}x^* - \mathrm{C}x^k, x^{k+1} - x^*\rangle \\
&+\langle M_1^k(x^k - x^{k+1}), x^{k+1} - x^*\rangle + c\langle \mathrm{L}x^{k+1} - z^{k+1}, z^{k+1} - \mathrm{L}x^*\rangle + \langle y^k - y^*, z^{k+1} - \mathrm{L}x^*\rangle \\
&+\langle M_2^k(z^k - z^{k+1}), z^{k+1} - \mathrm{L}x^*\rangle + \langle \mathrm{C}x^* - \mathrm{C}x^k, x^* - x^k\rangle - \eta^{-1}\|\mathrm{C}x^* - \mathrm{C}x^k\|^2 \geq 0. \qquad (29)
\end{aligned}
$$

According to (15), we also have

$$
\begin{aligned}
\langle y^* - y^k, \mathrm{L}x^{k+1} - \mathrm{L}x^*\rangle + \langle y^k - y^*, z^{k+1} - \mathrm{L}x^*\rangle &= \langle y^* - y^k, \mathrm{L}x^{k+1} - z^{k+1}\rangle \\
&= c^{-1}\langle y^* - y^k, y^{k+1} - y^k\rangle. \qquad (30)
\end{aligned}
$$

By expressing the inner products through norms, we further derive

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

By expressing $\mathrm{L}x^{k+1} - z^{k+1}$ through relation (15) and by taking into account that

$$
\begin{aligned}
&\langle \mathrm{C}x^* - \mathrm{C}x^k, x^{k+1} - x^k\rangle - \eta^{-1}\|\mathrm{C}x^* - \mathrm{C}x^k\|^2 = \\
&-\eta\left\|\eta^{-1}\left(\mathrm{C}x^* - \mathrm{C}x^k\right) + \frac{1}{2}\left(x^k - x^{k+1}\right)\right\|^2 + \frac{\eta}{4}\|x^k - x^{k+1}\|^2, \qquad (31)
\end{aligned}
$$

we obtain

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

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

$$\begin{aligned}
&\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-\mathrm{Lx}^*\|^2_{M_2^{k+1}+c\,\mathrm{Id}}+\frac{1}{2c}\|y^{k+1}-y^*\|^2\\
\leq\ &\frac{1}{2}\|x^{k}-x^*\|^2_{M_1^{k}}+\frac{1}{2}\|z^{k}-\mathrm{Lx}^*\|^2_{M_2^{k}+c\,\mathrm{Id}}+\frac{1}{2c}\|y^{k}-y^*\|^2\\
&-\frac{c}{2}\|z^k-\mathrm{Lx}^{k+1}\|^2-\frac{1}{2}\|x^k-x^{k+1}\|^2_{M_1^k-\frac{\eta}{2}\mathrm{Id}}-\frac{1}{2}\|z^k-z^{k+1}\|^2_{M_2^k}\\
&-\eta\left\|\eta^{-1}\left(\mathrm{Cx}^*-\mathrm{Cx}^k\right)+\frac{1}{2}\left(x^k-x^{k+1}\right)\right\|^2. \qquad (32)
\end{aligned}$$

In case $\eta=0$, similar arguments lead to the inequality

$$\begin{aligned}
&\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-\mathrm{Lx}^*\|^2_{M_2^{k+1}+c\,\mathrm{Id}}+\frac{1}{2c}\|y^{k+1}-y^*\|^2\\
\leq\ &\frac{1}{2}\|x^{k}-x^*\|^2_{M_1^{k}}+\frac{1}{2}\|z^{k}-\mathrm{Lx}^*\|^2_{M_2^{k}+c\,\mathrm{Id}}+\frac{1}{2c}\|y^{k}-y^*\|^2\\
&-\frac{c}{2}\|z^k-\mathrm{Lx}^{k+1}\|^2-\frac{1}{2}\|x^k-x^{k+1}\|^2_{M_1^k}-\frac{1}{2}\|z^k-z^{k+1}\|^2_{M_2^k}. \qquad (33)
\end{aligned}$$

By using telescoping arguments, one can easily see that both (32) and (33) imply

$$\sum_{k\geq 0}\|z^k-\mathrm{Lx}^{k+1}\|^2<+\infty,\ \sum_{k\geq 0}\|x^k-x^{k+1}\|^2_{M_1^k-\frac{\eta}{2}\mathrm{Id}}<+\infty,\ \sum_{k\geq 0}\|z^k-z^{k+1}\|^2_{M_2^k}<+\infty. \qquad (34)$$

Consider first the hypotheses in assumption (I).

Discarding the negative terms on the right-hand side of both (32) and (33), 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 $W^k := (M_1^k, M_2^k+c\,\mathrm{Id}, c^{-1}\,\mathrm{Id})$ for $k\geq 0$, and for S defined as in (28).

Since $M_1^k-\frac{\eta}{2}\mathrm{Id}\in\mathcal{P}_{\alpha_1}(\mathcal{H})$ for all $k\geq 0$ with $\alpha_1>0$, from (34), we get

$$x^k-x^{k+1}\to 0\ (k\to+\infty) \qquad (35)$$

and

$$z^k-\mathrm{Lx}^{k+1}\to 0\ (k\to+\infty). \qquad (36)$$

A direct consequence of (35) and (36) is

$$z^k-z^{k+1}\to 0\ (k\to+\infty). \qquad (37)$$

From (15), (36), and (37), we derive

$$y^k-y^{k+1}\to 0\ (k\to+\infty). \qquad (38)$$

The relations (35)–(38) will play an essential role when verifying assumption (ii) in the Opial Lemma for S taken as in (28). Let $(\overline{x},\overline{z},\overline{y})\in\mathcal{H}\times\mathcal{G}\times\mathcal{G}$ 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 $(\overline{x},\overline{z},\overline{y})$ (as $n\to+\infty$).

From (35), we obtain that $(\mathrm{L}x^{k_n+1})_{n\in\mathbb{N}}$ converges weakly to $L\overline{x}$ (as $n \to +\infty$), which combined with (36) yields $\overline{z} = L\overline{x}$. We use now the following notations for $n \geq 0$:

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

From (16), we have for all $n \geq 0$

$$a_n^* \in (A + C)(a_n). \tag{39}$$

Further, from (17) and (15), we have for all $n \geq 0$

$$b_n^* \in Bb_n. \tag{40}$$

Furthermore, from (35), we have

$$a_n \text{ converges weakly to } \overline{x} \text{ (as } n \to +\infty). \tag{41}$$

From (38) and (37), we obtain

$$b_n^* \text{ converges weakly to } \overline{y} \text{ (as } n \to +\infty). \tag{42}$$

Moreover, (15) and (38) yield

$$\mathrm{L}a_n - b_n \text{ converges strongly to } 0 \text{ (as } n \to +\infty). \tag{43}$$

Finally, we have

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

By using the fact that C is η-Lipschitz continuous, from (35)–(38), we get

$$a_n^* + L^*b_n^* \text{ converges strongly to } 0 \text{ (as } n \to +\infty). \tag{44}$$

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 $K : \mathcal{H} \times \mathcal{G} \to \mathcal{H} \times \mathcal{G}$ 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 [4]), $A + C$ is maximally monotone, too (see [4]); thus, T is maximally monotone. Since K is s skew operator, it is also maximally monotone (see [4]). Due to the fact that K has full domain, we conclude that

$$T + K \text{ is a maximally monotone operator.} \tag{45}$$

Moreover, from (39) and (40), we have

$$(a_n^* + L^*b_n^*, b_n - \mathrm{L}a_n) \in (T + K)(a_n, b_n^*) \ \forall n \geq 0. \tag{46}$$

Since the graph of a maximally monotone operator is sequentially closed with respect to the weak×strong topology (see [4, Proposition 20.33]), from (45), (46), (41), (42), (43), and (44), we derive that

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

The latter is nothing else than saying that $(\overline{x}, \overline{y})$ is a primal dual solution to (4)–(5), which combined with $\overline{z} = L\overline{x}$ 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 (4)–(5).

Consider now the hypotheses in assumption (II).

We start by showing that the relations (35)–(38) are fulfilled in this situation, too. Indeed, in this case, we derive from (34) that (36) and (37) hold. From (15), (36), and (37), we obtain (38). Finally, the inequalities

$$\begin{aligned}\alpha\|x^{k+1}-x^k\|^2 &\leq \|x^{k+1}-x^k\|^2_{M_1^k-\frac{\eta}{2}\operatorname{Id}} + \|\mathrm{L}x^{k+1}-\mathrm{L}x^k\|^2\\ &\leq \|x^{k+1}-x^k\|^2_{M_1^k-\frac{\eta}{2}\operatorname{Id}} + 2\|\mathrm{L}x^{k+1}-z^k\|^2 + 2\|z^k-\mathrm{L}x^k\|^2 \ \forall k\geq 0\end{aligned}$$

yield (35).

On the other hand, notice that both (32) and (33) yield

$$\exists \lim_{k\to+\infty}\left(\frac{1}{2}\|x^k-x^*\|^2_{M_1^k} + \frac{1}{2}\|z^k-Lx^*\|^2_{M_2^k+c\operatorname{Id}} + \frac{1}{2c}\|y^k-y^*\|^2\right); \tag{47}$$

hence, $(y^k)_{k\geq 0}$ and $(z^k)_{k\geq 0}$ are bounded. Combining this with (15) and the condition imposed on $M_1^k - \frac{\eta}{2}\operatorname{Id} + L^*L$, 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 proof of (I), 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 (28).

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 $(k_p)_{p\geq 0}$, $(k_q)_{q\geq 0}$, $k_p \to +\infty$ (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 shown above, (x_1, z_1, y_1) and (x_2, z_2, y_2) belong to the set S (see (28)); thus, $z_i = Lx_i$, $i \in \{1, 2\}$. From (47), which is true for every primal-dual solution to (4)–(5), we derive

$$\exists \lim_{k\to+\infty}\Big(E(x^k, z^k, y^k; x_1, \mathrm{L}x_1, y_1) - E(x^k, z^k, y^k; x_2, \mathrm{L}x_2, y_2)\Big), \tag{48}$$

where, for $(x^*, \mathrm{L}x^*, y^*)$ the expression $E(x^k, z^k, y^k; x^*, \mathrm{L}x^*, y^*)$ is defined as

$$E(x^k, z^k, y^k; x^*, \mathrm{L}x^*, y^*) = \frac{1}{2}\|x^k-x^*\|^2_{M_1^k} + \frac{1}{2}\|z^k-\mathrm{L}x^*\|^2_{M_2^k+c\operatorname{Id}} + \frac{1}{2c}\|y^k-y^*\|^2.$$

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

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

$$\begin{aligned}\frac{1}{2}\|z^k-\mathrm{L}x_1\|^2_{M_2^k+c\operatorname{Id}} - \frac{1}{2}\|z^k-\mathrm{L}x_2\|^2_{M_2^k+c\operatorname{Id}} = \frac{1}{2}\|\mathrm{L}x_2-\mathrm{L}x_1\|^2_{M_2^k+c\operatorname{Id}} + \langle z^k-\mathrm{L}x_2,\\ (M_2^k+c\operatorname{Id})(\mathrm{L}x_2-\mathrm{L}x_1)\rangle,\end{aligned}$$

and

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

Applying [36, Théorèème 104.1], there exists $M_1 \in \mathcal{S}_+(\mathcal{H})$ 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 $\sup_{k\geq 0}\|M_2^k + c\,\text{Id}\| < +\infty$. Thus, according to [17, Lemma 2.3], 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 in the strong topology (as $k \to +\infty$).

Taking the limit in (48) along the subsequences $(k_p)_{p\geq 0}$ and $(k_q)_{q\geq 0}$ and using the last three relations above, we obtain

$$\frac{1}{2}\|x_1 - x_2\|_{M_1}^2 + \langle x_1 - x_2, M_1(x_2 - x_1)\rangle + \frac{1}{2}\|\text{L}x_1 - \text{L}x_2\|_{M_2}^2 + \langle \text{L}x_1 - \text{L}x_2, M_2(\text{L}x_2 - \text{L}x_1)\rangle$$

$$+\frac{1}{2c}\|y_1 - y_2\|^2 + \frac{1}{c}\langle y_1 - y_2, y_2 - y_1\rangle = \frac{1}{2}\|x_1 - x_2\|_{M_1}^2 + \frac{1}{2}\|Lx_1 - \text{L}x_2\|_{M_2}^2 + \frac{1}{2c}\|y_1 - y_2\|^2,$$

hence,

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

thus, $\|x_1 - x_2\|_{M_1} = 0$, $\text{L}x_1 = \text{L}x_2$ and $y_1 = y_2$. Since,

$$\left(\alpha + \frac{\eta}{2}\right)\|x_1 - x_2\|^2 \leq \|x_1 - x_2\|_{M_1}^2 + \|\text{L}x_1 - \text{L}x_2\|^2,$$

we obtain that $x_1 = x_2$. In conclusion, $(x^k, z^k, y^k)_{k\geq 0}$ converges weakly to an element in S (see (28)).

Finally, consider the hypotheses in assumption (III). We start by refining the inequalities obtained in (32) and (33).

By considering the relation (18) 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 $y^{k+1} - y^k = c(\text{L}x^{k+1} - z^{k+1})$ in the last inequality, it follows:

$$\begin{aligned}&\|z^{k+1} - z^k\|_{M_2^k}^2 - \frac{1}{2}\|z^{k+1} - z^k\|_{M_2^{k-1}}^2 - \frac{1}{2}\|z^k - z^{k-1}\|_{M_2^{k-1}}^2 \\ &\leq \frac{c}{2}\left(\|z^k - \text{L}x^{k+1}\|^2 - \|z^{k+1} - z^k\|^2 - \|\text{L}x^{k+1} - z^{k+1}\|^2\right).\end{aligned} \quad (49)$$

In case $\eta > 0$, adding (49) and (32) leads to the following:

$$\begin{aligned}
&\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-Lx^*\|^2_{M_2^{k+1}+c\,\mathrm{Id}}+\frac{1}{2c}\|y^{k+1}-y^*\|^2+\frac{1}{2}\|z^{k+1}-z^k\|^2_{3M_2^k-M_2^{k-1}}\\
&\quad\le \frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-Lx^*\|^2_{M_2^k+c\,\mathrm{Id}}+\frac{1}{2c}\|y^k-y^*\|^2+\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}\\
&\quad-\eta\left\|\eta^{-1}\left(Cx^*-Cx^k\right)+\frac{1}{2}\left(x^k-x^{k+1}\right)\right\|^2-\frac{1}{2}\|x^{k+1}-x^k\|^2_{M_1^k-\frac{\eta}{2}\mathrm{Id}}\\
&\quad-\frac{c}{2}\|z^{k+1}-z^k\|^2-\frac{1}{2c}\|y^{k+1}-y^k\|^2.
\end{aligned}$$

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 $k\ge 1$ it holds

$$\begin{aligned}
&\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-Lx^*\|^2_{M_2^{k+1}+c\,\mathrm{Id}}+\frac{1}{2c}\|y^{k+1}-y^*\|^2+\frac{1}{2}\|z^{k+1}-z^k\|^2_{M_2^k}\\
\le\ &\frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-Lx^*\|^2_{M_2^k+c\,\mathrm{Id}}+\frac{1}{2c}\|y^k-y^*\|^2+\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}\\
&-\frac{1}{2}\|x^{k+1}-x^k\|^2_{M_1^k-\frac{\eta}{2}\mathrm{Id}}-\frac{c}{2}\|z^{k+1}-z^k\|^2-\frac{1}{2c}\|y^{k+1}-y^k\|^2. \qquad (50)
\end{aligned}$$

Similarly, we obtain in case $\eta=0$ for all $k\ge 1$, the inequality

$$\begin{aligned}
&\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-\mathrm{L}x^*\|^2_{M_2^{k+1}+c\,\mathrm{Id}}+\frac{1}{2c}\|y^{k+1}-y^*\|^2+\frac{1}{2}\|z^{k+1}-z^k\|^2_{M_2^k}\\
\le\ &\frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-Lx^*\|^2_{M_2^k+c\,\mathrm{Id}}+\frac{1}{2c}\|y^k-y^*\|^2+\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}\\
&-\frac{1}{2}\|x^{k+1}-x^k\|^2_{M_1^k}-\frac{c}{2}\|z^{k+1}-z^k\|^2-\frac{1}{2c}\|y^{k+1}-y^k\|^2. \qquad (51)
\end{aligned}$$

Using telescoping sum arguments, we obtain that $\|x^{k+1}-x^k\|_{M_1^k-\frac{\eta}{2}\mathrm{Id}}\to 0$, $y^k-y^{k+1}\to 0$ and $z^k-z^{k+1}\to 0$ as $k\to+\infty$. Using (15), it follows that $L(x^k-x^{k+1})\to 0$ as $k\to+\infty$, which, combined with $M_1^k-\frac{\eta}{2}\mathrm{Id}+L^*L\in\mathcal{P}_\alpha(\mathcal{H}), k\ge 0$, further implies that $x^k-x^{k+1}\to 0$ as $k\to+\infty$. Consequently, $z^k-\mathrm{L}x^{k+1}\to 0$ as $k\to+\infty$. Hence, the relations (35)–(38) are fulfilled. On the other hand, from both (50) and (51), we derive

$$\exists \lim_{k\to+\infty}\left(\frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-Lx^*\|^2_{M_2^k+c\,\mathrm{Id}}+\frac{1}{2c}\|y^k-y^*\|^2+\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}\right).$$

By using that

$$\|z^k-z^{k-1}\|^2_{M_2^{k-1}}\le\|z^k-z^{k-1}\|^2_{M_2^0}\le\|M_2^0\|\|z^k-z^{k-1}\|^2\ \forall k\ge 1,$$

it follows that $\lim_{k\to+\infty}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}=0$, which further implies that (47) holds. From here, the conclusion follows by arguing as in the proof provided above in the setting of assumption (II). □

Remark 8 (i) Choosing as in Remark 5 $M_1^k := \frac{1}{\tau_k}\,\mathrm{Id} - \mathrm{c}\mathrm{L}^*L$, with $(\tau_k)_{k\geq 0}$ a monotonically nondecreasing sequence of positive numbers and $\tau := \sup_{k\geq 0}\tau_k \in \mathbb{R}$, and $M_2^k := 0$ for all $k \geq 0$, we have

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

This shows that under the assumption $\frac{1}{\tau} - c\|L\|^2 > \frac{\eta}{2}$ (which recovers the one in Algorithm 3.2 and Theorem 3.1 in [20]) the operators $M_1^k - \frac{\eta}{2}\,\mathrm{Id}$ belong for all $k \geq 0$ to the class $\mathcal{P}_{\alpha_1}(\mathcal{H})$, with $\alpha_1 := \frac{1}{\tau} - c\|L\|^2 - \frac{\eta}{2} > 0$.

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

$$\exists\alpha > 0 \text{ such that } L^*L \in \mathcal{P}_\alpha(\mathcal{H}). \tag{52}$$

By taking into account [4, Fact 2.19], one can see that (52) holds if and only if L is injective and ran L^* is closed. This means that if ran L^* is closed, then (52) 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 $m \geq n \geq 1$, (52) 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.

Remark 9 In the finite dimensional variational case, the sequences generated by the classical ADMM algorithm, which corresponds to the iterative scheme (19)–(21) 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).

Remark 10 An anonymous referee asked whether it is possible to perform a similar analysis for a slight modification of Algorithm 3, in which (15) is replaced through

$$y^{k+1} = y^k + c\nu(\mathrm{L}\mathrm{x}^{k+1} - z^{k+1}), \tag{53}$$

where $\nu \in \left(0, \frac{\sqrt{5}+1}{2}\right)$. It has been noticed in [28] that the numerical performances of the classical ADMM algorithm for convex optimization problems, under the use of a relaxation parameter $\nu > 1$, outperform the ones obtained when $\nu = 1$.

In this remark, we give a positive answer to the question posed by the reviewer. To this end, we consider as follows Algorithm 3 with (15) replaced by (53), where $1 < \nu < \frac{\sqrt{5}+1}{2}$, and work under the hypotheses (III) of Theorem 7. We will prove that one can derive in this new setting inequalities, which are similar to (50) and (51), respectively.

Let $k \geq 0$ be fixed. Take first $\eta > 0$. We have relation (29), while instead of (30) we get

$$\begin{aligned}
&\langle y^* - y^k, \mathrm{L}\mathrm{x}^{k+1} - \mathrm{L}\mathrm{x}^*\rangle + \langle y^k - y^*, z^{k+1} - \mathrm{L}\mathrm{x}^*\rangle \\
&\quad = \langle y^* - y^k, \mathrm{L}\mathrm{x}^{k+1} - z^{k+1}\rangle = (c\nu)^{-1}\langle y^* - y^k, y^{k+1} - y^k\rangle \\
&\quad = (c\nu)^{-1}\langle y^* - y^{k+1}, y^{k+1} - y^k\rangle + c\nu\|\mathrm{L}\,\mathrm{x}^{k+1} - z^{k+1}\|^2.
\end{aligned} \tag{54}$$

Further, we have

$$\begin{aligned}
& c\langle z^k - \mathrm{Lx}^{k+1}, \mathrm{Lx}^{k+1} - \mathrm{Lx}^* \rangle + c\langle \mathrm{Lx}^{k+1} - z^{k+1}, z^{k+1} - \mathrm{Lx}^* \rangle \\
= \ & c\langle z^k - \mathrm{Lx}^{k+1}, \mathrm{Lx}^{k+1} - \mathrm{Lx}^* \rangle + c\langle \mathrm{Lx}^{k+1} - z^{k+1}, z^{k+1} - \mathrm{Lx}^{k+1} \rangle \\
& + c\langle \mathrm{Lx}^{k+1} - z^{k+1}, \mathrm{Lx}^{k+1} - \mathrm{Lx}^* \rangle \\
= \ & c\langle z^k - z^{k+1}, \mathrm{Lx}^{k+1} - \mathrm{Lx}^* \rangle - c\|\mathrm{Lx}^{k+1} - z^{k+1}\|^2 \\
= \ & c\langle z^k - z^{k+1}, \mathrm{Lx}^{k+1} - z^{k+1} \rangle + c\langle z^k - z^{k+1}, z^{k+1} - \mathrm{Lx}^* \rangle - c\|\mathrm{Lx}^{k+1} - z^{k+1}\|^2 \\
= \ & \frac{1}{\nu}\langle z^k - z^{k+1}, y^{k+1} - y^k \rangle + c\langle z^k - z^{k+1}, z^{k+1} - \mathrm{Lx}^* \rangle - c\|\mathrm{Lx}^{k+1} - z^{k+1}\|^2,
\end{aligned}$$

which, combined with (54) and (29), leads to

$$\begin{aligned}
& \frac{1}{\nu}\langle z^k - z^{k+1}, y^{k+1} - y^k \rangle + c\langle z^k - z^{k+1}, z^{k+1} - \mathrm{Lx}^* \rangle - c\|\mathrm{Lx}^{k+1} - z^{k+1}\|^2 \\
& + (c\nu)^{-1}\langle y^* - y^{k+1}, y^{k+1} - y^k \rangle + c\nu\|\mathrm{Lx}^{k+1} - z^{k+1}\|^2 \\
& + \langle M_1^k(x^k - x^{k+1}), x^{k+1} - x^* \rangle + \langle M_2^k(z^k - z^{k+1}), z^{k+1} - \mathrm{Lx}^* \rangle \\
& + \langle \mathrm{Cx}^* - \mathrm{Cx}^k, x^{k+1} - x^k \rangle - \eta^{-1}\|\mathrm{Cx}^* - \mathrm{Cx}^k\|^2 \geq 0. \qquad (55)
\end{aligned}$$

In order to estimate the term $\frac{1}{\nu}\langle z^k - z^{k+1}, y^{k+1} - y^k \rangle$, we use the monotonicity of B. Notice that (18) becomes in this case

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

From here, we obtain that for all $k \geq 1$

$$\begin{aligned}
& \langle z^{k+1} - z^k, y^k + \frac{1}{\nu}(y^{k+1} - y^k) + M_2^k(z^k - z^{k+1}) - y^{k-1} - \frac{1}{\nu}(y^k - y^{k-1}) \\
& - M_2^{k-1}(z^{k-1} - z^k)\rangle \geq 0;
\end{aligned}$$

hence,

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

For $\beta := c\nu^2$ (this choice for β will be clarified later), we have for all $k \geq 1$, the inequality

$$\beta\|z^{k+1} - z^k\|^2 + \beta^{-1}\|y^k - y^{k-1}\|^2 \geq 2\langle z^{k+1} - z^k, y^k - y^{k-1} \rangle.$$

By expressing the inner products through norms, we derive from (31) and (55) for all $k \geq 1$

$$
\begin{aligned}
&\frac{1}{2}\left(1-\frac{1}{\nu}\right)\left(\beta\|z^{k+1}-z^k\|^2+\frac{1}{\beta}\|y^k-y^{k-1}\|^2\right)-c(1-\nu)\|\mathrm{Lx}^{k+1}-z^{k+1}\|^2\\
&+\frac{c}{2}\left(\|z^k-\mathrm{Lx}^*\|^2-\|z^k-z^{k+1}\|^2-\|z^{k+1}-\mathrm{Lx}^*\|^2\right)\\
&+\frac{1}{2c\nu}\left(\|y^*-y^k\|^2-\|y^{k+1}-y^k\|^2-\|y^{k+1}-y^*\|^2\right)\\
&+\frac{1}{2}\left(\|x^k-x^*\|^2_{M_1^k}-\|x^k-x^{k+1}\|^2_{M_1^k}-\|x^{k+1}-x^*\|^2_{M_1^k}\right)\\
&+\frac{1}{2}\left(\|z^k-\mathrm{Lx}^*\|^2_{M_2^k}-\|z^k-z^{k+1}\|^2_{M_2^k}-\|z^{k+1}-\mathrm{Lx}^*\|^2_{M_2^k}\right)\\
&-\eta\left\|\eta^{-1}\left(\mathrm{Cx}^*-\mathrm{Cx}^k\right)+\frac{1}{2}\left(x^k-x^{k+1}\right)\right\|^2+\frac{\eta}{4}\|x^k-x^{k+1}\|^2\\
&-\|z^{k+1}-z^k\|^2_{M_2^k}+\frac{1}{2}\|z^{k+1}-z^k\|^2_{M_2^{k-1}}+\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}\geq 0.
\end{aligned}
$$

The coefficient of $\|z^{k+1}-z^k\|^2$ is $\frac{\beta}{2}\left(1-\frac{1}{\nu}\right)-\frac{c}{2}=-\frac{c}{2}(1+\nu-\nu^2)$. Taking into account (53), it yields that the coefficient of $\|\mathrm{Lx}^k-z^k\|^2$ is $\frac{1}{2}\frac{\nu-1}{\nu}\frac{1}{\beta}c^2\nu^2=\frac{c}{2}\left(1-\frac{1}{\nu}\right)$. On the other hand, the coefficient of $\|\mathrm{Lx}^{k+1}-z^{k+1}\|^2$ is $-c(1-\nu)-\frac{c\nu}{2}=c\frac{\nu-2}{2}$, while we have

$$
c\frac{\nu-2}{2}=-\frac{c}{2}\left(1-\frac{1}{\nu}\right)+\frac{c}{2}\nu^{-1}(-1-\nu+\nu^2).
$$

Taking into account the monotonicity of $(M_1^k)_{k\geq 0}$ and $(M_2^k)_{k\geq 0}$, and that $3M_2^k-M_2^{k-1}\succcurlyeq M_2^k$ for all $k\geq 1$, we finally obtain

$$
\begin{aligned}
&\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-\mathrm{Lx}^*\|^2_{M_2^{k+1}+c\,\mathrm{Id}}\\
&+\frac{1}{2c\nu}\|y^{k+1}-y^*\|^2+\frac{1}{2}\|z^{k+1}-z^k\|^2_{M_2^k}+\frac{c}{2}\left(1-\frac{1}{\nu}\right)\|\mathrm{Lx}^{k+1}-z^{k+1}\|^2\\
\leq\ &\frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-\mathrm{Lx}^*\|^2_{M_2^k+c\,\mathrm{Id}}+\frac{1}{2c\nu}\|y^k-y^*\|^2\\
&+\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}+\frac{c}{2}\left(1-\frac{1}{\nu}\right)\|\mathrm{Lx}^k-z^k\|^2\\
&-\frac{1}{2}\|x^{k+1}-x^k\|^2_{M_1^k-\frac{\eta}{2}\mathrm{Id}}-\frac{c}{2}(1+\nu-\nu^2)\|z^{k+1}\|z^{k+1}-z^k\|^2\\
&-z^k\|^2-\frac{c}{2}\nu^{-1}(1+\nu-\nu^2)\|y^{k+1}-y^k\|^2. \qquad (56)
\end{aligned}
$$

In a similar way, we obtain in case $\eta = 0$ for all $k \geq 1$, the inequality

$$\begin{aligned}
& \frac{1}{2}\|x^{k+1} - x^*\|^2_{M_1^{k+1}} + \frac{1}{2}\|z^{k+1} - \mathrm{L}x^*\|^2_{M_2^{k+1}+c\,\mathrm{Id}} + \frac{1}{2c\nu}\|y^{k+1} - y^*\|^2 \\
& + \frac{1}{2}\|z^{k+1} - z^k\|^2_{M_2^k} + \frac{c}{2}\left(1 - \frac{1}{\nu}\right)\|\mathrm{L}x^{k+1} - z^{k+1}\|^2 \\
\leq\ & \frac{1}{2}\|x^k - x^*\|^2_{M_1^k} + \frac{1}{2}\|z^k - \mathrm{L}x^*\|^2_{M_2^k+c\,\mathrm{Id}} + \frac{1}{2c\nu}\|y^k - y^*\|^2 \\
& + \frac{1}{2}\|z^k - z^{k-1}\|^2_{M_2^{k-1}} + \frac{c}{2}\left(1 - \frac{1}{\nu}\right)\|\mathrm{L}x^k - z^k\|^2 \\
& - \frac{1}{2}\|x^{k+1} - x^k\|^2_{M_1^k} - \frac{c}{2}(1 + \nu - \nu^2)\|z^{k+1} - z^k\|^2 \\
& - \frac{c}{2}\nu^{-1}(1 + \nu - \nu^2)\|y^{k+1} - y^k\|^2. \qquad (57)
\end{aligned}$$

In other words, we obtain (56) instead of (50) and (57) instead of (51), respectively. By using the same arguments as in the proof of Theorem 7, we obtain the convergence of the ADMM algorithm for monotone operators modified according to (53).

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-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 [13]). Notice that due to the celebrated Baillon-Haddad Theorem (see, for instance, [4, Corollary 8.16]), the gradient of a convex and Fréchet differentiable function is η-cocoercive if and only if it is η-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 (4) has at most one solution. Hence, if (x, v) is a primal-dual solution to the primal-dual pair (4)–(5), then x is the unique solution to (4). Notice that the problem (5) 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 $M_2^k : \mathcal{G} \to \mathcal{G}$ 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 $(x^0, z^0, y^0) \in \mathcal{H} \times \mathcal{H} \times \mathcal{G}$. For all $k \geq 0$, the generated the sequence $(x^k, z^k, y^k)_{k\geq 0}$ is as follows:*

$$y^{k+1} = \left(\tau_k LL^* + M_2^k + B^{-1}\right)^{-1} \left[-\tau_k L(z^k - \tau_k^{-1} x^k) + M_2^k y^k\right] \tag{58}$$

$$\begin{aligned} z^{k+1} &= \left(\frac{\theta_k}{\lambda} - 1\right) L^* y^{k+1} + \frac{\theta_k}{\lambda} C x^k + \frac{\theta_k}{\lambda} \left(\mathrm{Id} + \lambda \tau_{k+1}^{-1} A^{-1}\right)^{-1} \\ &\quad \times \left[-L^* y^{k+1} + \lambda \tau_{k+1}^{-1} x^k - C x^k\right] \end{aligned} \tag{59}$$

$$x^{k+1} = x^k + \frac{\tau_{k+1}}{\theta_k} \left(-L^* y^{k+1} - z^{k+1}\right), \tag{60}$$

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 (4) and (6) become

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + (L^* \circ B \circ L)x \tag{61}$$

and, respectively,

$$\text{find } v \in \mathcal{G} \text{ such that } \ 0 \in B^{-1} v + \Big((-L) \circ (A^{-1}) \circ (-L^*)\Big) v. \tag{62}$$

The two problems (61) and (62) are dual to each other in the sense of the Attouch-Théra duality (see [2]). By taking in (58)–(60) $\lambda = 1$, $\theta_k = 1$ (which corresponds to the limit case $\mu = 0$ and $\gamma = 0$ in the Eq. 66 below) and $\tau_k = c > 0$ for all $k \geq 0$, then the resulting iterative scheme reads as follows:

$$\begin{aligned} y^{k+1} &= \left(\mathrm{c\,LL}^* + M_2^k + B^{-1}\right)^{-1} \left[-\mathrm{c\,L}(z^k - c^{-1} x^k) + M_2^k y^k\right] \\ z^{k+1} &= \left(\mathrm{Id} + c^{-1} A^{-1}\right)^{-1} \left[-L^* y^{k+1} + c^{-1} x^k\right] \\ x^{k+1} &= x^k + c\left(-L^* y^{k+1} - z^{k+1}\right). \end{aligned}$$

This is nothing else than Algorithm 3 employed to the solving of the primal-dual system of monotone inclusions (62)–(61), that is, by treating (62) as the primal monotone inclusion and (61) as its dual monotone inclusion (notice that in this case we take in relation (14) of Algorithm 3 $M_2^k = 0$ for all $k \geq 0$).

We chose the parameters involved in Algorithm 14 such that

$$\mu \tau_1 < 2\gamma, \tag{63}$$

$$\lambda \geq \mu + 1, \tag{64}$$

$$\sigma_0 \tau_1 \|L\|^2 \leq 1, \tag{65}$$

and for all $k \geq 0$

$$\theta_k = \frac{1}{\sqrt{1 + \tau_{k+1}\lambda^{-1}(2\gamma - \mu\tau_{k+1})}} \tag{66}$$

$$\tau_{k+2} = \theta_k \tau_{k+1} \tag{67}$$

$$\sigma_{k+1} = \theta_k^{-1}\sigma_k \tag{68}$$

$$\tau_k LL^* + M_2^k \succcurlyeq \sigma_k^{-1} \operatorname{Id} \tag{69}$$

$$\frac{\tau_k}{\tau_{k+1}} LL^* + \frac{1}{\tau_{k+1}} M_2^k \succcurlyeq \frac{\tau_{k+1}}{\tau_{k+2}} LL^* + \frac{1}{\tau_{k+2}} M_2^{k+1}. \tag{70}$$

Remark 16 Fix an arbitrary $k \geq 1$. From (58), we have

$$-\tau_k L(z^k - \tau_k^{-1}x^k) + M_2^k y^k \in \widetilde{M_2}^k y^{k+1} + B^{-1}y^{k+1}, \tag{71}$$

where

$$\widetilde{M_2}^k := \tau_k \mathrm{LL}^* + M_2^k. \tag{72}$$

Due to (60), we have

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

which combined with (71) delivers

$$\widetilde{M_2}^k (y^k - y^{k+1}) + L\left[x^k + \theta_{k-1}(x^k - x^{k-1})\right] \in B^{-1}y^{k+1}. \tag{73}$$

Fix now an arbitrary $k \geq 0$. From (3) and (59), we have

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

By using (60), we obtain

$$x^{k+1} = J_{(\tau_{k+1}/\lambda)A}\left[x^k + \frac{\tau_{k+1}}{\lambda}\left(-L^*y^{k+1} - \mathrm{Cx}^k\right)\right]. \tag{74}$$

Finally, the definition of the resolvent yields the relation

$$\frac{\lambda}{\tau_{k+1}}\left(x^k - x^{k+1}\right) - L^*y^{k+1} + \mathrm{Cx}^{k+1} - \mathrm{Cx}^k \in (A + C)x^{k+1}. \tag{75}$$

Remark 17 Taking into consideration the above remark, in particular Eq. 74, one can notice that in Algorithm 14, the sequences $(x^k)_{k\geq 0}$ and $(y^k)_{k\geq 0}$ can be generated independently of the sequence $(z^k)_{k\geq 0}$. More precisely, for $(x^0, x^1, y^1) \in \mathcal{H} \times \mathcal{H} \times \mathcal{G}$ given starting points, one has for all $k \geq 1$

$$y^{k+1} = \left(\tau_k \mathrm{LL}^* + M_2^k + B^{-1}\right)^{-1}\left[L(x^k + \theta_{k-1}(x^k - x^{k-1})) + (\tau_k \mathrm{LL}^* + M_2^k)y^k\right] \tag{76}$$

$$x^{k+1} = J_{(\tau_{k+1}/\lambda)A}\left[x^k + \frac{\tau_{k+1}}{\lambda}\left(-L^*y^{k+1} - \mathrm{Cx}^k\right)\right]. \tag{77}$$

The sequence $(z^k)_{k\geq 1}$ can be then obtained by

$$z^k = \frac{\theta_{k-1}}{\tau_k}(x^k - x^{k+1}) - L^* y^k \ \forall k \geq 1.$$

Remark 18 The choice

$$\tau_k \mathrm{LL}^* + M_2^k = \sigma_k^{-1} \,\mathrm{Id} \ \forall k \geq 0 \tag{78}$$

leads to so-called accelerated versions of primal-dual algorithms that have been intensively studied in the literature. Indeed, in this setting, (76)–(77) becomes for all $k \geq 1$

$$\begin{aligned} y^{k+1} &= J_{\sigma_k B^{-1}} \left[y^k + \sigma_k L(x^k + \theta_{k-1}(x^k - x^{k-1})) \right] \\ x^{k+1} &= J_{(\tau_{k+1}/\lambda)A} \left[x^k + \frac{\tau_{k+1}}{\lambda} \left(-L^* y^{k+1} - \mathrm{C}x^k \right) \right], \end{aligned}$$

which is Algorithm 5 in [9]. Not least, in the variational case when $A = \partial f$ and $B = \partial g$, and for $C = 0$ and $\lambda = 1$, we obtain for all $k \geq 0$

$$\begin{aligned} y^{k+1} &= \mathrm{prox}_{\sigma_k g^*} \left[y^k + \sigma_k L \left(x^k + \theta_{k-1}(x^k - x^{k-1}) \right) \right] \\ x^{k+1} &= \mathrm{prox}_{\tau_{k+1} f} \left(x^k - \tau_{k+1} L^* y^{k+1} \right), \end{aligned}$$

which is the numerical scheme considered by Chambolle and Pock in [14, Algorithm 2].

We also notice that condition (78) guarantees the fulfillment of both (69) and (70), due to the fact that the sequence $(\tau_{k+1}\sigma_k)_{k\geq 0}$ is constant (see (67) and (68)).

Remark 19 Assume again that $C = 0$ and consider the variational case as described in Problem 2. From (71) and (72), 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

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

Algorithm 14 becomes in case $\lambda = 1$

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

which can be regarded as an accelerated version of the iterative scheme (19)–(21) from Remark 4.

We present the main theorem of this section.

Theorem 20 *Consider the setting of Problem 11 and let* (x, v) *be a primal-dual solution to the primal-dual system of monotone inclusions* (4)–(5). *Let* $(x^k, z^k, y^k)_{k\geq 0}$ *be the sequence generated by Algorithm 14 and assume that the relations* (63)–(70) *are fulfilled. Then, we have for all* $n \geq 2$

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

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

Proof Let $k \geq 1$ be fixed. From (73), the relation $\mathrm{Lx} \in B^{-1}v$ (see (7)) and the monotonicity of B^{-1} we obtain

$$\left\langle y^{k+1} - v, \widetilde{M_2}^k(y^k - y^{k+1}) + L\left[x^k + \theta_{k-1}(x^k - x^{k-1})\right] - \mathrm{Lx}\right\rangle \geq 0$$

or, equivalently,

$$\frac{1}{2}\|y^k - v\|^2_{\widetilde{M_2}^k} - \frac{1}{2}\|y^{k+1} - v\|^2_{\widetilde{M_2}^k} - \frac{1}{2}\|y^k - y^{k+1}\|^2_{\widetilde{M_2}^k}$$
$$\geq \left\langle y^{k+1} - v, Lx - L\left[x^k + \theta_{k-1}(x^k - x^{k-1})\right]\right\rangle. \quad (79)$$

Further, from (75), the relation $-L^*v \in (A + C)x$ (see (7)) and the γ-strong monotonicity of $A + C$ we obtain

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

or, equivalently,

$$\frac{\lambda}{2\tau_{k+1}}\|x^k - x\|^2 - \frac{\lambda}{2\tau_{k+1}}\|x^{k+1} - x\|^2 - \frac{\lambda}{2\tau_{k+1}}\|x^k - x^{k+1}\|^2$$
$$\geq \gamma\|x^{k+1} - x\|^2 + \langle x^{k+1} - x, \mathrm{Cx}^k - \mathrm{Cx}^{k+1}\rangle + \langle y^{k+1} - v, \mathrm{Lx}^{k+1} - \mathrm{Lx}\rangle. \quad (80)$$

Since C is μ-Lipschitz continuous, we have that

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

which combined with (80) implies

$$\frac{\lambda}{2\tau_{k+1}}\|x^k - x\|^2 \geq \left(\frac{\lambda}{2\tau_{k+1}} + \gamma - \frac{\mu\tau_{k+1}}{2}\right)\|x^{k+1} - x\|^2 + \frac{\lambda - \mu}{2\tau_{k+1}}\|x^{k+1} - x^k\|^2$$
$$+\langle y^{k+1} - v, \mathrm{Lx}^{k+1} - \mathrm{Lx}\rangle. \quad (81)$$

By adding the inequalities (79) and (81), we obtain

$$\begin{aligned}
\frac{1}{2}\|y^k-v\|^2_{\widetilde{M_2}^k}+\frac{\lambda}{2\tau_{k+1}}\|x^k-x\|^2 \geq\ & \frac{1}{2}\|y^{k+1}-v\|^2_{\widetilde{M_2}^k} \\
&+\left(\frac{\lambda}{2\tau_{k+1}}+\gamma-\frac{\mu\tau_{k+1}}{2}\right)\|x^{k+1}-x\|^2 \\
&+\frac{1}{2}\|y^k-y^{k+1}\|^2_{\widetilde{M_2}^k} \\
&+\frac{\lambda-\mu}{2\tau_{k+1}}\|x^{k+1}-x^k\|^2 \\
&+\left\langle y^{k+1}-v, L\left[x^{k+1}-x^k-\theta_{k-1}(x^k-x^{k-1})\right]\right\rangle. \quad (82)
\end{aligned}$$

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 (82), we obtain (after dividing by τ_{k+1})

$$\begin{aligned}
\frac{\|y^k-v\|^2_{\widetilde{M_2}^k}}{2\tau_{k+1}}+\frac{\lambda}{2\tau_{k+1}^2}\|x^k-x\|^2 \geq\ & \frac{\|y^{k+1}-v\|^2_{\widetilde{M_2}^k}}{2\tau_{k+1}} \\
&+\left(\frac{\lambda}{2\tau_{k+1}^2}+\frac{\gamma}{\tau_{k+1}}-\frac{\mu}{2}\right)\|x^{k+1}-x\|^2 \\
&+\frac{\|y^k-y^{k+1}\|^2_{\widetilde{M_2}^k}}{2\tau_{k+1}}-\frac{\|y^k-y^{k+1}\|^2}{2\tau_{k+1}\sigma_k} \quad (83) \\
&+\frac{\lambda-\mu}{2\tau_{k+1}^2}\|x^{k+1}-x^k\|^2 \\
&-\frac{\theta_{k-1}^2\|L\|^2\sigma_k}{2\tau_{k+1}}\|x^k-x^{k-1}\|^2 \\
&+\frac{1}{\tau_{k+1}}\langle L(x^{k+1}-x^k), y^{k+1}-v\rangle \\
&-\frac{\theta_{k-1}}{\tau_{k+1}}\langle L(x^k-x^{k-1}), y^k-v\rangle.
\end{aligned}$$

From (69) and (72), we have that the term in (83) is nonnegative. Further, noticing that (see (66), (67), (68), and (65))

$$\frac{\theta_{k-1}}{\tau_{k+1}} = \frac{1}{\tau_k}$$

$$\frac{\lambda}{2\tau_{k+1}^2} + \frac{\gamma}{\tau_{k+1}} - \frac{\mu}{2} = \frac{\lambda}{2\tau_{k+2}^2},$$

$$\tau_{k+1}\sigma_k = \tau_k\sigma_{k-1} = \ldots = \tau_1\sigma_0$$

and

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

we obtain (see also (64) and (70))

$$\begin{aligned}
\frac{\|y^k - v\|^2_{\widetilde{M_2}^k}}{2\tau_{k+1}} + \frac{\lambda}{2\tau_{k+1}^2}\|x^k - x\|^2 \ge\ & \frac{\|y^{k+1} - v\|^2_{\widetilde{M_2}^{k+1}}}{2\tau_{k+2}} + \frac{\lambda}{2\tau_{k+2}^2}\|x^{k+1} - x\|^2 \\
&+ \frac{1}{2\tau_{k+1}^2}\|x^{k+1} - x^k\|^2 - \frac{1}{2\tau_k^2}\|x^k - x^{k-1}\|^2 \\
&+ \frac{1}{\tau_{k+1}}\langle L(x^{k+1} - x^k), y^{k+1} - v\rangle \\
&- \frac{1}{\tau_k}\langle L(x^k - x^{k-1}), y^k - v\rangle.
\end{aligned}$$

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

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

The inequality in the statement of the theorem follows by combining this relation with (see (69))

$$\frac{\|y^n - v\|^2_{\widetilde{M_2}^n}}{2\tau_{n+1}} \ge \frac{\|y^n - v\|^2}{2\sigma_n\tau_{n+1}},$$

$$\frac{1}{2\tau_n^2}\|x^n - x^{n-1}\|^2 + \frac{1}{\tau_n}\langle L(x^n - x^{n-1}), y^n - v\rangle \ge -\frac{\|L\|^2}{2}\|y^n - v\|^2 \text{ and } \sigma_n\tau_{n+1} = \sigma_0\tau_1.$$

Finally, we notice that for any $n \ge 0$ (see (66) and (67))

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

From here, it follows that $\tau_{n+1} < \tau_n$ for all $n \geq 1$ and $\lim_{n\to+\infty} n\tau_n = \lambda/\gamma$ (see [9, page 261]). The proof is complete. □

Remark 21 In Remark 18, we provided an example of a family of linear, continuous and self-adjoint operators $(M_2^k)_{k\geq 0}$ for which the relations (69) and (70) 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

$$\mu\tau_1 \leq \gamma, \tag{84}$$

then $(\theta_k)_{k\geq 0}$ is monotonically increasing. In the examples below, we replace (63) with the stronger assumption (84).

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

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

Then (69) trivially holds, while (70), which can be equivalently written as

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

follows from the fact that $(\tau_{k+1}\sigma_k)_{k\geq 0}$ is constant (see (67) and (68)) and $(\theta_k)_{k\geq 0}$ is monotonically increasing.

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

$$M_2^k := 0.$$

Relation (70) holds since $(\theta_k)_{k\geq 0}$ is monotonically increasing. Condition (69) becomes in this setting

$$\sigma_k\tau_k\mathrm{LL}^* \succcurlyeq \mathrm{Id}\ \forall k \geq 0. \tag{85}$$

Since $\tau_k > \tau_{k+1}$ for all $k \geq 1$ and $(\tau_{k+1}\sigma_k)_{k\geq 0}$ is constant, (85) holds, if

$$\mathrm{LL}^* \in \mathcal{P}_{\frac{1}{\sigma_0\tau_1}}(\mathcal{G}). \tag{86}$$

Provided that $\mathcal{G}$ is finite dimensional, (86) holds if and only if $\sigma_0\tau_1\lambda_{\min}(\mathrm{LL}^*) \geq 1$, where $\lambda_{\min}(\mathrm{LL}^*)$ denotes the smallest eigenvalue of LL^*. Since $\sigma_0\tau_1\|L\|^2 \leq 1$, this is possible only in the particular case when $\mathrm{LL}^* = \dfrac{1}{\sigma_0\tau_1}\,\mathrm{Id}$. The resulting iterative scheme can be regarded as an accelerated version of the classical ADMM algorithm (see Remark 4).

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

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

Relation (70) holds, since $(\theta_k)_{k\geq 0}$ is monotonically increasing. On the other hand, condition (69) is equivalent to the following:

$$\sigma_k\tau_k(LL^* + \mathrm{Id}) \succcurlyeq \mathrm{Id}\,. \tag{87}$$

Since $\tau_k > \tau_{k+1}$ for all $k \geq 1$ and $(\tau_{k+1}\sigma_k)_{k\geq 0}$ is constant, (87) holds, if

$$\sigma_0\tau_1\mathrm{LL}^* \succcurlyeq (1 - \sigma_0\tau_1)\,\mathrm{Id}\,. \tag{88}$$

In case of $\sigma_0\tau_1 \geq 1$ (which is allowed according to (65) if $\|L\|^2 \leq 1$), this is obviously fulfilled. Otherwise (when $\sigma_0\tau_1 < 1$), in order to guarantee (88), we have to impose that

$$\mathrm{LL}^* \in \mathcal{P}_{\frac{1-\sigma_0\tau_1}{\sigma_0\tau_1}}(\mathcal{G}). \tag{89}$$

When $\mathcal{G}$ is finite dimensional, (89) holds if and only if $\sigma_0\tau_1(1+\lambda_{\min}(LL^*)) \geq 1$.

4 Numerical experiments

In this section, we will compare the performances of Algorithm 14, for different choices of the sequence of matrices $(M_2^k)_{k\geq 0}$, in the context of solving an image denoising problem. We considered in our numerical experiments the convex optimization problem

$$\inf_{x\in[0,255]^n}\left\{\frac{1}{2}\|x-b\|^2 + rTV_{\text{aniso}}(x)\right\}, \tag{90}$$

where $x \in \mathbb{R}^n$ stands for the vectorized colored image $X \in \mathbb{R}^{M\times N}$, $n = 3MN$, and $x_{i,j} = X_{i,j}$ represents the value of the pixel in the ith row and the jth column, for $1 \leq i \leq M$, $1 \leq j \leq N$. Further, $b \in \mathbb{R}^n$ denotes the observed noisy image, $r > 0$ a regularization parameter and $\mathrm{TV} : \mathbb{R}^n \to \mathbb{R}$, the discrete anisotropic total variation mapping.

Recall that the discrete anisotropic total variation mapping $\mathrm{TV}_{\text{aniso}} : \mathbb{R}^n \to \mathbb{R}$,

$$\begin{aligned}\mathrm{TV}_{\text{aniso}}(x) &= \sum_{i=1}^{M-1}\sum_{j=1}^{N-1} |x_{i+1,j}-x_{i,j}| + |x_{i,j+1}-x_{i,j}| \\ &\quad + \sum_{i=1}^{M-1}|x_{i+1,N}-x_{i,j}| + \sum_{j=1}^{N-1}|x_{M,j+1}-x_{M,j}|,\end{aligned}$$

can be written as follows:

$$\mathrm{TV}_{\text{aniso}}(x) = \|\mathrm{Lx}\|_1 \ \forall x \in \mathbb{R}^n,$$

where $L : \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$, $x_{i,j} \mapsto \big(L_1x_{i,j}, L_2x_{i,j}\big)$, with

$$L_1x_{i,j} = \begin{cases} x_{i+1,j}-x_{i,j}, & \text{if } i < M \\ 0, & \text{if } i = M \end{cases} \text{ and } L_2x_{i,j} = \begin{cases} x_{i,j+1}-x_{i,j}, & \text{if } j < N \\ 0, & \text{if } j = N \end{cases},$$

is a linear operator.

Let be $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $f = \delta_{[0,255]^n}$, the indicator function of $[0,255]^n$, $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $g(y_1,y_2) = r\|(y_1,y_2)\|_1$, and $h : \mathbb{R}^n \to \mathbb{R}$, $h(x) = \frac{1}{2}\|x-b\|^2$. Solving (90) means solving the monotone inclusion problem

$$\text{find } x \in \mathbb{R}^n \text{ such that } 0 \in Ax + (L^* \circ B \circ L)x + \mathrm{Cx}, \tag{91}$$

for $A = \partial f$, $B = \partial g$, $C = \nabla h$. Notice that C is 1-Lipschitz continuous and 1-strongly monotone.

We solved (91) with Algorithm 14 (actually by using the formulation (76)–(77)) for three different choices of the sequence of matrices $(M_2^k)_{k\geq 0}$, namely, (I) $M_2^k = \sigma_k^{-1}\,\mathrm{Id} - \tau_k L^*L$, $k \geq 0$, (see Remark 18); (II) $M_2^k = \sigma_k^{-1}\,\mathrm{Id}$, $k \geq 0$; and (III) $M_2^k = \tau_k\,\mathrm{Id}$, $k \geq 0$, (see Remark 21 (i) and (iii)).

In all three implementations, for updating the sequence $(x^k)_{k\geq 2}$, we used the closed form the proximal operator of the function $\lambda^{-1}\tau_{k+1} f$, which requires in every iteration nothing more than the calculation of the projection on the box $[0, 255]^n$. On the other hand, for updating the sequence $(y^k)_{k\geq 2}$, we used two different approaches. For the choice (I) of the sequence $(M_2^k)_{k\geq 0}$, the algorithm required only the closed formula of the proximal operator of $\sigma_k g^*$, which is the projection on the box $[-r, r]^{2n}$. For the choices (II)–(III), of the sequence $(M_2^k)_{k\geq 0}$, we determined in every iteration

$$y^{k+1} = \left(\tau_k \mathrm{LL}^* + M_2^k + \partial g^*\right)^{-1}\left[L(x^k + \theta_{k-1}(x^k - x^{k-1})) + (\tau_k \mathrm{LL}^* + M_2^k)y^k\right]$$

or, equivalently,

$$y^{k+1} = \underset{y\in\mathbb{R}^n\times\mathbb{R}^n}{\operatorname{argmin}}\left[g^*(y) - \langle y, L(x^k + \theta_{k-1}(x^k - x^{k-1}))\rangle + \frac{1}{2}\|y - y^k\|^2_{\tau_k \mathrm{LL}^* + M_2^k}\right] \tag{92}$$

by executing some steps of FISTA (see [5]).

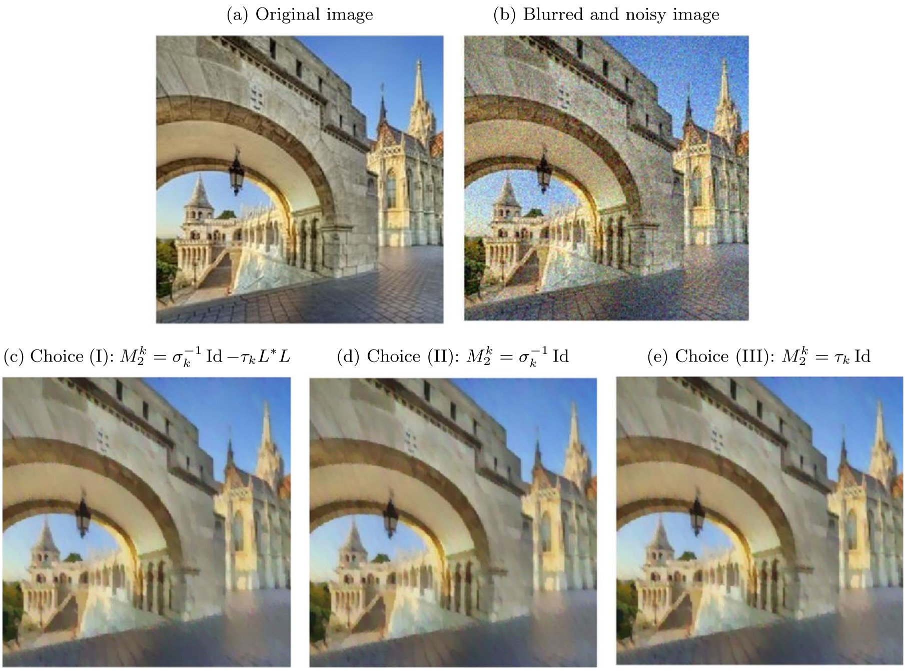

Fig. 1 The original image, the noisy image (corrupted with Gaussian noise with standard deviation $\sigma = 10$) and the obtained reconstructed images for the choices (I)–(III) and a tolerance error of $\varepsilon = 10^{-6}$

We used in the numerical experiments a 256 × 256 test image (see Fig. 1) corrupted with Gaussian noise with standard deviation $\sigma \in \{10, 20, 30\}$, took $\lambda = 2$ and as regularization parameter $r = 0.07$. We stopped the algorithm when the difference of two consecutive primal iterates was less than a given error tolerance $\varepsilon > 0$. In Fig. 1, we show the original image, the corrupted image, and the reconstructed images obtained for the four different choices (I)–(IV) for a tolerance error of $\varepsilon = 10^{-6}$.

In Table 1, we compare the performances of the three iterative schemes in case $\sigma = 10$ in terms of the number of iterations and CPU time in seconds needed to achieve two different tolerance errors. Prior to the comparisons we did for all schemes a parameter tuning in order to determine which choice of the initial step sizes σ_0 and τ_1 provides the highest value for the structural similarity index (SSIM). In Fig. 2 ,we show the dependence of the SSIM value on (σ_0, τ_1) for the cases (I) and (II) and on

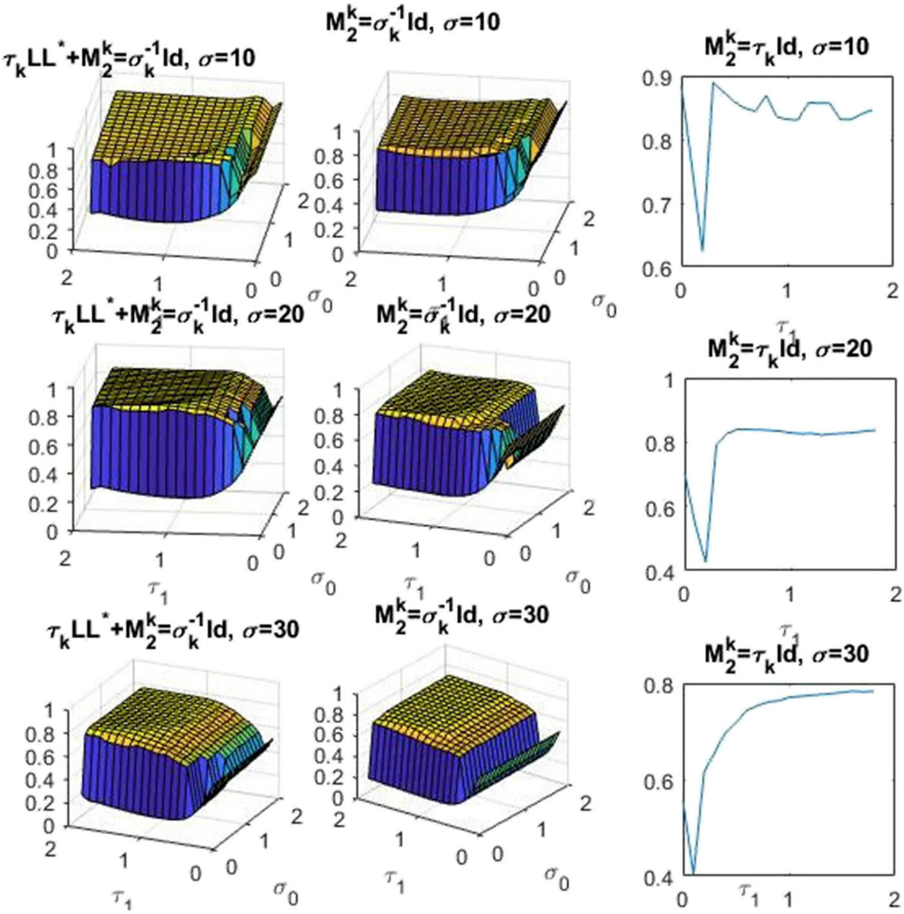

Fig. 2 The parameter tuning surfaces/curves generated by the SSIM values as functions of the initial step sizes

Table 1 Performance evaluation of Algorithm 14; the entries refer to the number of iterations and the CPU times in seconds

	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$
Choice (I): $M_2^k = \sigma_k^{-1}\,\mathrm{Id} - \tau_k L^* L$	14 (0.85s)	107 (5.02s)
Choice (II): $M_2^k = \sigma_k^{-1}\,\mathrm{Id}$	18 (3.21s)	115 (41.79s)
Choice (III): $M_2^k = \tau_k\,\mathrm{Id}$	16 (3.00s)	110 (42.42s)

τ_1 for case (III). In case (III), we took $\sigma_0 = \frac{1}{\tau_1(1+\lambda_{\min(LL^*)})}$, which proved to be the best choice.

The entries in Table 1 show that the iterative schemes that correspond to the choices (II) and (III) are, in terms of the number of iterates, as fast as the scheme that corresponds to (I). The differences in CPU time (which are substantial only for low tolerance errors) are caused by the fact that for the choices (II) and (III) inner loops are done in each iteration. One could possibly improve the CPU times in these two settings by solving (92) with numerical algorithms which are better adapted to outer loop.

Acknowledgements Open access funding provided by Austrian Science Fund (FWF). University of Vienna. We are thankful to two anonymous reviewers for comments and remarks which improved the quality of the paper. The numerical experiments have been carried out by Ulrik Hager-Roiser for a seminar paper at University of Vienna in the winter semester 2017/2018.

References

1. Arrow, K.J., Hurwicz, L., Uzawa, H.: Studies in Linear and Non-Linear Programming. Stanford University Press, Stanford (1958)
2. Attouch, H., Théra, M.: A general duality principle for the sum of two operators. J. Convex Anal. **3**(1), 1–24 (1996)
3. Banert, S., Boţ, R.I., Csetnek, E.R.: Fixing and extending some recent results on the ADMM algorithm, arXiv:1612.05057 (2016)
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. Beck, A., Teboulle, M.: A fast iterative shrinkage- thresholding algorithm for linear inverse problems. SIAM J. Imag. Sci. **2**(1), 183–202 (2009)
6. Borwein, J.M., Vanderwerff, J.D.: Convex Functions: Constructions, Characterizations and Counterexamples. Cambridge University Press, Cambridge (2010)
7. Boţ, R.I.: Conjugate Duality in Convex Optimization Lecture Notes in Economics and Mathematical Systems, vol. 637. Springer, Berlin (2010)
8. Boţ, R.I., Csetnek, E.R., Heinrich, A.: A primal-dual splitting algorithm for finding zeros of sums of maximal monotone operators. SIAM J. Optim. **23**(4), 2011–2036 (2013)

9. Boţ, R.I., Csetnek, E.R., Heinrich, A., Hendrich, C.: On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems. Math. Program. **150**(2), 251–279 (2015)
10. Boţ, R.I., Hendrich, C.: Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization. J. Math. Imag. Vis. **49**(3), 551–568 (2014)
11. Boţ, R.I., Hendrich, C.: A Douglas-Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators. SIAM J. Optim. **23**(4), 2541–2565 (2013)
12. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Found. Trends Mach. Learn. **3**, 1–12 (2010)
13. Briceño-Arias, L.M., Combettes, P.L.: A monotone + skew splitting model for composite monotone inclusions in duality. SIAM J. Optim. **21**(4), 1230–1250 (2011)
14. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imag. Vis. **40**(1), 120–145 (2011)
15. Chambolle, A., Pock, T.: On the ergodic convergence rates of a first-order primal-dual algorithm. Math. Program. **159**(1–2), 253–287 (2016)
16. 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 Variational Anal. **20**(2), 307–330 (2012)
17. Combettes, P.L., Vũ, B.C.: Variable metric quasi-Fejér monotonicity. Nonlinear Anal. **78**, 17–31 (2013)
18. Combettes, P.L., Vũ, B.C.: Variable metric forward-backward splitting with applications to monotone inclusions in duality. Optimization **63**(9), 1289–1318 (2014)
19. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. Multiscale Model. Simul. **4**(4), 1168–1200 (2005)
20. Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J. Optim. Theory Appl. **158**(2), 460–479 (2013)
21. Davis, D., Yin, W.: A three-operator splitting scheme and its optimization applications. Set-Valued Variational Anal. **25**(4), 829–858 (2017)
22. Eckstein, J.: Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results. Rutcor Research Report 32–2012 (2012)
23. Eckstein, J.: Some saddle-function splitting methods for convex programming. Optim. Methods Softw. **4**, 75–83 (1994)
24. Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**, 293–318 (1992)
25. Ekeland, I., Temam, R.: Convex Analysis and Variational Problems. North-Holland Publishing Company, Amsterdam (1976)
26. Esser, E., Zhang, X.Q., Chan, T.F.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imag. Sci. **3**(4), 1015–1046 (2010)
27. Fazel, M., Pong, T.K., Sun, D., Tseng, P.: Hankel matrix rank minimization with applications in system identification and realization. SIAM J. Matrix Anal. Appl. **34**, 946–977 (2013)
28. Fortin, M., Glowinski, R.: On decomposition-coordination methods using an augmented Lagrangian. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems. North-Holland, Amsterdam (1983)
29. Gabay, D.: Applications of the method of multipliers to variational inequalities. In: Fortin, M., Glowinski, R. (eds.) Augmented Lagrangian Methods: Applications to the Solution of Boundary-Value Problems. North-Holland, Amsterdam (1983)
30. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximations. Comput. Math. Appl. **2**, 17–40 (1976)
31. He, B.S., Yang, H., Wang, S.L.: Alternating direction method with self-adaptive penalty parameters for , monotone variational inequalities. J. Optim. Theory Appl. **106**(2), 337–356 (2000)
32. Yuan, B.S., Yuan, X.M.: Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIAM J. Imag. Sci. **5**(1), 119–149 (2012)
33. Korpelevich, G.M.: The extragradient method for finding saddle points and other problems. Matecon **12**, 747–756 (1976)
34. Malitsky, Y., Pock, T.: A first-order primal-dual algorithm with linesearch. SIAM J. Optim. **28**(1), 411–432 (2018)

35. Raguet, H., Landrieu, L.: Preconditioning of a generalized forward-backward splitting and application to optimization on graphs. SIAM J. Imag. Sci. **8**(4), 2706–2739 (2015)
36. Riesz, F., Sz.-Nagy, B. Leçons d'Analyse Fonctionnelle, 5th edn. Gauthier-Villars, Paris (1968)
37. Rockafellar, R.T.: On the maximal monotonicity of subdifferential mappings. Pac. J. Math. **33**(1), 209–216 (1970)
38. 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**(1), 269–297 (2014)
39. Simons, S.: From Hahn-Banach to Monotonicity. Springer-Verlag, Berlin (2008)
40. Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. SIAM J. Control. Optim. **38**(2), 431–446 (2000)
41. Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. Adv. Comput. Math. **38**(3), 667–681 (2013)
42. Zălinescu, C.: Convex Analysis in General Vector Spaces. World Scientific, Singapore (2002)