Fixing and extending some recent results on the ADMM algorithm

Sebastian Banert * Radu Ioan Boţ † Ernö Robert Csetnek ‡

April 12, 2020

Abstract. We investigate the techniques and ideas used in [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] in the convergence analysis of two proximal ADMM algorithms for solving convex optimization problems involving compositions with linear operators. Besides this, we formulate a variant of the ADMM algorithm that is able to handle convex optimization problems involving an additional smooth function in its objective, and which is evaluated through its gradient. Moreover, in each iteration we allow the use of variable metrics, while the investigations are carried out in the setting of infinite dimensional Hilbert spaces. This algorithmic scheme is investigated from the point of view of its convergence properties.

Key Words. ADMM algorithm, Lagrangian, saddle points, variable metrics, positive semidefinite operators

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

1 Introduction

One of the most popular numerical algorithms for solving optimization problems of the form

$$\inf_{x\in\mathbb{R}^n}\{f(x)+g(Ax)\}, \tag{1}$$

where $f:\mathbb{R}^n\to\overline{\mathbb{R}}:=\mathbb{R}\cup\{\pm\infty\}$ and $g:\mathbb{R}^m\to\overline{\mathbb{R}}$ are proper, convex, lower semicontinuous functions and $A:\mathbb{R}^n\to\mathbb{R}^m$ is a linear operator, is the alternating direction method of multipliers (ADMM). Here, the spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ are equipped with their usual inner products and induced norms, which we both denote by $\langle\cdot,\cdot\rangle$ and $\|\cdot\|$, respectively, as there is no risk of confusion.

*Lund University, Faculty of Engineering (LTH), Department of Automatic Control, Box 118, 221 00 Lund, Sweden, email: sebastian.banert@control.lth.se. Research supported by FWF (Austrian Science Fund), project I 2419-N32, and by Wallenberg AI, Autonomous Systems and Software Program (WASP).

†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), projects M 1682-N25 and P 29809-N32.

By introducing an auxiliary variable z one can rewrite (1) as

$$\inf_{\substack{(x,z)\in\mathbb{R}^n\times\mathbb{R}^m\\ Ax-z=0}}\{f(x)+g(z)\}. \tag{2}$$

The Lagrangian associated with problem (2) is

$$l:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^m\to\overline{\mathbb{R}},\ l(x,z,y)=f(x)+g(z)+\langle y,Ax-z\rangle,$$

and we say that $(x^*,z^*,y^*)\in\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^m$ is a saddle point of the Lagrangian, if

$$l(x^*,z^*,y)\leq l(x^*,z^*,y^*)\leq l(x,z,y^*)\ \forall(x,z,y)\in\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^m. \tag{3}$$

It is known that (x^*,z^*,y^*) is a saddle point of l if and only if $z^*=Ax^*$, (x^*,z^*) is an optimal solution of (2), y^* is an optimal solution of the Fenchel-Rockafellar dual problem (see [3–5,20,30]) to (1)

$$\sup_{y\in\mathbb{R}^m}\{-f^*(-A^Ty)-g^*(y)\}, \tag{4}$$

and the optimal objective values of (1) and (4) coincide. Notice that f^* and g^* are the conjugates of f and g, defined by $f^*(u)=\sup_{x\in\mathbb{R}^n}\{\langle u,x\rangle-f(x)\}$ for all $u\in\mathbb{R}^n$ and $g^*(y)=\sup_{z\in\mathbb{R}^m}\{\langle y,z\rangle-g(z)\}$ for all $y\in\mathbb{R}^m$, respectively.

If (1) has an optimal solution and $A(\mathrm{ri}(\mathrm{dom}\, f))\cap\mathrm{ri}\,\mathrm{dom}\, g\neq\emptyset$, then the set of saddle points of l is nonempty. Here, we denote by $\mathrm{ri}(S)$ the relative interior of a convex set S, which is the interior of S relative to its affine hull.

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

$$L_c:\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^m\to\overline{\mathbb{R}},\ L_c(x,z,y)=f(x)+g(z)+\langle y,Ax-z\rangle+\frac{c}{2}\|Ax-z\|^2.$$

The ADMM algorithm reads:

Algorithm 1 *Choose $(x^0,z^0,y^0)\in\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^m$ and $c>0$. For all $k\geq 0$ generate the sequence $(x^k,z^k,y^k)_{k\geq 0}$ as follows:*

$$x^{k+1} \in \operatorname*{argmin}_{x\in\mathbb{R}^n} L_c(x,z^k,y^k)=\operatorname*{argmin}_{x\in\mathbb{R}^n}\left\{f(x)+\frac{c}{2}\|Ax-z^k+c^{-1}y^k\|^2\right\} \tag{5}$$

$$z^{k+1} = \operatorname*{argmin}_{z\in\mathbb{R}^m} L_c(x^{k+1},z,y^k)=\operatorname*{argmin}_{z\in\mathbb{R}^m}\left\{g(z)+\frac{c}{2}\|Ax^{k+1}-z+c^{-1}y^k\|^2\right\} \tag{6}$$

$$y^{k+1} = y^k+c(Ax^{k+1}-z^{k+1}). \tag{7}$$

If A has full column rank, then the set of minimizers in (5) is a singleton, as is the set of minimizers in (6) without any further assumption, and the sequence $(x^k,z^k,y^k)_{k\geq 0}$ generated by Algorithm 1 converges to a saddle point of the Lagrangian l (see, for instance, [19]). The alternating direction method of multipliers was first introduced in [25] and [23]. Gabay has shown in [24] (see also [19]) that ADMM is nothing else than the Douglas-Rachford algorithm applied to the monotone inclusion problem

$$0\in\partial(f^*\circ(-A^T))(y)+\partial g^*(y)$$

For a function $k : \mathbb{R}^n \to \overline{\mathbb{R}}$, the set-valued operator defined by $\partial k(x) := \{u \in \mathbb{R}^n : k(t) - k(x) \geq \langle u, t - x\rangle \ \forall t \in \mathbb{R}^n\}$, for $k(x) \in \mathbb{R}$, and $\partial k(x) := \emptyset$, otherwise, denotes its (convex) subdifferential.

One of the limitations of the ADMM algorithm comes from the presence of the term Ax in the update rule of x^{k+1} (we refer to [14] for an approach to circumvent the limitations of ADMM). While in (6) a proximal step for the function g is taken, in (5) the function f and the operator A are not evaluated independently, which makes the ADMM algorithm less attractive for implementations than the primal-dual splitting algorithms (see, for instance, [8–10, 12, 13, 16, 29]). Despite of this fact, the ADMM algorithm has been widely used for solving convex optimization problems arising in real-life applications (see, for instance, [11, 21]). For a version of the ADMM algorithm with inertial and memory effects we refer the reader to [7].

In order to overcome the above-mentioned drawback of the classical ADMM method and to increase its flexibility, the following so-called proximal alternating direction proximal method of multipliers has been considered in [28] (see also [22, 26]):

Algorithm 2 *Choose* $(x^0, z^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ *and* $c > 0$. *For all* $k \geq 0$ *generate the sequence* $(x^k, z^k, y^k)_{k\geq 0}$ *as follows:*

$$x^{k+1} \in \operatorname*{argmin}_{x\in\mathbb{R}^n} \left\{ f(x) + \frac{c}{2}\|Ax - z^k + c^{-1}y^k\|^2 + \frac{1}{2}\|x - x^k\|^2_{M_1} \right\} \tag{8}$$

$$z^{k+1} = \operatorname*{argmin}_{z\in\mathbb{R}^m} \left\{ g(z) + \frac{c}{2}\|Ax^{k+1} - z + c^{-1}y^k\|^2 + \frac{1}{2}\|z - z^k\|^2_{M_2} \right\} \tag{9}$$

$$y^{k+1} = y^k + c(Ax^{k+1} - z^{k+1}). \tag{10}$$

Here, $M_1 \in \mathbb{R}^{n\times n}$ and $M_2 \in \mathbb{R}^{m\times m}$ are symmetric positive semidefinite matrices and $\|u\|^2_{M_i} = \langle u, M_i u\rangle$ denotes the squared seminorm induced by M_i, for $i \in \{1, 2\}$.

Indeed, for $M_1 = M_2 = 0$, Algorithm 2 becomes the classical ADMM method, while for $M_1 = \mu_1 \,\mathrm{Id}$ and $M_2 = \mu_2 \,\mathrm{Id}$ with $\mu_1, \mu_2 > 0$ and Id denoting the corresponding matrix, it becomes the algorithm proposed and investigated in [18]. Furthermore, if $M_1 = \tau^{-1}\,\mathrm{Id} - cA^TA$ with $\tau > 0$ such that $c\tau\|A\|^2 < 1$ and $M_2 = 0$, then one can show that Algorithm 2 is equivalent to one of the primal-dual algorithms formulated in [16].

The sequence $(z^k)_{k\geq 0}$ generated in Algorithm 2 is uniquely determined due to the fact that the objective function in (9) is lower semicontinuous and strongly convex. On the other hand, the set of minimizers in (8) is in general not a singleton and it can be even empty. However, if one imposes that $M_1 + A^*A$ is positive definite, then $(x^k)_{k\geq 0}$ will be uniquely determined, too.

Shefi and Teboulle provide in [28] in connection to Algorithm 2 an ergodic convergence rate result for a primal-dual gap function formulated in terms of the Lagrangian l, from which they deduce a global convergence rate result for the sequence of function values $(f(x^k)+g(Ax^k))_{k\geq 0}$ to the optimal objective value of (1), when g is Lipschitz continuous. Furthermore, they formulate a global convergence rate result for the sequence $(\|Ax^k - z^k\|)_{k\geq 0}$ to 0. Finally, Shefi and Teboulle prove the convergence of the sequence $(x^k, z^k, y^k)_{k\geq 0}$ to a saddle point of the Lagrangian l, provided that either $M_1 = 0$ and A has full column rank or M_1 is positive definite.

Algorithm 2 from [28] represents the starting point of our investigations. More precisely, in this paper:

• we point out some flaws in the proof of a statement in [28], which is fundamental for the derivation of the global convergence rate of $(\|Ax^k - z^k\|)_{k\geq 0}$ to 0 and of the convergence of the sequence $(x^k, z^k, y^k)_{k\geq 0}$;

• we show how the statement in cause can be proved by using different techniques;

• we formulate a variant of Algorithm 2 for solving convex optimization problems in infinite dimensional Hilbert spaces involving an additional smooth function in their objective, that we evaluate through its gradient, and which allows in each iteration the use of variable metrics;

• we prove an ergodic convergence rate result for this algorithm involving a primal-dual gap function formulated in terms of the associated Lagrangian l and a convergence result for the sequence of iterates to a saddle point of l.

2 Fixing some results from [28] related to the convergence analysis for Algorithm 2

In this section we point out several flaws that have been made in [28] when deriving a fundamental result for both the rate of convergence of the sequence $(\|Ax^k - z^k\|)_{k\geq 0}$ to 0 and the convergence of the sequence $(x^k, z^k, y^k)_{k\geq 0}$ to a saddle point of the Lagrangian l. We also show how these arguments can be fixed by relying on some of the building blocks of the analysis we will carry out in Section 3.

To proceed, we first recall some results from [28]. We start with a statement that follows from the variational characterization of the minimizers of (8)-(9).

Lemma 3 *(see [28, Lemma 4.2]) Let $(x^k, z^k, y^k)_{k\geq 0}$ be a sequence generated by Algorithm 2. Then for all $k \geq 0$ and for all $(x, z, y) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ it holds*

$$\begin{aligned} l(x^{k+1}, z^{k+1}, y) \leq\ & l(x, z, y^{k+1}) + c\langle z^{k+1} - z^k, A(x - x^{k+1})\rangle + \\ & + \frac{1}{2}\left(\|x - x^k\|^2_{M_1} - \|x - x^{k+1}\|^2_{M_1} + \|z - z^k\|^2_{M_2} - \|z - z^{k+1}\|^2_{M_2}\right) \\ & + \frac{1}{2}\left(c^{-1}\|y - y^k\|^2 - c^{-1}\|y - y^{k+1}\|^2\right) \\ & - \frac{1}{2}\left(\|x^{k+1} - x^k\|^2_{M_1} + \|z^{k+1} - z^k\|^2_{M_2} + c^{-1}\|y^{k+1} - y^k\|^2\right). \end{aligned}$$

Furthermore, by invoking the monotonicity of the convex subdifferential of g, in [28] the following estimation is derived.

Lemma 4 *(see [28, Proposition 5.3(b)]) Let $(x^k, z^k, y^k)_{k\geq 0}$ be a sequence generated by Algorithm 2. Then for all $k \geq 1$ and for all $(x, z) \in \mathbb{R}^n \times \mathbb{R}^m$ it holds*

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

By taking $(x, z, y) := (x^*, z^*, y^*)$ in Lemma 3, where (x^*, z^*, y^*) is a saddle point of the Lagrangian l, and by using the inequality (see (3))

$$l(x^{k+1}, z^{k+1}, y^*) \geq l(x^*, z^*, y^{k+1})\ \forall k \geq 0,$$

and the estimation in Lemma 4, one easily obtains the following result.

Lemma 5 *Let (x^*, z^*, y^*) be a saddle point of the Lagrangian l associated with (1), M_1, M_2 be symmetric positive semidefinite matrices and $c > 0$. Let $(x^k, z^k, y^k)_{k\geq 0}$ be a sequence generated by Algorithm 2. Then for all $k \geq 1$ the following inequality holds*

$$\|x^{k+1} - x^k\|^2_{M_1} + \|z^{k+1} - z^k\|^2_{M_2} + c^{-1}\|y^{k+1} - y^k\|^2 + \tag{11}$$

$$\|x^* - x^{k+1}\|^2_{M_1} + \|z^* - z^{k+1}\|^2_{M_2+cI_m} + c^{-1}\|y^* - y^{k+1}\|^2 + \|z^{k+1} - z^k\|^2_{M_2} \tag{12}$$

$$\leq \|x^* - x^k\|^2_{M_1} + \|z^* - z^k\|^2_{M_2+cI_m} + c^{-1}\|y^* - y^k\|^2 + \|z^k - z^{k-1}\|^2_{M_2}. \tag{13}$$

By using the notations from [28, Section 5.3], namely

$$v^{k+1} := \|x^{k+1} - x^k\|^2_{M_1} + \|z^{k+1} - z^k\|^2_{M_2+c\,\mathrm{Id}} + c^{-1}\|y^{k+1} - y^k\|^2 \ \forall k \geq 0$$

and

$$u^k := \|x^* - x^k\|^2_{M_1} + \|z^* - z^k\|^2_{M_2+c\,\mathrm{Id}} + c^{-1}\|y^* - y^k\|^2 + \|z^k - z^{k-1}\|^2_{M_2} \ \forall k \geq 1,$$

the inequality in Lemma 5 can be equivalently written as

$$v^{k+1} - c\|z^{k+1} - z^k\|^2 \leq u^k - u^{k+1} \ \forall k \geq 1. \tag{14}$$

In [28, Lemma 5.1, (5.37)], instead of (14), it is stated that

$$v^{k+1} \leq u^k - u^{k+1} \ \forall k \geq 1, \tag{15}$$

however, its proof, which follows the argument that goes through Lemma 3, Lemma 4 and Lemma 5, is not correct, since it leads to (14) instead of (15).

Since the sequence $(v^k)_{k\geq 0}$ is monotonically decreasing, statement (15), in combination with straightforward telescoping arguments, leads to the fact that $(v^k)_{k\geq 0}$ converges to zero with a rate of convergence of $\mathcal{O}(1/k)$. This implies that $(\|Ax^k - z^k\|)_{k\geq 0}$ converges to zero with a rate of convergence of $\mathcal{O}(1/\sqrt{k})$ (see [28, Theorem 5.4]). In addition, statement (15) is used in [28, Theorem 5.6] to prove the convergence of the sequence $(x^k, z^k, y^k)_{k\geq 0}$ to a saddle point of the Lagrangian l. However, the techniques used in [28], involving function values and the saddle point inequality, do not lead to (15), but to the weaker inequalitiy (14).

In the following we will show that one can in fact derive (15), however, to this end one needs to use different techniques. These are described in detail in the next section; here we will just show how do they lead to (15). We would like to notice that, differently from [28], in our analysis we will only use properties related to the fact that the convex subdifferential of a proper, convex and lower semicontinuous function is a maximally monotone set-valued operator.

We start our analysis with relation (40), which in case $h = 0$, $L = 0$, $M_1^k = M_1 \succcurlyeq 0$ and $M_2^k = M_2 \succcurlyeq 0$ for all $k \geq 0$ (see the setting of Section 3) reads

$$\begin{aligned} c\|z^k - Ax^{k+1}\|^2 + \|x^k - x^{k+1}\|^2_{M_1} + \|z^k - z^{k+1}\|^2_{M_2} \leq \\ \|x^k - x^*\|^2_{M_1} + \|z^k - Ax^*\|^2_{M_2+c\,\mathrm{Id}} + \frac{1}{c}\|y^k - y^*\|^2 \\ - \left(\|x^{k+1} - x^*\|^2_{M_1} + \|z^{k+1} - Ax^*\|^2_{M_2+c\,\mathrm{Id}} + \frac{1}{c}\|y^{k+1} - y^*\|^2\right). \end{aligned} \tag{16}$$

for all $k \geq 0$. Using that

$$c\|z^k - Ax^{k+1}\|^2 = c\left\|z^k - z^{k+1} - \frac{1}{c}(y^{k+1} - y^k)\right\|^2$$
$$= c\|z^k - z^{k+1}\|^2 + \frac{1}{c}\|y^{k+1} - y^k\|^2 + 2\langle z^{k+1} - z^k, y^{k+1} - y^k\rangle,$$

we obtain from (16) that

$$2\langle z^{k+1} - z^k, y^{k+1} - y^k\rangle + \|x^k - x^{k+1}\|^2_{M_1} + \|z^k - z^{k+1}\|^2_{M_2+c\,\mathrm{Id}} + \frac{1}{c}\|y^{k+1} - y^k\|^2 \leq$$
$$\|x^k - x^*\|^2_{M_1} + \|z^k - Ax^*\|^2_{M_2+c\,\mathrm{Id}} + \frac{1}{c}\|y^k - y^*\|^2$$
$$- \left(\|x^{k+1} - x^*\|^2_{M_1} + \|z^{k+1} - Ax^*\|^2_{M_2+c\,\mathrm{Id}} + \frac{1}{c}\|y^{k+1} - y^*\|^2\right) \tag{17}$$

for all $k \geq 0$. By taking into account that, according to (54),

$$\langle z^{k+1} - z^k, y^{k+1} - y^k\rangle \geq \frac{1}{2}\|z^{k+1} - z^k\|^2_{M_2} - \frac{1}{2}\|z^k - z^{k-1}\|^2_{M_2}$$

for all $k \geq 1$, it yields

$$\|x^k - x^{k+1}\|^2_{M_1} + \|z^k - z^{k+1}\|^2_{M_2+c\,\mathrm{Id}} + \frac{1}{c}\|y^{k+1} - y^k\|^2 \leq$$
$$\|x^k - x^*\|^2_{M_1} + \|z^k - Ax^*\|^2_{M_2+c\,\mathrm{Id}} + \frac{1}{c}\|y^k - y^*\|^2 + \|z^k - z^{k-1}\|^2_{M_2}$$
$$- \left(\|x^{k+1} - x^*\|^2_{M_1} + \|z^{k+1} - Ax^*\|^2_{M_2+c\,\mathrm{Id}} + \frac{1}{c}\|y^{k+1} - y^*\|^2 + \|z^{k+1} - z^k\|^2_{M_2}\right),$$

which is nothing else than (15).

From here, by using that $v^{k+1} \leq v^k$ for all $k \geq 0$ and straightforward telescoping arguments, it follows immediately that $(\|Ax^k - z^k\|)_{k\geq 0}$ converges to zero with a rate of $\mathcal{O}(1/\sqrt{k})$.

We will see in the following section that the inequality (40) will play an essential role also in the convergence analysis of the sequence of iterates. When applied to the particular context of the optimization problem (1) and Algorithm 2, Theorem 15 provides a rigorous formulation and a correct and clear proof of the convergence result stated in [28, Theorem 5.6].

3 A variant of the ADMM algorithm in the presence of a smooth function and by involving variable metrics

In this section we propose an extension of the ADMM algorithm considered in [28] that we also investigate from the perspective of its convergence properties. This extension is twofold: on the one hand, we consider an additional convex differentiable function in the objective of the optimization problem (1), which is evaluated in the algorithm through its gradient, and on the other hand, instead of fixed matrices M_1, M_2, we use different matrices in each iteration. Furthermore, we change the setting to infinite dimensional Hilbert spaces. We start by describing the problem under investigation:

Problem 6 Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces, $f : \mathcal{H} \to \overline{\mathbb{R}}$, $g : \mathcal{G} \to \overline{\mathbb{R}}$ be proper, convex and lower semicontinuous functions, $h : \mathcal{H} \to \mathbb{R}$ a convex and Fréchet differentiable function with L-Lipschitz continuous gradient (where $L \geq 0$) and $A : \mathcal{H} \to \mathcal{G}$ a linear continuous operator. The Lagrangian associated with the convex optimization problem

$$\inf_{x \in \mathcal{H}} \{f(x) + h(x) + g(Ax)\} \tag{18}$$

is

$$l : \mathcal{H} \times \mathcal{G} \times \mathcal{G} \to \overline{\mathbb{R}},\ l(x, z, y) = f(x) + h(x) + g(z) + \langle y, Ax - z \rangle.$$

We say that $(x^*, z^*, y^*) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}$ is a saddle point of the Lagrangian l, if the following inequalities hold

$$l(x^*, z^*, y) \leq l(x^*, z^*, y^*) \leq l(x, z, y^*)\ \forall (x, z, y) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}. \tag{19}$$

Notice that (x^*, z^*, y^*) is a saddle point if and only if $z^* = Ax^*$, x^* is an optimal solution of (18), y^* is an optimal solution of the Fenchel-Rockafellar dual problem to (18)

$$(D') \quad \sup_{y \in \mathcal{G}} \{-(f^* \,\square\, h^*)(-A^*y) - g^*(y)\}, \tag{20}$$

and the optimal objective values of (18) and (20) coincide, where $A^* : \mathcal{G} \to \mathcal{H}$ is the adjoint operator defined by $\langle A^*v, x \rangle = \langle v, Ax \rangle$ for all $(v, x) \in \mathcal{G} \times \mathcal{H}$. The infimal convolution $f^* \,\square\, h^* : \mathcal{H} \to \overline{\mathbb{R}}$ is defined by $(f^* \,\square\, h^*)(x) = \inf_{u \in \mathcal{H}} \{f^*(u) + h^*(x - u)\}$ for all $x \in \mathcal{H}$.

For the reader's convenience, we discuss some situations which lead to the existence of saddle points. This is for instance the case when (18) has an optimal solution and the Attouch-Brézis qualification condition

$$0 \in \operatorname{sri}(\operatorname{dom} g - A(\operatorname{dom} f)) \tag{21}$$

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

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

its strong relative interior. Notice that the classical interior is contained in the strong relative interior: $\operatorname{int} S \subseteq \operatorname{sri} S$, 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 $\operatorname{sri} S = \operatorname{ri} S$. Considering again the infinite dimensional setting, we remark that condition (21) is fulfilled if there exists $x' \in \operatorname{dom} f$ such that $Ax' \in \operatorname{dom} g$ and g is continuous at Ax'.

The optimality conditions for the primal-dual pair of optimization problems (18)-(20) read

$$-A^*y - \nabla h(x) \in \partial f(x) \text{ and } y \in \partial g(Ax). \tag{22}$$

This means that if (18) has an optimal solution $x^* \in \mathcal{H}$ and the qualification condition (21) is fulfilled, then there exists $y^* \in \mathcal{G}$, an optimal solution of (20), such that (22) holds and (x^*, Ax^*, y^*) is a saddle point of the Lagrangian l. Conversely, if the pair $(x^*, y^*) \in \mathcal{H} \times \mathcal{G}$ satisfies relation (22), then x^* is an optimal solution to (18), y^* is an optimal solution to (20) and (x^*, Ax^*, y^*) is a saddle point of the Lagrangian l. For further considerations on convex duality we invite the reader to consult [3–5, 20, 30].

Furthermore, we discuss some conditions ensuring that (18) has an optimal solution. Suppose that (18) is feasible, which means that its optimal objective value is not $+\infty$. The existence of optimal solutions to (18) is guaranteed if, for instance, $f+h$ is coercive (that is $\lim_{\|x\|\to\infty}(f+h)(x)=+\infty$) and g is bounded from below. Indeed, under these circumstances, the objective function of (18) is coercive and the statement follows via [3, Corollary 11.15]. On the other hand, if $f+h$ is strongly convex, then the objective function of (18) is strongly convex, too, thus (18) has a unique optimal solution (see [3, Corollary 11.16]).

Some more notations are in order before we state the algorithm for solving Problem 6. 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 semidefinite. For $U\in\mathcal{S}_+(\mathcal{H})$ we consider the semi-norm defined by

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

We also make use of the Loewner partial ordering 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

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

Algorithm 7 *Let $M_1^k\in\mathcal{S}_+(\mathcal{H})$ and $M_2^k\in\mathcal{S}_+(\mathcal{G})$ for all $k\geq0$. Choose $(x^0,z^0,y^0)\in\mathcal{H}\times\mathcal{G}\times\mathcal{G}$ and $c>0$. For all $k\geq0$ generate the sequence $(x^k,z^k,y^k)_{k\geq0}$ as follows:*

$$x^{k+1}\in\operatorname*{argmin}_{x\in\mathcal{H}}\left\{f(x)+\langle x-x^k,\nabla h(x^k)\rangle+\frac{c}{2}\|Ax-z^k+c^{-1}y^k\|^2+\frac{1}{2}\|x-x^k\|_{M_1^k}^2\right\} \tag{23}$$

$$z^{k+1}=\operatorname*{argmin}_{z\in\mathcal{G}}\left\{g(z)+\frac{c}{2}\|Ax^{k+1}-z+c^{-1}y^k\|^2+\frac{1}{2}\|z-z^k\|_{M_2^k}^2\right\} \tag{24}$$

$$y^{k+1}=y^k+c(Ax^{k+1}-z^{k+1}). \tag{25}$$

Remark 8 (i) If $h=0$ and $M_1^k=M_1$, $M_2^k=M_2$ are constant in each iteration, then Algorithm 7 becomes Algorithm 2, which has been investigated in [28].

(ii) In order to ensure that the sequence $(x^k)_{k\geq0}$ is uniquely determined one can assume that for all $k\geq0$ there exists $\alpha_1^k>0$ such that $M_1^k+cA^*A\in\mathcal{P}_{\alpha_1^k}(\mathcal{H})$.

This is in particular the case when

$$\exists\alpha>0\text{ such that }A^*A\in\mathcal{P}_\alpha(\mathcal{H}). \tag{26}$$

Relying on [3, Fact 2.19], on can see that (26) holds if and only if A is injective and $\operatorname{ran}A^*$ is closed. Hence, in finite dimensional spaces, namely, if $\mathcal{H}=\mathbb{R}^n$ and $\mathcal{G}=\mathbb{R}^m$, with $m\geq n\geq1$, (26) is nothing else than saying that A has full column rank.

(iii) One of the pioneering works addressing proximal ADMM algorithms in Hilbert spaces, in the particular case when $h=0$ and M_1^k and M_2^k are equal for all $k\geq0$ to the corresponding identity operators, is the paper by Attouch and Soueycatt [2]. We also refer the reader to [22,26] for versions of the proximal ADMM algorithm stated in finite-dimensional spaces and with proximal terms induced by constant linear operators.

Remark 9 We show that the particular choices $M_1^k = \frac{1}{\tau_k}\,\mathrm{Id} - cA^*A$, for $\tau_k > 0$, and $M_2^k = 0$ for all $k \geq 0$ lead to a primal-dual algorithm introduced in [16]. Here $\mathrm{Id} : \mathcal{H} \to \mathcal{H}$ denotes the identity operator on $\mathcal{H}$. Let $k \geq 0$ be fixed. The optimality condition for (23) reads (for x^{k+2}):

$$\begin{aligned} 0 &\in \partial f(x^{k+2}) + cA^*(Ax^{k+2} - z^{k+1} + c^{-1}y^{k+1}) + M_1^{k+1}(x^{k+2} - x^{k+1}) + \nabla h(x^{k+1}) \\ &= \partial f(x^{k+2}) + (cA^*A + M_1^{k+1})x^{k+2} + cA^*(-z^{k+1} + c^{-1}y^{k+1}) - M_1^{k+1}x^{k+1} + \nabla h(x^{k+1}). \end{aligned}$$

From (25) we have

$$cA^*(-z^{k+1} + c^{-1}y^{k+1}) = A^*(2y^{k+1} - y^k) - cA^*Ax^{k+1},$$

hence

$$0 \in \partial f(x^{k+2}) + (cA^*A + M_1^{k+1})(x^{k+2} - x^{k+1}) + A^*(2y^{k+1} - y^k) + \nabla h(x^{k+1}). \tag{27}$$

By taking into account the special choice of M_1^k we obtain

$$0 \in \partial f(x^{k+2}) + \frac{1}{\tau_{k+1}}\left(x^{k+2} - x^{k+1}\right) + A^*(2y^{k+1} - y^k) + \nabla h(x^{k+1}),$$

thus,

$$\begin{aligned} x^{k+2} &= (\mathrm{Id} + \tau_{k+1}\partial f)^{-1}\left(x^{k+1} - \tau_{k+1}\nabla h(x^{k+1}) - \tau_{k+1}A^*(2y^{k+1} - y^k)\right) \\ &= \operatorname*{argmin}_{x\in\mathcal{H}}\left\{f(x) + \frac{1}{2\tau_{k+1}}\left\|x - \left(x^{k+1} - \tau_{k+1}\nabla h(x^{k+1}) - \tau_{k+1}A^*(2y^{k+1} - y^k)\right)\right\|^2\right\}. \end{aligned} \tag{28}$$

Furthermore, from the optimality condition for (24) we obtain

$$c(Ax^{k+1} - z^{k+1} + c^{-1}y^k) + M_2^k(z^k - z^{k+1}) \in \partial g(z^{k+1}), \tag{29}$$

which combined with (25) gives

$$y^{k+1} + M_2^k(z^k - z^{k+1}) \in \partial g(z^{k+1}). \tag{30}$$

Using that $M_2^k = 0$ and again (25), it further follows

$$\begin{aligned} 0 &\in \partial g^*(y^{k+1}) - z^{k+1} \\ &= \partial g^*(y^{k+1}) + c^{-1}(y^{k+1} - y^k - cAx^{k+1}), \end{aligned}$$

which is equivalent to

$$\begin{aligned} y^{k+1} &= (\mathrm{Id} + c\partial g^*)^{-1}\left(y^k + cAx^{k+1}\right) \\ &= \operatorname*{argmin}_{z\in\mathcal{G}}\left\{g^*(z) + \frac{1}{2c}\left\|z - \left(y^k + cAx^{k+1}\right)\right\|^2\right\}. \end{aligned} \tag{31}$$

The iterative scheme obtained in (31) and (28) generates, for a given starting point $(x^1, y^0) \in \mathcal{H} \times \mathcal{G}$ and $c > 0$, the sequence $(x^k, y^k)_{k\geq 1}$ for all $k \geq 0$ as follows

$$\begin{aligned} y^{k+1} &= \operatorname*{argmin}_{z\in\mathcal{G}}\left\{g^*(z) + \frac{1}{2c}\left\|z - \left(y^k + cAx^{k+1}\right)\right\|^2\right\} \\ x^{k+2} &= \operatorname*{argmin}_{x\in\mathcal{H}}\left\{f(x) + \frac{1}{2\tau_{k+1}}\left\|x - \left(x^{k+1} - \tau_{k+1}\nabla h(x^{k+1}) - \tau_{k+1}A^*(2y^{k+1} - y^k)\right)\right\|^2\right\}. \end{aligned}$$

For $\tau_k = \tau > 0$ for all $k \geq 1$ one recovers a primal-dual algorithm from [16] that has been investigated under the assumption $\frac{1}{\tau} - c\|A\|^2 > \frac{L}{2}$ (see Algorithm 3.2 and Theorem 3.1 in [16]). We invite the reader to consult [8, 9, 13, 29] for more insights into primal-dual algorithms and their highlights. Primal-dual algorithms with dynamic step sizes have been investigated in [13] and [9], where it has been shown that clever strategies in the choice of the step sizes can improve the convergence behavior.

3.1 Ergodic convergence rates for the primal-dual gap

In this section we will provide a convergence rate result for a primal-dual gap function formulated in terms of the associated Lagrangian l. We start by proving a technical statement (see also [28]).

Lemma 10 *In the context of Problem 6, let $(x^k, z^k, y^k)_{k\geq 0}$ be a sequence generated by Algorithm 7. Then for all $k \geq 0$ and all $(x, z, y) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}$ the following inequality holds*

$$\begin{aligned} l(x^{k+1}, z^{k+1}, y) \leq\ & l(x, z, y^{k+1}) + c\langle z^{k+1} - z^k, A(x - x^{k+1})\rangle \\ & + \frac{1}{2}\left(\|x - x^k\|^2_{M_1^k} + \|z - z^k\|^2_{M_2^k} + c^{-1}\|y - y^k\|^2\right) \\ & - \frac{1}{2}\left(\|x - x^{k+1}\|^2_{M_1^k} + \|z - z^{k+1}\|^2_{M_2^k} + c^{-1}\|y - y^{k+1}\|^2\right) \\ & - \frac{1}{2}\left(\|x^{k+1} - x^k\|^2_{M_1^k} - L\|x^{k+1} - x^k\|^2 + \|z^{k+1} - z^k\|^2_{M_2^k} + c^{-1}\|y^{k+1} - y^k\|^2\right). \end{aligned}$$

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

$$c\langle z^{k+1} - z^k, A(x - x^{k+1})\rangle \leq \frac{c}{2}\left(\|Ax - z^k\|^2 - \|Ax - z^{k+1}\|^2\right) + \frac{1}{2c}\|y^{k+1} - y^k\|^2.$$

Proof. We fix $k \geq 0$ and $(x, z, y) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}$. Writing the optimality conditions for (23) we obtain

$$-\nabla h(x^k) + cA^*(z^k - c^{-1}y^k - Ax^{k+1}) + M_1^k(x^k - x^{k+1}) \in \partial f(x^{k+1}). \tag{32}$$

From the definition of the convex subdifferential we derive

$$\begin{aligned} f(x^{k+1}) - f(x) &\leq \langle \nabla h(x^k) + cA^*(-z^k + c^{-1}y^k + Ax^{k+1}) + M_1^k(-x^k + x^{k+1}), x - x^{k+1}\rangle \\ &= \langle \nabla h(x^k), x - x^{k+1}\rangle + \langle y^{k+1}, A(x - x^{k+1})\rangle - c\langle z^k - z^{k+1}, A(x - x^{k+1})\rangle \\ &\quad + \langle M_1^k(x^{k+1} - x^k), x - x^{k+1}\rangle, \end{aligned} \tag{33}$$

where for the last equality we used (25).

Furthermore, we claim that

$$h(x^{k+1}) - h(x) \leq -\langle \nabla h(x^k), x - x^{k+1}\rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2. \tag{34}$$

Indeed, this follows by applying the convexity of h and the descent lemma (see [3, Theorem 18.15 (iii)]):

$$\begin{aligned} h(x) - h(x^{k+1}) - \langle \nabla h(x^k), x - x^{k+1}\rangle \geq \\ h(x^k) + \langle \nabla h(x^k), x - x^k\rangle - h(x^{k+1}) - \langle \nabla h(x^k), x - x^{k+1}\rangle = \\ h(x^k) - h(x^{k+1}) + \langle \nabla h(x^k), x^{k+1} - x^k\rangle \geq -\frac{L}{2}\|x^{k+1} - x^k\|^2. \end{aligned}$$

By combining (33) and (34) we obtain

$$(f+h)(x^{k+1}) \leq (f+h)(x) + \langle y^{k+1}, A(x-x^{k+1})\rangle - c\langle z^k - z^{k+1}, A(x-x^{k+1})\rangle \\ + \frac{1}{2}\|x-x^k\|^2_{M_1^k} - \frac{1}{2}\|x-x^{k+1}\|^2_{M_1^k} - \frac{1}{2}\|x^{k+1}-x^k\|^2_{M_1^k} + \frac{L}{2}\|x^{k+1}-x^k\|^2. \tag{35}$$

From the optimality condition for (24) we obtain

$$c(Ax^{k+1} - z^{k+1} + c^{-1}y^k) + M_2^k(z^k - z^{k+1}) \in \partial g(z^{k+1}), \tag{36}$$

which, combined with (25), gives

$$y^{k+1} + M_2^k(z^k - z^{k+1}) \in \partial g(z^{k+1}). \tag{37}$$

From here we derive the inequality

$$g(z^{k+1}) - g(z) \leq \langle -y^{k+1} + M_2^k(z^{k+1}-z^k), z - z^{k+1}\rangle \\ = -\langle y^{k+1}, z - z^{k+1}\rangle + \frac{1}{2}\|z-z^k\|^2_{M_2^k} - \frac{1}{2}\|z-z^{k+1}\|^2_{M_2^k} - \frac{1}{2}\|z^{k+1}-z^k\|^2_{M_2^k}. \tag{38}$$

The first statement of the lemma follows by combining the inequalities (35) and (38) with the identity (see (25))

$$\langle y, Ax^{k+1} - z^{k+1}\rangle = \langle y^{k+1}, Ax^{k+1} - z^{k+1}\rangle + \frac{1}{2c}\left(\|y-y^k\|^2 - \|y-y^{k+1}\|^2 - \|y^{k+1}-y^k\|^2\right).$$

The second statement follows easily from the arithmetic-geometric mean inequality in Hilbert spaces (see [28, Proposition 5.3(a)]). ■

A direct consequence of the two inequalities in Lemma 10 is the following result.

Lemma 11 *In the context of Problem 6, assume that $M_1^k - L\,\mathrm{Id} \in \mathcal{S}_+(\mathcal{H}), M_1^k \succcurlyeq M_1^{k+1}$, $M_2^k \in \mathcal{S}_+(\mathcal{G}), M_2^k \succcurlyeq M_2^{k+1}$ for all $k \geq 0$, and let $(x^k, z^k, y^k)_{k\geq 0}$ be the sequence generated by Algorithm 7. Then for all $k \geq 0$ and all $(x, z, y) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}$ the following inequality holds*

$$\begin{aligned} l(x^{k+1}, z^{k+1}, y) \;\leq\; & l(x, z, y^{k+1}) + \frac{c}{2}\left(\|Ax - z^k\|^2 - \|Ax - z^{k+1}\|^2\right) \\ & + \frac{1}{2}\left(\|x-x^k\|^2_{M_1^k} - \|x-x^{k+1}\|^2_{M_1^{k+1}} + \|z-z^k\|^2_{M_2^k} - \|z-z^{k+1}\|^2_{M_2^{k+1}}\right) \\ & + \frac{1}{2c}\left(\|y-y^k\|^2 - \|y-y^{k+1}\|^2\right). \end{aligned}$$

We can now state the main result of this subsection.

Theorem 12 *In the context of Problem 6, assume that $M_1^k - L\,\mathrm{Id} \in \mathcal{S}_+(\mathcal{H}), M_1^k \succcurlyeq M_1^{k+1}$, $M_2^k \in \mathcal{S}_+(\mathcal{G}), M_2^k \succcurlyeq M_2^{k+1}$ for all $k \geq 0$, and let $(x^k, z^k, y^k)_{k\geq 0}$ be the sequence generated by Algorithm 7. For all $k \geq 1$ define the ergodic sequences*

$$\overline{x}^k := \frac{1}{k}\sum_{i=1}^k x^i,\ \overline{z}^k := \frac{1}{k}\sum_{i=1}^k z^i,\ \overline{y}^k := \frac{1}{k}\sum_{i=1}^k y^i.$$

Then for all $k \geq 1$ and all $(x, z, y) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}$ it holds

$$l(\overline{x}^k, \overline{z}^k, y) - l(x, z, \overline{y}^k) \leq \frac{\gamma(x, z, y)}{k},$$

where $\gamma(x, z, y) := \frac{c}{2}\|Ax - z^0\|^2 + \frac{1}{2}\left(\|x - x^0\|^2_{M_1^0} + \|z - z^0\|^2_{M_2^0}\right) + \frac{1}{2c}\|y - y^0\|^2$.

Proof. We fix $k \geq 1$ and $(x, z, y) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}$. Summing up the inequalities in Lemma 11 for $i = 0, ..., k-1$ and using classical arguments for telescoping sums, we obtain

$$\sum_{i=0}^{k-1} l(x^{k+1}, z^{k+1}, y) \leq \sum_{i=0}^{k-1} l(x, z, y^{k+1}) + \gamma(x, z, y).$$

Since l is convex in (x, z) and linear in y, the conclusion follows from the definition of the ergodic sequences. ■

Remark 13 Let (x^*, z^*, y^*) be a saddle point for the Lagrangian l. By taking $(x, z, y) := (x^*, z^*, y^*)$ in the above theorem it yields

$$(f + h)(\overline{x}^k) + g(\overline{z}^k) + \langle y^*, A\overline{x}^k - \overline{z}^k\rangle - \big(f(x^*) + h(x^*) + g(Ax^*)\big) \leq \frac{\gamma(x^*, z^*, y^*)}{k} \ \forall k \geq 1,$$

where $f(x^*) + h(x^*) + g(Ax^*)$ is the optimal objective value of the problem (18). Hence, if we suppose that the set of optimal solutions of the dual problem (20) is contained in a bounded set, there exists $R > 0$ such that for all $k \geq 1$

$$(f + h)(\overline{x}^k) + g(\overline{z}^k) + R\|A\overline{x}^k - \overline{z}^k\| - \big(f(x^*) + h(x^*) + g(Ax^*)\big) \leq$$
$$\frac{1}{k}\left(\frac{c}{2}\|Ax^* - z^0\|^2 + \frac{1}{2}\|x^* - x^0\|^2_{M_1^0} + \frac{1}{2}\|z^* - z^0\|^2_{M_2^0} + \frac{1}{c}(R^2 + \|y^0\|^2)\right).$$

The set of dual optimal solutions of (20) is equal to the convex subdifferential of the infimal value function of the problem (18)

$$\psi : \mathcal{G} \to \overline{\mathbb{R}},\ \psi(y) = \inf_{x \in \mathcal{H}} \left(f(x) + h(x) + g(Ax + y)\right),$$

at 0. This set is weakly compact, thus bounded, if $0 \in \operatorname{int}(\operatorname{dom}\psi) = \operatorname{int}(A(\operatorname{dom} f) - \operatorname{dom} g)$ (see [3, 5, 30]).

3.2 Convergence of the sequence of generated iterates

In this subsection we will address the convergence of the sequence of iterates generated by Algorithm 7 (see also [6, Theorem 7]). One of the important tools for the proof of the convergence result will be [15, Theorem 3.3], which we recall below.

Lemma 14 *(see [15, Theorem 3.3]) Let S be a nonempty subset of $\mathcal{H}$ and $(x^k)_{k\geq 0}$ 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.

The proof of the convergence result relies on techniques specific to monotone operator theory and does not make use of the values of the objective function or of the Lagrangian l. This makes it different from the proofs in [28] and from the majority of other conventional convergence proofs for ADMM methods. To the few exceptions belong [2] and [19].

Theorem 15 *In the context of Problem 6, assume that the set of saddle points of the Lagrangian l is nonempty and that $M_1^k - \frac{L}{2}\,\mathrm{Id} \in \mathcal{S}_+(\mathcal{H}), M_1^k \succcurlyeq M_1^{k+1}$, $M_2^k \in \mathcal{S}_+(\mathcal{G}), M_2^k \succcurlyeq M_2^{k+1}$ for all $k \geq 0$, and let $(x^k, z^k, y^k)_{k\geq 0}$ be the sequence generated by Algorithm 7. If one of the following assumptions*

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

(II) there exists $\alpha, \alpha_2 > 0$ such that $M_1^k - \frac{L}{2}\,\mathrm{Id} + A^*A \in \mathcal{P}_\alpha(\mathcal{H})$ and $M_2^k \in \mathcal{P}_{\alpha_2}(\mathcal{G})$ for all $k \geq 0$;

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

is fulfilled, then $(x^k, z^k, y^k)_{k\geq 0}$ converges weakly to a saddle point of the Lagrangian l. This means that $(x^k)_{k\geq 0}$ converges weakly to an optimal solution of problem (18), and $(y^k)_{k\geq 0}$ converges weakly to an optimal solution of its dual problem (20).

Proof. Let $S \subseteq \mathcal{H}\times\mathcal{G}\times\mathcal{G}$ denote the set of the saddle points of the Lagrangian l and (x^*, z^*, y^*) be a fixed element in S. Then $z^* = Ax^*$ and the optimality conditions hold

$$-A^*y^* - \nabla h(x^*) \in \partial f(x^*),\ y^* \in \partial g(Ax^*).$$

Let $k \geq 0$ be fixed. Taking into account (32), (36) and the monotonicity of ∂f and ∂g, we obtain

$$\langle cA^*(z^k - Ax^{k+1} - c^{-1}y^k) + M_1^k(x^k - x^{k+1}) - \nabla h(x^k) + A^*y^* + \nabla h(x^*), x^{k+1} - x^*\rangle \geq 0$$

and

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

We consider first the case $L > 0$. By the Baillon-Haddad Theorem (see [3, Corollary 18.16]), the gradient of h is L^{-1}-cocoercive, hence the following inequality holds

$$\langle \nabla h(x^*) - \nabla h(x^k), x^* - x^k\rangle \geq L^{-1}\|\nabla h(x^*) - \nabla h(x^k)\|^2.$$

Summing up the three inequalities obtained above we get

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

Further, by taking into account (25), it holds

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

By using some expressions of the inner products in terms of norms, we obtain

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

By using again relation (25) for expressing $Ax^{k+1} - z^{k+1}$ and by taking into account that

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

it yields

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

and from here, by using the monotonicity assumptions on $(M_1^k)_{k\geq 0}$ and $(M_2^k)_{k\geq 0}$, we finally get

$$\begin{aligned}
&\frac{1}{2}\|x^{k+1} - x^*\|^2_{M_1^{k+1}} + \frac{1}{2}\|z^{k+1} - Ax^*\|^2_{M_2^{k+1} + c\,\mathrm{Id}} + \frac{1}{2c}\|y^{k+1} - y^*\|^2 \leq\\
&\frac{1}{2}\|x^k - x^*\|^2_{M_1^k} + \frac{1}{2}\|z^k - Ax^*\|^2_{M_2^k + c\,\mathrm{Id}} + \frac{1}{2c}\|y^k - y^*\|^2\\
&-\frac{c}{2}\|z^k - Ax^{k+1}\|^2 - \frac{1}{2}\|x^k - x^{k+1}\|^2_{M_1^k - \frac{L}{2}\,\mathrm{Id}} - \frac{1}{2}\|z^k - z^{k+1}\|^2_{M_2^k}\\
&-L\left\|L^{-1}\left(\nabla h(x^*) - \nabla h(x^k)\right) + \frac{1}{2}\left(x^k - x^{k+1}\right)\right\|^2.
\end{aligned} \tag{39}$$

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

$$\begin{aligned}
&\frac{1}{2}\|x^{k+1} - x^*\|^2_{M_1^{k+1}} + \frac{1}{2}\|z^{k+1} - Ax^*\|^2_{M_2^{k+1} + c\,\mathrm{Id}} + \frac{1}{2c}\|y^{k+1} - y^*\|^2 \leq\\
&\frac{1}{2}\|x^k - x^*\|^2_{M_1^k} + \frac{1}{2}\|z^k - Ax^*\|^2_{M_2^k + c\,\mathrm{Id}} + \frac{1}{2c}\|y^k - y^*\|^2\\
&-\frac{c}{2}\|z^k - Ax^{k+1}\|^2 - \frac{1}{2}\|x^k - x^{k+1}\|^2_{M_1^k} - \frac{1}{2}\|z^k - z^{k+1}\|^2_{M_2^k}.
\end{aligned} \tag{40}$$

It is easy to see, by using arguments invoking telescoping sums, that, in both cases, (39) and (40) yield

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

The case when Assumption (I) is valid.

By neglecting the negative terms from the right-hand side of both (39) and (40), it follows that the first assumption in Lemma 14 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 $S \subseteq \mathcal{H}\times\mathcal{G}\times\mathcal{G}$ the set of saddle points of the Lagrangian l.

From (41) we get

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

since $M_1^k - \frac{L}{2}\,\mathrm{Id} \in \mathcal{P}_{\alpha_1}(\mathcal{H})$ for all $k \geq 0$ with $\alpha_1 > 0$, and

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

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

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

From (25), (43) and (44) we derive

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

The relations (42)-(45) will play an essential role in the verification of the second assumption in Lemma 14. 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 (42) we obtain that $(Ax^{k_n+1})_{n\in\mathbb{N}}$ converges weakly to $A\overline{x}$ (as $n \to +\infty$), which combined with (43) yields $\overline{z} = A\overline{x}$. We use now the following notations for all $n \geq 0$

$$\begin{aligned}
a_n^* &:= cA^*(z^{k_n} - Ax^{k_n+1} - c^{-1}y^{k_n}) + M_1^{k_n}(x^{k_n} - x^{k_n+1}) + \nabla h(x^{k_n+1}) - \nabla h(x^{k_n})\\
a_n &:= x^{k_n+1}\\
b_n^* &:= y^{k_n+1} + M_2^{k_n}(z^{k_n} - z^{k_n+1})\\
b_n &:= z^{k_n+1}.
\end{aligned}$$

From (32) and (37) we have for all $n \geq 0$

$$a_n^* \in \partial(f+h)(a_n) \quad (46)$$

and

$$b_n^* \in \partial g(b_n). \quad (47)$$

Furthermore, from (42) we have

$$a_n \text{ converges weakly to } \overline{x} \text{ (as } n \to +\infty). \quad (48)$$

From (45) and (44) we obtain

$$b_n^* \text{ converges weakly to } \overline{y} \text{ (as } n \to +\infty). \quad (49)$$

Moreover, (25) and (45) yield

$$Aa_n - b_n \text{ converges strongly to } 0 \text{ (as } n \to +\infty). \tag{50}$$

Finally, we have

$$a_n^* + A^*b_n^* = cA^*(z^{k_n} - Ax^{k_n+1}) + A^*(y^{k_n+1} - y^{k_n}) + M_1^{k_n}(x^{k_n} - x^{k_n+1}) + A^*M_2^{k_n}(z^{k_n} - z^{k_n+1}) + \nabla h(x^{k_n+1}) - \nabla h(x^{k_n}).$$

By using the fact that ∇h is Lipschitz continuous, from (42)-(45) we get

$$a_n^* + A^*b_n^* \text{ converges strongly to } 0 \text{ (as } n \to +\infty). \tag{51}$$

Taking into account the relations (46)-(51) and applying [1, Proposition 2.4] to the operators $\partial(f+h)$ and ∂g, we conclude that

$$-A^*\overline{y} \in \partial(f+h)(\overline{x}) = \partial f(\overline{x}) + \nabla h(\overline{x}) \text{ and } \overline{y} \in \partial g(A\overline{x}),$$

hence $(\overline{x}, \overline{z}, \overline{y}) = (\overline{x}, A\overline{x}, \overline{y})$ is a saddle point of the Lagrangian l, thus the second assumption of the Lemma 14 is verified, too. In conclusion, $(x^k, z^k, y^k)_{k\geq 0}$ converges weakly to a saddle point of the Lagrangian l.

The case when Assumption (II) is valid.

We show that the relations (42)-(45) are fulfilled also in this case. Indeed, Assumption (II) allows to derive from (41) that (43) and (44) hold. From (25), (43) and (44) we obtain (45). Finally, the inequalities

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

yield (42).

On the other hand, notice that both (39) and (40) yield

$$\exists \lim_{k\to+\infty} \left(\frac{1}{2}\|x^k - x^*\|^2_{M_1^k} + \frac{1}{2}\|z^k - z^*\|^2_{M_2^k + c\,\mathrm{Id}} + \frac{1}{2c}\|y^k - y^*\|^2\right), \tag{52}$$

hence $(y^k)_{k\geq 0}$ and $(z^k)_{k\geq 0}$ are bounded. Combining this with (25) and the condition imposed on $M_1^k - \frac{L}{2}\,\mathrm{Id} + A^*A$, we derive that $(x^k)_{k\geq 0}$ is bounded, too. Hence there exists a weakly convergent subsequence of $(x^k, z^k, y^k)_{k\geq 0}$. By using the same arguments as in the proof of (I), it follows that every weak sequential cluster point of $(x^k, z^k, y^k)_{k\geq 0}$ is a saddle point of the Lagrangian l.

Now we show that the set of weak sequential cluster points of $(x^k, z^k, y^k)_{k\geq 0}$ is a singleton. Let $(x_1, z_1, y_1), (x_2, z_2, y_2)$ be two such weak sequential cluster points. Then there exist $(k_p)_{p\geq 0}, (k_q)_{q\geq 0}$, $k_p \to +\infty$ (as $p \to +\infty$), $k_q \to +\infty$ (as $q \to +\infty$), a subsequence $(x^{k_p}, z^{k_p}, y^{k_p})_{p\geq 0}$ which converges weakly to (x_1, z_1, y_1) (as $p \to +\infty$), and a subsequence $(x^{k_q}, z^{k_q}, y^{k_q})_{q\geq 0}$ which converges weakly to (x_2, z_2, y_2) (as $q \to +\infty$). As seen, (x_1, z_1, y_1) and (x_2, z_2, y_2) are saddle points of the Lagrangian l and $z_i = Ax_i$ for $i \in \{1, 2\}$. From (52), which is true for every saddle point of the Lagrangian l, we derive

$$\exists \lim_{k\to+\infty} \Big(E(x^k, z^k, y^k; x_1, z_1, y_1) - E(x^k, z^k, y^k; x_2, z_2, y_2)\Big), \tag{53}$$

where, for (x^*, z^*, y^*) the expression $E(x^k, z^k, y^k; x^*, z^*, y^*)$ is defined as

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

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

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

$$\frac{1}{2}\|z^k - z_1\|^2_{M_2^k + c\,\mathrm{Id}} - \frac{1}{2}\|z^k - z_2\|^2_{M_2^k + c\,\mathrm{Id}} = \frac{1}{2}\|z_2 - z_1\|^2_{M_2^k + c\,\mathrm{Id}} + \langle z^k - z_2, (M_2^k + c\,\mathrm{Id})(z_2 - z_1)\rangle,$$

and

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

Applying [27, Théorème 104.1], there exists $M_1 \in \mathcal{S}_+(\mathcal{H})$ such that $(M_1^k)_{k\geq 0}$ converges to M_1 in the strong operator topology, i.e., $\|M_1^k x - M_1 x\| \to 0$ for all $x \in \mathcal{H}$ (as $k \to +\infty$). Similarly, the monotonicity condition imposed on $(M_2^k)_{k\geq 0}$ implies that $\sup_{k\geq 0}\|M_2^k + c\,\mathrm{Id}\| < +\infty$. Thus, according to [15, Lemma 2.3], there exists $\alpha' > 0$ and $M_2 \in \mathcal{P}_{\alpha'}(\mathcal{G})$ such that $(M_2^k + c\,\mathrm{Id})_{k\geq 0}$ converges to M_2 in the strong operator topology (as $k \to +\infty$).

Taking the limit in (53) along the subsequences $(k_p)_{p\geq 0}$ and $(k_q)_{q\geq 0}$ and using the last three relations above we obtain

$$\frac{1}{2}\|x_1 - x_2\|^2_{M_1} + \langle x_1 - x_2, M_1(x_2 - x_1)\rangle + \frac{1}{2}\|z_1 - z_2\|^2_{M_2} + \langle z_1 - z_2, M_2(z_2 - z_1)\rangle$$

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

hence

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

From here we get $\|x_1 - x_2\|_{M_1} = 0$, $z_1 = z_2$ and $y_1 = y_2$. Since

$$\left(\alpha + \frac{L}{2}\right)\|x_1 - x_2\|^2 \leq \|x_1 - x_2\|^2_{M_1} + \|Ax_1 - Ax_2\|^2,$$

we obtain that $x_1 = x_2$. In conclusion, $(x^k, z^k, y^k)_{k\geq 0}$ converges weakly to a saddle point of the Lagrangian l.

The case when Assumption (III) is valid.

Under Assumption (III) we can further refine the inequalities in (39) and (40). Let $k \geq 1$ be fixed. By considering the relation (37) for consecutive iterates and by taking into account the monotonicity of ∂g 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

$$\begin{aligned}\langle z^{k+1} - z^k, y^{k+1} - y^k\rangle &\geq \|z^{k+1} - z^k\|^2_{M_2^k} + \langle z^{k+1} - z^k, M_2^{k-1}(z^{k-1} - z^k)\rangle \\ &\geq \|z^{k+1} - z^k\|^2_{M_2^k} - \frac{1}{2}\|z^{k+1} - z^k\|^2_{M_2^{k-1}} - \frac{1}{2}\|z^k - z^{k-1}\|^2_{M_2^{k-1}}. \qquad (54)\end{aligned}$$

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

$$\|z^{k+1}-z^k\|^2_{M_2^k}-\frac{1}{2}\|z^{k+1}-z^k\|^2_{M_2^{k-1}}-\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}\leq \\ \frac{c}{2}\left(\|z^k-Ax^{k+1}\|^2-\|z^{k+1}-z^k\|^2-\|Ax^{k+1}-z^{k+1}\|^2\right). \tag{55}$$

In case $L>0$, adding (55) and (39) leads to

$$\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-Ax^*\|^2_{M_2^{k+1}+c\,\mathrm{Id}}+\frac{1}{2c}\|y^{k+1}-y^*\|^2+\frac{1}{2}\|z^{k+1}-z^k\|^2_{3M_2^k-M_2^{k-1}}\leq \\ \frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-Ax^*\|^2_{M_2^k+c\,\mathrm{Id}}+\frac{1}{2c}\|y^k-y^*\|^2+\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}} \\ -\frac{1}{2}\|x^{k+1}-x^k\|^2_{M_1^k-\frac{L}{2}\,\mathrm{Id}}-\frac{c}{2}\|z^{k+1}-z^k\|^2-\frac{1}{2c}\|y^{k+1}-y^k\|^2 \\ -L\left\|L^{-1}\left(\nabla h(x^*)-\nabla h(x^k)\right)+\frac{1}{2}\left(x^k-x^{k+1}\right)\right\|^2.$$

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

$$\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-Ax^*\|^2_{M_2^{k+1}+c\,\mathrm{Id}}+\frac{1}{2c}\|y^{k+1}-y^*\|^2+\frac{1}{2}\|z^{k+1}-z^k\|^2_{M_2^k}\leq \\ \frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-Ax^*\|^2_{M_2^k+c\,\mathrm{Id}}+\frac{1}{2c}\|y^k-y^*\|^2+\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}} \\ -\frac{1}{2}\|x^{k+1}-x^k\|^2_{M_1^k-\frac{L}{2}\,\mathrm{Id}}-\frac{c}{2}\|z^{k+1}-z^k\|^2-\frac{1}{2c}\|y^{k+1}-y^k\|^2. \tag{56}$$

Similarly, in case $L=0$ we obtain

$$\frac{1}{2}\|x^{k+1}-x^*\|^2_{M_1^{k+1}}+\frac{1}{2}\|z^{k+1}-Ax^*\|^2_{M_2^{k+1}+c\,\mathrm{Id}}+\frac{1}{2c}\|y^{k+1}-y^*\|^2+\frac{1}{2}\|z^{k+1}-z^k\|^2_{M_2^k}\leq \\ \frac{1}{2}\|x^k-x^*\|^2_{M_1^k}+\frac{1}{2}\|z^k-Ax^*\|^2_{M_2^k+c\,\mathrm{Id}}+\frac{1}{2c}\|y^k-y^*\|^2+\frac{1}{2}\|z^k-z^{k-1}\|^2_{M_2^{k-1}} \\ -\frac{1}{2}\|x^{k+1}-x^k\|^2_{M_1^k}-\frac{c}{2}\|z^{k+1}-z^k\|^2-\frac{1}{2c}\|y^{k+1}-y^k\|^2. \tag{57}$$

Using telescoping sum arguments, we obtain that $\|x^{k+1}-x^k\|_{M_1^k-\frac{L}{2}\,\mathrm{Id}}\to 0$, $y^k-y^{k+1}\to 0$ and $z^k-z^{k+1}\to 0$ as $k\to+\infty$. Using (25), it follows that $A(x^k-x^{k+1})\to 0$ as $k\to+\infty$, which, combined with the fact that $M_1^k-\frac{L}{2}\,\mathrm{Id}+A^*A\in\mathcal{P}_\alpha(\mathcal{H})$, for all $k\geq 0$, yields $x^k-x^{k+1}\to 0$ as $k\to+\infty$. Consequently, $z^k-Ax^{k+1}\to 0$ as $k\to+\infty$. Hence, the relations (42)-(45) are fulfilled. On the other hand, from both (56) and (57) we derive

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

By using that

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

it follows that $\lim_{k\to+\infty}\|z^k-z^{k-1}\|^2_{M_2^{k-1}}=0$, which further implies that (52) holds. From here the conclusion follows by arguing as in the proof provided in the setting of Assumption (II). ■

Remark 16 Choosing as in Remark 9, $M_1^k = \frac{1}{\tau_k}\mathrm{Id} - cA^*A$, with $\tau_k > 0$ and such that $\tau := \sup_{k\geq 0}\tau_k \in \mathbb{R}$, and $M_2^k = 0$ for all $k \geq 0$, we have

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

which means that under the assumption $\frac{1}{\tau} - c\|A\|^2 > \frac{L}{2}$ (which recovers the one in Algorithm 3.2 and Theorem 3.1 in [16]), the operators $M_1^k - \frac{L}{2}\mathrm{Id}$ belong for all $k \geq 0$ to the class $\mathcal{P}_{\alpha_1}(\mathcal{H})$, with $\alpha_1 := \frac{1}{\tau} - c\|A\|^2 - \frac{L}{2} > 0$.

Remark 17 By taking $h = 0$ and $L = 0$, and in each iteration constant operators $M_1^k = M_1 \succcurlyeq 0$ and $M_2^k = M_2 \succcurlyeq 0$ for all $k \geq 0$, Theorem 15 in the context of Assumption (I) covers the first situation investigated in [28, Theorem 5.6], where in finite dimensional spaces the matrix M_1 was assumed to be positive definite and the matrix M_2 to be positive semidefinite.

The arguments used in [28, Theorem 5.6] for proving convergence in the case when $M_1 = 0$ and A has full column rank contain flaws and rely on incorrect statements. Theorem 15 provides in the context of Assumption (III) (for $h = 0$, $L = 0$, $M_1^k = 0$ and $M_2^k = M_2 \succcurlyeq 0$ for all $k \geq 0$) the correct proof of this result.

Finally, we notice that the convergence theorem for the iterates of the classical ADMM algorithm (which corresponds to the situation when $h = 0$, $L = 0$, $M_1 = M_2 = 0$ and A has full column rank, see for example [19]) is covered by Theorem 15 in the context of Assumption (III).

Acknowledgements. The authors are thankful to two anonymous reviewers and the handling editor for comments and remarks which have improved the quality of the paper.

References

[1] A. Alotaibi, P.L. Combettes, N. Shahzad, *Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn-Tucker set*, SIAM Journal on Optimization 24(4), 2076–2095, 2014

[2] H. Attouch, M. Soueycatt, *Augmented Lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. Applications to games, PDE's and control*, Pacific Journal of Optimization 5, 17–37, 2009

[3] H.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, second edition, CMS Books in Mathematics, Springer, New York, 2017

[4] J.M. Borwein and J.D. Vanderwerff, *Convex Functions: Constructions, Characterizations and Counterexamples*, Cambridge University Press, Cambridge, 2010

[5] R.I. Boţ, *Conjugate Duality in Convex Optimization*, Lecture Notes in Economics and Mathematical Systems, Vol. 637, Springer, Berlin Heidelberg, 2010

[6] R.I. Boţ, E.R. Csetnek, *ADMM for monotone operators: convergence analysis and rates, Advances in Computational Mathematics 45, 327–359, 2019*

[7] R.I. Boţ, E.R. Csetnek, *An inertial alternating direction method of multipliers*, Minimax Theory and its Applications 1(1), 29–49, 2016

[8] R.I. Boţ, 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

[9] R.I. Boţ, 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

[10] R.I. Boţ, 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] M.N. Bùi, P.L. Combettes, *Warped proximal iterations for monotone inclusions*, arXiv: 1908.07077, 2019

[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, *Monotone operator theory in convex optimization*, Mathematical Programming 170(1), 177–206, 2018

[15] P.L. Combettes, B.C. Vũ, *Variable metric quasi-Fejér monotonicity*, Nonlinear Analysis 78, 17–31, 2013

[16] 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

[17] J. Eckstein, *Augmented Lagrangian and alternating direction methods for convex optimization: a tutorial and some illustrative computational results*, Rutcor Research Report 32-2012, 2012

[18] J. Eckstein, *Some saddle-function splitting methods for convex programming*, Optimization Methods and Software 4, 75–83, 1994

[19] 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

[20] I. Ekeland, R. Temam, *Convex Analysis and Variational Problems*, North-Holland Publishing Company, Amsterdam, 1976

[21] E. Esser, *Applications of Lagrangian-based alternating direction methods and connections to split Bregman*, CAM Reports 09-31, UCLA, Center for Applied Mathematics, 2009

[22] M. Fazel, T.K. Pong, D.F. Sun, P. Tseng, *Hankel matrix rankg minimization with applications to system identification and realization*, SIAM Journal onf Matrix Analysis and Applications 534, 946–977, 2013

[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] M. Li, D.F. Sun, K.C. Toh, *A majorized ADMM with indefinite proximal terms for linearly constrained convex composite optimization*, SIAM Journal on Optimization 26, 922–950, 2016

[27] F. Riesz, B.Sz. Nagy, *Leçons d'Analyse Fonctionnelle*, fifth edition, Gauthier-Villars, Paris, 1968

[28] 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

[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