Fixing and extending some recent results on the ADMM algorithm

Sebastian Banert [*] Radu Ioan Boţ [†] Ernö Robert Csetnek [‡]

December 15, 2016

Abstract. We first point out several flaws in the recent paper *[R. Shefi, M. Teboulle: Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization, SIAM J. Optim. 24, 269–297, 2014]* that proposes two ADMM-type algorithms for solving convex optimization problems involving compositions with linear operators and show how some of the considered arguments can be fixed. 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 point of view of its convergence properties.

Key Words. ADMM algorithm, Lagrangian, saddle points, subdifferential, convex optimization, Fenchel duality

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). 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.

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

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

[*]University of Vienna, Faculty of Mathematics, Oskar-Morgenstern-Platz 1, A-1090 Vienna, Austria, email: sebastian.banert@univie.ac.at. Research 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: 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), Lise Meitner Programme, project M 1682-N25.

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 dual problem to (1)

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

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.

Notice that in case (1) has an optimal solution and $A(\operatorname{ri}(\operatorname{dom} f)) \cap \operatorname{ri}\operatorname{dom} g \neq \emptyset$, the set of saddle points of l is nonempty. Here, we denote by $\operatorname{ri}(S)$ 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 $(z^0, y^0) \in \mathbb{R}^m \times \mathbb{R}^m$ 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. The alternating direction method of multipliers was first introduced in [20] and [18]. Gabay has shown in [19] (see also [15]) 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 proper 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}. 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, [6–8, 10, 12, 22]). Despite of this fact, the ADMM algorithm has been widely used for solving convex optimization problems arising in real-life applications (see, for instance, [9, 17]). For a version of the ADMM algorithm with inertial and memory effects we refer the reader to [5].

In order to overcome the above-mentioned drawback of the classical ADMM method and to increase its flexibility, Shefi and Teboulle proposed in [21] the following so-called alternating direction proximal method of multipliers (AD-PMM):

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:*

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

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

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

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 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 $M_1 = \mu_1 I_n$ and $M_2 = \mu_2 I_m$ with $\mu_1, \mu_2 > 0$ and I_n and I_m denoting the identity $n \times n$ and $m \times m$ matrices, respectively, one recovers the algorithm from [14]. Furthermore, when $M_1 = \tau^{-1} I_n - cA^T A$ 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 [12].

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 nonempty, and in general is not a singleton. However, if one imposes that either A has full column rank or M_1 is positive definite, then $(x^k)_{k\geq 0}$ will be uniquely determined, too.

Shefi and Teboulle provide in [21] 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 functions value $(f(x^k)+g(Ax^k))_{k\geq 0}$ 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 [21] represents the starting point of the investigations that we will carry out as follows. More precisely, in this paper:

- we point out several flaws in [21], which influence the validity of the arguments used in the proof of the global convergence rate result for the sequence $(\|Ax^k - z^k\|)_{k\geq 0}$ to 0 and of the convergence result for the sequence $(x^k, z^k, y^k)_{k\geq 0}$;
- we show how some of the arguments used in the two statements mentioned above can be fixed under not very restrictive assumptions;
- 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 [21] related to the convergence analysis for Algorithm 2

In this section we point out several flaws in [21] in connection to the convergence analysis made for the sequence $(x^k, z^k, y^k)_{k\geq 0}$ generated by Algorithm 2. The statements in discussion influence the validity of the arguments used in the proof of the global convergence rate result for the sequence $(\|Ax^k - z^k\|)_{k\geq 0}$ to 0 and of the convergence result for the sequence $(x^k, z^k, y^k)_{k\geq 0}$. We also show how these arguments can be fixed under not very restrictive assumptions.

To proceed, we first recall some results from [21]. We start with a statement that follows from the variational characterization of the minimizers of (8)-(9) by means of the convex subdifferential.

Lemma 3 *(see [21, 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 $(x, z, y) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ the following inequality holds:*

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

Furthermore, by invoking the monotonicity of the convex subdifferential of g, the following estimation is derived in [21].

Lemma 4 *(see [21, 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 $(x, z) \in \mathbb{R}^n \times \mathbb{R}^m$ the following inequality holds:*

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

By taking $(x, z, y) := (x^*, z^*, y^*)$ in Lemma 3, where (x^*, z^*, y^*) is a saddle point of the Lagrangian l, 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 can derive 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:*

$$\begin{aligned}&\|x^{k+1}-x^k\|_{M_1}^2 + \|z^{k+1}-z^k\|_{M_2}^2 + c^{-1}\|y^{k+1}-y^k\|^2 + \\ &\|x^*-x^{k+1}\|_{M_1}^2 + \|z^*-z^{k+1}\|_{M_2+cI_m}^2 + c^{-1}\|y^*-y^{k+1}\|^2 + \|z^{k+1}-z^k\|_{M_2}^2 \\ \leq\ &\|x^*-x^k\|_{M_1}^2 + \|z^*-z^k\|_{M_2+cI_m}^2 + c^{-1}\|y^*-y^k\|^2 + \|z^k-z^{k-1}\|_{M_2}^2.\end{aligned}$$

By using the notations from [21, Section 5.3], namely

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

and

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

the inequality stated in Lemma 5 can be equivalently written as

$$v^{k+1} - c\|z^{k+1}-z^k\|^2 \leq u^k - u^{k+1} \ \forall k \geq 1. \tag{11}$$

Our first observation related to the validity of the results in [21] is that we doubt that in general the inequality

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

as stated in [21, Lemma 5.1, (5.37)], holds. In our opinion, the correct version of [21, Lemma 5.1, (5.37)] is the inequality (11) above.

Consequently, the arguments used for proving [21, Theorem 5.4], concerning the $\mathcal{O}(1/\sqrt{k})$ rate of convergence of the sequence $(\|Ax^k - z^k\|)_{k\geq 0}$, and [21, Theorem 5.6], concerning the convergence of the iterates generated by Algorithm 2, are not valid, either.

However, the proof given for [21, Theorem 5.4] in the context of proving the $\mathcal{O}(1/\sqrt{k})$ rate of convergence of the sequence $(\|Ax^k - z^k\|)_{k\geq 0}$ can be fixed, if one of M_1 or M_2 is positive definite.

Assume first that M_2 is positive definite (and M_1 remains positive semidefinite). From the inequality in Lemma 5 it follows that $\sum_{k\geq 1}\|z^{k+1}-z^k\|_{M_2}^2 < +\infty$, hence $\sum_{k\geq 1}\|z^{k+1}-z^k\|^2 < +\infty$. The $\mathcal{O}(1/\sqrt{k})$ rate of convergence for $(\|Ax^k - z^k\|)_{k\geq 0} = (c^{-1}\|y^k - y^{k-1}\|)_{k\geq 1}$ is obtained by summing up the inequalities (11) and by using the fact that $(v^k)_{k\geq 1}$ is nonincreasing (see (5.38) in [21, Lemma 5.1]). We notice that for the above arguments the fact that

$$\sum_{k\geq 1}\|z^{k+1}-z^k\|^2 < +\infty \tag{12}$$

is essential. In order to conclude that $\sum_{k\geq 1}\|z^{k+1}-z^k\|_{M_2}^2 < +\infty$ implies (12), one cannot avoid imposing that M_2 is positive definite.

Assume now that M_1 is positive definite (and M_2 remains positive semidefinite). In this case, by using again the inequality in Lemma 5, it follows that

$$\sum_{k\geq 1}\|x^{k+1}-x^k\|^2 < +\infty$$

and

$$\sum_{k\geq 1}\|y^{k+1}-y^k\|^2<+\infty,$$

which combined with (10) deliver (12). The $\mathcal{O}(1/\sqrt{k})$ rate of convergence for $(\|Ax^k-z^k\|)_{k\geq 0}=(c^{-1}\|y^k-y^{k-1}\|)_{k\geq 1}$ is obtained again by summing up the inequalities (11) and by using the fact that $(v^k)_{k\geq 1}$ is nonincreasing.

We come now to [21, Theorem 5.6], which addresses the convergence of the iterates generated by Algorithm 2 to a saddle point of the Lagrangian l. The result, which also assumes that a saddle point (x^*,z^*,y^*) of the Lagrangian l is given, has two parts. The first one considers the case when M_1 positive definite, a context in which the proof can be fixed. We will give the corresponding details in the next section, where we will address a more general setting.

The second case in [21, Theorem 5.6] concerns the situation when $M_1=0$ and A has full column rank. According to (11), $\left(\|z^k-z^*\|^2_{M_2+cI_m}+c^{-1}\|y^k-y^*\|^2+\|z^k-z^{k-1}\|^2_{M_2}\right)_{k\geq 1}$ is a nonincreasing sequence, which implies that both sequences $(z^k)_{k\geq 1}$ and $(y^k)_{k\geq 1}$ are bounded. Furthermore, $\lim_{k\rightarrow+\infty}c^{-1}\|y^k-y^{k-1}\|=\lim_{k\rightarrow+\infty}\|Ax^k-z^k\|=0$ and $\lim_{k\rightarrow+\infty}\|z^k-z^{k-1}\|_{M_2}=0$, while one does not obtain that $\lim_{k\rightarrow+\infty}\|z^k-z^{k-1}\|_{M_2+cI_m}=0$, as written in [21]. This means that one cannot deduce from here that $\lim_{k\rightarrow+\infty}\|z^k-z^{k-1}\|=0$. This fact has a decisive impact on the proof of the fact that every limit point of the sequence generated by Algorithm 2 is a saddle point of the Lagrangian l. The main argument for showing this relies on the key inequality in [21, Lemma 4.2], which is applied to a subsequence converging to such a limit point. Without knowing that $\lim_{k\rightarrow+\infty}\|z^k-z^{k-1}\|=0$, it is not clear if the middle summand in the right-hand side of the key inequality on such a convergent subsequence will converge to 0, a fact which basically questions the correctness of the statement in [21, Theorem 5.6] when $M_1=$ and A has full column rank.

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 [21] 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, $f:\mathcal{H}\rightarrow\overline{\mathbb{R}}$, $g:\mathcal{G}\rightarrow\overline{\mathbb{R}}$ be proper, convex and lower semicontinuous functions, $h:\mathcal{H}\rightarrow\mathbb{R}$ a convex and Fréchet differentiable function with L-Lipschitz continuous gradient (where $L>0$) and $A:\mathcal{H}\rightarrow\mathcal{G}$ a linear continuous operator. The Lagrangian associated with the convex optimization problem

$$\inf_{x\in\mathcal{H}}\{f(x)+h(x)+g(Ax)\} \tag{13}$$

is

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

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

$$l(x^*, z^*, y) \leq l(x^*, z^*, y^*) \leq l(x, z, y^*) \ \forall (x, z, y) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}. \tag{14}$$

Notice that (x^*, z^*, y^*) is a saddle point if and only if $z^* = Ax^*$, x^* is an optimal solution of (13), y^* is an optimal solution of the Fenchel dual problem to (13)

$$(D') \quad \sup_{y \in \mathcal{G}} \{-(f^* \,\Box\, h^*)(-A^*y) - g^*(y)\}, \tag{15}$$

and the optimal objective values of (13) and (15) 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 $f^* \,\Box\, h^* : \mathcal{H} \to \overline{\mathbb{R}}$ is defined by $(f^* \,\Box\, h^*)(x) = \inf_{u \in \mathcal{H}} \{f^*(u) + h^*(x - u)\}$ 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 (13) has an optimal solution and the Attouch-Brézis qualification condition:

$$0 \in \operatorname{sqri}(\operatorname{dom} g - A(\operatorname{dom} f)) \tag{16}$$

holds. Here, for a convex set $S \subseteq \mathcal{G}$, 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. Notice that the classical 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 $\operatorname{sqri} S = \operatorname{ri} S$. Considering again the infinite dimensional setting, we remark that condition (16) 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 (13)-(15) read:

$$-A^*y - \nabla h(x) \in \partial f(x) \text{ and } y \in \partial g(Ax). \tag{17}$$

This means that if (13) has an optimal solution $x \in \mathcal{H}$ and the qualification condition (16) is fulfilled, then there exists $y \in \mathcal{G}$, an optimal solution of (15), such that (17) 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 (17), then x is an optimal solution to (13), y is an optimal solution to (13) and (x, Ax, y) is a saddle point of the Lagrangian l. For further considerations on convex duality we invite the reader to consult [2–4, 16, 23].

Furthermore, we discuss some conditions ensuring that (13) has an optimal solution. Suppose that (13) is feasible, which means that its optimal objective value is not $+\infty$. The existence of optimal solutions to (13) 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 (13) is coercive and the statement follows via [2, Corollary 11.15]. On the other hand, when $f + h$ is strongly convex, then the objective function of (13) is strongly convex, too, thus (13) has a unique optimal solution (see [2, 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

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

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

$$x^{k+1} \in \underset{x\in\mathcal{H}}{\operatorname{argmin}} \left\{ f(x) + \langle x - x^k, \nabla h(x^k)\rangle + \frac{c}{2}\|Ax - z^k + c^{-1}y^k\|^2 + \frac{1}{2}\|x - x^k\|_{M_1^k}^2 \right\} \quad (18)$$

$$z^{k+1} = \underset{z\in\mathcal{G}}{\operatorname{argmin}} \left\{ g(z) + \frac{c}{2}\|Ax^{k+1} - z + c^{-1}y^k\|^2 + \frac{1}{2}\|z - z^k\|_{M_2^k}^2 \right\} \quad (19)$$

$$y^{k+1} = y^k + c(Ax^{k+1} - z^{k+1}). \quad (20)$$

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 [21].

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

An alternative is to assume that

$$\exists \alpha > 0 \text{ such that } A^*A \in \mathcal{P}_\alpha(\mathcal{H}). \quad (21)$$

This condition guarantees that the objective function in (18) is strongly convex, hence $(x^k)_{k\geq 0}$ is well-defined (see [2, Corollary 11.16]). Notice that if A is injective and $\operatorname{ran} A^*$ is closed, then (21) holds (see [2, Fact 2.19]). Moreover, (21) implies that A is injective. This means that if $\operatorname{ran} A^*$ is closed, then (21) is equivalent to A being injective. Hence, in finite dimensional spaces, namely, if $\mathcal{H} = \mathbb{R}^n$ and $\mathcal{G} = \mathbb{R}^m$, with $m \geq n \geq 1$, (21) is nothing else than saying that A has full column rank.

Remark 9 Let us now show that the particular choices $M_1^k = \frac{1}{\tau_k}\operatorname{Id} - cA^*A$, for $\tau_k > 0$, and $M_2^k = 0$ for all $k \geq 0$ lead to a primal-dual algorithm introduced in [12]. Here $\operatorname{Id} : \mathcal{H} \to \mathcal{H}$ denotes the identity operator on $\mathcal{H}$. Let $k \geq 0$ be fixed. The optimality condition for (18) 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 (20) we have

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

hence

$$0 \in \partial f(x^{k+2}) + (cA^*A + M_1^{k+1})(x^{k+2} - x^{k+1}) + A^*(2y^{k+1} - y^k) + \nabla h(x^{k+1}). \quad (22)$$

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,

$$\begin{aligned} x^{k+2} &= (\mathrm{Id} + \tau_{k+1}\partial f)^{-1}\left(x^{k+1} - \tau_{k+1}\nabla h(x^{k+1}) - \tau_{k+1}A^*(2y^{k+1} - y^k)\right) \\ &= \underset{x\in\mathcal{H}}{\operatorname{argmin}}\left\{ f(x) + \frac{1}{2\tau_{k+1}}\left\| x - \left(x^{k+1} - \tau_{k+1}\nabla h(x^{k+1}) - \tau_{k+1}A^*(2y^{k+1} - y^k)\right)\right\|^2 \right\}. \end{aligned} \tag{23}$$

Furthermore, from the optimality condition for (19) we obtain

$$c(Ax^{k+1} - z^{k+1} + c^{-1}y^k) + M_2^k(z^k - z^{k+1}) \in \partial g(z^{k+1}), \tag{24}$$

which combined with (20) gives

$$y^{k+1} + M_2^k(z^k - z^{k+1}) \in \partial g(z^{k+1}). \tag{25}$$

Using that $M_2^k = 0$ and again (20), 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

$$\begin{aligned} y^{k+1} &= (\mathrm{Id} + c\partial g^*)^{-1}\left(y^k + cAx^{k+1}\right) \\ &= \underset{z\in\mathcal{G}}{\operatorname{argmin}}\left\{ g^*(z) + \frac{1}{2c}\left\| z - \left(y^k + cAx^{k+1}\right)\right\|^2 \right\}. \end{aligned} \tag{26}$$

The iterative scheme obtained in (26) and (23) 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:

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

For $\tau_k = \tau > 0$ for all $k \geq 1$ one recovers a primal-dual algorithm from [12] 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 [12]). We invite the reader to consult [6, 7, 10, 22] for more insights into primal-dual algorithms and their highlights. Primal-dual algorithms with dynamic step sizes have been investigated in [10] and [7], where it has been shown that clever strategies in the choice of the step sizes can improve the convergence behaviour.

3.1 Ergodic convergence rates for the primal-dual gap

Our first convergence result related to Algorithm 7 provides a convergence rate for a primal-dual gap function formulated in terms of the associated Lagrangian l. We start by proving a technical result (see also [21]).

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 (18) we obtain

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

From the definition of the convex subdifferential we derive

$$\begin{aligned} f(x^{k+1}) - f(x) \ &\leq\ \langle \nabla h(x^k) + cA^*(-z^k + c^{-1}y^k + Ax^{k+1}) + M_1^k(-x^k + x^{k+1}), x - x^{k+1}\rangle \\ &=\ \langle \nabla h(x^k), x - x^{k+1}\rangle + \langle y^{k+1}, A(x - x^{k+1})\rangle - c\langle z^k - z^{k+1}, A(x - x^{k+1})\rangle \\ &\quad + \langle M_1^k(x^{k+1} - x^k), x - x^{k+1}\rangle, \end{aligned} \tag{28}$$

where for the last equality we used (20).

Furthermore, we claim that

$$h(x^{k+1}) - h(x) \leq -\langle \nabla h(x^k), x - x^{k+1}\rangle + \frac{L}{2}\|x^{k+1} - x^k\|^2. \tag{29}$$

Indeed, this follows by applying the convexity of h and the Descent Lemma (see [2, 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 (28) and (29) we obtain

$$(f + h)(x^{k+1}) \leq (f + h)(x) + \langle y^{k+1}, A(x - x^{k+1})\rangle - c\langle z^k - z^{k+1}, A(x - x^{k+1})\rangle \tag{30}$$

$$+ \frac{1}{2}\|x - x^k\|^2_{M_1^k} - \frac{1}{2}\|x - x^{k+1}\|^2_{M_1^k} - \frac{1}{2}\|x^{k+1} - x^k\|^2_{M_1^k} + \frac{L}{2}\|x^{k+1} - x^k\|^2. \tag{31}$$

From the optimality condition written for (19) we obtain

$$c(Ax^{k+1} - z^{k+1} + c^{-1}y^k) + M_2^k(z^k - z^{k+1}) \in \partial g(z^{k+1}), \tag{32}$$

which combined with (20) gives

$$y^{k+1} + M_2^k(z^k - z^{k+1}) \in \partial g(z^{k+1}). \tag{33}$$

From here we derive the inequality

$$\begin{aligned} g(z^{k+1}) - g(z) &\leq \langle -y^{k+1} + M_2^k(z^{k+1} - z^k), z - z^{k+1}\rangle \\ &= -\langle y^{k+1}, z - z^{k+1}\rangle + \frac{1}{2}\|z - z^k\|^2_{M_2^k} - \frac{1}{2}\|z - z^{k+1}\|^2_{M_2^k} - \frac{1}{2}\|z^{k+1} - z^k\|^2_{M_2^k}. \end{aligned} \tag{34}$$

The first statement of the lemma follows by combining the inequalities (30), (31), (34) with the identity (see (20)):

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

The second statement follows easily from the arithmetic-geometric mean inequality in Hilbert spaces (see [21, Proposition 5.3(a)]). ■

A direct consequence of the two inequalities from the above lemma 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:*

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

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:*

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

Then for all $k \geq 1$ *and all* $(x, z, y) \in \mathcal{H} \times \mathcal{G} \times \mathcal{G}$ *the following inequality 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 related to 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, we derive the inequality

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

where $f(x^*) + h(x^*) + g(Ax^*)$ is the optimal objective value of the problem (13). Hence, if we suppose that the set of optimal solutions of the dual problem (15) is contained in a bounded set, we obtain that there exists $R > 0$ such that

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

The set of dual optimal solutions of (15) is equal to the convex subdifferential of the infimal value function of the problem (13)

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

at 0. This set is weakly compact, thus bounded, if $0 \in \operatorname{int}(\operatorname{dom} \psi) = \operatorname{int}(A(\operatorname{dom} g) - \operatorname{dom} f)$ (see [2, 4, 23]).

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. One of the important tools for the proof of the convergence result will be the following version of the Opial Lemma formulated in the context of variable metrics (see [11, Theorem 3.3]).

Lemma 14 *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.

The proof of the first 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 [21] and from the other conventional convergence proofs for ADMM methods.

Theorem 15 *In the context of Problem 6, assume that the set of saddle points of the Lagrangian l is nonempty and that $M_1^k - \frac{L}{2}\operatorname{Id} \in \mathcal{S}_+(\mathcal{H}), M_1^k \succcurlyeq M_1^{k+1}$, $M_2^k \in \mathcal{S}_+(\mathcal{G}), M_2^k \succcurlyeq M_2^{k+1}$ for all $k \geq 0$, 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 $M_1^k - \frac{L}{2}\operatorname{Id} \in \mathcal{P}_{\alpha_1}(\mathcal{H})$ for all $k \geq 0$;

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

is fulfilled, then $(x^k, z^k, y^k)_{k\geq 0}$ converges weakly to a saddle point of the Lagrangian l.

Proof. Let (x^*, z^*, y^*) be a fixed saddle point of the Lagrangian l. 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 (27), (32) and the monotonicity of ∂f and ∂g, we obtain the inequalities

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

and

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

By the Baillon-Haddad Theorem (see [2, 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

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

Notice that, by taking into account (20), 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 through the norm, we derive the following inequality:

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

By using again relation (20) for expressing $Ax^{k+1} - z^{k+1}$ and by taking into account that

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

we obtain

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

and from here, by using the monotonicity assumptions on $(M_1^k)_{k\ge0}$ and $(M_2^k)_{k\ge0}$, it yields

$$\begin{aligned} \frac{1}{2}\|x^{k+1} - x^*\|^2_{M_1^{k+1}} + \frac{1}{2}\|z^{k+1} - Ax^*\|^2_{M_2^{k+1} + c\,\mathrm{Id}} + \frac{1}{2c}\|y^{k+1} - y^*\|^2 \le \\ \frac{1}{2}\|x^k - x^*\|^2_{M_1^k} + \frac{1}{2}\|z^k - Ax^*\|^2_{M_2^k + c\,\mathrm{Id}} + \frac{1}{2c}\|y^k - y^*\|^2 \\ -\frac{c}{2}\|z^k - Ax^{k+1}\|^2 - \frac{1}{2}\|x^k - x^{k+1}\|^2_{M_1^k - \frac{L}{2}\mathrm{Id}} - \frac{1}{2}\|z^k - z^{k+1}\|^2_{M_2^k} \\ -L \left\| L^{-1} \left(\nabla h(x^*) - \nabla h(x^k) \right) + \frac{1}{2} \left(x^k - x^{k+1} \right) \right\|^2. \end{aligned} \tag{35}$$

By neglecting the negative terms (notice that $M_1^k - \frac{L}{2}\mathrm{Id} \in \mathcal{S}_+(\mathcal{H})$ for all $k \ge 0$) from the above inequality it follows that the first assumption in the Opial Lemma (Lemma 14) holds, when applied in the product space $\mathcal{H} \times \mathcal{G} \times \mathcal{G}$, for the sequence $(x^k, z^k, y^k)_{k\ge0}$, for $W^k := (M_1^k, M_2^k + c\,\mathrm{Id}, c^{-1}\,\mathrm{Id})$ for $k \ge 0$, and for $S \subseteq \mathcal{H} \times \mathcal{G} \times \mathcal{G}$ the set of saddle points of the Lagrangian l.

Furthermore, by using arguments invoking telescoping sums, (35) yields

$$\sum_{k\ge0} \|z^k - Ax^{k+1}\|^2 < +\infty,\ \sum_{k\ge0} \|x^k - x^{k+1}\|^2_{M_1^k - \frac{L}{2}\mathrm{Id}} < +\infty,\ \sum_{k\ge0} \|z^k - z^{k+1}\|^2_{M_2^k} < +\infty. \tag{36}$$

Assume that the condition (I) holds. Since $M_1^k - \frac{L}{2}\mathrm{Id} \in \mathcal{P}_{\alpha_1}(\mathcal{H})$ for all $k \ge 0$ with $\alpha_1 > 0$, we get

$$x^k - x^{k+1} \to 0 \ (k \to +\infty) \tag{37}$$

and

$$z^k - Ax^{k+1} \to 0 \ (k \to +\infty). \tag{38}$$

A direct consequence of (37) and (38) is

$$z^k - z^{k+1} \to 0 \ (k \to +\infty). \tag{39}$$

From (20), (38) and (39) we derive

$$y^k - y^{k+1} \to 0 \ (k \to +\infty). \tag{40}$$

We show now that the relations (37)-(40) are fulfilled also under assumption (II). Indeed, in this situation we derive from (36) that (38) and (39) hold. From (20), (38) and (39) we obtain (40). Finally, the inequalities

$$\alpha\|x^{k+1} - x^k\|^2 \leq \|Ax^{k+1} - Ax^k\|^2 \leq 2\|Ax^{k+1} - z^k\|^2 + 2\|z^k - Ax^k\|^2 \ \forall k \geq 0 \tag{41}$$

yield (37).

The relations (37)-(40) will play an essential role when verifying the second assumption in the Opial Lemma for S taken as the set of saddle points of the Lagrangian l. 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 (37) and the linearity of A we obtain that $(Ax^{k_n+1})_{n\in\mathbb{N}}$ converges weakly to $A\overline{x}$ (as $n \to +\infty$), which combined with (38) yields $\overline{z} = A\overline{x}$. We use now the following notations for $n \geq 0$:

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

From (27) and (33) we have for all $n \geq 0$

$$a_n^* \in \partial(f + h)(a_n) \tag{42}$$

and

$$b_n^* \in \partial g(b_n). \tag{43}$$

Furthermore, from (37) we have

$$a_n \text{ converges weakly to } \overline{x} \text{ (as } n \to +\infty). \tag{44}$$

From (40) and (39) we obtain

$$b_n^* \text{ converges weakly to } \overline{y} \text{ (as } n \to +\infty). \tag{45}$$

Moreover, (20) and (40) yield

$$Aa_n - b_n \text{ converges strongly to } 0 \text{ (as } n \to +\infty). \tag{46}$$

Finally, we have

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

By using the fact that ∇h is Lipschitz continuous, from (37)-(40) we get

$$a_n^* + A^* b_n^* \text{ converges strongly to } 0 \text{ (as } n \to +\infty). \tag{47}$$

Taking into account the relations (42)-(47) 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 Opial Lemma is verified, too. In conclusion, $(x^k, z^k, y^k)_{k\geq 0}$ converges weakly to a saddle point of the Lagrangian l. ■

Remark 16 Choosing as in Remark 9, $M_1^k = \frac{1}{\tau_k}\operatorname{Id} - cA^*A$, 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

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

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 [12]), the operators $M_1^k - \frac{L}{2}\operatorname{Id}$ belong for all $k \geq 0$ to the class $\mathcal{P}_{\alpha_1}(\mathcal{H})$, with $\alpha_1 := \frac{1}{\tau} - c\|A\|^2 - \frac{L}{2} > 0$.

In the second convergence result of this subsection we consider the case when h is identically 0. We provide an extension of the correct part of [21, Theorem 5.6] and an adjustment of the part which we doubt to be correct.

Theorem 17 *In the context of Problem 6, assume that $h = 0$, that the set of saddle points of the Lagrangian l is nonempty and that $M_1^k \in \mathcal{S}_+(\mathcal{H}), M_1^k \succcurlyeq M_1^{k+1}$, $M_2^k \in \mathcal{S}_+(\mathcal{G}), M_2^k \succcurlyeq M_2^{k+1}$ for all $k \geq 0$, 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 $M_1^k \in \mathcal{P}_{\alpha_1}(\mathcal{H})$ for all $k \geq 0$;

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

is fulfilled, then $(x^k, z^k, y^k)_{k\geq 0}$ converges weakly to a saddle point of the Lagrangian l.

Proof. Let (x^*, z^*, y^*) be a saddle point of the Lagrangian l and $k \geq 0$ be fixed. As in the proof of Theorem 15, we can derive the inequality

$$\begin{aligned}\frac{1}{2}\|x^{k+1} - x^*\|^2_{M_1^{k+1}} + \frac{1}{2}\|z^{k+1} - Ax^*\|^2_{M_2^{k+1}+c\operatorname{Id}} + \frac{1}{2c}\|y^{k+1} - y^*\|^2 \leq \\ \frac{1}{2}\|x^k - x^*\|^2_{M_1^k} + \frac{1}{2}\|z^k - Ax^*\|^2_{M_2^k+c\operatorname{Id}} + \frac{1}{2c}\|y^k - y^*\|^2 \\ -\frac{c}{2}\|z^k - Ax^{k+1}\|^2 - \frac{1}{2}\|x^k - x^{k+1}\|^2_{M_1^k} - \frac{1}{2}\|z^k - z^{k+1}\|^2_{M_2^k}.\end{aligned} \tag{48}$$

Under assumption (I) the conclusion follows as in the proof of Theorem 15 by applying the Opial Lemma.

However, under assumption (II), as the norms quantifying the difference $(x^k - x^*)_{k\geq 0}$ are not available anymore, we need other arguments.

In this case (48) becomes

$$\frac{1}{2}\|z^{k+1} - Ax^*\|^2_{M_2^{k+1}+c\,\mathrm{Id}} + \frac{1}{2c}\|y^{k+1} - y^*\|^2 \leq$$
$$\frac{1}{2}\|z^k - Ax^*\|^2_{M_2^k+c\,\mathrm{Id}} + \frac{1}{2c}\|y^k - y^*\|^2 - \frac{c}{2}\|z^k - Ax^{k+1}\|^2 - \frac{1}{2}\|z^k - z^{k+1}\|^2_{M_2^k}. \tag{49}$$

From here it follows that (38) and (39) hold. From (20), (38) and (39) we obtain (40). Finally, by using again the inequality (41), relation (37) holds, too.

On the other hand, (49) yields that

$$\exists \lim_{k\to+\infty}\left(\frac{1}{2}\|z^k - z^*\|^2_{M_2^k+c\,\mathrm{Id}} + \frac{1}{2c}\|y^k - y^*\|^2\right), \tag{50}$$

hence $(y^k)_{k\geq 0}$ and $(z^k)_{k\geq 0}$ are bounded. Combining this with the condition imposed on 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 second part of the proof of Theorem 15, one can see that every weak sequential cluster point of $(x^k, z^k, y^k)_{k\geq 0}$ is a saddle point of the Lagrangian l.

In the remainder of the proof 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 exists $(k_p)_{p\geq 0}, (k_q)_{q\geq 0}$, $k_p \to +\infty$ (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) are saddle points of the Lagrangian l and $z_i = Ax_i$, $i \in \{1, 2\}$. From (50),which is true for every saddle point of the Lagrangian l, we derive

$$\exists \lim_{k\to+\infty}\left(\frac{1}{2}\|z^k - z_1\|^2_{M_2^k+c\,\mathrm{Id}} - \frac{1}{2}\|z^k - z_2\|^2_{M_2^k+c\,\mathrm{Id}} + \frac{1}{2c}\|y^k - y_1\|^2 - \frac{1}{2c}\|y^k - y_2\|^2\right). \tag{51}$$

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

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

and

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

Now, 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 [11], 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 pointwise to M_2 (as $k \to +\infty$).

Taking the limit in (51) along the subsequences $(k_p)_{p\geq 0}$ and $(k_q)_{q\geq 0}$ and using the last two relations above we obtain

$$\frac{1}{2}\|z_1 - z_2\|^2_{M_2} + \langle z_1 - z_2, M_2(z_2 - z_1)\rangle + \frac{1}{2}\|y_1 - y_2\|^2 + \langle y_1 - y_2, y_2 - y_1\rangle$$
$$= \frac{1}{2}\|z_1 - z_2\|^2_{M_2} + \frac{1}{2}\|y_1 - y_2\|^2,$$

hence

$$-\|z_1 - z_2\|^2_{M_2} - \|y_1 - y_2\|^2 = 0,$$

thus $z_1 = z_2$ and $y_1 = y_2$. The relations $z_i = Ax_i$, $i \in \{1, 2\}$, imply that $x_1 = x_2$ (due to (II)). In conclusion, $(x^k, z^k, y^k)_{k\geq 0}$ converges weakly to a saddle point of the Lagrangian l ■

Remark 18 By taking in each iteration constant operators $M_1^k = M_1$ and $M_2^k = M_2$ for all $k \geq 0$, Theorem 17 covers the first situation investigated in [21, Theorem 5.6], where in finite dimensional spaces the matrix M_1 was assumed to be positive definite and M_2 to be positive semidefinite.

As pointed out in Section 2, the arguments used in [21, Theorem 5.6] for the second situation, namely when $M_1 = 0$ and A has full column rank, do not seem not to be valid. As the proof of Theorem 17 shows, this result becomes valid under the supplementary assumption that M_2 is positive definite.

However, it is known from the literature that the sequence generated by the ADMM method converges also when $M_1 = M_2 = 0$ and A has full column rank (see [15]).

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.H. Bauschke, P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York, 2011

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

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

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

[6] 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

[7] 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

[8] 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

[9] 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

[10] 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

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

[12] 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

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

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

[15] 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

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

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

[18] 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

[19] 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

[20] 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

[21] 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

[22] B.C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Advances in Computational Mathematics 38(3), 667–681, 2013

[23] C. Zălinescu, *Convex Analysis in General Vector Spaces*, World Scientific, Singapore, 2002