The proximal alternating direction method of multipliers in the nonconvex setting: convergence analysis and rates

Radu Ioan Boţ[*][†]   Dang-Khoa Nguyen [‡]

March 22, 2019

Abstract. We propose two numerical algorithms in the fully nonconvex setting for the minimization of the sum of a smooth function and the composition of a nonsmooth function with a linear operator. The iterative schemes are formulated in the spirit of the proximal alternating direction method of multipliers and its linearized variant, respectively. The proximal terms are introduced via variable metrics, a fact which allows us to derive new proximal splitting algorithms for nonconvex structured optimization problems, as particular instances of the general schemes. Under mild conditions on the sequence of variable metrics and by assuming that a regularization of the associated augmented Lagrangian has the Kurdyka-Łojasiewicz property, we prove that the iterates converge to a KKT point of the objective function. By assuming that the augmented Lagrangian has the Łojasiewicz property, we also derive convergence rates for both the augmented Lagrangian and the iterates.

Keywords. nonconvex complexly structured optimization problems, alternating direction method of multipliers, proximal splitting algorithms, variable metric, convergence analysis, convergence rates, Kurdyka-Łojasiewicz property, Łojasiewicz exponent

AMS subject classification. 47H05, 65K05, 90C26

1 Introduction

1.1 Problem formulation and motivation

In this paper we are interested in solving optimization problems of the form

$$\min_{x \in \mathbb{R}^n} \left\{ g\left(Ax\right) + h\left(x\right) \right\}, \tag{1}$$

where $g \colon \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is a proper and lower semicontinuous function, $h \colon \mathbb{R}^n \to \mathbb{R}$ is a Fréchet differentiable function with L-Lipschitz continuous gradient and $A \colon \mathbb{R}^n \to \mathbb{R}^m$ is a linear operator. The spaces $\mathbb{R}^n$ and $\mathbb{R}^m$ are equipped with Euclidean inner products $\langle \cdot, \cdot \rangle$ and associated norms $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$, which are both denoted in the same way, as there is no risk of confusion.

We start by briefly describing the Alternating Direction Method of Multipliers (ADMM) designed to solve optimization problems of the form

$$\min_{x \in \mathbb{R}^n} \left\{ f\left(x\right) + g\left(Ax\right) + h\left(x\right) \right\}, \tag{2}$$

where g and h are assumed to be also *convex* and $f \colon \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is another proper, convex and lower semicontinuous function. By introducing an auxiliary variable, one can rewrite problem (2) as

$$\min_{\substack{(x,z) \in \mathbb{R}^n \times \mathbb{R}^m \\ Ax - z = \mathbf{0}}} \left\{ f\left(x\right) + g\left(z\right) + h\left(x\right) \right\}. \tag{3}$$

For a fixed real number $r > 0$, the *augmented Lagrangian* associated with problem (3) reads

$$\mathcal{L}_r : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}, \ \mathcal{L}_r(x, z, y) = f(x) + g(z) + h(x) + \langle y, Ax - z \rangle + \frac{r}{2} \|Ax - z\|^2.$$

Given a starting vector $\left(x^0, z^0, y^0\right) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ and $\{\mathbf{M}_1^k\}_{k \geqslant 0} \subseteq \mathbb{R}^{n \times n}, \{\mathbf{M}_2^k\}_{k \geqslant 0} \subseteq \mathbb{R}^{m \times m}$, two sequences of symmetric and positive semidefinite matrices, the following *proximal ADMM algorithm formulated in the presence of a smooth function and involving variable metrics* has been proposed and investigated in [5]: generate the sequence $\{(x^k, z^k, y^k)\}_{k \geqslant 0}$ for every $k \geqslant 0$ as

$$x^{k+1} \in \arg\min_{x \in \mathbb{R}^n} \left\{ f(x) + \langle x - x^k, \nabla h(x^k) \rangle + \frac{r}{2} \left\| Ax - z^k + \frac{1}{r} y^k \right\|^2 + \frac{1}{2} \|x - x^k\|_{\mathbf{M}_1^k}^2 \right\}, \qquad (4a)$$

$$z^{k+1} = \arg\min_{z \in \mathbb{R}^m} \left\{ g(z) + \frac{r}{2} \left\| Ax^{k+1} - z + \frac{1}{r} y^k \right\|^2 + \frac{1}{2} \|z - z^k\|_{\mathbf{M}_2^k}^2 \right\}, \qquad (4b)$$

$$y^{k+1} = y^k + \rho r \left(Ax^{k+1} - z^{k+1} \right). \qquad (4c)$$

It has been proved in [5] that, if $\rho = 1$ and the set of the saddle points of the Lagrangian associated with (3) (which is nothing else than $\mathcal{L}_r$ when $r = 0$) is nonempty, and the two matrix sequences and the operator A fulfill mild additional assumptions, then the sequence $\{(x^k, z^k, y^k)\}_{k \geqslant 0}$ converges to a saddle point of the Lagrangian associated with problem (3) and provides in this way both an optimal solution of (1) and an optimal solution of its Fenchel dual problem. Furthermore, an ergodic primal-dual gap convergence rate result has been proved.

In case $h = 0$, the above iterative scheme encompasses as special cases different numerical algorithms considered in the literature. If $\mathbf{M}_1^k = \mathbf{M}_2^k = 0$ for all $k \geqslant 0$, then (4a)-(4c) becomes the *classical ADMM algorithm* ([17, 24, 26, 27]), which lately gained a huge popularity in the optimization community, despite its poor implementation properties caused by the fact that, in general, the calculation of the sequence of primal variables $\{x^k\}_{k \geqslant 0}$ does not correspond to a proximal step. For an *inertial version* of the classical ADMM algorithm we refer the reader to [12]. On the other hand, if $\mathbf{M}_1^k = \mathbf{M}_1$ and $\mathbf{M}_2^k = \mathbf{M}_2$ for all $k \geqslant 0$, then (4a)-(4c) recovers the *proximal ADMM algorithm* investigated by Shefi and Teboulle in [43] (see also [22, 23]). It has been pointed out in [43] that, for suitable choices of the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, the proximal ADMM algorithm becomes a primal-dual splitting algorithm in the sense of those considered in [15, 18, 21, 45], and which, due to its full splitting character, overcomes the drawbacks of the classical ADMM algorithm. Recently, in [14] it has been shown that, if f is strongly convex, then suitable choices of the non-constant sequences $\{\mathbf{M}_1^k\}_{k \geqslant 0}$ and $\{\mathbf{M}_2^k\}_{k \geqslant 0}$ lead to a rate of convergence of $\mathcal{O}(1/k)$ for the sequence of primal iterates.

In this paper, we propose a proximal ADMM (P-ADMM) algorithm and a proximal linearized ADMM (PL-ADMM) algorithm for solving the optimization problem (1) and carry out a convergence analysis for both algorithms. We first prove, under not very restrictive assumptions on the problem data, that the sequence of generated iterates $\{(x^k, z^k, y^k)\}_{k \geqslant 0}$ is bounded. Given these premises we show that the cluster points of $\{(x^k, z^k, y^k)\}_{k \geqslant 0}$ are *KKT points* of the problem (1). Provided that a regularization of the augmented Lagrangian satisfies the Kurdyka-Łojasiewicz property, we show global convergence of the generated sequence of iterates. Provided this regularization of the augmented Lagrangian has the Łojasiewicz property, we derive rates of convergence for the sequence of iterates. To the best of our knowledge, these are the first results in the literature that deal with convergence rates for the nonconvex ADMM.

In the following we will comment on previous works addressing the ADMM algorithm in the non-convex setting. None of the papers which have addressed nonconvex optimization problems involving compositions with linear operators propose and investigate iterative schemes designed in the spirit of full splitting algorithms. In [34], the convergence of the ADMM algorithm for solving the problem (1) is studied under the assumption that h is twice continuously differentiable with bounded Hessian. In [30], the ADMM algorithm is used to minimize the sum of finitely many smooth nonconvex functions and a nonsmooth convex function, by rewriting it as an general consensus problem. No linear operator occurs in the formulation of the optimization problem under investigation. In [1], the ADMM algorithm is used to solve a DC optimization problem over the unit ball which occurs in the penalized zero-variance linear discriminant analysis. In [46], a nonconvex ADMM algorithm involving proximal terms induced via Bregman distances is introduced and investigated, however, without addressing the question of the boundedness of the generated iterates. On the other hand, in [28], in order to guarantee boundedness of the iterates a strong assumption on g is made, which is proved to hold for the normed-squared function.

In [47], a lot of efforts are made to guarantee boundedness for the generated iterates of the nonconvex ADMM algorithm, which is an essential component of the convergence analysis, however, this is done by assuming that the objective function is continuous and *coercive over the feasible set*, while its non-smooth part is either *restricted prox-regular* or *piecewise linear*. Similar ingredients are used in [36] in the convergence analysis of a nonconvex linearized ADMM algorithm.

Recently, Bolte, Sabach and Teboulle have proposed in [11] a generic iterative scheme for solving a general optimization problem of the form (1), but by replacing the linear operator A with a general continuously differentiable operator. A global convergence analysis relying on the use of the Kurdyka-Łojasiewicz property is carried out. It is also shown that the generic iterative scheme encompasses several Lagrangian based algorithms, including the proximal alternating direction method of multipliers and the proximal alternating linearized minimization method. The latter is analysed into detail in the particular case when g is composed with a linear operator, which coincides with the one in this paper. The two algorithms we propose in this paper are formulated in the same spirit, however, they lead for some particular choices of the variable metrics to full splitting algorithms. In addition, we carefully address the issue of the boundedness of the sequence of generated iterates and complement the convergence analysis with the derivation of convergence rates.

The major strengths of our paper are:

1. We prove under quite general assumptions that the sequence $\{(x^k, z^k, y^k)\}_{k \geqslant 0}$ is bounded. In the nonconvex setting, the boundedness of the sequence of generated iterates plays a central role in the convergence analysis. In fact, the reason, why we assume in this paper that the function g is smooth, is exclusively given by the fact that only in this setting we can prove boundedness of this sequence under general assumptions.

2. We prove convergence for relaxed variants of the nonconvex ADMM algorithms, which allow to chose in the update of the dual sequence $\rho \in (0, 2)$. We notice that $\rho = 1$ is the standard choice in the literature ([1, 5, 14, 34, 43, 47]). Gabay and Mercier proved in [27] in the convex setting that ρ may be chosen in $(0, 2)$, however, the majority of the extensions of the convex relaxed ADMM algorithm assume that $\rho \in \left(0, \frac{1+\sqrt{5}}{2}\right)$ (see [22, 23, 26, 44, 48, 49]) or ask for a particular choice of ρ, which is interpreted as a step size (see [31]). In [49], an alternating minimization algorithm for the minimization of the sum of a simple nonsmooth function and a smooth function in the nonconvex setting, which allows for a parameter ρ different from 1, has been proposed.

3. By appropriate choices of the matrix sequences, we derive from the proposed iterative schemes full splitting algorithms for solving the nonconvex complexly structured optimization problem (1). More precisely, (P-ADMM) gives rise to an iterative scheme formulated only in terms of proximal steps for the functions g and h and of forward evaluations of the matrix A, while (PL-ADMM) gives rise to an iterative scheme in which the function h is performed via a gradient step. Exact formulas for proximal operators are available not only for large classes of convex functions ([6, 20]), but also of nonconvex functions ([3, 29, 33]). The fruitful idea to linearize the step involving the smooth term has been used in the past in the context of ADMM algorithms mostly in the convex setting (see [35, 40, 41, 48, 50]), but also in the nonconvex setting (see [11, 36]).

1.2 Notations and preliminaries

Let N be a strictly positive integer. We denote by $\mathbb{1} := (1, \ldots, 1) \in \mathbb{R}^N$ and write for $x := (x_1, \ldots, x_N)$, $y := (y_1, \ldots, y_N) \in \mathbb{R}^N$

$$x < y \quad \text{if and only if} \quad x_i < y_i \ \forall i = 1, \ldots, N.$$

We endow the Cartesian product $\mathbb{R}^{N_1} \times \mathbb{R}^{N_2} \times \ldots \times \mathbb{R}^{N_p}$, where p is a strictly positive integer, with *inner product* and associated *norm* defined for $u := (u_1, \ldots, u_p), u' := (u'_1, \ldots, u'_p) \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2} \times \ldots \times \mathbb{R}^{N_p}$ by

$$\langle\!\langle u, u' \rangle\!\rangle = \sum_{i=1}^{p} \langle u_i, u'_i \rangle \quad \text{and} \quad |\!|\!| u |\!|\!| = \sqrt{\sum_{i=1}^{p} \|u_i\|^2},$$

respectively. For every $u := (u_1, \ldots, u_p), u' := (u'_1, \ldots, u'_p) \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2} \times \ldots \times \mathbb{R}^{N_p}$ we have

$$\frac{1}{\sqrt{p}} \sum_{i=1}^{p} \|u_i\| \leqslant |\!|\!| u |\!|\!| = \sqrt{\sum_{i=1}^{p} \|u_i\|^2} \leqslant \sum_{i=1}^{p} \|u_i\|. \tag{5}$$

We denote by $\mathbb{S}_+^N$ the family of symmetric and positive semidefinite matrices $M \in \mathbb{R}^{N \times N}$. Every $M \in \mathbb{S}_+^N$ induces a *semi-norm* defined by

$$\|x\|_M^2 := \langle Mx, x \rangle \quad \forall x \in \mathbb{R}^N.$$

The *Loewner partial ordering* on $\mathbb{S}_+^N$ is defined for $M, M' \in \mathbb{S}_+^N$ as

$$M \succcurlyeq M' \Leftrightarrow \|x\|_M^2 \geqslant \|x\|_{M'}^2 \quad \forall x \in \mathbb{R}^N.$$

Thus $M \in \mathbb{S}_+^N$ is nothing else than $M \succcurlyeq \mathbf{0}$. For $\alpha > 0$ we set

$$\mathcal{P}_\alpha^N := \left\{ M \in \mathbb{S}_+^N : M \succcurlyeq \alpha \mathbf{Id} \right\},$$

where $\mathbf{Id}$ denotes the identity matrix in $\mathbb{R}^{N \times N}$. If $M \in \mathcal{P}_\alpha^N$, then the semi-norm $\|\cdot\|_M$ becomes a norm.

The linear operator A is *surjective* if and only if its associated matrix has full row rank. This assumption is further equivalent to the fact that the matrix associated to AA^*, where A^* denotes the *adjoint operator* of A, is positively definite. Since

$$\lambda_{\min}(AA^*)\|y\|^2 \leqslant \|y\|_{AA^*}^2 = \langle AA^*y, y \rangle = \|A^*y\|^2 \quad \forall y \in \mathbb{R}^m,$$

this is further equivalent to $\lambda_{\min}(AA^*) > 0$ (and $AA^* \in \mathcal{P}_{\lambda_{\min}(AA^*)}^n$), where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of a matrix. Similarly, A is injective if and only if $\lambda_{\min}(A^*A) > 0$ (and $A^*A \in \mathcal{P}_{\lambda_{\min}(A^*A)}^m$).

Proposition 1. *Let $\Psi \colon \mathbb{R}^N \to \mathbb{R}$ be Fréchet differentiable such that its gradient is Lipschitz continuous with constant $L > 0$. Then the following statements are true:*

1. *For every $x, y \in \mathbb{R}^N$ and every $z \in [x, y] = \{(1-t)x + ty : t \in [0,1]\}$ it holds*

$$\Psi(y) \leqslant \Psi(x) + \langle \nabla\Psi(z), y - x \rangle + \frac{L}{2}\|y - x\|^2; \tag{6}$$

2. *If Ψ is bounded from below, then for every $\sigma > 0$ it holds*

$$\inf_{x \in \mathbb{R}^N} \left\{ \Psi(x) - \left(\frac{1}{\sigma} - \frac{L}{2\sigma^2} \right) \|\nabla\Psi(x)\|^2 \right\} > -\infty.$$

Proof. 1. Let be $x, y \in \mathbb{R}^N$ and $z := (1-t)x + ty$ for $t \in [0,1]$. By the fundamental theorem for line integrals we have

$$\Psi(y) - \Psi(x) = \int_0^1 \langle \nabla\Psi((1-s)x + sy), y - x \rangle \, ds$$

$$= \int_0^1 \langle \nabla\Psi((1-s)x + sy) - \nabla\Psi(z), y - x \rangle \, ds + \langle \nabla\Psi(z), y - x \rangle. \tag{7}$$

Since

$$\left| \int_0^1 \langle \nabla\Psi((1-s)x + sy) - \nabla\Psi(z), y - x \rangle \, ds \right|$$

$$\leqslant \int_0^1 \|\nabla\Psi((1-s)x + sy) - \nabla\Psi(z)\| \cdot \|y - x\| \, ds \leqslant L \|x - y\|^2 \int_0^1 |s - t| \, ds$$

$$= L \|x - y\|^2 \left(\int_0^t (-s + t) \, ds + \int_t^1 (s - t) \, ds \right) = L \left(\frac{1}{2} - t(1-t) \right) \|x - y\|^2. \tag{8}$$

The inequality in (6) follows by combining (7) and (8) and by using that $0 \leqslant t \leqslant 1$.

2. The inequality in (6) gives for every $x \in \mathbb{R}^N$

$$-\infty < \inf_{y \in \mathbb{R}^N} \Psi(y) \leqslant \Psi\left(x - \frac{1}{\sigma}\nabla\Psi(x) \right)$$

$$\leqslant \Psi(x) + \left\langle \left(x - \frac{1}{\sigma}\nabla\Psi(x) \right) - x, \nabla\Psi(x) \right\rangle + \frac{L}{2}\left\| \left(x - \frac{1}{\sigma}\nabla\Psi(x) \right) - x \right\|^2$$

$$= \Psi(x) - \left(\frac{1}{\sigma} - \frac{L}{2\sigma^2} \right) \|\nabla\Psi(x)\|^2,$$

which leads to the desired conclusion. $\qquad\square$

Remark 1. The so-called Descent Lemma, which says that for a Fréchet differentiable function $\Psi \colon \mathbb{R}^N \to \mathbb{R}$ having Lipschitz continuous gradient with constant $L > 0$ it holds

$$\Psi\left(y\right) \leqslant \Psi\left(x\right) + \left\langle \nabla\Psi\left(x\right), y - x \right\rangle + \frac{L}{2}\left\|y - x\right\|^2 \quad \forall x, y \in \mathbb{R}^N,$$

follows from statement (i) of the above proposition for $z := x$.

Moreover, for $z := y$ we have that

$$\Psi\left(x\right) \geqslant \Psi\left(y\right) + \left\langle \nabla\Psi\left(y\right), x - y \right\rangle - \frac{L}{2}\left\|x - y\right\|^2 \quad \forall x, y \in \mathbb{R}^N,$$

which is equivalent to the fact that $\Psi + \dfrac{L}{2}\left\|\cdot\right\|^2$ is a convex function, in other words, Ψ is a L-semiconvex function ([9]). It follows from the previous result that a Fréchet differentiable function with L-Lipschitz continuous gradient is L-semiconvex.

The *limiting subdifferential* will play an important role in the convergence analysis we are going to carry out for the nonconvex ADMM algorithm. Let $\Psi \colon \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. For $x \in \text{dom}\,\Psi := \left\{x \in \mathbb{R}^N : \Psi\left(x\right) < +\infty\right\}$, the *Fréchet (viscosity) subdifferential* of Ψ at x is

$$\widehat{\partial}\Psi\left(x\right) := \left\{d \in \mathbb{R}^N : \liminf_{y \to x} \frac{\Psi\left(y\right) - \Psi\left(x\right) - \left\langle d, y - x \right\rangle}{\left\|y - x\right\|} \geqslant 0\right\}$$

and the *limiting (Mordukhovich) subdifferential* of Ψ at x is

$$\partial\Psi\left(x\right) := \{d \in \mathbb{R}^N : \text{exist sequences } x^k \to x \text{ and } d^k \to d \text{ as } k \to +\infty$$
$$\text{such that } \Psi\left(x^k\right) \to \Psi\left(x\right) \text{ as } k \to +\infty \text{ and } d^k \in \widehat{\partial}\Psi\left(x^k\right) \text{ for all } k \geqslant 0\}.$$

For $x \notin \text{dom}\left(\Psi\right)$, we set $\widehat{\partial}\Psi\left(x\right) = \partial\Psi\left(x\right) := \varnothing$.

The inclusion $\widehat{\partial}\Psi\left(x\right) \subseteq \Psi\left(x\right)$ holds for every $x \in \mathbb{R}^N$. If Ψ is convex, then these two subdifferentials coincide with the *convex subdifferential*, in other words

$$\widehat{\partial}\Psi\left(x\right) = \partial\Psi\left(x\right) = \left\{d \in \mathbb{R}^N : \Psi\left(y\right) \geqslant \Psi\left(x\right) + \left\langle d, y - x \right\rangle \; \forall y \in \mathbb{R}^N\right\} \text{ for all } x \in \text{dom}\,\Psi.$$

If $x \in \mathbb{R}^N$ is a local minimum of Ψ, then $0 \in \partial\Psi\left(x\right)$. We denote by $\text{crit}(\Psi) = \{x \in \mathbb{R}^N : 0 \in \partial\Psi\left(x\right)\}$ the set of *critical points* of Ψ. The limiting subdifferential fulfills the *closedness criterion*: if $\left\{x^k\right\}_{k \geqslant 0}$ and $\{d^k\}_{k \geqslant 0}$ are sequence in $\mathbb{R}^N$ such that $d^k \in \partial\Psi\left(x^k\right)$ for all $k \geqslant 0$, and $\left(x^k, d^k\right) \to (x, d)$ and $\Psi\left(x^k\right) \to \Psi\left(x\right)$ as $k \to +\infty$, then $d \in \partial\Psi\left(x\right)$. We have the following *subdifferential sum rule* holds ([38, Proposition 1.107], [42, Exercise 8.8]): if $\Phi \colon \mathbb{R}^N \to \mathbb{R}$ is a continuously differentiable function, then $\partial\left(\Psi + \Phi\right)\left(x\right) = \partial\Psi\left(x\right) + \nabla\Phi\left(x\right)$ for all $x \in \mathbb{R}^N$; and the following *subdifferential rule for the composition with a linear operator* $A \colon \mathbb{R}^{N'} \to \mathbb{R}^N$ ([38, Proposition 1.112], [42, Exercise 10.7]): if $x \in \text{dom}\,\Psi$ and A is injective, then $\partial\left(\Psi \circ A\right)\left(x\right) = A^*\partial\Psi\left(Ax\right)$.

We close this section by presenting two convergence results for real sequences that will be used in the sequel in the convergence analysis. The next lemma is often used in the literature when proving convergence of numerical algorithms relying on Fejér monotonicity techniques (see, for instance, [13, Lemma 2.2], [16, Lemma 2]).

Lemma 2. *Let* $\{b_k\}_{k \geqslant 0}$ *be a sequence in* $\mathbb{R}$ *and* $\{\xi_k\}_{k \geqslant 0}$ *a sequence in* $\mathbb{R}_+$. *Assume that* $\{b_k\}_{k \geqslant 0}$ *is bounded from below and that for every* $k \geqslant 0$

$$b_{k+1} + \xi_k \leqslant b_k.$$

Then the following statements hold:

1. the sequence $\{\xi_k\}_{k \geqslant 0}$ *is summable, namely* $\displaystyle\sum_{k \geqslant 0} \xi_k < +\infty$;

2. the sequence $\{b_k\}_{k \geqslant 0}$ *is monotonically decreasing and convergent.*

The following lemma, which is an extension of [13, Lemma 2.3] (see, also [16, Lemma 3]), is of interest by its own.

Lemma 3. *Let* $\left\{a^k := \left(a_1^k, a_2^k, \ldots, a_N^k\right)\right\}_{k \geqslant 0}$ *be a sequence in* $\mathbb{R}_+^N$ *and* $\{\delta_k\}_{k \geqslant 0}$ *a sequence in* $\mathbb{R}$ *such that*

$$\left\langle \mathbb{1}, a^{k+1} \right\rangle \leqslant \left\langle c_0, a^k \right\rangle + \left\langle c_1, a^{k-1} \right\rangle + \left\langle c_2, a^{k-2} \right\rangle + \delta_k \ \forall k \geqslant 2, \tag{9}$$

where $c_0 := (c_{0,1}, c_{0,2}, \ldots, c_{0,N}) \in \mathbb{R}^N$, $c_1 := (c_{1,1}, c_{1,2}, \ldots, c_{1,N}) \in \mathbb{R}_+^N$ *and* $c_2 := (c_{2,1}, c_{2,2}, \ldots, c_{2,N}) \in \mathbb{R}_+^N$ *fulfill* $c_0 + c_1 + c_2 < \mathbb{1}$. *Assume further that there exists* $\bar{\delta} \geqslant 0$ *such that for every* $\overline{K} \geqslant \underline{K} \geqslant 2$

$$\sum_{k=\underline{K}}^{\overline{K}} \delta_k \leqslant \bar{\delta}.$$

Then, for every $i = 1, \ldots, N$, *it holds*

$$\sum_{k \geqslant 0} a_i^k < +\infty.$$

In particular, for every $i = 1, \ldots, N$ *and every* $\overline{K} \geqslant \underline{K} \geqslant 2$, *it holds*

$$\sum_{k=\underline{K}}^{\overline{K}} a_i^k \leqslant \frac{\sum_{j=1}^N \left[(1 - c_{0,j} - c_{1,j}) \, a_j^{\underline{K}} + (1 - c_{0,j}) \, a_j^{\underline{K}+1} + a_j^{\underline{K}+2} \right] + \bar{\delta}}{1 - c_{0,i} - c_{1,i} - c_{2,i}}. \tag{10}$$

Proof. Fix $\overline{K} \geqslant \underline{K} \geqslant 2$. If $\overline{K} = \underline{K}$ or $\overline{K} = \underline{K} + 1$, then (10) holds automatically. Assume now that $\overline{K} \geqslant \underline{K} + 2$. Summing up the inequality in (9) for $k = \underline{K} + 2, \cdots, \overline{K}$, we obtain

$$\left\langle \mathbb{1}, \sum_{k=\underline{K}+2}^{\overline{K}} a^{k+1} \right\rangle \leqslant \left\langle c_0, \sum_{k=\underline{K}+2}^{\overline{K}} a^k \right\rangle + \left\langle c_1, \sum_{k=\underline{K}+2}^{\overline{K}} a^{k-1} \right\rangle + \left\langle c_2, \sum_{k=\underline{K}+2}^{\overline{K}} a^{k-2} \right\rangle + \sum_{k=\underline{K}+2}^{\overline{K}} \delta_k. \tag{11}$$

Since

$$\sum_{k=\underline{K}+2}^{\overline{K}} a^{k+1} = \sum_{k=\underline{K}+3}^{\overline{K}+1} a^k = \sum_{k=\underline{K}}^{\overline{K}} a^k + a^{\overline{K}+1} - a^{\underline{K}} - a^{\underline{K}+1} - a^{\underline{K}+2}$$

$$\sum_{k=\underline{K}+2}^{\overline{K}} a^k = \sum_{k=\underline{K}}^{\overline{K}} a^k - \left(a^{\underline{K}} + a^{\underline{K}+1} \right)$$

$$\sum_{k=\underline{K}+2}^{\overline{K}} a^{k-1} = \sum_{k=\underline{K}+1}^{\overline{K}-1} a^k = \sum_{k=\underline{K}}^{\overline{K}} a^k - \left(a^{\underline{K}} + a^{\overline{K}} \right)$$

$$\sum_{k=\underline{K}+2}^{\overline{K}} a^{k-2} = \sum_{k=\underline{K}}^{\overline{K}-2} a^k = \sum_{k=\underline{K}}^{\overline{K}} a^k - \left(a^{\overline{K}-1} + a^{\overline{K}} \right),$$

the inequality in (11) can be rewritten as

$$\left\langle \mathbb{1}, \sum_{k=\underline{K}}^{\overline{K}} a^k \right\rangle + \left\langle \mathbb{1}, a^{\overline{K}+1} - a^{\underline{K}} - a^{\underline{K}+1} - a^{\underline{K}+2} \right\rangle \leqslant \left\langle c_0, \sum_{k=\underline{K}}^{\overline{K}} a^k \right\rangle - \left\langle c_0, a^{\underline{K}} + a^{\underline{K}+1} \right\rangle$$

$$+ \left\langle c_1, \sum_{k=\underline{K}}^{\overline{K}} a^k \right\rangle - \left\langle c_1, a^{\underline{K}} + a^{\overline{K}} \right\rangle + \left\langle c_2, \sum_{k=\underline{K}}^{\overline{K}} a^k \right\rangle - \left\langle c_2, a^{\overline{K}-1} + a^{\overline{K}} \right\rangle + \sum_{k=\underline{K}+2}^{\overline{K}} \delta_k,$$

which further implies

$$\sum_{j=1}^N \left[(1 - c_{0,j} - c_{1,j} - c_{2,j}) \sum_{k=\underline{K}}^{\overline{K}} a_j^k \right] = \left\langle \mathbb{1} - c_0 - c_1 - c_2, \sum_{k=\underline{K}}^{\overline{K}} a^k \right\rangle$$

$$\leqslant \left\langle \mathbb{1} - c_0 - c_1, a^{\underline{K}} \right\rangle + \left\langle \mathbb{1} - c_0, a^{\underline{K}+1} \right\rangle + \left\langle \mathbb{1}, a^{\underline{K}+2} \right\rangle + \sum_{k=\underline{K}+2}^{\overline{K}} \delta_k$$

$$= \sum_{j=1}^N \left[(1 - c_{0,j} - c_{1,j}) \, a_j^{\underline{K}} + (1 - c_{0,j}) \, a_j^{\underline{K}+1} + a_j^{\underline{K}+2} \right] + \sum_{k=\underline{K}+2}^{\overline{K}} \delta_k.$$

Hence, for every $i = 1, \ldots, N$, it holds

$$(1 - c_{0,i} - c_{1,i} - c_{2,i}) \sum_{k=\underline{K}}^{\overline{K}} a_i^k \leqslant \sum_{j=1}^{N} \left[(1 - c_{0,j} - c_{1,j}) a_j^{\underline{K}} + (1 - c_{0,j}) a_j^{\underline{K}+1} + a_j^{\underline{K}+2} \right] + \overline{\delta}$$

and the conclusion follows by taking into consideration that $c_0 + c_1 + c_2 < \mathbb{1}$. $\qquad\square$

2 A proximal ADMM algorithm and a proximal linearized ADMM algorithm in the nonconvex setting

In this section we propose two proximal ADMM algorithms for solving the optimization problem (1) and study their convergence behaviour. A central role will be played by the augmented Lagrangian associated with problem (1), which is defined for every $r > 0$ as

$$\mathcal{L}_r \colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}, \ \mathcal{L}_r(x, z, y) = g(z) + h(x) + \langle y, Ax - z \rangle + \frac{r}{2} \|Ax - z\|^2.$$

2.1 General formulations and full proximal splitting algorithms as particular instances

Algorithm 1. *Let be the matrix sequences* $\{\mathbf{M}_1^k\}_{k \geqslant 0} \in \mathbb{S}_+^n$, $\{\mathbf{M}_2^k\}_{k \geqslant 0} \in \mathbb{S}_+^m$, $r > 0$ *and* $0 < \rho < 2$. *For a given starting vector* $(x^0, z^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$, *generate the sequence* $\{(x^k, z^k, y^k)\}_{k \geqslant 0}$ *for every* $k \geqslant 0$ *as:*

$$\begin{aligned}
z^{k+1} &\in \arg \min_{z \in \mathbb{R}^m} \left\{ \mathcal{L}_r(x^k, z, y^k) + \frac{1}{2} \|z - z^k\|_{\mathbf{M}_2^k}^2 \right\} \\
&= \arg \min_{z \in \mathbb{R}^m} \left\{ g(z) + \langle y^k, Ax^k - z \rangle + \frac{r}{2} \|Ax^k - z\|^2 + \frac{1}{2} \|z - z^k\|_{\mathbf{M}_2^k}^2 \right\},
\end{aligned} \tag{12a}$$

$$\begin{aligned}
x^{k+1} &\in \arg \min_{x \in \mathbb{R}^n} \left\{ \mathcal{L}_r(x, z^{k+1}, y^k) + \frac{1}{2} \|x - x^k\|_{\mathbf{M}_1^k}^2 \right\} \\
&= \arg \min_{x \in \mathbb{R}^n} \left\{ h(x) + \langle y^k, Ax - z^{k+1} \rangle + \frac{r}{2} \|Ax - z^{k+1}\|^2 + \frac{1}{2} \|x - x^k\|_{\mathbf{M}_1^k}^2 \right\},
\end{aligned} \tag{12b}$$

$$y^{k+1} := y^k + \rho r \left(Ax^{k+1} - z^{k+1} \right). \tag{12c}$$

Let $\{t_k\}_{k \geqslant 0}$ be a sequence of positive real numbers such that $t_k \geqslant r \|A\|^2$, and $\mathbf{M}_1^k := t_k \mathbf{Id} - rA^*A$ and $\mathbf{M}_2^k := \mathbf{0}$ for every $k \geqslant 0$. In this particular case Algorithm 1 becomes an iterative scheme which generates a sequence $\{(x^k, z^k, y^k)\}_{k \geqslant 0}$ for every $k \geqslant 0$ as:

$$z^{k+1} \in \arg \min_{z \in \mathbb{R}^m} \left\{ g(z) + \frac{r}{2} \left\| z - Ax^k - \frac{1}{r} y^k \right\|^2 \right\},$$

$$x^{k+1} \in \arg \min_{x \in \mathbb{R}^n} \left\{ h(x) + \frac{t_k}{2} \left\| x - x^k + \frac{1}{t_k} A^* \left[y^k + r \left(Ax^k - z^{k+1} \right) \right] \right\|^2 \right\},$$

$$y^{k+1} := y^k + \rho r \left(Ax^{k+1} - z^{k+1} \right).$$

Recall that the *proximal point operator with parameter* $\gamma > 0$ of a proper and lower semicontinuous function $\Psi \colon \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is the set-valued operator defined as ([39])

$$\mathrm{prox}_{\gamma \Psi} \colon \mathbb{R}^N \mapsto 2^{\mathbb{R}^N}, \quad \mathrm{prox}_{\gamma \Psi}(x) = \arg \min_{y \in \mathbb{R}^N} \left\{ \Psi(y) + \frac{1}{2\gamma} \|x - y\|^2 \right\}.$$

The above particular instance of Algorithm 1 is an iterative scheme formulated in the spirit of full splitting numerical methods; in other words, the functions g and h are evaluated by their proximal operators, while the linear operator A and its adjoint operator are evaluated by simple forward steps. Exact formulas for the proximal operator are available not only for large classes of convex functions ([6, 20]), but also for many nonconvex functions occurring in applications ([3, 29, 33]).

The second algorithm that we propose in this paper replaces for every $k \geqslant 0$ the function h in the definition of x^{k+1} by its linearization at x^k.

Algorithm 2. *Let be the matrix sequences $\left\{\mathbf{M}_1^k\right\}_{k\geqslant 0} \in \mathbb{S}_+^n$, $\left\{\mathbf{M}_2^k\right\}_{k\geqslant 0} \in \mathbb{S}_+^m$, $r > 0$ and $0 < \rho < 2$. For a given starting vector $\left(x^0, z^0, y^0\right) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$, generate the sequence $\left\{\left(x^k, z^k, y^k\right)\right\}_{k\geqslant 0}$ for every $k \geqslant 0$ as:*

$$z^{k+1} \in \arg\min_{z\in\mathbb{R}^m} \left\{g\left(z\right) + \left\langle y^k, Ax^k - z\right\rangle + \frac{r}{2}\left\|Ax^k - z\right\|^2 + \frac{1}{2}\left\|z - z^k\right\|_{\mathbf{M}_2^k}^2\right\}, \tag{13a}$$

$$x^{k+1} \in \arg\min_{x\in\mathbb{R}^n} \left\{\left\langle x - x^k, \nabla h\left(x^k\right)\right\rangle + \left\langle y^k, Ax - z^{k+1}\right\rangle + \frac{r}{2}\left\|Ax - z^{k+1}\right\|^2 + \frac{1}{2}\left\|x - x^k\right\|_{\mathbf{M}_1^k}^2\right\}, \tag{13b}$$

$$y^{k+1} := y^k + \rho r\left(Ax^{k+1} - z^{k+1}\right). \tag{13c}$$

Due to the presence of the variable metric inducing matrix sequences, Algorithm 2 represents a unifying scheme for several linearized ADMM algorithms from the literature (see [35, 36, 40, 41, 48, 50]). By choosing as above $\mathbf{M}_1^k := t_k\mathbf{Id} - rA^*A$, where t_k is positive such that $t_k \geqslant r\left\|A\right\|^2$, and $\mathbf{M}_2^k := \mathbf{0}$, for every $k \geqslant 0$, Algorithm 2 translates for every $k \geqslant 0$ into:

$$z^{k+1} \in \arg\min_{z\in\mathbb{R}^m} \left\{g\left(z\right) + \frac{r}{2}\left\|z - Ax^k - \frac{1}{r}y^k\right\|^2\right\},$$

$$x^{k+1} := x^k - \frac{1}{t_k}\left(\nabla h\left(x^k\right) + A^*\left[y^k + r\left(Ax^k - z^{k+1}\right)\right]\right),$$

$$y^{k+1} := y^k + \rho r\left(Ax^{k+1} - z^{k+1}\right).$$

In this iterative scheme the smooth term is evaluated via a gradient step, which is an improvement with respect to other nonconvex ADMM algorithms, such as [47, 49], where the smooth function is involved in a subproblem, which may be difficult to solve, unless it can be reformulated as a proximal step (see [34]).

We will carry out a parallel convergence analysis for Algorithm 1 and Algorithm 2 in the following setting.

Assumption 1. *We assume that*

1. *g and h are bounded from below;*

2. *A is surjective and thus the constant*

$$T_0 := \begin{cases} \dfrac{1}{\lambda_{\min}(AA^*)\rho}, & \text{if } 0 < \rho \leqslant 1, \\ \dfrac{\rho}{\lambda_{\min}(AA^*)\left(2 - \rho\right)^2}, & \text{if } 1 < \rho < 2, \end{cases}$$

   is well-defined;

3. *$\mu_1 := \sup\limits_{k\geqslant 0}\left\|\mathbf{M}_1^k\right\| < +\infty$ and $\mu_2 := \sup\limits_{k\geqslant 0}\left\|\mathbf{M}_2^k\right\| < +\infty$;*

4. *$r > 0, \rho \in (0, 2)$ and $\mu_1 \geqslant 0$ are such that*

$$r \geqslant 4T_0L > 0 \tag{14}$$

and

$$2\mathbf{M}_1^k + rA^*A \succcurlyeq \left(L + \frac{C_\mathbf{M}}{r}\right)\mathbf{Id} \quad \forall k \geqslant 0, \tag{15}$$

where

$$C_\mathbf{M} := \begin{cases} \left(6\mu_1^2 + 4\left(L + \mu_1\right)^2\right)T_0, & \text{for Algorithm 1}, \\ \left(4\mu_1^2 + 6\left(L + \mu_1\right)^2\right)T_0, & \text{for Algorithm 2}. \end{cases}$$

Remark 2. 1. It has been noticed also by other authors (see, for instance, [11]) that the surjectivity of the linear operator is an assumption which at this moment cannot be omitted when aiming to prove convergence for nonconvex Lagrangian based algorithms.

2. In the following we discuss possible choices of the matrix sequence $\left\{\mathbf{M}_1^k\right\}_{k\geqslant 0}$ which fulfil Assumption 1:

(a) If $\sup_{k \geqslant 0} \left\| \mathbf{M}_1^k \right\| = \mu_1 > \dfrac{L}{2}$, then, for every

$$r \geqslant \max \left\{ 4T_0 L, \frac{C_{\mathbf{M}}}{2\mu_1 - L} \right\} > 0,$$

there exists $\alpha_1 > 0$ such that

$$\mu_1 \geqslant \alpha_1 \geqslant \frac{1}{2} \left(L + \frac{C_{\mathbf{M}}}{r} \right) > 0.$$

The inequality in (15) is ensured for $\mathbf{M}_1^k$ chosen such that

$$\mu_1 \mathbf{Id} \geqslant \mathbf{M}_1^k \geqslant \alpha_1 \mathbf{Id} \quad \forall k \geqslant 0.$$

(b) If A is assumed to be also injective, then $\lambda_{\min} (A^* A) > 0$. By choosing

$$r \geqslant \max \left\{ 4T_0 L, \frac{L + \sqrt{L^2 + 4\lambda_{\min} (A^* A) C_{\mathbf{M}}}}{2\lambda_{\min} (A^* A)} \right\} > 0,$$

it follows that $r^2 \lambda_{\min} (A^* A) - rL - C_{\mathbf{M}} \geqslant 0$. Thus,

$$r A^* A - \left(L + r^{-1} C_{\mathbf{M}} \right) \mathbf{Id} \geqslant 0,$$

and (15) holds for an arbitrary sequence of symmetric and positive semidefinite matrices $\left\{ \mathbf{M}_1^k \right\}_{k \geqslant 0}$. A possible choice is $\mathbf{M}_1^k = 0$, which, together $\mathbf{M}_2^k = 0$, for every $k \geqslant 0$, allows us to recover the classical ADMM algorithm and the linearized ADMM algorithm as particular instances of our iterative schemes.

(c) For $t > 0$, we take $\mathbf{M}_1^k := t\mathbf{Id} - r A^* A$ for every $k \geqslant 0$ (see also Section 6.3 in [11]). Then

$$\mu_1 = \left\| t\mathbf{Id} - r A^* A \right\| = \lambda_{\max} (t\mathbf{Id} - r A^* A) = t - r \lambda_{\min} (A^* A).$$

Condition (15) is equivalent to

$$2t - r \left\| A \right\|^2 - \left(L + \frac{C_{\mathbf{M}}}{r} \right) \geqslant 0$$

and is guaranteed for both algorithms when

$$2t - r \left\| A \right\|^2 - \left(L + \frac{\left(4\mu_1^2 + 6 \left(L + \mu_1 \right)^2 \right) T_0}{r} \right) \geqslant 0$$

or, equivalently,

$$10T_0 \mu_1^2 - 2 \left(r - 6T_0 L \right) \mu_1 + 6T_0 L^2 + r^2 \left(\left\| A \right\|^2 - 2\lambda_{\min} (A^* A) \right) - Lr \leqslant 0.$$

This quadratic inequality in $\mu_1 \geqslant 0$ has nonnegative solutions if, for instance, $r \geqslant 6T_0 L$ (thus (14) holds) and the reduced discriminant

$$\Delta := \left(r - 6T_0 L \right)^2 - 60T_0^2 L^2 - 10T_0 r^2 \left(\left\| A \right\|^2 - 2\lambda_{\min} (A^* A) \right) + 10T_0 Lr$$

$$= \left[1 + 10T_0 \left(2\lambda_{\min} (A^* A) - \left\| A \right\|^2 \right) \right] r^2 - 2T_0 Lr - 24T_0^2 L^2$$

is nonnegative. This holds true if the condition number of the matrix $A^* A$ fulfils

$$\kappa (A^* A) := \frac{\lambda_{\max} (A^* A)}{\lambda_{\min} (A^* A)} = \frac{\left\| A \right\|^2}{\lambda_{\min} (A^* A)} \leqslant 2.$$

In conclusions, if the latter is given, then we can chose an arbitrary

$$r \geqslant 6T_0 L$$

and t such that

$$r \lambda_{\min} (A^* A) \leqslant t \leqslant r \lambda_{\min} (A^* A) + \frac{1}{10T_0} \left(r - 6T_0 L + \sqrt{\Delta} \right).$$

For a similar choice for the

3. When proving convergence and deriving convergence rates for variable metric algorithms designed for convex optimization problems one usually assumes monotonicity for the matrix sequences inducing the variable metrics (see, for instance, [19, 5, 14]). It is worth to mention that the convergence analysis for both Algorithm 1 and Algorithm 2 does not require monotonicity assumptions on $\left\{\mathbf{M}_1^k\right\}_{k \geqslant 0}$ or $\left\{\mathbf{M}_2^k\right\}_{k \geqslant 0}$.

2.2 Preliminaries of the convergence analysis

Within the setting of Assumption 1 we will make use of the following constants:

$$
C_0 := \begin{cases} L + \dfrac{4T_0\left(L + \mu_1\right)^2}{r}, & \text{for Algorithm 1,} \\ L + \dfrac{4T_0\mu_1^2}{r}, & \text{for Algorithm 2,} \end{cases}
\qquad
C_1 := \begin{cases} \dfrac{4T_0\mu_1^2}{r}, & \text{for Algorithm 1,} \\ \dfrac{4T_0\left(L + \mu_1\right)^2}{r}, & \text{for Algorithm 2,} \end{cases}
$$

and

$$
T_1 := \begin{cases} \dfrac{1 - \rho}{\lambda_{\min}\left(AA^*\right)\rho^2 r}, & \text{if } 0 < \rho \leqslant 1, \\ \dfrac{\rho - 1}{\lambda_{\min}\left(AA^*\right)\left(2 - \rho\right)\rho r}, & \text{if } 1 < \rho < 2, \end{cases}
$$

and we will denote for every $k \geqslant 0$

$$
\mathbf{M}_3^k := 2\mathbf{M}_1^k + rA^*A - C_0\mathbf{Id}.
$$

The following monotonicity result will play a fundamental role in our convergence analysis.

Lemma 4. *Suppose that Assumption 1 holds true and let $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ be a sequence generated by Algorithm 1 or Algorithm 2. Then for every $k \geqslant 1$ it holds:*

$$
\mathcal{L}_r\left(x^{k+1}, z^{k+1}, y^{k+1}\right) + T_1\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2 + \frac{1}{2}\left\|x^{k+1} - x^k\right\|_{\mathbf{M}_3^k}^2 + \frac{1}{2}\left\|z^{k+1} - z^k\right\|_{\mathbf{M}_2^k}^2
$$
$$
\leqslant \mathcal{L}_r\left(x^k, z^k, y^k\right) + T_1\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 + \frac{C_1}{2}\left\|x^k - x^{k-1}\right\|^2. \tag{16}
$$

Proof. Let $k \geqslant 1$ be fixed. In both cases the proof builds on showing that the following inequality

$$
\mathcal{L}_r\left(x^{k+1}, z^{k+1}, y^{k+1}\right) + \frac{1}{2}\left\|x^{k+1} - x^k\right\|_{2\mathbf{M}_1^k + rA^*A}^2 - \frac{L}{2}\left\|x^{k+1} - x^k\right\|^2 + \frac{1}{2}\left\|z^{k+1} - z^k\right\|_{\mathbf{M}_2^k}^2
$$
$$
\leqslant \mathcal{L}_r\left(x^k, z^k, y^k\right) + \frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2 \tag{17}
$$

is true and on providing afterwards an upper bound for $\dfrac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2$.

1. For *Algorithm 1*: From (12a) we have

$$
g\left(z^{k+1}\right) + \left\langle y^k, Ax^k - z^{k+1}\right\rangle + \frac{r}{2}\left\|Ax^k - z^{k+1}\right\|^2 + \frac{1}{2}\left\|z^{k+1} - z^k\right\|_{\mathbf{M}_2^k}^2
$$
$$
\leqslant g\left(z^k\right) + \left\langle y^k, Ax^k - z^k\right\rangle + \frac{r}{2}\left\|Ax^k - z^k\right\|^2. \tag{18}
$$

The optimality criterion of (12b) is

$$
\nabla h\left(x^{k+1}\right) = -A^*y^k - rA^*\left(Ax^{k+1} - z^{k+1}\right) + \mathbf{M}_1^k\left(x^k - x^{k+1}\right). \tag{19}
$$

From (6) (applied for $z := x^{k+1}$) we get

$$
h\left(x^{k+1}\right) \leqslant h\left(x^k\right) + \left\langle y^k, Ax^k - Ax^{k+1}\right\rangle + r\left\langle Ax^{k+1} - z^{k+1}, Ax^k - Ax^{k+1}\right\rangle
$$
$$
- \left\|x^{k+1} - x^k\right\|_{\mathbf{M}_1^k}^2 + \frac{L}{2}\left\|x^{k+1} - x^k\right\|^2. \tag{20}
$$

By combining (12c), (18) and (20), after some rearrangements, we obtain (17).

By using the notation

$$
u_1^l := -\nabla h\left(x^l\right) + \mathbf{M}_1^{l-1}\left(x^{l-1} - x^l\right) \quad \forall l \geqslant 1 \tag{21}
$$

and by taking into consideration (12c), we can rewrite (19) as

$$
A^*y^{l+1} = \rho u_1^{l+1} + \left(1 - \rho\right)A^*y^l \quad \forall l \geqslant 0. \tag{22}
$$

- *The case $0 < \rho \leqslant 1$.* We have

$$A^* \left(y^{k+1} - y^k\right) = \rho \left(u_1^{k+1} - u_1^k\right) + (1 - \rho) A^* \left(y^k - y^{k-1}\right).$$

Since $0 < \rho \leqslant 1$, the convexity of $\|\cdot\|^2$ gives

$$\left\|A^* \left(y^{k+1} - y^k\right)\right\|^2 \leqslant \rho \left\|u_1^{k+1} - u_1^k\right\|^2 + (1 - \rho) \left\|A^* \left(y^k - y^{k-1}\right)\right\|^2$$

and from here we get

$$\lambda_{\min}(AA^*)\rho \left\|y^{k+1} - y^k\right\|^2 \leqslant \rho \left\|A^* \left(y^{k+1} - y^k\right)\right\|^2$$
$$\leqslant \rho \left\|u_1^{k+1} - u_1^k\right\|^2 + (1 - \rho) \left\|A^* \left(y^k - y^{k-1}\right)\right\|^2 - (1 - \rho) \left\|A^* \left(y^{k+1} - y^k\right)\right\|^2. \tag{23}$$

By using the Lipschitz continuity of ∇h we have

$$\left\|u_1^{k+1} - u_1^k\right\| \leqslant (L + \mu_1) \left\|x^{k+1} - x^k\right\| + \mu_1 \left\|x^k - x^{k-1}\right\|, \tag{24}$$

thus

$$\left\|u_1^{k+1} - u_1^k\right\|^2 \leqslant 2 (L + \mu_1)^2 \left\|x^{k+1} - x^k\right\|^2 + 2\mu_1^2 \left\|x^k - x^{k-1}\right\|^2. \tag{25}$$

After plugging (25) into (23) we get

$$\frac{1}{\rho r} \left\|y^{k+1} - y^k\right\|^2 \leqslant \frac{2 (L + \mu_1)^2}{\lambda_{\min}(AA^*)\rho r} \left\|x^{k+1} - x^k\right\|^2 + \frac{2\mu_1^2}{\lambda_{\min}(AA^*)\rho r} \left\|x^k - x^{k-1}\right\|^2$$
$$+ \frac{(1 - \rho)}{\lambda_{\min}(AA^*)\rho^2 r} \left\|A^* \left(y^k - y^{k-1}\right)\right\|^2 - \frac{(1 - \rho)}{\lambda_{\min}(AA^*)\rho^2 r} \left\|A^* \left(y^{k+1} - y^k\right)\right\|^2, \tag{26}$$

which, combined with (17), provides (16).

- *The case $1 < \rho < 2$.* This time we have from (22) that

$$A^* \left(y^{k+1} - y^k\right) = (2 - \rho) \frac{\rho}{2 - \rho} \left(u_1^{k+1} - u_1^k\right) + (\rho - 1) A^* \left(y^{k-1} - y^k\right).$$

As $1 < \rho < 2$, the convexity of $\|\cdot\|^2$ gives

$$\left\|A^* \left(y^{k+1} - y^k\right)\right\|^2 \leqslant \frac{\rho^2}{2 - \rho} \left\|u_1^{k+1} - u_1^k\right\|^2 + (\rho - 1) \left\|A^* \left(y^k - y^{k-1}\right)\right\|^2$$

and from here it follows

$$\lambda_{\min}(AA^*) (2 - \rho) \left\|y^{k+1} - y^k\right\|^2 \leqslant (2 - \rho) \left\|A^* \left(y^{k+1} - y^k\right)\right\|^2$$
$$\leqslant \frac{\rho^2}{2 - \rho} \left\|u_1^{k+1} - u_1^k\right\|^2 + (\rho - 1) \left\|A^* \left(y^k - y^{k-1}\right)\right\|^2 - (\rho - 1) \left\|A^* \left(y^{k+1} - y^k\right)\right\|^2. \tag{27}$$

After plugging (25) into (27) we get

$$\frac{1}{\rho r} \left\|y^{k+1} - y^k\right\|^2 \leqslant \frac{2\rho (L + \mu_1)^2}{\lambda_{\min}(AA^*) (2 - \rho)^2 r} \left\|x^{k+1} - x^k\right\|^2 + \frac{2\rho\mu_1^2}{\lambda_{\min}(AA^*) (2 - \rho)^2 r} \left\|x^k - x^{k-1}\right\|^2$$
$$+ \frac{(\rho - 1)}{\lambda_{\min}(AA^*) (2 - \rho) \rho r} \left\|A^* \left(y^k - y^{k-1}\right)\right\|^2$$
$$- \frac{(\rho - 1)}{\lambda_{\min}(AA^*) (2 - \rho) \rho r} \left\|A^* \left(y^{k+1} - y^k\right)\right\|^2, \tag{28}$$

which, combined with (17), provides (16).

2. For *Algorithm 2*: The optimality criterion of (13b) is

$$\nabla h \left(x^k\right) = -A^* y^k - r A^* \left(A x^{k+1} - z^{k+1}\right) + \mathbf{M}_1^k \left(x^k - x^{k+1}\right). \tag{29}$$

11

From (6) (applied for $z := x^k$) we get

$$h\left(x^{k+1}\right) \leqslant h\left(x^k\right) + \left\langle y^k, Ax^k - Ax^{k+1}\right\rangle + r\left\langle Ax^{k+1} - z^{k+1}, Ax^k - Ax^{k+1}\right\rangle$$
$$- \left\|x^{k+1} - x^k\right\|_{\mathbf{M}_1^k}^2 + \frac{L}{2}\left\|x^{k+1} - x^k\right\|^2. \tag{30}$$

Since the definition of z^{k+1} in (13a) leads also to (18), by combining this inequality with (30) and (13c), after some rearrangments, (17) follows. By using this time the notation

$$u_2^l := -\nabla h\left(x^{l-1}\right) + \mathbf{M}_1^{l-1}\left(x^{l-1} - x^l\right) \ \forall l \geqslant 1 \tag{31}$$

and by taking into consideration (13c), we can rewrite (29) as

$$A^* y^{l+1} = \rho u_2^{l+1} + (1-\rho) A^* y^l \ \forall l \geqslant 0. \tag{32}$$

- *The case $0 < \rho \leqslant 1$.* As in (23) we obtain

$$\lambda_{\min}(AA^*)\rho \left\|y^{k+1} - y^k\right\|^2 \leqslant \rho \left\|A^*\left(y^{k+1} - y^k\right)\right\|^2$$
$$\leqslant \rho \left\|u_2^{k+1} - u_2^k\right\|^2 + (1-\rho)\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 - (1-\rho)\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2. \tag{33}$$

By using the Lipschitz continuity of ∇h we have

$$\left\|u_2^{k+1} - u_2^k\right\| \leqslant \mu_1 \left\|x^{k+1} - x^k\right\| + (L + \mu_1)\left\|x^k - x^{k-1}\right\|, \tag{34}$$

thus

$$\left\|u_2^{k+1} - u_2^k\right\|^2 \leqslant 2\mu_1^2 \left\|x^{k+1} - x^k\right\|^2 + 2(L + \mu_1)^2 \left\|x^k - x^{k-1}\right\|^2. \tag{35}$$

After plugging (35) into (33) it follows

$$\frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2 \leqslant \frac{2\mu_1^2}{\lambda_{\min}(AA^*)\rho r}\left\|x^{k+1} - x^k\right\|^2 + \frac{2(L+\mu_1)^2}{\lambda_{\min}(AA^*)\rho r}\left\|x^k - x^{k-1}\right\|^2$$
$$+ \frac{(1-\rho)}{\lambda_{\min}(AA^*)\rho^2 r}\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 - \frac{(1-\rho)}{\lambda_{\min}(AA^*)\rho^2 r}\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2, \tag{36}$$

which, combined with (17), provides (16).

- *The case $1 < \rho < 2$.* As in (27) we obtain

$$\lambda_{\min}(AA^*)(2-\rho)\left\|y^{k+1} - y^k\right\|^2 \leqslant (2-\rho)\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2$$
$$\leqslant \frac{\rho^2}{2-\rho}\left\|u_2^{k+1} - u_2^k\right\|^2 + (\rho-1)\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 - (\rho-1)\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2. \tag{37}$$

After plugging (35) into (37) it follows

$$\frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2 \leqslant \frac{2\rho\mu_1^2}{\lambda_{\min}(AA^*)(2-\rho)^2 r}\left\|x^{k+1} - x^k\right\|^2 + \frac{2\rho(L+\mu_1)^2}{\lambda_{\min}(AA^*)(2-\rho)^2 r}\left\|x^k - x^{k-1}\right\|^2$$
$$+ \frac{(\rho-1)}{\lambda_{\min}(AA^*)(2-\rho)\rho r}\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2$$
$$- \frac{(\rho-1)}{\lambda_{\min}(AA^*)(2-\rho)\rho r}\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2. \tag{38}$$

which, combined with (17), provides (16).

This concludes the proof. □

The following three estimates will be useful in the sequel.

Lemma 5. *Suppose that Assumption 1 holds true and let $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ be a sequence generated by Algorithm 1 or Algorithm 2. Then the following statements are true:*

(i) for every $k \geqslant 1$

$$\left\| z^{k+1} - z^k \right\| \leqslant \|A\| \cdot \left\| x^{k+1} - x^k \right\| + \left\| Ax^{k+1} - z^{k+1} \right\| + \left\| Ax^k - z^k \right\|$$

$$= \|A\| \cdot \left\| x^{k+1} - x^k \right\| + \frac{1}{\rho r} \left\| y^{k+1} - y^k \right\| + \frac{1}{\rho r} \left\| y^k - y^{k-1} \right\|; \tag{39}$$

(ii) for every $k \geqslant 0$

$$\frac{1}{2r} \left\| y^{k+1} \right\|^2 \leqslant \frac{T_1}{2} \left\| A^* \left(y^{k+1} - y^k \right) \right\|^2 + \frac{T_0}{r} \left\| \nabla h \left(x^{k+1} \right) \right\|^2 + \frac{C_1}{4} \left\| x^{k+1} - x^k \right\|^2; \tag{40}$$

(iii) for every $k \geqslant 1$

$$\left\| y^{k+1} - y^k \right\| \leqslant C_3 \left\| x^{k+1} - x^k \right\| + C_4 \left\| x^k - x^{k-1} \right\| + T_2 \left(\left\| A^* \left(y^k - y^{k-1} \right) \right\| - \left\| A^* \left(y^{k+1} - y^k \right) \right\| \right), \tag{41}$$

where

$$C_3 := \begin{cases} \dfrac{\rho \left(L + \mu_1 \right)}{\sqrt{\lambda_{\min}(AA^*)} \left(1 - |1 - \rho| \right)}, & \text{for Algorithm 1,} \\[4mm] \dfrac{\rho \mu_1}{\sqrt{\lambda_{\min}(AA^*)} \left(1 - |1 - \rho| \right)}, & \text{for Algorithm 2,} \end{cases}$$

$$C_4 := \begin{cases} \dfrac{\rho \mu_1}{\sqrt{\lambda_{\min}(AA^*)} \left(1 - |1 - \rho| \right)}, & \text{for Algorithm 1,} \\[4mm] \dfrac{\rho \left(L + \mu_1 \right)}{\sqrt{\lambda_{\min}(AA^*)} \left(1 - |1 - \rho| \right)}, & \text{for Algorithm 2,} \end{cases}$$

$$T_2 := \frac{|1 - \rho|}{\sqrt{\lambda_{\min}(AA^*)} \left(1 - |1 - \rho| \right)}.$$

Proof. The statement in (39) is straightforward.

From (22) and (32) we have for every $k \geqslant 0$

$$A^* y^{k+1} = \rho u^{k+1} + (1 - \rho) A^* y^k$$

or, equivalently,

$$\rho A^* y^{k+1} = \rho u^{k+1} + (1 - \rho) A^* \left(y^k - y^{k+1} \right),$$

where u^{k+1} is defined as being equal to u_1^{k+1} in (21), for Algorithm 1, and, respectively, to u_2^{k+1} in (31), for Algorithm 2.

For $0 < \rho \leqslant 1$ we have

$$\lambda_{\min}(AA^*) \rho^2 \left\| y^{k+1} \right\|^2 \leqslant \rho^2 \left\| A^* y^{k+1} \right\|^2 \leqslant \rho \left\| u^{k+1} \right\|^2 + (1 - \rho) \left\| A^* \left(y^{k+1} - y^k \right) \right\|^2, \tag{42}$$

while, for $1 < \rho < 2$, we have

$$\lambda_{\min}(AA^*) \rho^2 \left\| y^{k+1} \right\|^2 \leqslant \rho^2 \left\| A^* y^{k+1} \right\|^2 \leqslant \frac{\rho^2}{2 - \rho} \left\| u^{k+1} \right\|^2 + (\rho - 1) \left\| A^* \left(y^{k+1} - y^k \right) \right\|^2. \tag{43}$$

Notice further that for $1 < \rho < 2$ we have $\dfrac{1}{\rho} < 1$ and $1 < \dfrac{\rho}{2 - \rho}$.

In case u^{k+1} is defined as in (21) it holds

$$\left\| u^{k+1} \right\|^2 = \left\| u_1^{k+1} \right\|^2 \leqslant 2 \left\| \nabla h \left(x^{k+1} \right) \right\|^2 + 2\mu_1^2 \left\| x^{k+1} - x^k \right\|^2 \ \forall k \geqslant 0, \tag{44}$$

while, in case u_2^{k+1} is defined as in (31), it holds

$$\left\| u^{k+1} \right\|^2 = \left\| u_2^{k+1} \right\|^2 \leqslant 2 \left\| \nabla h \left(x^{k+1} \right) \right\|^2 + 2 \left(L + \mu_1 \right)^2 \left\| x^{k+1} - x^k \right\|^2 \ \forall k \geqslant 0. \tag{45}$$

13

We divide (42) and (43) by $2\lambda_{\min}(AA^*)\rho^2 r > 0$ and plug (44) and, respectively, (45) into the resulting inequalities. This gives us (40).

Finally, in order to prove (41), we notice that for every $k \geqslant 1$ it holds

$$\left\| A^* \left(y^{k+1} - y^k \right) \right\| \leqslant \rho \left\| u^{k+1} - u^k \right\| + |1 - \rho| \left\| A^* \left(y^k - y^{k-1} \right) \right\|,$$

so,

$$\sqrt{\lambda_{\min}(AA^*)} \left(1 - |1 - \rho| \right) \left\| y^{k+1} - y^k \right\| \leqslant \left(1 - |1 - \rho| \right) \left\| A^* \left(y^{k+1} - y^k \right) \right\|$$
$$\leqslant \rho \left\| u^{k+1} - u^k \right\| + |1 - \rho| \left\| A^* \left(y^k - y^{k-1} \right) \right\| - |1 - \rho| \left\| A^* \left(y^{k+1} - y^k \right) \right\|. \tag{46}$$

We plug into (46) the estimates for $\left\| u^{k+1} - u^k \right\|$ derived in (24) and, respectively, (34) and divide the resulting inequality by $\sqrt{\lambda_{\min}(AA^*)}\left(1 - |1 - \rho|\right) > 0$. This furnishes the desired statement. $\qquad \square$

The following regularization of the augmented Lagrangian will play an important role in the convergence analysis of the nonconvex proximal ADMM algorithms

$$\mathcal{F}_r \colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\},$$

$$\mathcal{F}_r(x, z, y, x', y') = \mathcal{L}_r(x, z, y) + T_1 \left\| A^* \left(y - y' \right) \right\|^2 + \frac{C_1}{2} \left\| x - x' \right\|^2,$$

where T_1 and C_1 are defined in Assumption 1. For every $k \geqslant 1$ we denote

$$\mathcal{F}_k := \mathcal{F}_r \left(x^k, z^k, y^k, x^{k-1}, y^{k-1} \right) = \mathcal{L}_r \left(x^k, z^k, y^k \right) + T_1 \left\| A^* \left(y^k - y^{k-1} \right) \right\|^2 + \frac{C_1}{2} \left\| x^k - x^{k-1} \right\|^2. \tag{47}$$

Since the convergence analysis will rely on the fact that the set of cluster points of the sequence $\left\{ \left(x^k, z^k, y^k \right) \right\}_{k \geqslant 0}$ is nonempty, we will present first two situations which guarantee that this sequence is bounded. They make use of standard coercivity assumptions for the functions g and h, respectively. Recall that a function $\Psi : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ is called *coercive*, if $\lim_{\|x\| \to +\infty} \Psi(x) = +\infty$.

Theorem 6. *Suppose that Assumption 1 holds true and let* $\left\{ \left(x^k, z^k, y^k \right) \right\}_{k \geqslant 0}$ *be a sequence generated by Algorithm 1 or Algorithm 2. Suppose that one of the following conditions holds:*

(B-I) A is invertible and g is coercive;

(B-II) h is coercive.

Then the sequence $\left\{ \left(x^k, z^k, y^k \right) \right\}_{k \geqslant 0}$ *is bounded.*

Proof. From Lemma 4 we have that for every $k \geqslant 1$

$$\mathcal{F}_{k+1} + \frac{1}{2} \left\| x^{k+1} - x^k \right\|^2_{\mathbf{M}_3^k - C_1 \mathbf{Id}} + \frac{1}{2} \left\| z^{k+1} - z^k \right\|^2_{\mathbf{M}_2^k} \leqslant \mathcal{F}_k \tag{48}$$

which shows, according to (15), that $\{\mathcal{F}_k\}_{k \geqslant 1}$ is monotonically decreasing. Consequently, for every $k \geqslant 1$ we have

$$\mathcal{F}_1 \geqslant \mathcal{F}_{k+1} + \frac{1}{2} \left\| x^{k+1} - x^k \right\|^2_{\mathbf{M}_3^k - C_1 \mathbf{Id}} + \frac{1}{2} \left\| z^{k+1} - z^k \right\|^2_{\mathbf{M}_2^k}$$
$$= h \left(x^{k+1} \right) + g \left(z^{k+1} \right) - \frac{1}{2r} \left\| y^{k+1} \right\|^2 + \frac{r}{2} \left\| A x^{k+1} - z^{k+1} + \frac{1}{r} y^{k+1} \right\|^2$$
$$+ T_1 \left\| A^* \left(y^{k+1} - y^k \right) \right\|^2 + \frac{1}{2} \left\| x^{k+1} - x^k \right\|^2_{\mathbf{M}_3^k - C_0 \mathbf{Id}} + \frac{1}{2} \left\| z^{k+1} - z^k \right\|^2_{\mathbf{M}_2^k} + \frac{C_1}{2} \| x^{k+1} - x^k \|,$$

which, thanks to (40), leads to

$$\mathcal{F}_1 \geqslant h \left(x^{k+1} \right) + g \left(z^{k+1} \right) - \frac{T_0}{r} \left\| \nabla h \left(x^{k+1} \right) \right\|^2 + \frac{r}{2} \left\| A x^{k+1} - z^{k+1} + \frac{1}{r} y^{k+1} \right\|^2$$
$$+ \frac{T_1}{2} \left\| A^* \left(y^{k+1} - y^k \right) \right\|^2 + \frac{1}{2} \left\| x^{k+1} - x^k \right\|^2_{\mathbf{M}_3^k - C_1 \mathbf{Id}} + \frac{1}{2} \left\| z^{k+1} - z^k \right\|^2_{\mathbf{M}_2^k} + \frac{C_1}{4} \| x^{k+1} - x^k \|^2. \tag{49}$$

Next we will prove the boundedness of $\left\{ \left(x^k, z^k, y^k \right) \right\}_{k \geqslant 0}$ under each of the two scenarios.

(B-I) Since $r \geqslant 4T_0L$, there exists $\sigma > 0$ such that

$$\frac{1}{\sigma} - \frac{L}{2\sigma^2} = \frac{T_0}{r}.$$

From Proposition 1 and the relation (49) we see that for every $k \geqslant 1$

$$g\left(z^{k+1}\right) + \frac{r}{2}\left\|Ax^{k+1} - z^{k+1} + \frac{1}{r}y^{k+1}\right\|^2 + \frac{C_1}{4}\|x^{k+1} - x^k\|^2$$
$$\leqslant \mathcal{F}_1 - \inf_{x \in \mathbb{R}^n}\left\{h\left(x\right) - \frac{T_0}{r}\left\|\nabla h\left(x\right)\right\|^2\right\} < +\infty.$$

Since g is coercive, it follows that the sequence $\left\{z^k\right\}_{k \geqslant 0}$ is bounded. On the other hand, since g is bounded from below, it follows that the sequences $\left\{Ax^k - z^k + r^{-1}y^k\right\}_{k \geqslant 0}$ and $\left\{x^{k+1} - x^k\right\}_{k \geqslant 0}$ are bounded as well. In addition, since for every $k \geqslant 0$ it holds

$$\left\|A\left(x^{k+1} - x^k\right) - \left(z^{k+1} - z^k\right)\right\| \leqslant \|A\| \cdot \left\|x^{k+1} - x^k\right\| + \left\|z^{k+1}\right\| + \left\|z^k\right\|$$

it follows that $\left\{A\left(x^{k+1} - x^k\right) - \left(z^{k+1} - z^k\right)\right\}_{k \geqslant 0}$ is bounded, thus $\left\{r^{-1}\left(y^{k+1} - y^k\right)\right\}_{k \geqslant 0}$ is bounded. According to the third update in the iterative scheme we obtain that $\left\{Ax^k - z^k\right\}_{k \geqslant 0}$ is bounded and from here that $\left\{y^k\right\}_{k \geqslant 0}$ is also bounded. This implies the boundedness of $\left\{Ax^k\right\}_{k \geqslant 0}$ and, finally, since A is invertible, the boundedness of $\left\{x^k\right\}_{k \geqslant 0}$.

(B-II) Again thanks to (14) there exists $\sigma > 0$ such that

$$\frac{1}{\sigma} - \frac{L}{2\sigma^2} = \frac{3T_0}{2r}.$$

We assume first that $\rho \neq 1$ or, equivalently, $T_1 \neq 0$. From Proposition 1 and (49) we see that for every $k \geqslant 1$

$$\frac{1}{2}h\left(x^{k+1}\right) + \frac{T_0}{4r}\left\|\nabla h\left(x^{k+1}\right)\right\|^2 + \frac{r}{2}\left\|Ax^{k+1} - z^{k+1} + \frac{1}{r}y^{k+1}\right\|^2 + \frac{T_1}{2}\left\|A^*\left(y^{k+1} - y^k\right)\right\|$$
$$\leqslant \mathcal{F}_1 - \inf_{z \in \mathbb{R}^m} g\left(z\right) - \frac{1}{2}\inf_{x \in \mathbb{R}^n}\left\{h\left(x\right) - \frac{3T_0}{2r}\left\|\nabla h\left(x\right)\right\|^2\right\} < +\infty.$$

Since h is coercive and bounded from below, we obtain that $\left\{x^k\right\}_{k \geqslant 0}$, $\left\{Ax^k - z^k + r^{-1}y^k\right\}_{k \geqslant 0}$ and $\left\{A^*\left(y^{k+1} - y^k\right)\right\}_{k \geqslant 0}$ are bounded. For every $k \geqslant 0$ we have that

$$\lambda_{\min}\left(A^*A\right)\rho^2 r^2\left\|Ax^{k+1} - z^{k+1}\right\|^2 = \lambda_{\min}\left(A^*A\right)\left\|y^{k+1} - y^k\right\|^2 \leqslant \left\|A^*\left(y^{k+1} - y^k\right)\right\|^2,$$

thus $\left\{Ax^k - z^k\right\}_{k \geqslant 0}$ is bounded. Consequently, $\left\{y^k\right\}_{k \geqslant 0}$ and $\left\{z^k\right\}_{k \geqslant 0}$ are bounded.

In case $\rho = 1$ or, equivalently, $T_1 = 0$, we have that for every $k \geqslant 1$

$$\frac{1}{2}h\left(x^{k+1}\right) + \frac{T_0}{4r}\left\|\nabla h\left(x^{k+1}\right)\right\|^2 + \frac{r}{2}\left\|Ax^{k+1} - z^{k+1} + \frac{1}{r}y^{k+1}\right\|^2$$
$$\leqslant \mathcal{F}_1 - \inf_{z \in \mathbb{R}^m} g\left(z\right) - \frac{1}{2}\inf_{x \in \mathbb{R}^n}\left\{h\left(x\right) - \frac{3T_0}{2r}\left\|\nabla h\left(x\right)\right\|^2\right\} < +\infty$$

from which we deduce that $\left\{x^k\right\}_{k \geqslant 0}$ and $\left\{Ax^k - z^k + r^{-1}y^k\right\}_{k \geqslant 0}$ are bounded. From Lemma 5 (iii), which now reads

$$\left\|y^{k+1} - y^k\right\| \leqslant C_3\left\|x^{k+1} - x^k\right\| + C_4\left\|x^k - x^{k-1}\right\| \quad \forall k \geqslant 1,$$

it yields that $\left\{y^{k+1} - y^k\right\}_{k \geqslant 0}$ is bounded, thus, $\left\{Ax^k - z^k\right\}_{k \geqslant 0}$ is bounded. Consequently, $\left\{y^k\right\}_{k \geqslant 0}$ and $\left\{z^k\right\}_{k \geqslant 0}$ are bounded.

Both considered scenarios lead to the conclusion that the sequence $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ is bounded. $\quad\square$

We state now the first convergence result of this paper.

Theorem 7. *Suppose that Assumption 1 holds true and let $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. The following statements are true:*

(i) *for every $k \geqslant 1$ it holds*

$$\mathcal{F}_{k+1} + \frac{C_1}{4} \left\| x^{k+1} - x^k \right\|^2 + \frac{1}{2} \left\| z^{k+1} - z^k \right\|_{\mathbf{M}_2^k}^2 \leqslant \mathcal{F}_k; \tag{50}$$

(ii) *the sequence $\{\mathcal{F}_k\}_{k\geqslant 0}$ is bounded from below and convergent. In addition,*

$$x^{k+1} - x^k \to 0, \ z^{k+1} - z^k \to 0 \ and \ y^{k+1} - y^k \to 0 \ as \ k \to +\infty; \tag{51}$$

(iii) *the sequences $\{\mathcal{F}_k\}_{k\geqslant 0}$, $\{\mathcal{L}_r\left(x^k, z^k, y^k\right)\}_{k\geqslant 0}$ and $\{h\left(x^k\right) + g\left(z^k\right)\}_{k\geqslant 0}$ have the same limit, which we denote by $\mathcal{F}_* \in \mathbb{R}$.*

Proof.   (i) According to (15) we have that $\mathbf{M}_3^k - C_1 \mathbf{Id} \in \mathcal{P}_{\frac{C_1}{2}}^n$ and thus (48) implies (50).

(ii) We will show that $\{\mathcal{L}_r\left(x^k, z^k, y^k\right)\}_{k\geqslant 0}$ is bounded from below, which will imply that $\{\mathcal{F}_k\}_{k\geqslant 0}$ is bounded from below as well. Assuming the contrary, as $\{\left(x^k, z^k, y^k\right)\}_{k\geqslant 0}$ is bounded, there exists a subsequence $\{\left(x^{k_q}, z^{k_q}, y^{k_q}\right)\}_{q\geqslant 0}$ converging to an element $(\hat{x}, \hat{z}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ such that $\{\mathcal{L}_r\left(x^{k_q}, z^{k_q}, y^{k_q}\right)\}_{q\geqslant 0}$ converges to $-\infty$ as $q \to +\infty$. However, using the lower semicontinuity of g and the continuity of h, we obtain

$$\liminf_{q \to +\infty} \mathcal{L}_r\left(x^{k_q}, z^{k_q}, y^{k_q}\right) \geqslant h\left(\hat{x}\right) + g\left(\hat{z}\right) + \langle \hat{y}, A\hat{x} - \hat{z} \rangle + \frac{r}{2} \left\| A\hat{x} - \hat{z} \right\|^2,$$

which leads to a contradiction. From Lemma 2 we conclude that $\{\mathcal{F}_k\}_{k\geqslant 1}$ is convergent and

$$\sum_{k\geqslant 0} \left\| x^{k+1} - x^k \right\|^2 < +\infty,$$

thus $x^{k+1} - x^k \to 0$ as $k \to +\infty$.

We proved in (26), (28), (36) and (38) that for every $k \geqslant 1$

$$\frac{1}{\rho r} \left\| y^{k+1} - y^k \right\|^2 \leqslant \frac{C_0 - L}{2} \left\| x^{k+1} - x^k \right\|^2 + \frac{C_1}{2} \left\| x^k - x^{k-1} \right\|^2$$
$$+ T_1 \left\| A^* \left(y^k - y^{k-1} \right) \right\|^2 - T_1 \left\| A^* \left(y^{k+1} - y^k \right) \right\|^2.$$

Summing up the above inequality for $k = 1, \ldots, K$, for $K > 1$, we get

$$\frac{1}{\rho r} \sum_{k=1}^{K} \left\| y^{k+1} - y^k \right\|^2 \leqslant \frac{C_0 - L}{2} \sum_{k=1}^{K} \left\| x^{k+1} - x^k \right\|^2 + \frac{C_1}{2} \sum_{k=1}^{K} \left\| x^k - x^{k-1} \right\|^2$$
$$+ T_1 \left\| A^* \left(y^1 - y^0 \right) \right\|^2 - T_1 \left\| A^* \left(y^{K+1} - y^K \right) \right\|^2$$
$$\leqslant \frac{C_0 - L}{2} \sum_{k=1}^{K} \left\| x^{k+1} - x^k \right\|^2 + \frac{C_1}{2} \sum_{k=1}^{K} \left\| x^k - x^{k-1} \right\|^2 + T_1 \left\| A^* \left(y^1 - y^0 \right) \right\|^2.$$

We let K converge to $+\infty$ and conclude

$$\rho r \sum_{k\geqslant 0} \left\| Ax^{k+1} - z^{k+1} \right\|^2 = \frac{1}{\rho r} \sum_{k\geqslant 0} \left\| y^{k+1} - y^k \right\|^2 < +\infty,$$

thus $Ax^{k+1} - z^{k+1} \to 0$ and $y^{k+1} - y^k \to 0$ as $k \to +\infty$. Since $x^{k+1} - x^k \to 0$ as $k \to +\infty$, it follows that $z^{k+1} - z^k \to 0$ as $k \to +\infty$.

(iii) By using (51) and the fact that $\{y^k\}_{k\geqslant 0}$ is bounded, it follows

$$\mathcal{F}_* = \lim_{k \to +\infty} \mathcal{F}_k = \lim_{k \to +\infty} \mathcal{L}_r\left(x^k, z^k, y^k\right) = \lim_{k \to +\infty} \{h\left(x^k\right) + g\left(z^k\right)\},$$

which is the desired statement. $\qquad\qquad\square$

The following lemmas provides upper estimates in terms of the iterates for limiting subgradients of the augmented Lagrangian and the regularized augmented Lagrangian $\mathcal{F}_r$, respectively.

Lemma 8. *Suppose that Assumption 1 holds true and let $\left\{(x^k, z^k, y^k)\right\}_{k \geqslant 0}$ be a sequence generated by Algorithm 1 or Algorithm 2. For every $k \geqslant 0$ we have*

$$d^{k+1} := \left(d_x^{k+1}, d_z^{k+1}, d_y^{k+1}\right) \in \partial \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1}\right), \tag{52}$$

where

$$d_x^{k+1} := C_2 \left(\nabla h \left(x^{k+1}\right) - \nabla h \left(x^k\right)\right) + A^* \left(y^{k+1} - y^k\right) + \mathbf{M}_1^k \left(x^k - x^{k+1}\right), \tag{53a}$$

$$d_z^{k+1} := y^k - y^{k+1} + rA \left(x^k - x^{k+1}\right) + \mathbf{M}_2^k \left(z^k - z^{k+1}\right), \tag{53b}$$

$$d_y^{k+1} := \frac{1}{\rho r} \left(y^{k+1} - y^k\right). \tag{53c}$$

and

$$C_2 := \begin{cases} 0, & \text{for Algorithm 1,} \\ 1, & \text{for Algorithm 2.} \end{cases}$$

Moreover, for every $k \geqslant 0$ it holds

$$\left\|\left\|d^{k+1}\right\|\right\| \leqslant C_5 \left\|x^{k+1} - x^k\right\| + C_6 \left\|z^{k+1} - z^k\right\| + C_7 \left\|y^{k+1} - y^k\right\|, \tag{54}$$

where

$$C_5 := C_2 L + \mu_1 + r \left\|A\right\|, \quad C_6 := \mu_2, \quad C_7 := 1 + \left\|A\right\| + \frac{1}{\rho r}.$$

Proof. Let $k \geqslant 0$ be fixed. Applying the calculus rules of the limiting subdifferential, we obtain

$$\nabla_x \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1}\right) = \nabla h \left(x^{k+1}\right) + A^* y^{k+1} + rA^* \left(Ax^{k+1} - z^{k+1}\right), \tag{55a}$$

$$\partial_z \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1}\right) = \partial g \left(z^{k+1}\right) - y^{k+1} - r \left(Ax^{k+1} - z^{k+1}\right), \tag{55b}$$

$$\nabla_y \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1}\right) = Ax^{k+1} - z^{k+1}. \tag{55c}$$

Then (53c) follows directly from (55c) and (12c), respectively, (13c), while (53b) follows from

$$y^k + r(Ax^k - z^{k+1}) + \mathbf{M}_2^k \left(z^k - z^{k+1}\right) \in \partial g \left(z^{k+1}\right),$$

which is a consequence of the optimality criterion of (12a) and (13a), respectively. In order to derive (53a), let us notice that for Algorithm 1 we have (see (19))

$$-A^* y^k + \mathbf{M}_1^k \left(x^k - x^{k+1}\right) = \nabla h \left(x^{k+1}\right) + rA^* \left(Ax^{k+1} - z^{k+1}\right), \tag{56}$$

while for Algorithm 2 we have (see (29))

$$-\nabla h \left(x^k\right) - A^* y^k + \mathbf{M}_1^k \left(x^k - x^{k+1}\right) = rA^* \left(Ax^{k+1} - z^{k+1}\right). \tag{57}$$

By using (55a) we get the desired statement.

Relation (54) follows by combining the inequalities

$$\left\|d_x^{k+1}\right\| \leqslant \left(C_2 L + \mu_1\right) \left\|x^{k+1} - x^k\right\| + \left\|A\right\| \cdot \left\|y^{k+1} - y^k\right\|,$$

$$\left\|d_z^{k+1}\right\| \leqslant \left\|y^k - y^{k+1}\right\| + r \left\|A\right\| \cdot \left\|x^{k+1} - x^k\right\| + \mu_2 \left\|z^{k+1} - z^k\right\|$$

and (5). □

Lemma 9. *Suppose that Assumption 1 holds true and let $\left\{(x^k, z^k, y^k)\right\}_{k \geqslant 0}$ be a sequence generated by Algorithm 1 or Algorithm 2. For every $k \geqslant 0$ we have*

$$D^{k+1} := \left(D_x^{k+1}, D_z^{k+1}, D_y^{k+1}, D_{x'}^{k+1}, D_{y'}^{k+1}\right) \in \partial \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k\right) \tag{58}$$

where

$$D_x^{k+1} := d_x^{k+1} + C_1 \left(x^{k+1} - x^k\right), \quad D_z^{k+1} := d_z^{k+1}, \quad D_y^{k+1} := d_y^{k+1} + 2T_1 AA^* \left(y^{k+1} - y^k\right),$$

$$D_{x'}^{k+1} := -C_1 \left(x^{k+1} - x^k\right), \quad D_{y'}^{k+1} := -2T_1 AA^* \left(y^{k+1} - y^k\right). \tag{59}$$

Moreover, for every $k \geqslant 0$ it holds

$$\left\|\left\|D^{k+1}\right\|\right\| \leqslant C_8 \left\|x^{k+1} - x^k\right\| + C_9 \left\|z^{k+1} - z^k\right\| + C_{10} \left\|y^{k+1} - y^k\right\|, \tag{60}$$

where

$$C_8 := 2C_1 + C_5, \quad C_9 := C_6, \quad C_{10} := C_7 + 4T_1 \left\|A\right\|^2.$$

17

Proof. Let $k \geqslant 0$ be fixed. Applying the calculus rules of the limiting subdifferential it follows

$$\nabla_x \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k\right) := \nabla_x \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1}\right) + C_1 \left(x^{k+1} - x^k\right), \tag{61a}$$

$$\partial_z \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k\right) := \partial_z \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1}\right) \tag{61b}$$

$$\nabla_y \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k\right) := \nabla_y \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1}\right) + 2T_1 A A^* \left(y^{k+1} - y^k\right), \tag{61c}$$

$$\nabla_{x'} \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k\right) := -C_1 \left(x^{k+1} - x^k\right), \tag{61d}$$

$$\nabla_{y'} \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k\right) := -2T_1 A A^* \left(y^{k+1} - y^k\right). \tag{61e}$$

Then (58) follows directly from the above relations and (52). Inequality (60) follows by combining

$$\left\| D_x^{k+1} \right\| \leqslant \left\| d_x^{k+1} \right\| + C_1 \left\| x^{k+1} - x^k \right\|,$$
$$\left\| D_y^{k+1} \right\| \leqslant \left\| d_y^{k+1} \right\| + 2T_1 \left\| A \right\|^2 \cdot \left\| y^{k+1} - y^k \right\|.$$

and (5). $\qquad\square$

The following result is a straightforward consequence of Lemma 5 and Lemma 9.

Corollary 10. *Suppose that Assumption 1 holds true and let $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ be a sequence generated by Algorithm 1 or Algorithm 2. Then the norm of the element $D^{k+1} \in \partial \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k\right)$ defined in the previous lemma verifies for every $k \geqslant 2$ the following estimate*

$$\begin{aligned} \left\|\!\left\| D^{k+1} \right\|\!\right\| \leqslant \; & C_{11} \left(\left\| x^{k+1} - x^k \right\| + \left\| x^k - x^{k-1} \right\| + \left\| x^{k-1} - x^{k-2} \right\| \right) \\ & + C_{12} \left(\left\| A^* \left(y^k - y^{k-1}\right) \right\| - \left\| A^* \left(y^{k+1} - y^k\right) \right\| \right) \\ & + C_{13} \left(\left\| A^* \left(y^{k-1} - y^{k-2}\right) \right\| - \left\| A^* \left(y^k - y^{k-1}\right) \right\| \right), \end{aligned} \tag{62}$$

where

$$C_{11} := \max \left\{ C_8 + C_9 \left\| A \right\| + C_3 C_{10} + \frac{C_3 C_9}{\rho r}, C_4 C_{10} + \frac{C_3 C_9}{\rho r}, \frac{C_4 C_9}{\rho r} \right\},$$
$$C_{12} := \left(C_{10} + \frac{C_9}{\rho r} \right) T_2, \qquad C_{13} := \frac{C_9 T_2}{\rho r}.$$

In the following, we denote by $\omega \left(\left\{u^k\right\}_{k \geqslant 0} \right)$ the set of *cluster points* of the sequence $\left\{u^k\right\}_{k \geqslant 0} \subseteq \mathbb{R}^N$.

Lemma 11. *Suppose that Assumption 1 holds true and let $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. The following statements are true:*

(i) *if $\left\{\left(x^{k_q}, z^{k_q}, y^{k_q}\right)\right\}_{q \geqslant 0}$ is a subsequence of $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ which converges to $(\widehat{x}, \widehat{z}, \widehat{y})$ as $q \rightarrow +\infty$, then*

$$\lim_{q \rightarrow \infty} \mathcal{L}_r \left(x^{k_q}, z^{k_q}, y^{k_q}\right) = \mathcal{L}_r \left(\widehat{x}, \widehat{z}, \widehat{y}\right);$$

(ii) *it holds*

$$\begin{aligned} \omega \left(\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0} \right) &\subseteq \operatorname{crit} \left(\mathcal{L}_r\right) \\ &\subseteq \left\{ (\widehat{x}, \widehat{z}, \widehat{y}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \colon -A^* \widehat{y} = \nabla h\left(\widehat{x}\right), \widehat{y} \in \partial g\left(\widehat{z}\right), \widehat{z} = A\widehat{x} \right\}; \end{aligned}$$

(iii) *we have $\displaystyle\lim_{k \rightarrow +\infty} \operatorname{dist} \left[\left(x^k, z^k, y^k\right), \omega \left(\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0} \right) \right] = 0$;*

(iv) *the set $\omega \left(\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0} \right)$ is nonempty, connected and compact;*

(v) *the function $\mathcal{L}_r$ takes on $\omega \left(\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0} \right)$ the value $\mathcal{F}_* = \lim_{k \rightarrow +\infty} \mathcal{L}_r \left(x^k, z^k, y^k\right)$, as the objective function $g \circ A + h$ does on the projection of the set $\omega \left(\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0} \right)$ onto the space $\mathbb{R}^n$ corresponding to the first component.*

Proof. Let $(\widehat{x}, \widehat{z}, \widehat{y}) \in \omega \left(\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0} \right)$ and $\left\{\left(x^{k_q}, z^{k_q}, y^{k_q}\right)\right\}_{q \geqslant 0}$ be a subsequence of $\left\{x^k, z^k, y^k\right\}_{k \geqslant 0}$ converging to $(\widehat{x}, \widehat{z}, \widehat{y})$ as $q \rightarrow +\infty$.

(i) From either (12a) or (13a) we obtain for every $q \geqslant 1$

$$g\left(z^{k_q}\right) + \left\langle y^{k_q-1}, Ax^{k_q-1} - z^{k_q}\right\rangle + \frac{r}{2}\left\|Ax^{k_q-1} - z^{k_q}\right\|^2 + \frac{1}{2}\left\|z^{k_q} - z^{k_q-1}\right\|^2_{\mathbf{M}_2^{k_q-1}}$$

$$\leqslant g\left(\widehat{z}\right) + \left\langle y^{k_q-1}, Ax^{k_q-1} - \widehat{z}\right\rangle + \frac{r}{2}\left\|Ax^{k_q-1} - \widehat{z}\right\|^2 + \frac{1}{2}\left\|\widehat{z} - z^{k_q-1}\right\|^2_{\mathbf{M}_2^{k_q-1}}.$$

Taking the limit superior on both sides of the above inequalities we get

$$\limsup_{q\to+\infty} g\left(z^{k_q}\right) \leqslant g\left(\widehat{z}\right),$$

which, combined with the lower semicontinuity of g, leads to

$$\lim_{q\to+\infty} g\left(z^{k_q}\right) = g\left(\widehat{z}\right).$$

Since h is continuous, we further obtain

$$\lim_{q\to+\infty} \mathcal{L}_r\left(x^{k_q}, z^{k_q}, y^{k_q}\right) = \lim_{q\to+\infty}\left[g\left(z^{k_q}\right) + h\left(x^{k_q}\right) + \left\langle y^{k_q}, Ax^{k_q} - z^{k_q}\right\rangle + \frac{r}{2}\left\|Ax^{k_q} - z^{k_q}\right\|^2\right]$$

$$= g\left(\widehat{z}\right) + h\left(\widehat{x}\right) + \left\langle \widehat{y}, A\widehat{x} - \widehat{z}\right\rangle + \frac{r}{2}\left\|A\widehat{x} - \widehat{z}\right\|^2 = \mathcal{L}_r\left(\widehat{x}, \widehat{z}, \widehat{y}\right).$$

(ii) For the sequence $\left\{d^k\right\}_{k\geqslant 0}$ defined in (53a)-(53c) we have that $d^{k_q} \in \partial\mathcal{L}_r(x^{k_q}, z^{k_q}, y^{k_q})$ for every $q \geqslant 1$ and $d^{k_q} \to 0$ as $q \to +\infty$, while $\left(x^{k_q}, z^{k_q}, y^{k_q}\right) \to (\widehat{x}, \widehat{z}, \widehat{y})$ and $\mathcal{L}_r\left(x^{k_q}, z^{k_q}, y^{k_q}\right) \to \mathcal{L}_r\left(\widehat{x}, \widehat{z}, \widehat{y}\right)$ as $q \to +\infty$. The closedness criterion of the limiting subdifferential guarantees that $0 \in \partial\mathcal{L}_r\left(\widehat{x}, \widehat{z}, \widehat{y}\right)$ or, in other words, $(\widehat{x}, \widehat{z}, \widehat{y}) \in \mathrm{crit}\left(\mathcal{L}_r\right)$. Choosing now an element $(\widehat{x}, \widehat{z}, \widehat{y}) \in \mathrm{crit}\left(\mathcal{L}_r\right)$ it holds

$$\begin{aligned}
0 &= \nabla h\left(\widehat{x}\right) + A^*\widehat{y} + rA^*\left(A\widehat{x} - \widehat{z}\right)\\
0 &\in \partial g\left(\widehat{z}\right) - \widehat{y} - r\left(A\widehat{x} - \widehat{z}\right)\\
0 &= A\widehat{x} - \widehat{z},
\end{aligned}$$

which is further equivalent to

$$-A^*\widehat{y} = \nabla h\left(\widehat{x}\right), \ \widehat{y} \in \partial g\left(\widehat{z}\right), \ \widehat{z} = A\widehat{x}.$$

(iii)-(iv) The proof follows in the lines of the proof of Theorem 5 (ii)-(iii) in [10], also by taking into consideration [10, Remark 5], according to which the properties in (iii) and (iv) are generic for sequences satisfying $\left(x^{k+1}, z^{k+1}, y^{k+1}\right) - \left(x^k, z^k, y^k\right) \to 0$ as $k \to +\infty$, which is indeed the case due to (51).

(v) The conclusion follows according to the first two statements of this theorem and of the third statement of Theorem 7. $\qquad\square$

Remark 3. An element $(\widehat{x}, \widehat{z}, \widehat{y}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ fulfilling

$$-A^*\widehat{y} = \nabla h\left(\widehat{x}\right), \ \widehat{y} \in \partial g\left(\widehat{z}\right), \ \widehat{z} = A\widehat{x}$$

is a so-called *KKT point* of the optimization problem (1). For such a KKT point we have

$$0 = A^*\partial g\left(A\widehat{x}\right) + \nabla h\left(\widehat{x}\right). \tag{63}$$

When A is injective this is further equivalent to

$$0 \in \partial(g \circ A)(\widehat{x}) + \nabla h\left(\widehat{x}\right) = \partial\left(g \circ A + h\right)\left(\widehat{x}\right), \tag{64}$$

in other words, $\widehat{x}$ is a *critical point* of the optimization problem (1).

If the functions g and h are convex, then (63) and (64) are equivalent, which means that $\widehat{x}$ is a *global optimal solution* of the optimization problem (1). In this case, $\widehat{y}$ is a *global optimal solution* of the Fenchel dual problem of (1).

By combining Lemma 9, Theorem 7 and Lemma 11, one obtains the following result.

Lemma 12. *Suppose that Assumption 1 holds true and let* $\left\{\left(x^k, z^k, y^k\right)\right\}_{k\geqslant 0}$ *be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. Let* $\Omega := \omega\left(\left\{\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right\}_{k\geqslant 1}\right)$. *The following statements are true:*

19

(i) it holds
$$\Omega \subseteq \{(\widehat{x}, \widehat{z}, \widehat{y}, \widehat{x}, \widehat{y}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \colon (\widehat{x}, \widehat{z}, \widehat{y}) \in \mathrm{crit}\,(\mathcal{L}_r)\};$$

(ii) we have
$$\lim_{k \to +\infty} \mathrm{dist}\left[\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right), \Omega\right] = 0;$$

(iii) the set Ω is nonempty, connected and compact;

(iv) the regularized augmented Lagrangian $\mathcal{F}_r$ takes on Ω the value $\mathcal{F}_* = \lim_{k \to +\infty} \mathcal{F}_k$, as the objective function $g \circ A + h$ does on the projection of the set Ω onto the space $\mathbb{R}^n$ corresponding to the first component.

2.3 Convergence analysis under Kurdyka-Łojasiewicz assumptions

In this subsection we will prove global convergence for the sequence $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ generated by the two nonconvex proximal ADMM algorithms in the context of *KL property*. The origins of this notion go back to the pioneering work of Kurdyka who introduced in [32] a general form of the Łojasiewicz inequality (see [37]). A further extension to the nonsmooth setting has been proposed and studied in [7, 8, 9].

We recall that the *distance function* of a given set $\Omega \subseteq \mathbb{R}^N$ is defined for every x by $\mathrm{dist}\,(x, \Omega) := \inf\{\|x - y\| : y \in \Omega\}$.

Definition 1. Let $\eta \in (0, +\infty]$. We denote by Φ_η the set of all concave and continuous functions $\varphi \colon [0, \eta) \to [0, +\infty)$ which satisfy the following conditions:

1. $\varphi(0) = 0$;

2. φ is $\mathcal{C}^1$ on $(0, \eta)$ and continuous at 0;

3. for all $s \in (0, \eta) : \varphi'(s) > 0$.

Definition 2. Let $\Psi \colon \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ be proper and lower semicontinuous.

1. The function Ψ is said to have the Kurdyka-Łojasiewicz (KŁ) property at a point $\widehat{u} \in \mathrm{dom}\,\partial\Psi := \left\{u \in \mathbb{R}^N \colon \partial\Psi(u) \neq \varnothing\right\}$, if there exists $\eta \in (0, +\infty]$, a neighborhood U of $\widehat{u}$ and a function $\varphi \in \Phi_\eta$ such that for every
$$u \in U \cap [\Psi(\widehat{u}) < \Psi(u) < \Psi(\widehat{u}) + \eta]$$
it holds
$$\varphi'\left(\Psi(u) - \Psi(\widehat{u})\right) \cdot \mathrm{dist}\,(\mathbf{0}, \partial\Psi(u)) \geqslant 1.$$

2. If Ψ satisfies the KŁ property at each point of $\mathrm{dom}\,\partial\Psi$, then Ψ is called KŁ function.

The functions φ belonging to the set Φ_η for $\eta \in (0, +\infty]$ are called *desingularization functions*. The KŁ property reveals the possibility to reparameterize the values of Ψ in order to avoid flatness around the critical points. To the class of KŁ functions belong semialgebraic, real subanalytic, uniformly convex functions and convex functions satisfying a growth condition. We refer the reader to [2, 3, 4, 7, 8, 9, 10] and to the references therein for more properties of KŁ functions and illustrating examples.

The following result, taken from [10, Lemma 6], will be crucial in our convergence analysis.

Lemma 13. *(Uniformized KŁ property) Let Ω be a compact set and $\Psi \colon \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. Assume that Ψ is constant on Ω and satisfies the KŁ property at each point of Ω. Then there exist $\varepsilon > 0, \eta > 0$ and $\varphi \in \Phi_\eta$ such that for every $\widehat{u} \in \Omega$ and every element u in the intersection*
$$\left\{u \in \mathbb{R}^N \colon \mathrm{dist}\,(u, \Omega) < \varepsilon\right\} \cap [\Psi(\widehat{u}) < \Psi(u) < \Psi(\widehat{u}) + \eta]$$
it holds
$$\varphi'\left(\Psi(u) - \Psi(\widehat{u})\right) \cdot \mathrm{dist}\,(\mathbf{0}, \partial\Psi(u)) \geqslant 1.$$

Working in the hypotheses of Lemma 12 we define for every $k \geqslant 1$

$$\mathcal{E}_k := \mathcal{F}\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right) - \mathcal{F}_* = \mathcal{F}_k - \mathcal{F}_* \geqslant 0,$$

where $\mathcal{F}_*$ is the limit of $\{\mathcal{F}_k\}_{k \geqslant 1}$ as $k \to +\infty$. The sequence $\{\mathcal{E}_k\}_{k \geqslant 1}$ is monotonically decreasing and it converges to 0 as $k \to +\infty$.

The next result shows that, if the regularization of the augmented Lagrangian $\mathcal{F}_r$ is a KŁ function, then the sequence $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ converges to a KKT point of the optimization problem (1).

Theorem 14. *Suppose that Assumption 1 holds true and let $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. If $\mathcal{F}_r$ is a KŁ function, then the following statements are true:*

(i) *the sequence $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ has finite length, namely,*

$$\sum_{k \geqslant 0} \left\|x^{k+1} - x^k\right\| < +\infty, \qquad \sum_{k \geqslant 0} \left\|z^{k+1} - z^k\right\| < +\infty, \qquad \sum_{k \geqslant 0} \left\|y^{k+1} - y^k\right\| < +\infty; \qquad (65)$$

(ii) *the sequence $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ converges to a KKT point of the optimization problem (1).*

Proof. As in Lemma 12, we denote by $\Omega := \omega\left(\left\{\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right\}_{k \geqslant 1}\right)$, which is a nonempty set. Let be $(\widehat{x}, \widehat{z}, \widehat{y}, \widehat{x}, \widehat{y}) \in \Omega$, thus $\mathcal{F}_r\left(\widehat{x}, \widehat{z}, \widehat{y}, \widehat{x}, \widehat{y}\right) = \mathcal{F}_*$. We have seen that $\{\mathcal{E}_k = \mathcal{F}_k - \mathcal{F}^*\}_{k \geqslant 1}$ converges to 0 as $k \to +\infty$ and will consider, consequently, two cases.

We assume first that there exists an integer $k' \geqslant 0$ such that $\mathcal{E}_{k'} = 0$ or, equivalently, $\mathcal{F}_{k'} = \mathcal{F}_*$. Due to the monotonicity of $\{\mathcal{E}_k\}_{k \geqslant 1}$ it follows that $\mathcal{E}_k = 0$ or, equivalently, $\mathcal{F}_k = \mathcal{F}_*$ for all $k \geqslant k'$. Combining the inequality in (50) with Lemma 5, it yields that $x^{k+1} - x^k = 0$ for all $k \geqslant k' + 1$. Using Lemma 5 (iii) and telescoping sum arguments, it yields $\sum_{k \geqslant 0} \left\|y^{k+1} - y^k\right\| < +\infty$. Finally, by using Lemma 5 (i), we obtain that $\sum_{k \geqslant 0} \left\|z^{k+1} - z^k\right\| < +\infty$.

Consider now the case when $\mathcal{E}_k > 0$ or, equivalently, $\mathcal{F}_k > \mathcal{F}_*$ for every $k \geqslant 1$. According to Lemma 13, there exist $\varepsilon > 0$, $\eta > 0$ and a desingularization function φ such that for every element u in the intersection

$$\{u \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \colon \mathrm{dist}\,(u, \Omega) < \varepsilon\} \cap$$
$$\{u \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \colon \mathcal{F}_* < \mathcal{F}_r\,(u) < \mathcal{F}_* + \eta\} \qquad (66)$$

it holds

$$\varphi'\left(\mathcal{F}_r\,(u) - \mathcal{F}_*\right) \cdot \mathrm{dist}\,(\mathbf{0}, \partial \mathcal{F}_r\,(u)) \geqslant 1.$$

Let be $k_1 \geqslant 1$ such that for every $k \geqslant k_1$

$$\mathcal{F}_* < \mathcal{F}_k < \mathcal{F}_* + \eta.$$

Since $\lim_{k \to +\infty} \mathrm{dist}\left[\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right), \Omega\right] = 0$, see Lemma 12 (ii), there exists $k_2 \geqslant 1$ such that for every $k \geqslant k_2$

$$\mathrm{dist}\left[\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right), \Omega\right] < \varepsilon.$$

Thus, $\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)$ belongs to the intersection in (66) for every $k \geqslant k_0 := \max\{k_1, k_2, 3\}$, which further implies

$$\varphi'\left(\mathcal{F}_k - \mathcal{F}_*\right) \cdot \mathrm{dist}\left(\mathbf{0}, \partial \mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right) = \varphi'\left(\mathcal{E}_k\right) \cdot \mathrm{dist}\left(\mathbf{0}, \partial \mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right) \geqslant 1. \qquad (67)$$

Define for two arbitrary nonnegative integers p and q

$$\Delta_{p,q} := \varphi\left(\mathcal{F}_p - \mathcal{F}_*\right) - \varphi\left(\mathcal{F}_q - \mathcal{F}_*\right) = \varphi\left(\mathcal{E}_p\right) - \varphi\left(\mathcal{E}_q\right).$$

For every $K \geqslant k_0 \geqslant 1$ it holds

$$\sum_{k=k_0}^{K} \Delta_{k,k+1} = \Delta_{k_0, K+1} = \varphi\left(\mathcal{E}_{k_0}\right) - \varphi\left(\mathcal{E}_{K+1}\right) \leqslant \varphi\left(\mathcal{E}_{k_0}\right),$$

21

from which we get $\sum_{k \geqslant 1} \Delta_{k,k+1} < +\infty$.

By combining Theorem 7 (i) with the concavity of φ we obtain for every $k \geqslant 1$

$$\Delta_{k,k+1} = \varphi\left(\mathcal{E}_k\right) - \varphi\left(\mathcal{E}_{k+1}\right) \geqslant \varphi'\left(\mathcal{E}_k\right)\left[\mathcal{E}_k - \mathcal{E}_{k+1}\right] = \varphi'\left(\mathcal{E}_k\right)\left[\mathcal{F}_k - \mathcal{F}_{k+1}\right] \geqslant \varphi'\left(\mathcal{E}_k\right)\frac{C_1}{4}\left\|x^{k+1} - x^k\right\|^2. \tag{68}$$

The last relation combined with (67) imply

$$\begin{aligned}
\left\|x^{k+1} - x^k\right\|^2 &\leqslant \varphi'\left(\mathcal{E}_k\right) \cdot \operatorname{dist}\left(\mathbf{0}, \partial\mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right)\left\|x^{k+1} - x^k\right\|^2 \\
&\leqslant \frac{4}{C_1}\Delta_{k,k+1} \cdot \operatorname{dist}\left(\mathbf{0}, \partial\mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right) \quad \forall k \geqslant k_0.
\end{aligned}$$

By the arithmetic mean-geometric mean inequality and Corollary 10 we have that for every $k \geqslant k_0$ and every $\beta > 0$

$$\begin{aligned}
\left\|x^{k+1} - x^k\right\| &\leqslant \sqrt{\frac{4}{C_1}\Delta_{k,k+1} \cdot \operatorname{dist}\left(\mathbf{0}, \partial\mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right)} \\
&\leqslant \frac{\beta}{C_1}\Delta_{k,k+1} + \frac{1}{\beta}\operatorname{dist}\left(\mathbf{0}, \partial\mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right) \\
&\leqslant \frac{\beta}{C_1}\Delta_{k,k+1} + \frac{C_{11}}{\beta}\left(\left\|x^k - x^{k-1}\right\| + \left\|x^{k-1} - x^{k-2}\right\| + \left\|x^{k-2} - x^{k-3}\right\|\right) \\
&\quad + \frac{C_{12}}{\beta}\left(\left\|A^*\left(y^{k-1} - y^{k-2}\right)\right\| - \left\|A^*\left(y^k - y^{k-1}\right)\right\|\right) \\
&\quad + \frac{C_{13}}{\beta}\left(\left\|A^*\left(y^{k-2} - y^{k-3}\right)\right\| - \left\|A^*\left(y^{k-1} - y^{k-2}\right)\right\|\right). \tag{69}
\end{aligned}$$

We denote for every $k \geqslant 3$

$$\begin{aligned}
a^k &:= \left\|x^k - x^{k-1}\right\| \geqslant 0, \\
\delta_k &:= \frac{\beta}{C_1}\Delta_{k,k+1} + \frac{C_{12}}{\beta}\left(\left\|A^*\left(y^{k-1} - y^{k-2}\right)\right\| - \left\|A^*\left(y^k - y^{k-1}\right)\right\|\right) \\
&\quad + \frac{C_{13}}{\beta}\left(\left\|A^*\left(y^{k-2} - y^{k-3}\right)\right\| - \left\|A^*\left(y^{k-1} - y^{k-2}\right)\right\|\right).
\end{aligned}$$

The inequality (69) is nothing than (9) with $c_0 = c_1 = c_2 := \dfrac{C_{11}}{\beta}$. Observe that for every $K \geqslant k_0$ we have

$$\sum_{k=k_0}^K \delta_k \leqslant \frac{\beta}{C_1}\varphi\left(\mathcal{E}_{k_0}\right) + \frac{C_{12}}{\beta}\left\|A^*\left(y^{k_0-1} - y^{k_0-2}\right)\right\| + \frac{C_{13}}{\beta}\left\|A^*\left(y^{k_0-2} - y^{k_0-3}\right)\right\|$$

and thus, by choosing $\beta > 3C_{11}$, we can use Lemma 3 to conclude that

$$\sum_{k \geqslant 0}\left\|x^{k+1} - x^k\right\| < +\infty.$$

The other two statements in (65) follow from Lemma 5. This means that the sequence $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ is Cauchy, thus it converges to an element $(\hat{x}, \hat{z}, \hat{y})$ which is, according to Lemmas 11, a KKT point of the optimization problem (1). $\qquad\square$

Remark 4. The function $\mathcal{F}_r$ is a KL function if, for instance, the objective function of (1) is semi-algebraic, which is the case when the functions g and h are semi-algebraic.

3 Convergence rates under Łojasiewicz assumptions

In this section we derive convergence rates for the sequence $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ generated by Algorithm 1 or Algorithm 2 as well as for the regularized augmented Lagrangian function $\mathcal{F}_r$ along this sequence, provided that the latter satisfies the Łojasiewicz property.

3.1 Łojasiewicz property and a technical lemma

We recall the following definition from [2] (see, also, [37]).

Definition 3. Let $\Psi\colon \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ be proper and lower semicontinuous. Then Ψ satisfies the Łojasiewicz property if for any critical point $\widehat{u}$ of Ψ, there exists $C_L > 0$, $\theta \in [0,1)$ and $\varepsilon > 0$ such that

$$\left| \Psi\left(u\right) - \Psi\left(\widehat{u}\right)\right|^{\theta} \leqslant C_L \cdot \operatorname{dist}(0, \partial\Psi(u)) \ \forall u \in \operatorname{Ball}\left(\widehat{u}, \varepsilon\right),$$

where $\operatorname{Ball}\left(\widehat{u}, \varepsilon\right)$ denotes the open ball with centre $\widehat{u}$ and radius ε.

If Assumption 1 is fulfilled and $\left\{\left(x^k, z^k, y^k\right)\right\}_{k\geqslant 0}$ is the sequence generated by Algorithm 1 or Algorithm 2, assumed to be bounded, then, as seen in Lemma 12, the set of cluster points $\Omega = \omega\left(\left\{\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right\}_{k\geqslant 0}\right)$ is nonempty, compact and connected and $\mathcal{F}_r$ takes on Ω the value $\mathcal{F}_*$; in addition, for every $(\widehat{x}, \widehat{z}, \widehat{y}, \widehat{x}, \widehat{y}) \in \Omega$, $(\widehat{x}, \widehat{z}, \widehat{y})$ belongs to $\operatorname{crit}(\mathcal{L}_r)$. According to [2, Lemma 1], if $\mathcal{F}_r$ has the Łojasiewicz property, then there exist $C_L > 0$, $\theta \in [0,1)$ and $\varepsilon > 0$ such that for every

$$\left(x, z, y, x', y'\right) \in \{u \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \colon \operatorname{dist}\left(u, \Omega\right) < \varepsilon\}$$

it holds

$$\left| \mathcal{F}_r\left(x, z, y, x', y'\right) - \mathcal{F}_*\right|^{\theta} \leqslant C_L \cdot \operatorname{dist}\left(0, \partial\mathcal{F}_r\left(x, z, y, x', y'\right)\right).$$

Obviously, $\mathcal{F}_r$ is a KŁ function with desingularization function $\varphi : [0, +\infty) \to [0, +\infty)$, $\varphi\left(s\right) := \frac{1}{1-\theta}C_L s^{1-\theta}$, which, according to Theorem 14, means that Ω contains a single element $(\widehat{x}, \widehat{z}, \widehat{y}, \widehat{x}, \widehat{y})$, namely, the limit of $\left\{\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right\}_{k\geqslant 0}$ as $k \to +\infty$. In other words, if $\mathcal{F}_r$ has the Łojasiewicz property, then there exist $C_L > 0$, $\theta \in [0,1)$ and $\varepsilon > 0$ such that

$$\left| \mathcal{F}_r\left(x, z, y, x', y'\right) - \mathcal{F}_*\right|^{\theta} \leqslant C_L \cdot \operatorname{dist}\left(0, \partial\mathcal{F}_r\left(x, z, y, x', y'\right)\right) \ \forall \left(x, z, y, x', y'\right) \in \operatorname{Ball}\left((\widehat{x}, \widehat{z}, \widehat{y}, \widehat{x}, \widehat{y}), \varepsilon\right). \tag{70}$$

In this case, $\mathcal{F}_r$ is said to satisfy the Łojasiewicz property with *Lojasiewicz constant* $C_L > 0$ and *Lojasiewicz exponent* $\theta \in [0,1)$.

The following lemma will convergence rates for a particular class of monotonically decreasing sequences converging to 0.

Lemma 15. *Let $\{e_k\}_{k\geqslant 0}$ be a monotonically decreasing sequence in $\mathbb{R}_+$ converging 0. Assume further that there exists natural numbers $k_0 \geqslant l_0 \geqslant 1$ such that for every $k \geqslant k_0$*

$$e_{k-l_0} - e_k \geqslant C_e e_k^{2\theta}, \tag{71}$$

where $C_e > 0$ is some constant and $\theta \in [0,1)$. Then following statements are true:

   (i) if $\theta = 0$, then $\{e_k\}_{k\geqslant 0}$ converges in finite time;

   (ii) if $\theta \in (0, 1/2]$, then there exists $C_{e,0} > 0$ and $Q \in [0,1)$ such that for every $k \geqslant k_0$

$$0 \leqslant e_k \leqslant C_{e,0} Q^k;$$

   (iii) if $\theta \in (1/2, 1)$, then there exists $C_{e,1} > 0$ such that for every $k \geqslant k_0 + l_0$

$$0 \leqslant e_k \leqslant C_{e,1}\left(k - l_0 + 1\right)^{-\frac{1}{2\theta-1}}.$$

Proof. Fix an integer $k \geqslant k_0$. Since $k_0 \geqslant l_0 \geqslant 0$, the recurrence inequality (71) is well defined for every $k \geqslant k_0$.

(i) The case when $\theta = 0$. We assume that $e_k > 0$ for every $k \geqslant 0$. From (71) we get

$$e_{k-l_0} - e_k \geqslant C_e > 0$$

for every $k \geqslant k_0$, which actually contradicts the fact that $\{e_k\}_{k\geqslant 0}$ converges to 0 as $k \to +\infty$. Consequently, there exists $k' \geqslant 0$ such that $e_{k'} = 0$ for every $k \geqslant k'$ and thus the conclusion follows.

For the proof of (ii) and (iii) we can assume that $e_k > 0$ for every $k \geqslant 0$. Otherwise, as $\{e_k\}_{k\geqslant 0}$ is monotonically decreasing and convergent to 0, the sequence is constant beginning with a given index, which means that both statements are true.

(ii) The case when $\theta \in (0, 1/2]$. We have $e_k \leqslant e_0$, which leads to $e_{k-l_0} - e_k \geqslant C_e e_k^{2\theta} \geqslant C_e e_0^{2\theta-1} e_k$ for every $k \geqslant k_0$. Therefore,

$$e_k \leqslant \left(\frac{1}{C_e e_0^{2\theta-1} + 1} \right)^{\frac{k}{l_0} - \frac{k_0}{l_0} - 1} e_0 = e_0 \left(C_e e_0^{2\theta-1} + 1 \right)^{\frac{k_0}{l_0} + 1} \left(\frac{1}{\sqrt[l_0]{C_e e_0^{2\theta-1} + 1}} \right)^k.$$

(iii) The case when $\theta \in (1/2, 1)$. From (71) we get

$$C_e \leqslant (e_{k-l_0} - e_k) e_k^{-2\theta}. \tag{72}$$

Define $\zeta \colon (0, +\infty) \to \mathbb{R}, \zeta(s) = s^{-2\theta}$. We have that

$$\frac{d}{ds} \left(\frac{1}{1 - 2\theta} s^{1-2\theta} \right) = s^{-2\theta} = \zeta(s) \text{ and } \zeta'(s) = -2\theta s^{-2\theta-1} < 0 \; \forall s \in (0, +\infty).$$

Consequently, $\zeta(e_{k-l_0}) \leqslant \zeta(s)$ for all $s \in [e_k, e_{k-l_0}]$.

- *Assume that $\zeta(e_k) \leqslant 2\zeta(e_{k-l_0})$.* Then (72) gives

$$C_e \leqslant 2\zeta(e_{k-l_0}) \int_{e_k}^{e_{k-l_0}} 1 ds \leqslant 2 \int_{e_k}^{e_{k-l_0}} \zeta(s) \, ds = \frac{2}{2\theta - 1} \left(e_k^{1-2\theta} - e_{k-l_0}^{1-2\theta} \right)$$

  or, equivalently,

$$e_k^{1-2\theta} - e_{k-l_0}^{1-2\theta} \geqslant C_1', \quad \text{where } C_1' := \frac{(2\theta - 1) C_e}{2} > 0. \tag{73}$$

- *Assume that $\zeta(e_k) > 2\zeta(e_{k-l_0})$.* For $\nu := 2^{-\frac{1}{2\theta}} \in (0, 1)$ this is equivalent to $\left(\nu^{1-2\theta} - 1 \right) e_{k-l_0}^{1-2\theta} \leqslant e_k^{1-2\theta} - e_{k-l_0}^{1-2\theta}$, thus,

$$e_k^{1-2\theta} - e_{k-l_0}^{1-2\theta} \geqslant \left(\nu^{1-2\theta} - 1 \right) e_{k-l_0}^{1-2\theta} \geqslant C_2', \quad \text{where } C_2' := \left(\nu^{1-2\theta} - 1 \right) e_0^{2\theta-1} > 0. \tag{74}$$

In both situations we get for every $i \geqslant k_0$

$$e_i^{1-2\theta} - e_{i-l_0}^{1-2\theta} \geqslant C' := \min\left\{ C_1', C_2' \right\} > 0, \tag{75}$$

where C_1' and C_2' are defined as in (73) and (74), respectively. For every $k \geqslant k_0 + 2l_0$, by summing up the inequalities (75) for $i = k_0 + l_0, \cdots, k$, we get

$$\sum_{j=0}^{l_0-1} \left(e_{k-j}^{1-2\theta} - e_{k_0+j}^{1-2\theta} \right) \geqslant (k - k_0 - l_0 + 1) C' > 0.$$

Since

$$l_0 \left(e_k^{1-2\theta} - e_{k_0}^{1-2\theta} \right) \geqslant \sum_{j=0}^{l_0-1} \left(e_{k-j}^{1-2\theta} - e_{k_0+j}^{1-2\theta} \right) \geqslant (k - k_0 - l_0 + 1) C',$$

we have

$$e_k^{1-2\theta} \geqslant e_{k_0}^{1-2\theta} + \frac{k - k_0 - l_0 + 1}{l_0} C'. \tag{76}$$

We obtain from (75) that

$$e_{k_0}^{1-2\theta} \geqslant \left\lfloor \frac{k_0 + l_0}{l_0} \right\rfloor C' \geqslant \left(\frac{k_0 + l_0}{l_0} - 1 \right) C' = \frac{k_0}{l_0} C', \tag{77}$$

where $\lfloor p \rfloor$ denotes the greatest integer that is less than or equal to the real number p. By plugging (77) into (76) we obtain

$$e_k^{1-2\theta} \geqslant \frac{k - l_0 + 1}{l_0} C',$$

which implies

$$e_k \leqslant \left(\frac{C'}{l_0} \right)^{-\frac{1}{2\theta-1}} (k - l_0 + 1)^{-\frac{1}{2\theta-1}}. \tag{78}$$

This concludes the proof. $\qquad\square$

Remark 5. The inequality in Lemma 15 (iii) can be writen for k large enough in terms of k instead of $k - l_0 + 1$. If, for instance, $k \geqslant 2(l_0 + 1)$, then $k - l_0 + 1 \geqslant \frac{1}{2}k$ and thus from (78) we get

$$
e_k \leqslant \left(\frac{C'}{l_0}\right)^{-\frac{1}{2\theta-1}} (k - l_0 + 1)^{-\frac{1}{2\theta-1}} \leqslant \left(\frac{C'}{2l_0}\right)^{-\frac{1}{2\theta-1}} k^{-\frac{1}{2\theta-1}}.
$$

3.2 Convergence rates

In this subsection we will address convergence rates for Algorithm 1 and Algorithm 2 in the context of an assumption which is slightly more restricitve than Assumption 1.

Assumption 2. *We work in the hypotheses of Assumption 1 except for* (15) *which is replaced by*

$$
2\mathbf{M}_1^k + rA^*A \succcurlyeq \left(L + \frac{C'_\mathbf{M}}{r}\right)\mathbf{Id} \quad \forall k \geqslant 0, \tag{79}
$$

where

$$
C'_\mathbf{M} := \begin{cases} \left(10\mu_1^2 + 8(L + \mu_1)^2\right)T_0, & \text{for Algorithm 1}, \\ \left(8\mu_1^2 + 10(L + \mu_1)^2\right)T_0, & \text{for Algorithm 2}. \end{cases}
$$

The condition (79) is nothing else than (15) after replacing $C_\mathbf{M}$ by the bigger constant $C'_\mathbf{M}$.

The examples in Remark 2 can be all adapted to the new setting and one can provide different settings which guarantee Assumption 2. The scenarios which ensure Assumption 2 evidently satisfy Assumption 1, too, therefore the results investigated in Section 2 remain valid in this setting. As follows we will provide improvements of the statements used in the convergence analysis which follow thanks to Assumption 2.

Lemma 16. *Suppose that Assumption 2 holds true and let* $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ *be a sequence generated by Algorithm 1 or Algorithm 2. Then for every* $k \geqslant 1$ *it holds*

$$
\mathcal{L}_r\left(x^{k+1}, z^{k+1}, y^{k+1}\right) + 2T_1 \left\|A^*\left(y^{k+1} - y^k\right)\right\|^2 + \frac{1}{2}\left\|x^{k+1} - x^k\right\|_{\mathbf{M}_3^k}^2 + \frac{1}{2}\left\|z^{k+1} - z^k\right\|_{\mathbf{M}_2^k}^2
$$
$$
+ \frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2
$$
$$
\leqslant \mathcal{L}_r\left(x^k, z^k, y^k\right) + 2T_1\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 + C_1\left\|x^k - x^{k-1}\right\|^2. \tag{80}
$$

Proof. Let $k \geqslant 1$ be fixed. By the same arguments as in Lemma 4, we have that (see (17))

$$
\mathcal{L}_r\left(x^{k+1}, z^{k+1}, y^{k+1}\right) + \frac{1}{2}\left\|x^{k+1} - x^k\right\|_{2\mathbf{M}_1^k + rA^*A}^2 - \frac{L}{2}\left\|x^{k+1} - x^k\right\|^2 + \frac{1}{2}\left\|z^{k+1} - z^k\right\|_{\mathbf{M}_2^k}^2
$$
$$
\leqslant \mathcal{L}_r\left(x^k, z^k, y^k\right) + \frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2. \tag{81}
$$

From (26), (28), (36) and (38) it follows that

$$
\frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2 \leqslant \frac{C_0 - L}{2}\left\|x^{k+1} - x^k\right\|^2 + \frac{C_1}{2}\left\|x^k - x^{k-1}\right\|^2 +
$$
$$
T_1\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 - T_1\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2. \tag{82}
$$

By multiplying (82) by 2 and by adding the resulting inequality to (81) we obtain (80). □

We replace T_1 with $2T_1$ in the definition of the regularized augmented Lagrangian $\mathcal{F}_r$, thus, the sequence $\{\mathcal{F}_k\}_{k \geqslant 1}$ in (47) becomes

$$
\mathcal{F}_k := \mathcal{L}_r\left(x^k, z^k, y^k\right) + 2T_1\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 + C_1\left\|x^k - x^{k-1}\right\|^2 \quad \forall k \geqslant 1.
$$

In this new context the inequality (80) reads for every $k \geqslant 1$

$$
\mathcal{F}_{k+1} + \frac{C_1}{4}\left\|x^{k+1} - x^k\right\|^2 + \frac{1}{2}\left\|z^{k+1} - z^k\right\|_{\mathbf{M}_2^k}^2 + \frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2 \leqslant \mathcal{F}_k \tag{83}
$$

and provides an inequality which is tighter than relation (50) in Theorem 7. Furthermore, for a subgradient D^{k+1} of $\mathcal{F}_r$ at $(x^{k+1}, z^{k+1}, y^{k+1}, x^k, z^k)$ defined as in (59) (again by replacing T_1 by $2T_1$) we obtain for every $k \geqslant 2$ the following estimate, which is simpler than (62) in Corollary 10

$$\left\| \! \left\| D^{k+1} \right\| \! \right\| \leqslant C_{14} \left\| x^{k+1} - x^k \right\| + C_{15} \left\| y^{k+1} - y^k \right\| + C_{16} \left\| y^k - y^{k-1} \right\|,$$

where

$$C_{14} := C_8 + C_9 \left\| A \right\|, \quad C_{15} := C_{10} + \frac{C_9}{\rho r}, \quad C_{16} := \frac{C_9}{\rho r}.$$

This improvement provides, instead of inequality (68) in the proof of Theorem 14, the following very useful estimate

$$\Delta_{k,k+1} = \varphi(\mathcal{E}_k) - \varphi(\mathcal{E}_{k+1}) \geqslant \varphi'(\mathcal{E}_k) \min\left\{ \frac{C_1}{4}, \frac{1}{\rho r} \right\} \left(\left\| x^{k+1} - x^k \right\|^2 + \left\| y^{k+1} - y^k \right\|^2 \right)$$

$$\geqslant C_{17} \varphi'(\mathcal{E}_k) \left(\left\| x^{k+1} - x^k \right\| + \left\| y^{k+1} - y^k \right\| \right)^2,$$

where

$$C_{17} := \frac{1}{2} \min\left\{ \frac{C_1}{4}, \frac{1}{\rho r} \right\}.$$

The last relation together with (67) imply that for every $k \geqslant k_0$

$$\left(\left\| x^{k+1} - x^k \right\| + \left\| y^{k+1} - y^k \right\| \right)^2 \leqslant \frac{\Delta_{k,k+1}}{C_{17}} \cdot \mathrm{dist}\left(\mathbf{0}, \partial \mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1} \right) \right)$$

and from here, for arbitrary $\beta > 0$,

$$\begin{aligned} &\left\| x^{k+1} - x^k \right\| + \left\| y^{k+1} - y^k \right\| \\ &\leqslant \frac{\beta \Delta_{k,k+1}}{4C_{17}} + \frac{\max\{C_{14}, C_{15}\}}{\beta} \left(\left\| x^k - x^{k-1} \right\| + \left\| y^k - y^{k-1} \right\| + \left\| y^{k-1} - y^{k-2} \right\| \right). \end{aligned} \tag{84}$$

By denoting

$$a^k := \left(\left\| x^k - x^{k-1} \right\|, \left\| y^k - y^{k-1} \right\| \right) \in \mathbb{R}_+^2 \quad \text{and} \quad \delta_k := \frac{\beta \Delta_{k,k+1}}{4C_{17}},$$

inequality (84) can be rewritten for every $k \geqslant k_0$ as

$$\left\langle \mathbb{1}, a^{k+1} \right\rangle \leqslant \left\langle c_0, a^k \right\rangle + \left\langle c_1, a^{k-1} \right\rangle + \delta_k, \tag{85}$$

where

$$c_0 := \frac{\max\{C_{14}, C_{15}\}}{\beta} (1, 1) \quad \text{and} \quad c_1 := \frac{\max\{C_{14}, C_{15}\}}{\beta} (0, 1).$$

Choosing $\beta > 2 \max\{C_{14}, C_{15}\}$, Lemma 3 and Lemma 5 imply that $\left\{ (x^k, z^k, y^k) \right\}_{k \geqslant 0}$ has finite length (see (65)).

Next we prove a recurrence inequality for the sequence $\{\mathcal{E}_k\}_{k \geqslant 0}$.

Lemma 17. *Suppose that Assumption 2 holds true and let* $\left\{ (x^k, z^k, y^k) \right\}_{k \geqslant 0}$ *be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. If $\mathcal{F}_r$ satisfies the Łojasiewicz property with Łojasiewicz constant $C_L > 0$ and Łojasiewicz exponent $\theta \in [0, 1)$, then there exists $k_0 \geqslant 1$ such that the following estimate holds for every $k \geqslant k_0$*

$$\mathcal{E}_{k-1} - \mathcal{E}_{k+1} \geqslant C_{19} \mathcal{E}_{k+1}^{2\theta}, \quad \text{where} \quad C_{19} := \frac{\min\left\{ \dfrac{C_1}{4}, \dfrac{1}{\rho r} \right\}}{3 C_L^2 \max\{C_{14}, C_{15}\}^2}. \tag{86}$$

Proof. For every $k \geqslant 2$ we obtain from (83)

$$\begin{aligned} \mathcal{E}_{k-1} - \mathcal{E}_{k+1} &= \mathcal{F}_{k-1} - \mathcal{F}_k + \mathcal{F}_k - \mathcal{F}_{k+1} \\ &\geqslant \min\left\{ \frac{C_1}{4}, \frac{1}{\rho r} \right\} \left(\left\| x^{k+1} - x^k \right\|^2 + \left\| y^{k+1} - y^k \right\|^2 + \left\| y^k - y^{k-1} \right\|^2 \right) \\ &\geqslant \frac{1}{3} \min\left\{ \frac{C_1}{4}, \frac{1}{\rho r} \right\} \left(\left\| x^{k+1} - x^k \right\| + \left\| y^{k+1} - y^k \right\| + \left\| y^k - y^{k-1} \right\| \right)^2 \\ &\geqslant C_{19} C_L^2 \left\| \! \left\| D^{k+1} \right\| \! \right\|^2. \end{aligned}$$

Let $\varepsilon > 0$ be such that (70) is fulfilled and choose $k_0 \geqslant 1$ such that $(x^{k+1}, z^{k+1}, y^{k+1})$ belongs to $\mathrm{Ball}\left((\hat{x}, \hat{z}, \hat{y}), \varepsilon \right)$ for every $k \geqslant k_0$. Then (70) implies (86) for every $k \geqslant k_0$. $\qquad \square$

The following convergence rates follow by combining Lemma 15 with Lemma 17.

Theorem 18. *Suppose that Assumption 2 holds true and let $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. If $\mathcal{F}_r$ satisfies the Lojasiewicz property with Lojasiewicz constant $C_L > 0$ and Lojasiewicz exponent $\theta \in [0, 1)$, then the following statements are true:*

(i) if $\theta = 0$, then $\{\mathcal{F}_k\}_{k \geqslant 1}$ converges in finite time;

(ii) if $\theta \in (0, 1/2]$, then there exist $k_0 \geqslant 1$, $\widehat{C}_0 > 0$ and $Q \in [0, 1)$ such that for every $k \geqslant k_0$

$$0 \leqslant \mathcal{F}_k - \mathcal{F}_* \leqslant \widehat{C}_0 Q^k;$$

(iii) if $\theta \in (1/2, 1)$, then there exist $k_0 \geqslant 3$ and $\widehat{C}_1 > 0$ such that for every $k \geqslant k_0$

$$0 \leqslant \mathcal{F}_k - \mathcal{F}_* \leqslant \widehat{C}_1 \left(k - 1\right)^{-\frac{1}{2\theta - 1}}.$$

The next lemma will play an importat role when transferring the convergence rates for $\{\mathcal{F}_k\}_{k \geqslant 0}$ to the sequence of iterates $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ (see [25] for a similar statement).

Lemma 19. *Suppose that Assumption 2 holds true and let $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. Suppose further that $\mathcal{F}_r$ satisfies the Lojasiewicz property with Lojasiewicz constant $C_L > 0$, Lojasiewicz exponent $\theta \in [0, 1)$ and desingularization function $\varphi : [0, +\infty) \to [0, +\infty), \varphi(s) := \dfrac{1}{1 - \theta} C_L s^{1 - \theta}$. Let $(\widehat{x}, \widehat{z}, \widehat{y})$ be the KKT point of the optimization problem (1) to which $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ converges as $k \to +\infty$. Then there exists $k_0 \geqslant 2$ such that the following estimates hold for every $k \geqslant k_0$*

$$\left\| x^k - \widehat{x} \right\| \leqslant C_{20} \max\left\{\sqrt{\mathcal{E}_k}, \varphi\left(\mathcal{E}_k\right)\right\}, \quad \text{where} \quad C_{20} := \frac{7}{\sqrt{C_{17}}} + \frac{1}{C_{17}}, \tag{87a}$$

$$\left\| y^k - \widehat{y} \right\| \leqslant C_{21} \max\left\{\sqrt{\mathcal{E}_k}, \varphi\left(\mathcal{E}_k\right)\right\}, \quad \text{where} \quad C_{21} := \frac{7}{2\sqrt{C_{17}}} + \frac{1}{2C_{17}}, \tag{87b}$$

$$\left\| z^k - \widehat{z} \right\| \leqslant C_{22} \max\left\{\sqrt{\mathcal{E}_{k-1}}, \varphi\left(\mathcal{E}_{k-1}\right)\right\}, \quad \text{where} \quad C_{22} := C_{20} \|A\| + \frac{2C_{21}}{\rho r}. \tag{87c}$$

Proof. We assume that $\mathcal{E}_k > 0$ for every $k \geqslant 0$. Otherwise, beginning with a given index, the sequence $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ becomes identical to $(\widehat{x}, \widehat{z}, \widehat{y})$ and the conclusion follows as in the proof of Theorem 14. Let $\varepsilon > 0$ be such that (70) is fulfilled and $k_0 \geqslant 2$ such that $\left(x^{k+1}, z^{k+1}, y^{k+1}\right)$ belongs to $\mathrm{Ball}\left((\widehat{x}, \widehat{z}, \widehat{y}), \varepsilon\right)$ for every $k \geqslant k_0$. We fix $k \geqslant k_0$. One can easily notice that

$$\left\| x^k - \widehat{x} \right\| \leqslant \left\| x^{k+1} - x^k \right\| + \left\| x^{k+1} - \widehat{x} \right\| \leqslant \cdots \leqslant \sum_{l \geqslant k} \left\| x^{l+1} - x^l \right\| \tag{88a}$$

and, similarly,

$$\left\| z^k - \widehat{z} \right\| \leqslant \sum_{l \geqslant k} \left\| z^{l+1} - z^l \right\| \quad \text{and} \quad \left\| y^k - \widehat{y} \right\| \leqslant \sum_{l \geqslant k} \left\| y^{l+1} - y^l \right\|. \tag{88b}$$

Recall that the inequality (84) can be rewritten as (85). For $\beta := 3 \max\{C_{14}, C_{15}\} > 2 \max\{C_{14}, C_{15}\}$, thanks to Lemma 3 and the estimate (83), we have that

$$\sum_{l \geqslant k} \left\| x^{l+1} - x^l \right\| = \sum_{l \geqslant k} a_1^{l+1} = \sum_{l \geqslant k+1} a_1^l$$

$$\leqslant \left\| x^{k+1} - x^k \right\| + 2 \left\| x^{k+2} - x^{k+1} \right\| + 3 \left\| x^{k+3} - x^{k+2} \right\| + 2 \left\| y^{k+1} - y^k \right\|$$

$$+ 2 \left\| y^{k+2} - y^{k+1} \right\| + 3 \left\| y^{k+3} - y^{k+2} \right\| + \frac{\varphi\left(\mathcal{E}_k\right)}{C_{17}}$$

$$\leqslant \frac{2}{\sqrt{C_{17}}} \sqrt{\mathcal{F}_k - \mathcal{F}_{k+1}} + \frac{2}{\sqrt{C_{17}}} \sqrt{\mathcal{F}_{k+1} - \mathcal{F}_{k+2}} + \frac{3}{\sqrt{C_{17}}} \sqrt{\mathcal{F}_{k+2} - \mathcal{F}_{k+3}} + \frac{\varphi\left(\mathcal{E}_k\right)}{C_{17}}$$

$$\leqslant \frac{2}{\sqrt{C_{17}}} \sqrt{\mathcal{E}_k} + \frac{2}{\sqrt{C_{17}}} \sqrt{\mathcal{E}_{k+1}} + \frac{3}{\sqrt{C_{17}}} \sqrt{\mathcal{E}_{k+2}} + \frac{\varphi\left(\mathcal{E}_k\right)}{C_{17}}$$

and, similarly,

$$\sum_{l \geqslant k} \left\| y^{l+1} - y^l \right\| \leqslant \frac{1}{\sqrt{C_{17}}} \sqrt{\mathcal{E}_k} + \frac{1}{\sqrt{C_{17}}} \sqrt{\mathcal{E}_{k+1}} + \frac{3}{2\sqrt{C_{17}}} \sqrt{\mathcal{E}_{k+2}} + \frac{\varphi(\mathcal{E}_k)}{2C_{17}}.$$

By taking into account the relations above, (88a)-(88b) as well as

$$\sqrt{\mathcal{E}_{k+2}} \leqslant \sqrt{\mathcal{E}_{k+1}} \leqslant \sqrt{\mathcal{E}_k} \qquad \text{and} \qquad \varphi(\mathcal{E}_{k+1}) \leqslant \varphi(\mathcal{E}_k) \ \forall k \geqslant 1,$$

the estimates (87a) and (87b) follow. Statement (87c) follows from Lemma 5 and by considering (88b). $\qquad\square$

We provide now convergence rates for the sequence $\left\{ (x^k, z^k, y^k) \right\}_{k \geqslant 0}$.

Theorem 20. *Suppose that Assumption 2 holds true and let $\left\{ (x^k, z^k, y^k) \right\}_{k \geqslant 0}$ be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. Suppose further that $\mathcal{F}_r$ satisfies the Łojasiewicz property with Łojasiewicz constant $C_L > 0$ and Łojasiewicz exponent $\theta \in [0, 1)$. Let $(\widehat{x}, \widehat{z}, \widehat{y})$ be the KKT point of the optimization problem (1) to which $\left\{ (x^k, z^k, y^k) \right\}_{k \geqslant 0}$ converges as $k \to +\infty$. Then the following statements are true:*

(i) if $\theta = 0$, then the algorithms converge in finite time;

(ii) if $\theta \in (0, 1/2]$, then there exist $k_0 \geqslant 1$, $\widehat{C}_{0,1}, \widehat{C}_{0,2}, \widehat{C}_{0,3} > 0$ and $\widehat{Q} \in [0, 1)$ such that for every $k \geqslant k_0$

$$\left\| x^k - \widehat{x} \right\| \leqslant \widehat{C}_{0,1} \widehat{Q}^k, \quad \left\| y^k - \widehat{y} \right\| \leqslant \widehat{C}_{0,2} \widehat{Q}^k, \quad \left\| z^k - \widehat{z} \right\| \leqslant \widehat{C}_{0,3} \widehat{Q}^k;$$

(iii) if $\theta \in (1/2, 1)$, then there exist $k_0 \geqslant 3$ and $\widehat{C}_{1,1}, \widehat{C}_{1,2}, \widehat{C}_{1,3} > 0$ such that for every $k \geqslant k_0$

$$\left\| x^k - \widehat{x} \right\| \leqslant \widehat{C}_{1,1} (k-1)^{-\frac{1-\theta}{2\theta-1}}, \quad \left\| y^k - \widehat{y} \right\| \leqslant \widehat{C}_{1,2} (k-1)^{-\frac{1-\theta}{2\theta-1}}, \quad \left\| z^k - \widehat{z} \right\| \leqslant \widehat{C}_{1,3} (k-2)^{-\frac{1-\theta}{2\theta-1}}.$$

Proof. By denoting $\varphi : [0, +\infty) \to [0, +\infty), \varphi(s) := \frac{1}{1-\theta} C_L s^{1-\theta}$, the desingularization function, there exist $k_0' \geqslant 2$ such that for every $k \geqslant k_0'$ the inequalities (87a)-(87c) in Lemma 19 and $\mathcal{E}_k \leqslant \left(\frac{1}{1-\theta} C_L \right)^{\frac{2}{2\theta-1}}$ hold.

(i) If $\theta = 0$, then $\{\mathcal{F}_k\}_{k \geqslant 1}$ converges in finite time. According to (83), the sequences $\left\{ (x^k) \right\}_{k \geqslant 0}$ and $\left\{ (y^k) \right\}_{k \geqslant 0}$ converge also in finite time. Further, by Lemma 5, it follows that $\left\{ (z^k) \right\}_{k \geqslant 0}$ converges in finite time, too. In other words, starting from a given index, the sequence $\left\{ (x^k, z^k, y^k) \right\}_{k \geqslant 0}$ becomes identical to $(\widehat{x}, \widehat{z}, \widehat{y})$ and the conclusion follows.

(ii) If $\theta \in (0, 1/2]$, then $\frac{1}{1-\theta} C_L \mathcal{E}_k^{1-\theta} \leqslant \sqrt{\mathcal{E}_k}$, for every $k \geqslant k_0'$, which implies that $\max \left\{ \sqrt{\mathcal{E}_k}, \varphi(\mathcal{E}_k) \right\} = \sqrt{\mathcal{E}_k}$. By Theorem 18, there exist $k_0'' \geqslant 1$, $\widehat{C}_0 > 0$ and $Q \in [0, 1)$ such that for $\widehat{Q} := Q^{\frac{1}{2}}$ and every $k \geqslant k_0''$ it holds

$$\sqrt{\mathcal{E}_k} \leqslant \sqrt{\widehat{C}_0} Q^{\frac{k}{2}} = \sqrt{\widehat{C}_0} \widehat{Q}^k.$$

The conclusion follows from Lemma 19 for $k_0 := \max\{k_0', k_0''\}$, by noticing that

$$\sqrt{\mathcal{E}_{k-1}} \leqslant \sqrt{\widehat{C}_0} Q^{\frac{k-1}{2}} = \sqrt{\frac{\widehat{C}_0}{Q}} \widehat{Q}^k \quad \text{and} \quad \sqrt{\mathcal{E}_{k-2}} \leqslant \sqrt{\widehat{C}_0} Q^{\frac{k-2}{2}} = \frac{\sqrt{\widehat{C}_0}}{Q} \widehat{Q}^k \forall k \geqslant k_0.$$

(iii) If $\theta \in (1/2, 1)$, then $\mathcal{E}_k^{\frac{1}{2}} \leqslant \frac{1}{1-\theta} C_L \mathcal{E}_k^{1-\theta}$, for every $k \geqslant k_0'$, which implies that $\max \left\{ \sqrt{\mathcal{E}_k}, \varphi(\mathcal{E}_k) \right\} = \varphi(\mathcal{E}_k) = \frac{1}{1-\theta} C_L \mathcal{E}_k^{1-\theta}$. By Theorem 18, there exist $k_0'' \geqslant 3$ and $\widehat{C}_1 > 0$ such that for all $k \geqslant k_0''$

$$\frac{1}{1-\theta} C_L \mathcal{E}_k^{1-\theta} \leqslant \frac{1}{1-\theta} C_L \widehat{C}_1^{1-\theta} (k-2)^{-\frac{1-\theta}{2\theta-1}}.$$

The conclusion follows again for $k_0 := \max\{k_0', k_0''\}$ from Lemma 19. $\qquad\square$

Remark 6. For $\rho = 1$ the same convergence rates can be obtained under the original Assumption 1. Indeed, when $\rho = 1$ we have that $T_1 = 0$ and, as a consequence, the sequence $\{\mathcal{F}_k\}_{k \geqslant 1}$ defined in (47) becomes

$$\mathcal{F}_k = \mathcal{L}_r \left(x^k, z^k, y^k \right) + C_1 \left\| x^k - x^{k-1} \right\|^2 \ \forall k \geqslant 1.$$

In addition, the inequality (41) simplifies to

$$\left\| y^{k+1} - y^k \right\| \leqslant C_3 \left\| x^{k+1} - x^k \right\| + C_4 \left\| x^k - x^{k-1} \right\| \ \forall k \geqslant 1,$$

as T_2 is equal to 0. Combining this inequality with (39) and, by taking into account Lemma 9, we obtain (instead of (62))

$$\left\| \! \left\| D^{k+1} \right\| \! \right\| \leqslant \ C_{11} \left(\left\| x^{k+1} - x^k \right\| + \left\| x^k - x^{k-1} \right\| + \left\| x^{k-1} - x^{k-2} \right\| \right) \ \forall k \geqslant 2.$$

Consequently, for every $k \geqslant 3$ we have that

$$\begin{aligned}
\mathcal{E}_{k-2} - \mathcal{E}_{k+1} &= \mathcal{F}_{k-2} - \mathcal{F}_{k-1} + \mathcal{F}_{k-1} - \mathcal{F}_k + \mathcal{F}_k - \mathcal{F}_{k+1} \\
&\geqslant \frac{C_1}{4} \left(\left\| x^{k-1} - x^{k-2} \right\|^2 + \left\| x^k - x^{k-1} \right\|^2 + \left\| x^{k+1} - x^k \right\|^2 \right) \\
&\geqslant \frac{C_1}{12} \left(\left\| x^{k-1} - x^{k-2} \right\| + \left\| x^k - x^{k-1} \right\| + \left\| x^{k+1} - x^k \right\| \right)^2 \\
&\geqslant \frac{C_1}{12 C_{11}^2} \left\| \! \left\| D^{k+1} \right\| \! \right\|^2.
\end{aligned}$$

Let $\varepsilon > 0$ be such that (70) is fulfilled and $k_0 \geqslant 3$ such that $\left(x^{k+1}, z^{k+1}, y^{k+1} \right)$ belongs to the open ball $\mathrm{Ball}\left(\left(\widehat{x}, \widehat{z}, \widehat{y} \right), \varepsilon \right)$ for every $k \geqslant k_0$. Then (70) implies that for every $k \geqslant k_0$

$$\mathcal{E}_{k-2} - \mathcal{E}_{k+1} \geqslant C_{23} \mathcal{E}_{k+1}, \quad \text{where} \quad C_{23} := \frac{C_1}{12 C_L^2 C_{11}^2},$$

which is the key inequality for deriving convergence rates, as we have seen above.

References

[1] **B. Ames, M. Hong**. *Alternating direction method of multipliers for penalized zero-variance discriminant analysis.* Computational Optimization and Applications 64(3), 725–754 (2016)

[2] **H. Attouch, J. Bolte**. *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features.* Mathematical Programming 116(1), 5–16 (2009)

[3] **H. Attouch, J. Bolte, P. Redont, A. Soubeyran**. *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Łojasiewicz inequality.* Mathematics of Operations Research 35(2), 438–457 (2010)

[4] **H. Attouch, J. Bolte, B. F. Svaiter**. *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods.* Mathematical Programming 137(1–2), 91–129 (2013)

[5] **S. Banert, R. I. Boţ, E. R. Csetnek**. *Fixing and extending some recent results on the ADMM algorithm.* arXiv:1612.05057 (2016)

[6] **A. Beck**. *First-Order Methods in Optimization.* MOS-SIAM Series on Optimization. SIAM, Philadelphia (2017)

[7] **J. Bolte, A. Daniilidis, A. Lewis**. *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems.* SIAM Journal on Optimization 17(4), 1205–1223 (2006)

[8] **J. Bolte, A. Daniilidis, A. Lewis, M. Shiota**. *Clarke subgradients of stratifiable functions.* SIAM Journal on Optimization 18(2), 556–572 (2007)

[9] **J. Bolte, A. Daniilidis, O. Ley, L. Mazet**. *Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity.* Transactions of the American Mathematical Society 362(6), 3319–3363 (2010)

[10] **J. Bolte, S. Sabach, M. Teboulle**. *Proximal alternating linearized minimization for nonconvex and nonsmooth problems.* Mathematical Programming 146(1), 459–494 (2014)

[11] **J. Bolte, S. Sabach, M. Teboulle**. *Nonconvex Lagrangian-based optimization: monitoring schemes and global convergence.* Mathematics of Operations Research 43(4), 1210–1232 (2018)

[12] **R. I. Boţ, E. R. Csetnek**. *An inertial alternating direction method of multipliers.* Minimax Theory and its Applications 71(3), 29–49 (2016)

[13] **R. I. Boţ, E. R. Csetnek**. *An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems.* Journal of Optimization Theory and Applications 171(2), 600–616 (2016)

[14] **R. I. Boţ, E. R. Csetnek**. *ADMM for monotone operators: convergence analysis and rates.* Advances in Computational Mathematics 45(1), 327–359 (2019)

[15] **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)

[16] **R. I. Boţ, E. R. Csetnek, S. C. Laszlo**. *An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions.* EURO Journal on Computational Optimization 4(1), 3–25 (2016)

[17] **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), 1–122 (2010)

[18] **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)

[19] **P. L. Combettes, B. C. Vũ**. *Variable metric quasi-Fejér monotonicity.* Nonlinear Analysis: Theory, Methods and Applications 78, 17–31 (2014)

[20] **P. L. Combettes, V. R. Wajs**. *Signal recovery by proximal forward-backward splitting.* Multiscale Modeling and Simulation 4(4), 1168–1200 (2005)

[21] **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)

[22] **Y. Cui, X.D. Li, D.F. Sun, K.C. Toh**. *On the convergence properties of a majorized ADMM for linearly constrained convex optimization problems with coupled objective functions.* Journal of Optimization Theory and Applications 169, 1013–1041 (2016)

[23] **M. Fazel, T.K. Pong, D.F. Sun, P. Tseng**. *Hankel matrix rank minimization with applications to system identification and realization.* SIAM Journal on Matrix Analysis and Applications 34, 946–977 (2013)

[24] **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)

[25] **P. Frankel, G. Garrigos, J. Peypouquet**. *Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates.* Journal of Optimization Theory and Applications 165(3), 874–900 (2015)

[26] **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)

[27] **D. Gabay, B. Mercier**. *A dual algorithm for the solution of nonlinear variational problems via finite element approximation.* Computers and Mathematics with Applications 2(1), 17–40 (1976)

[28] **K. Guo, D.R. Han, T.T. Wu**. *Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints.* International Journal of Computer Mathematics 94(8), 1653-1669 (2017)

[29] **W. Hare, C. Sagastizábal**. *Computing proximal points of nonconvex functions.* Mathematical Programming 116(1-2), 221–258 (2009)

[30] **M. Hong, Z.Q. Luo, M. Razaviyayn**. *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems.* SIAM Journal on Optimization 26(1), 337–364 (2016)

[31] **M. Hong, Z.-Q. Luo**. *On the linear convergence of the alternating direction method of multipliers.* Mathematical Programming 162, 165–199 (2017)

[32] **K. Kurdyka**. *On gradients of functions definable in o-minimal structures.* Annales de l'Institut Fourier 48, 769–783 (1998)

[33] **A. Lewis, J. Malick.** *Alternating projection on manifolds.* Mathematics of Operations Research 33(1), 216–234 (2008)

[34] **G. Li, T. K. Pong**. *Global convergence of splitting methods for nonconvex composite optimization.* SIAM Journal on Optimization 25(4), 2434–2460 (2015)

[35] **Z. Lin, R. Liu, H. Li**. *Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning.* Machine Learning 99(2), 287–325 (2015)

[36] **Q. Liu, X. Shen, Y. Gu**. *Linearized ADMM for non-convex non-smooth optimization with convergence analysis.* arXiv:1705.02502 (2017)

[37] **S. Łojasiewicz**. *Une propriété topologique des sous-ensembles analytiques réels.* in: Colloques internationaux du C.N.R.S.: Les équations aux dérivées partielles, Paris (1962), Éditions du Centre National de la Recherche Scientifique, Paris, 87–89 (1963)

[38] **B. Mordukhovich**. *Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications..* Springer, Berlin (2006)

[39] **J. Moreau**. *Fonctions convexes duales et points proximaux dans un espace hilbertien.* Comptes Rendus de l'Académie des Sciences (Paris), Série A 255, 2897–2899 (1962)

[40] **Y. Ouyang, Y. Chen, G. Lan, E. Pasiliao, Jr.**. *An accelerated linearized alternating direction method of multipliers.* SIAM Journal on Imaging Sciences 8(1), 644–681 (2015)

[41] **X. Ren, Z. Lin**. *Linearized alternating Ddirection method with adaptive penalty and warm starts for fast solving transform invariant low-Rank textures.* International Journal of Computer Vision 104(1), 1–14 (2013)

[42] **R. T. Rockafellar, R. J.-B. Wets**. *Variational Analysis,* Fundamental Principles of Mathematical Sciences 317, Springer, Berlin (1998)

[43] **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)

[44] **D. Sun, K.-C. Toh, L. Yang**. *A convergent 3-block semi-proximal alternating direction method of multipliers for conic programming with 4-type constraints.* SIAM Journal on Optimization 25(2), 882–915 (2015)

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

[46] **Y. Wang, Z. Xu, H.-K. Xu**. *Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems.* UCLA CAM Report 15-62, UCLA (2015)

[47] **Y. Wang, W. Yin, J. Zeng**. *Global convergence of ADMM in nonconvex nonsmooth optimization.* Journal of Scientific Computing 78(1), 29–63 (2019)

[48] **M.H. Xu and T. Wu**. *A class of linearized proximal alternating direction methods.* Journal of Optimization Theory and Applications 151(2), 321–337 (2011)

[49] **L. Yang, T. K. Pong, X. Chen**. *Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction.* SIAM Journal on Imaging Sciences, 10(1), 74–110 (2017)

[50] **J. Yang, X. Yuan**. *Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization.* Mathematics of Computation 82, 301–329 (2013)