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

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

January 6, 2018

Abstract. We propose two numerical algorithms for minimizing the sum of a smooth function and the composition of a nonsmooth function with a linear operator in the fully nonconvex setting. The iterative schemes are formulated in the spirit of the proximal and, respectively, proximal linearized alternating direction method of multipliers. The proximal terms are introduced through variable metrics, which facilitates the derivation of proximal splitting algorithms for nonconvex complexly structured optimization problems as particular instances of the general schemes. Convergence of the iterates to a KKT point of the objective function is proved 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. If the augmented Lagrangian has the Łojasiewicz property, then convergence rates of both augmented Lagrangian and iterates are derived.

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 address the solving of the optimization problem

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

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) in the context of solving the more general problem

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

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. We rewrite the problem (2), by introducing an auxiliary variable, as

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

[*]Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, e-mail: `radu.bot@univie.ac.at`. Research partially supported by FWF (Austrian Science Fund), project I 2419-N32.

[†]Invited Associate Professor, Babeş-Bolyai University, Faculty of Mathematics and Computer Sciences, str. Mihail Kogălniceanu 1, 400084 Cluj-Napoca, Romania

[‡]Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, e-mail: `dang-khoa.nguyen@univie.ac.at`. The author gratefully acknowledges the financial support of the Doctoral Programme *Vienna Graduate School on Computational Optimization (VGSCO)* which is funded by Austrian Science Fund (FWF, project W1260-N35).

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 $(x^0, z^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ 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]: for all $k \geqslant 0$ generate the sequence $\{(x^k, z^k, y^k)\}_{k\geqslant 0}$ by

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

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

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

In case $\rho = 1$, it has been proved in [5] that when the set 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 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) (which is nothing else than $\mathcal{L}_r$ when $r = 0$) 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 expressed in terms of the Lagrangian has been shown.

In case $h = 0$, the above iterative scheme encompasses different numerical algorithms considered in the literature. When $\mathbf{M}_1^k = \mathbf{M}_2^k = 0$ for all $k \geqslant 0$, (4a)-(4c) becomes the *classical ADMM algorithm* ([15, 22, 25, 26]), which has a huge popularity in the optimization community. And this 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 to [10]. When $\mathbf{M}_1^k = \mathbf{M}_1$ and $\mathbf{M}_2^k = \mathbf{M}_2$ for all $k \geqslant 0$, (4a)-(4c) recovers the *proximal ADMM algorithm* investigated by Shefi and Teboulle in [39] (see also [20, 21]). It has been pointed out in [39] that, for suitable choices of the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$, this proximal ADMM algorithm becomes a primal-dual splitting algorithm in the sense of those considered in [13, 16, 19, 41], and which, due to their full splitting character, overcome the drawbacks of the classical ADMM algorithm. Recently, in [12] it has been shown that, when f is strongly convex, suitable choices of the non-constant sequences $\{\mathbf{M}_1^k\}_{k\geqslant 0}$ and $\{\mathbf{M}_2^k\}_{k\geqslant 0}$ may lead to a rate of convergence for the sequence of primal iterates of $\mathcal{O}(1/k)$.

The reason why we address in this paper the slightly less general optimization problem (1) is exclusively given by the fact that in this setting we can provide sufficient conditions which guarantee that the sequence generated by the ADMM algorithm is bounded. In the nonconvex setting, the boundedness of the sequence $\{(x^k, z^k, y^k)\}_{k\geqslant 0}$ plays a central role the convergence analysis.

The contributions of the paper are as follows:

1. We propose a proximal ADMM (P-ADMM) algorithm and a proximal linearized ADMM (PL-ADMM) algorithm for solving (1) and carry out a convergence analysis in parallel for both algorithms. We first prove under certain assumptions on the matrix sequences boundedness for the sequence of generated iterates $\{(x^k, z^k, y^k)\}_{k\geqslant 0}$. Under 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). Global convergence of the sequence is shown provide that a regularization of the augmented Lagrangian satisfies the Kurdyka-Łojasiewicz property. In case 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 addressing convergence rates for the nonconvex ADMM.

2. The two ADMM algorithms under investigation are of *relaxed type*, namely, we allow $\rho \in (0, 2)$. We notice that $\rho = 1$ is the standard choice in the literature ([1, 5, 12, 30, 39, 42]). Gabay and Mercier proved in [26] 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 [20, 21, 25, 40, 43, 44] or ask for a particular choice of ρ, which is interpreted as a step size, see [27].

   The only work in the nonconvex setting dealing with an alternating minimization algorithm, however for the minimization of the sum of a simple nonsmooth with a smooth function, and which allows a relaxed parameter ρ different from 1 is [44].

3. Particular outcomes of the proposed algorithms will be full splitting algorithms for solving the nonconvex complexly structured optimization (1), which we will obtain by an appropriate choice of the matrix sequences. (P-ADMM) will give rise to an iterative scheme formulated only in terms of proximal steps for the function g and h and of forward evaluations of the matrix A, while (PL-ADMM) will give rise to an iterative scheme in which the function h will be performed via a gradient step. Exact formulas for proximal operators are available not only for large classes of convex ([18]), but also of nonconvex functions ([3, 23, 29]). 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 [31, 36, 37, 43, 45]; the paper [32] being the only exception in the nonconvex setting.

For previous works addressing the ADMM algorithm in the nonconvex setting we mention: [30], where (1) is studied by assuming that h is twice continuously differentiable with bounded Hessian; [24], where the convergence is studied in the context of solving a very particular nonconvex consensus and sharing problems; and [1], where the ADMM algorithm is used in the penalized zero-variance discriminant analysis. In [42] and [32], the investigations of the ADMM algorithm are carried out in very restrictive settings generated by the strong assumptions on the nonsmooth functions and linear operators.

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 $x := (x_1, \ldots, x_N)$, $y := (y_1, \ldots, y_N) \in \mathbb{R}^N$

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

The Cartesian product $\mathbb{R}^{N_1} \times \mathbb{R}^{N_2} \times \ldots \times \mathbb{R}^{N_p}$ with some strictly positive integer p will be endowed with *inner product* and associated *norm* defined for $u := (u_1, \ldots, u_p), u' := \left(u_1', \ldots, u_p'\right) \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2} \times \ldots \times \mathbb{R}^{N_p}$ by

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

respectively. Moreover, for every $u := (u_1, \ldots, u_p), u' := \left(u_1', \ldots, u_p'\right) \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2} \times \ldots \times \mathbb{R}^{N_p}$ 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

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

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

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

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. If $M \in \mathcal{P}_\alpha^N$, then the semi-norm $\|\cdot\|_M$ obviously 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^* is positively definite. Since

$$\lambda_{\min}(AA^*) \|y\|^2 \leqslant \|y\|_{AA^*}^2 = \langle AA^* y, y \rangle = \|A^* y\|^2 \;\; \forall y \in \mathbb{R}^m, \tag{6}$$

this is further equivalent to $\lambda_{\min}(AA^*) > 0$ (and $AA^* \in \mathcal{P}^n_{\lambda_{\min}(AA^*)}$), 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 $A^*A \in \mathcal{P}^n_{\lambda_{\min}(A^*A)}$).

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*

$$\Psi(y) \leqslant \Psi(x) + \langle \nabla \Psi(z), y - x \rangle + \frac{L}{2} \|y - x\|^2; \tag{7}$$

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

$$\inf_{x \in \mathbb{R}^N} \left\{ \Psi(x) - \left(\frac{1}{\sigma} - \frac{L}{2\sigma^2} \right) \|\nabla \Psi(x)\|^2 \right\} > -\infty. \tag{8}$$

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 get

$$\begin{aligned} \Psi(y) - \Psi(x) &= \int_0^1 \langle \nabla \Psi((1-s)x + sy), y - x \rangle ds \\ &= \int_0^1 \langle \nabla \Psi((1-s)x + sy) - \nabla \Psi(z), y - x \rangle ds + \langle \nabla \Psi(z), y - x \rangle. \end{aligned} \tag{9}$$

Since

$$\begin{aligned} &\left| \int_0^1 \langle \nabla \Psi((1-s)x + sy) - \nabla \Psi(z), y - x \rangle ds \right| \\ \leqslant &\int_0^1 \|\nabla \Psi((1-s)x + sy) - \nabla \Psi(z)\| \cdot \|y - x\| ds \leqslant L \|x - y\|^2 \int_0^1 |s - t| ds \\ = &\ L \|x - y\|^2 \left(\int_0^t (-s + t) ds + \int_t^1 (s - t) ds \right) = L \left(\frac{1}{2} - t(1-t) \right) \|x - y\|^2. \end{aligned} \tag{10}$$

The inequality (7) is obtained by combining (9) and (10) and by using that $0 \leqslant t \leqslant 1$.

2. The inequality (7) gives for every $x \in \mathbb{R}^N$

$$\begin{aligned} -\infty < \inf_{y \in \mathbb{R}^N} \Psi(y) &\leqslant \Psi\left(x - \frac{1}{\sigma} \nabla \Psi(x) \right) \\ &\leqslant \Psi(x) + \left\langle \left(x - \frac{1}{\sigma} \nabla \Psi(x) \right) - x, \nabla \Psi(x) \right\rangle + \frac{L}{2} \left\| \left(x - \frac{1}{\sigma} \nabla \Psi(x) \right) - x \right\|^2 \\ &= \Psi(x) - \left(\frac{1}{\sigma} - \frac{L}{2\sigma^2} \right) \|\nabla \Psi(x)\|^2, \end{aligned}$$

which leads to the desired conclusion. □

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

$$\Psi(y) \leqslant \Psi(x) + \langle \nabla \Psi(x), y - x \rangle + \frac{L}{2} \|y - x\|^2 \quad \forall x, y \in \mathbb{R}^N, \tag{11}$$

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

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

$$\Psi(x) \geqslant \Psi(y) + \langle \nabla \Psi(y), x - y \rangle - \frac{L}{2} \|x - y\|^2 \quad \forall x, y \in \mathbb{R}^N, \tag{12}$$

which is equivalent to the fact that $\Psi + \frac{L}{2} \|\cdot\|^2$ is a convex function, in other words, Ψ is a L-semiconvex function ([8]). It follows from the previous result that a Fréchet differentiable function with L-Lipschitz continuous gradient is L-semiconvex.

Further, we will recall the definition and some properties of the *limiting subdifferential*, a notion which 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 any $x \in \operatorname{dom}\Psi := \left\{x \in \mathbb{R}^N : \Psi(x) < +\infty\right\}$, the *Fréchet (viscosity) subdifferential* of Ψ at x is

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

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

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

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

The inclusion $\widehat{\partial}\Psi(x) \subseteq \Psi(x)$ holds for each $x \in \mathbb{R}^N$ in general. In case Ψ is convex, these two subdifferential notions coincide with the *convex subdifferential*, thus

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

If $x \in \mathbb{R}^N$ is a local minimum of Ψ, then $0 \in \partial\Psi(x)$. We denote by $\operatorname{crit}(\Psi) = \{x \in \mathbb{R}^N : 0 \in \partial\Psi(x)\}$ 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 $d^k \in \Psi\left(x^k\right)$ for all $k \geqslant 0$ and $\left(x^k, d^k\right) \to (x, d)$ and $\Psi\left(x^k\right) \to \Psi(x)$ as $k \to +\infty$, then $d \in \Psi(x)$. We also have the following *subdifferential sum rule* ([34, Proposition 1.107], [38, Exercise 8.8]): if $\Phi\colon \mathbb{R}^N \to \mathbb{R}$ is a continuously differentiable function, then $\partial(\Psi + \Phi)(x) = \partial\Psi(x) + \nabla\Phi(x)$ for all $x \in \mathbb{R}^N$; and the following formula for the *subdifferential of the composition with a linear operator* $A\colon \mathbb{R}^{N'} \to \mathbb{R}^N$ ([34, Proposition 1.112], [38, Exercise 10.7]): if $x \in \operatorname{dom}\Psi$ and A is injective, then $\partial(\Psi \circ A)(x) = A^*\partial\Psi(Ax)$.

We close this section by presenting some 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, [11, Lemma 2.2], [14, 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 $\sum_{k\geqslant 0} \xi_k < +\infty$.*
2. *the sequence $\{b_k\}_{k\geqslant 0}$ is monotonically decreasing and convergent.*

The following lemma, which is an extension of [11, Lemma 2.3] (see, also [14, 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 $\mathbb{R}$ satisfy*

$$\left\langle \mathbb{1}, a^{k+1}\right\rangle \leqslant \left\langle c_0, a^k\right\rangle + \left\langle c_1, a^{k-1}\right\rangle + \left\langle c_2, a^{k-2}\right\rangle + \delta_k \ \forall k \geqslant 2, \tag{13}$$

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 $\overline{\delta} \geqslant 0$ such that for every $\overline{K} \geqslant \underline{K} \geqslant 2$

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

Then for every $i = 1, \ldots, N$ we have

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

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

$$\sum_{k=\underline{K}}^{\overline{K}} a_i^k \leqslant \frac{\sum_{j=1}^{N} \left[(1 - c_{0,j} - c_{1,j})\, a_j^{\underline{K}} + (1 - c_{0,j})\, a_j^{\underline{K}+1} + a_j^{\underline{K}+2}\right] + \overline{\delta}}{1 - c_{0,i} - c_{1,i} - c_{2,i}}. \tag{14}$$

Proof. Fix $\overline{K} \geqslant \underline{K} \geqslant 2$. If $\overline{K} = \underline{K}$ or $\overline{K} = \underline{K} + 1$ then (14) holds automatically. Consider now the case when $\overline{K} \geqslant \underline{K} + 2$. Summing up the inequality(13) for $k = \underline{K} + 2, \cdots, \overline{K}$, we obtain

$$\left\langle \mathbb{1}, \sum_{k=\underline{K}+2}^{\overline{K}} a^{k+1} \right\rangle \leqslant \left\langle c_0, \sum_{k=\underline{K}+2}^{\overline{K}} a^k \right\rangle + \left\langle c_1, \sum_{k=\underline{K}+2}^{\overline{K}} a^{k-1} \right\rangle + \left\langle c_2, \sum_{k=\underline{K}+2}^{\overline{K}} a^{k-2} \right\rangle + \sum_{k=\underline{K}+2}^{\overline{K}} \delta_k. \tag{15}$$

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 (15) 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

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

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 $c_0 + c_1 + c_2 < \mathbb{1}$. $\square$

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

In this section we will propose two proximal ADMM algorithms for solving the optimization problem (1) and we will study their convergence behaviour. In this context, 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 particular instances written in the spirit of full proximal splitting algorithms

Algorithm 1. *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:*

$$z^{k+1} \in \arg\min_{z\in\mathbb{R}^m} \left\{ g(z) + \left\langle y^k, Ax^k - z \right\rangle + \frac{r}{2}\left\|Ax^k - z\right\|^2 + \frac{1}{2}\left\|z - z^k\right\|_{\mathbf{M}_2^k}^2 \right\}, \tag{16a}$$

$$x^{k+1} \in \arg\min_{x\in\mathbb{R}^n} \left\{ h(x) + \left\langle y^k, Ax - z^{k+1} \right\rangle + \frac{r}{2}\left\|Ax - z^{k+1}\right\|^2 + \frac{1}{2}\left\|x - x^k\right\|_{\mathbf{M}_1^k}^2 \right\}, \tag{16b}$$

$$y^{k+1} := y^k + \rho r\left(Ax^{k+1} - z^{k+1}\right). \tag{16c}$$

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

$$\begin{aligned}
z^{k+1} &\in \arg\min_{z\in\mathbb{R}^m} \left\{ g(z) + \frac{r}{2}\left\| z - Ax^k - \frac{1}{r}y^k \right\|^2 \right\},\\
x^{k+1} &\in \arg\min_{x\in\mathbb{R}^n} \left\{ h(x) + \frac{1}{2t_k}\left\| x - x^k + t_k A^*\left[y^k + r\left(Ax^k - z^{k+1}\right)\right]\right\|^2 \right\},\\
y^{k+1} &:= y^k + \rho r\left(Ax^{k+1} - z^{k+1}\right).
\end{aligned}$$

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 ([35])

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

The above particular instance of Algorithm 1 is an iterative scheme formulated in the spirit of full splitting numerical methods, namely, the functions g and h are evaluated by their proximal operators, while the linear operator A and its adjoint are evaluated by simple forward steps. Exact formulas for the proximal operator are available not only for large classes of convex functions ([18]), but also for many nonconvex functions appearing in applications ([3, 23, 29]).

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

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:*

$$z^{k+1} \in \arg\min_{z\in\mathbb{R}^m} \left\{ g(z) + \left\langle y^k, Ax^k - z \right\rangle + \frac{r}{2}\left\|Ax^k - z\right\|^2 + \frac{1}{2}\left\|z - z^k\right\|_{\mathbf{M}_2^k}^2 \right\}, \tag{17a}$$

$$x^{k+1} \in \arg\min_{x\in\mathbb{R}^n} \left\{ \left\langle x - x^k, \nabla h\left(x^k\right) \right\rangle + \left\langle y^k, Ax - z^{k+1} \right\rangle + \frac{r}{2}\left\|Ax - z^{k+1}\right\|^2 + \frac{1}{2}\left\|x - x^k\right\|_{\mathbf{M}_1^k}^2 \right\}, \tag{17b}$$

$$y^{k+1} := y^k + \rho r\left(Ax^{k+1} - z^{k+1}\right). \tag{17c}$$

Due to the presence of the variable metric inducing matrix sequences we can thus provide a unifying scheme for several linearized ADMM algorithms discussed in the literature (see [31, 32, 36, 37, 43, 45]), which can be recovered for specific choices of the variable metrics. When taking as for Algorithm 1 $\mathbf{M}_1^k := \frac{1}{t_k}\mathbf{Id} - rA^*A$, where $t_k r\|A\|^2 \leqslant 1$, and $\mathbf{M}_2^k := \mathbf{0}$, for every $k \geqslant 0$, then Algorithm 2 translates for every $k \geqslant 0$ into:

$$\begin{aligned}
z^{k+1} &\in \arg\min_{z\in\mathbb{R}^m} \left\{ g(z) + \frac{r}{2}\left\| z - Ax^k - \frac{1}{r}y^k \right\|^2 \right\},\\
x^{k+1} &:= x^k - t_k\left(\nabla h\left(x^k\right) + A^*\left[y^k + r\left(Ax^k - z^{k+1}\right)\right]\right),\\
y^{k+1} &:= y^k + \rho r\left(Ax^{k+1} - z^{k+1}\right).
\end{aligned}$$

This iterative scheme has the remarkable property that the smooth term is evaluated via a gradient step. This is an improvement with respect to other nonconvex ADMM algorithms, such as [42, 44], where the smooth function is involved in a subproblem, which can be in general difficult to solve, unless it can be reformulated as a proximal step (see [30]).

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

Assumption 1. *Assume that*

$$A \text{ is surjective}$$

and $r > 0$, $\rho \in (0,2)$, $\mu_1 := \sup_{k \geqslant 0} \left\| \mathbf{M}_1^k \right\| < +\infty$ *and* $\mu_2 := \sup_{k \geqslant 0} \left\| \mathbf{M}_2^k \right\| < +\infty$ *are such that there exists* $\gamma > 1$ *with*

$$r \geqslant (2+\gamma) T_1 L > 0 \tag{18}$$

and

$$\mathbf{M}_3^k := 2\mathbf{M}_1^k + rA^*A - C_1 \mathbf{Id} \geqslant \frac{3}{2} C_0 \mathbf{Id} \quad \forall k \geqslant 0, \tag{19}$$

where

$$T_0 := \begin{cases} \dfrac{1-\rho}{\lambda_{\min}(AA^*)\rho^2 r} & \text{if } 0 < \rho \leqslant 1, \\ \dfrac{\rho - 1}{\lambda_{\min}(AA^*)(2-\rho)\rho r} & \text{if } 1 < \rho < 2, \end{cases} \qquad T_1 := \begin{cases} \dfrac{1}{\lambda_{\min}(AA^*)\rho}, & \text{if } 0 < \rho \leqslant 1, \\ \dfrac{\rho}{\lambda_{\min}(AA^*)(2-\rho)^2}, & \text{if } 1 < \rho < 2, \end{cases}$$

and

$$C_0 := \begin{cases} \dfrac{4T_1\mu_1^2}{r}, & \text{for Algorithm 1}, \\ \dfrac{4T_1(L+\mu_1)^2}{r}, & \text{for Algorithm 2}, \end{cases} \qquad C_1 := \begin{cases} L + \dfrac{4T_1(L+\mu_1)^2}{r}, & \text{for Algorithm 1}, \\ L + \dfrac{4T_1\mu_1^2}{r}, & \text{for Algorithm 2}. \end{cases}$$

Remark 2. Notice that (19) can be equivalently written as

$$2\mathbf{M}_1^k + rA^*A - \left(L + r^{-1}C_{\mathbf{M}}\right)\mathbf{Id} \geqslant 0 \ \forall k \geqslant 0, \text{ where } C_{\mathbf{M}} := \begin{cases} \left(6\mu_1^2 + 4(L+\mu_1)^2\right) T_1, & \text{for Algorithm 1}, \\ \left(4\mu_1^2 + 6(L+\mu_1)^2\right) T_1, & \text{for Algorithm 2}. \end{cases} \tag{20}$$

In the following we present some possible choices of the matrix sequences $\left\{\mathbf{M}_1^k\right\}_{k \geqslant 0}$ and $\left\{\mathbf{M}_2^k\right\}_{k \geqslant 0}$ which fulfill Assumption 1.

1. Since $rA^*A \in \mathbb{S}_+^n$, when $\sup_{k \geqslant 0} \left\| \mathbf{M}_1^k \right\| = \mu_1 > \frac{L}{2}$, by choosing

   $$r \geqslant \max\left\{(2+\gamma) T_1 L, \frac{C_{\mathbf{M}}}{2\mu_1 - L}\right\} > 0,$$

   there exists $\alpha_1 > 0$ such that

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

   Thus (18) is verified, while (20) is ensured when choosing $\mathbf{M}_1^k$ such that $\mu_1 \mathbf{Id} \geqslant M_1^k \geqslant \alpha_1 \mathbf{Id}$ for every $k \geqslant 0$.

2. Let $\mathbf{M}_1^k := \frac{1}{t}\mathbf{Id} - rA^*A$ for every $k \geqslant 0$, where $0 < t < \min\left\{\frac{1}{r\|A\|^2}, \frac{1}{L}\right\}$. Then the relation (20) becomes

   $$\frac{2}{t}\mathbf{Id} - rA^*A - \left(L + r^{-1}C_{\mathbf{M}}\right)\mathbf{Id} \geqslant 0,$$

   which automatically holds (as also (18) does), if

   $$r \geqslant \max\left\{(2+\gamma) T_1 L, \frac{tC_{\mathbf{M}}}{1-tL}\right\} > 0.$$

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

$$r \geqslant \max\left\{(2+\gamma)T_1L, \frac{L+\sqrt{L^2+4\lambda_{\min}(A^*A)C_{\mathbf{M}}}}{2\lambda_{\min}(A^*A)}\right\} > 0,$$

it follows that

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

thus, (18) and (20) hold for an arbitrary sequence of symmetric and positive semidefinite matrices $\left\{\mathbf{M}_1^k\right\}_{k\geqslant 0}$. A possible choice is $\mathbf{M}_1^k = 0$ and $\mathbf{M}_2^k = 0$ for every $k \geqslant 0$, which allows us to recover the classical ADMM.

When proving convergence for variable metric algorithms designed for convex optimization problems one usually assumes monotonicity for the matrix sequences inducing the variable metrics (see, for instance, [17, 5]). It is worth to mention that in this paper we manage to perform the convergence analysis for both Algorithm 1 and Algorithm 2 without any monotonicity assumption on $\left\{\mathbf{M}_1^k\right\}_{k\geqslant 0}$ and $\left\{\mathbf{M}_2^k\right\}_{k\geqslant 0}$.

2.2 Preliminaries of the convergence analysis

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

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

$$\begin{aligned}
&\mathcal{L}_r\left(x^{k+1}, z^{k+1}, y^{k+1}\right) + T_0\left\|A^*\left(y^{k+1}-y^k\right)\right\|^2 + \frac{1}{2}\left\|x^{k+1}-x^k\right\|^2_{\mathbf{M}_3^k} + \frac{1}{2}\left\|z^{k+1}-z^k\right\|^2_{\mathbf{M}_2^k}\\
&\leqslant \mathcal{L}_r\left(x^k, z^k, y^k\right) + T_0\left\|A^*\left(y^k-y^{k-1}\right)\right\|^2 + \frac{C_0}{2}\left\|x^k-x^{k-1}\right\|^2. \qquad (21)
\end{aligned}$$

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

$$\begin{aligned}
&\mathcal{L}_r\left(x^{k+1}, z^{k+1}, y^{k+1}\right) + \frac{1}{2}\left\|x^{k+1}-x^k\right\|^2_{2\mathbf{M}_1^k+rA^*A} - \frac{L}{2}\left\|x^{k+1}-x^k\right\|^2 + \frac{1}{2}\left\|z^{k+1}-z^k\right\|^2_{\mathbf{M}_2^k}\\
&\leqslant \mathcal{L}_r\left(x^k, z^k, y^k\right) + \frac{1}{\rho r}\left\|y^{k+1}-y^k\right\|^2 \qquad (22)
\end{aligned}$$

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

1. For *Algorithm 1*: From (16a) we have

$$\begin{aligned}
&g\left(z^{k+1}\right) + \left\langle y^k, Ax^k - z^{k+1}\right\rangle + \frac{r}{2}\left\|Ax^k - z^{k+1}\right\|^2 + \frac{1}{2}\left\|z^{k+1}-z^k\right\|^2_{\mathbf{M}_2^k}\\
&\leqslant g\left(z^k\right) + \left\langle y^k, Ax^k - z^k\right\rangle + \frac{r}{2}\left\|Ax^k - z^k\right\|^2. \qquad (23)
\end{aligned}$$

The optimality criterion of (16b) is

$$\nabla h\left(x^{k+1}\right) = -A^*y^k - rA^*\left(Ax^{k+1}-z^{k+1}\right) + \mathbf{M}_1^k\left(x^k - x^{k+1}\right). \qquad (24)$$

From (7) (applied for $z := x^{k+1}$) we get

$$\begin{aligned}
h\left(x^{k+1}\right) \leqslant{} & h\left(x^k\right) + \left\langle y^k, Ax^k - Ax^{k+1}\right\rangle + r\left\langle Ax^{k+1}-z^{k+1}, Ax^k - Ax^{k+1}\right\rangle\\
& - \left\|x^{k+1}-x^k\right\|^2_{\mathbf{M}_1^k} + \frac{L}{2}\left\|x^{k+1}-x^k\right\|^2. \qquad (25)
\end{aligned}$$

By combining (23), (25) and (16c), after some rearrangements, we obtain (22).

By using the notation

$$u_1^l := -\nabla h\left(x^l\right) + \mathbf{M}_1^{l-1}\left(x^{l-1}-x^l\right) \ \forall l \geqslant 1 \qquad (26)$$

and by taking into consideration (16c), we can rewrite (24) as

$$A^*y^{l+1} = \rho u_1^{l+1} + (1-\rho)A^*y^l \ \forall l \geqslant 0. \qquad (27)$$

- *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

$$\begin{aligned} &\lambda_{\min}(AA^*)\rho \left\|y^{k+1} - y^k\right\|^2 \leqslant \rho \left\|A^* \left(y^{k+1} - y^k\right)\right\|^2 \\ &\leqslant \rho \left\|u_1^{k+1} - u_1^k\right\|^2 + (1-\rho) \left\|A^* \left(y^k - y^{k-1}\right)\right\|^2 - (1-\rho) \left\|A^* \left(y^{k+1} - y^k\right)\right\|^2, \end{aligned} \tag{28}$$

By using the Lipschitz continuity of ∇h, we have

$$\left\|u_1^{k+1} - u_1^k\right\| \leqslant (L + \mu_1) \left\|x^{k+1} - x^k\right\| + \mu_1 \left\|x^k - x^{k-1}\right\|, \tag{29}$$

thus

$$\left\|u_1^{k+1} - u_1^k\right\|^2 \leqslant 2 (L + \mu_1)^2 \left\|x^{k+1} - x^k\right\|^2 + 2\mu_1^2 \left\|x^k - x^{k-1}\right\|^2. \tag{30}$$

After plugging (30) into (28), we get

$$\begin{aligned} \frac{1}{\rho r} \left\|y^{k+1} - y^k\right\|^2 \leqslant &\frac{2 (L + \mu_1)^2}{\lambda_{\min}(AA^*)\rho r} \left\|x^{k+1} - x^k\right\|^2 + \frac{2\mu_1^2}{\lambda_{\min}(AA^*)\rho r} \left\|x^k - x^{k-1}\right\|^2 \\ &+ \frac{(1-\rho)}{\lambda_{\min}(AA^*)\rho^2 r} \left\|A^* \left(y^k - y^{k-1}\right)\right\|^2 - \frac{(1-\rho)}{\lambda_{\min}(AA^*)\rho^2 r} \left\|A^* \left(y^{k+1} - y^k\right)\right\|^2, \end{aligned} \tag{31}$$

which, combined with (22), provides (21).

- *The case* $1 < \rho < 2$. This time we have from (27) 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

$$\begin{aligned} &\lambda_{\min}(AA^*) (2-\rho) \left\|y^{k+1} - y^k\right\|^2 \leqslant (2-\rho) \left\|A^* \left(y^{k+1} - y^k\right)\right\|^2 \\ &\leqslant \frac{\rho^2}{2-\rho} \left\|u_1^{k+1} - u_1^k\right\|^2 + (\rho - 1) \left\|A^* \left(y^k - y^{k-1}\right)\right\|^2 - (\rho - 1) \left\|A^* \left(y^{k+1} - y^k\right)\right\|^2, \end{aligned} \tag{32}$$

After plugging (30) into (32), we get

$$\begin{aligned} \frac{1}{\rho r} \left\|y^{k+1} - y^k\right\|^2 \leqslant &\frac{2\rho (L + \mu_1)^2}{\lambda_{\min}(AA^*) (2-\rho)^2 r} \left\|x^{k+1} - x^k\right\|^2 + \frac{2\rho\mu_1^2}{\lambda_{\min}(AA^*) (2-\rho)^2 r} \left\|x^k - x^{k-1}\right\|^2 \\ &+ \frac{(\rho - 1)}{\lambda_{\min}(AA^*) (2-\rho) \rho r} \left\|A^* \left(y^k - y^{k-1}\right)\right\|^2 \\ &- \frac{(\rho - 1)}{\lambda_{\min}(AA^*) (2-\rho) \rho r} \left\|A^* \left(y^{k+1} - y^k\right)\right\|^2, \end{aligned} \tag{33}$$

which, combined with (22), provides (21).

2. For *Algorithm 2*: The optimality criterion of (17b) is

$$\nabla h \left(x^k\right) = -A^* y^k - r A^* \left(A x^{k+1} - z^{k+1}\right) + \mathbf{M}_1^k \left(x^k - x^{k+1}\right). \tag{34}$$

From (7) (applied for $z := x^k$) we get

$$h\left(x^{k+1}\right) \leqslant h\left(x^k\right) + \left\langle y^k, Ax^k - Ax^{k+1}\right\rangle + r\left\langle Ax^{k+1} - z^{k+1}, Ax^k - Ax^{k+1}\right\rangle \\ - \left\|x^{k+1} - x^k\right\|_{\mathbf{M}_1^k}^2 + \frac{L}{2}\left\|x^{k+1} - x^k\right\|^2. \tag{35}$$

Since the definition of z^{k+1} in (17a) leads also to (23), by combining this inequality with (35) and (17c), after some rearrangments, (22) follows. By using this time the notation

$$u_2^l := -\nabla h\left(x^{l-1}\right) + \mathbf{M}_1^{l-1}\left(x^{l-1} - x^l\right) \ \forall l \geqslant 1 \tag{36}$$

and by taking into consideration (17c), we can rewrite (34) as

$$A^* y^{l+1} = \rho u_2^{l+1} + (1-\rho) A^* y^l \ \forall l \geqslant 0. \tag{37}$$

- *The case* $0 < \rho \leqslant 1$. As in (28), we obtain

$$\lambda_{\min}(AA^*)\rho\left\|y^{k+1} - y^k\right\|^2 \leqslant \rho\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2 \\ \leqslant \rho\left\|u_2^{k+1} - u_2^k\right\|^2 + (1-\rho)\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 - (1-\rho)\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2. \tag{38}$$

  By using the Lipschitz continuity of ∇h, we have

$$\left\|u_2^{k+1} - u_2^k\right\| \leqslant \mu_1\left\|x^{k+1} - x^k\right\| + (L + \mu_1)\left\|x^k - x^{k-1}\right\|, \tag{39}$$

  thus

$$\left\|u_2^{k+1} - u_2^k\right\|^2 \leqslant 2\mu_1^2\left\|x^{k+1} - x^k\right\|^2 + 2(L + \mu_1)^2\left\|x^k - x^{k-1}\right\|^2. \tag{40}$$

  After plugging (40) into (38), it follows

$$\frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2 \leqslant \frac{2\mu_1^2}{\lambda_{\min}(AA^*)\rho r}\left\|x^{k+1} - x^k\right\|^2 + \frac{2(L + \mu_1)^2}{\lambda_{\min}(AA^*)\rho r}\left\|x^k - x^{k-1}\right\|^2 \\ + \frac{(1-\rho)}{\lambda_{\min}(AA^*)\rho^2 r}\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 - \frac{(1-\rho)}{\lambda_{\min}(AA^*)\rho^2 r}\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2, \tag{41}$$

  which, combined with (22), provides (21).

- *The case* $1 < \rho < 2$. As in (32), we obtain

$$\lambda_{\min}(AA^*)(2-\rho)\left\|y^{k+1} - y^k\right\|^2 \leqslant (2-\rho)\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2 \\ \leqslant \frac{\rho^2}{2-\rho}\left\|u_2^{k+1} - u_2^k\right\|^2 + (\rho-1)\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 - (\rho-1)\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2. \tag{42}$$

  After plugging (40) into (42), it follows

$$\frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2 \leqslant \frac{2\rho\mu_1^2}{\lambda_{\min}(AA^*)(2-\rho)^2 r}\left\|x^{k+1} - x^k\right\|^2 + \frac{2\rho(L + \mu_1)^2}{\lambda_{\min}(AA^*)(2-\rho)^2 r}\left\|x^k - x^{k-1}\right\|^2 \\ + \frac{(\rho-1)}{\lambda_{\min}(AA^*)(2-\rho)\rho r}\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 \\ - \frac{(\rho-1)}{\lambda_{\min}(AA^*)(2-\rho)\rho r}\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2. \tag{43}$$

  which, combined with (22), provides (21).

This concludes the proof. □

The following three estimates will be useful in the sequel.

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

(i)

$$\left\|z^{k+1}-z^k\right\| \leqslant \|A\| \cdot \left\|x^{k+1}-x^k\right\| + \left\|Ax^{k+1}-z^{k+1}\right\| + \left\|Ax^k-z^k\right\|$$
$$= \|A\| \cdot \left\|x^{k+1}-x^k\right\| + \frac{1}{\rho r}\left\|y^{k+1}-y^k\right\| + \frac{1}{\rho r}\left\|y^k-y^{k-1}\right\| \quad \forall k \geqslant 1; \tag{44}$$

(ii)

$$\frac{1}{2r}\left\|y^{k+1}\right\|^2 \leqslant \frac{T_0}{2}\left\|A^*\left(y^{k+1}-y^k\right)\right\|^2 + \frac{T_1}{r}\left\|\nabla h\left(x^{k+1}\right)\right\|^2 + \frac{C_0}{4}\left\|x^{k+1}-x^k\right\|^2 \quad \forall k \geqslant 0; \tag{45}$$

(iii)

$$\left\|y^{k+1}-y^k\right\| \leqslant C_3\left\|x^{k+1}-x^k\right\| + C_4\left\|x^k-x^{k-1}\right\| + T_2\left(\left\|A^*\left(y^k-y^{k-1}\right)\right\| - \left\|A^*\left(y^{k+1}-y^k\right)\right\|\right) \quad \forall k \geqslant 1, \tag{46}$$

where

$$C_3 := \begin{cases} \dfrac{\rho\left(L+\mu_1\right)}{\sqrt{\lambda_{\min}(AA^*)}\left(1-|1-\rho|\right)}, & \text{for Algorithm 1,} \\[2ex] \dfrac{\rho\mu_1}{\sqrt{\lambda_{\min}(AA^*)}\left(1-|1-\rho|\right)}, & \text{for Algorithm 2,} \end{cases}$$
$$C_4 := \begin{cases} \dfrac{\rho\mu_1}{\sqrt{\lambda_{\min}(AA^*)}\left(1-|1-\rho|\right)}, & \text{for Algorithm 1,} \\[2ex] \dfrac{\rho\left(L+\mu_1\right)}{\sqrt{\lambda_{\min}(AA^*)}\left(1-|1-\rho|\right)}, & \text{for Algorithm 2,} \end{cases}$$
$$T_2 := \frac{|1-\rho|}{\sqrt{\lambda_{\min}(AA^*)}\left(1-|1-\rho|\right)}. \tag{47}$$

Proof. The statement in (44) is straightforward.

From (27) and (37) 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 (26), for Algorithm 1, and, respectively, to u_2^{k+1} in (36), for Algorithm 2.

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

$$\lambda_{\min}(AA^*)\rho^2\left\|y^{k+1}\right\|^2 \leqslant \rho^2\left\|A^*y^{k+1}\right\|^2 \leqslant \rho\left\|u^{k+1}\right\|^2 + (1-\rho)\left\|A^*\left(y^{k+1}-y^k\right)\right\|^2, \tag{48}$$

while when $1 < \rho < 2$ we have

$$\lambda_{\min}(AA^*)\rho^2\left\|y^{k+1}\right\|^2 \leqslant \rho^2\left\|A^*y^{k+1}\right\|^2 \leqslant \frac{\rho^2}{2-\rho}\left\|u^{k+1}\right\|^2 + (\rho-1)\left\|A^*\left(y^{k+1}-y^k\right)\right\|^2. \tag{49}$$

Notice further that when $1 < \rho < 2$ we have $1/\rho < 1$ and $1 < \rho/(2-\rho)$.

When u^{k+1} is defined as in (26), it holds

$$\left\|u^{k+1}\right\|^2 = \left\|u_1^{k+1}\right\|^2 \leqslant 2\left\|\nabla h\left(x^{k+1}\right)\right\|^2 + 2\mu_1^2\left\|x^{k+1}-x^k\right\|^2 \quad \forall k \geqslant 0, \tag{50}$$

while, when u_2^{k+1} is defined as in (36), it holds

$$\left\|u^{k+1}\right\|^2 = \left\|u_2^{k+1}\right\|^2 \leqslant 2\left\|\nabla h\left(x^{k+1}\right)\right\|^2 + 2\left(L+\mu_1\right)^2\left\|x^{k+1}-x^k\right\|^2 \quad \forall k \geqslant 0. \tag{51}$$

We divide (48) and (49) by $2\lambda_{\min}(AA^*)\rho^2 r > 0$ and plug (50) and, respectively, (51) into the resulting inequalities. This gives us (45).

Finally, in order to prove (46), 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,

$$\begin{aligned} & \sqrt{\lambda_{\min}(AA^*)}\,(1-|1-\rho|)\left\|y^{k+1}-y^k\right\| \leqslant (1-|1-\rho|)\left\|A^*\left(y^{k+1}-y^k\right)\right\| \\ \leqslant\ & \rho\left\|u^{k+1}-u^k\right\|+|1-\rho|\left\|A^*\left(y^k-y^{k-1}\right)\right\|-|1-\rho|\left\|A^*\left(y^{k+1}-y^k\right)\right\|. \end{aligned} \tag{52}$$

We plug into (52) the estimates for $\left\|u^{k+1}-u^k\right\|$ derived in (29) and, respectively, (39) and divide the resulting inequality by $\sqrt{\lambda_{\min}(AA^*)}\,(1-|1-\rho|)>0$. This furnishes the desired statement. □

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

$$\begin{aligned} & \mathcal{F}_r\colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}, \\ & \mathcal{F}_r(x,z,y,x',y') = \mathcal{L}_r(x,z,y)+T_0\left\|A^*\left(y-y'\right)\right\|^2+\frac{C_0}{2}\left\|x-x'\right\|^2, \end{aligned} \tag{53}$$

where T_0 and C_0 are defined in Assumption 1. For every $k \geqslant 1$, we denote

$$\mathcal{F}_k := \mathcal{F}_r\left(x^k,z^k,y^k,x^{k-1},y^{k-1}\right) = \mathcal{L}_r\left(x^k,z^k,y^k\right)+T_0\left\|A^*\left(y^k-y^{k-1}\right)\right\|^2+\frac{C_0}{2}\left\|x^k-x^{k-1}\right\|^2. \tag{54}$$

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 $\lim\limits_{\|x\|\to+\infty}\Psi(x)=+\infty$.

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

(B-I) The operator A is invertible, g is coercive and h is bounded from below;

(B-II) The function h is coercive and g and h are bounded from below.

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 $k \geqslant 1$

$$\mathcal{F}_{k+1}+\frac{1}{2}\left\|x^{k+1}-x^k\right\|^2_{\mathbf{M}_3^k-C_0\mathbf{Id}}+\frac{1}{2}\left\|z^{k+1}-z^k\right\|^2_{\mathbf{M}_2^k} \leqslant \mathcal{F}_k \tag{55}$$

which shows, according to (19), that $\{\mathcal{F}_k\}_{k\geqslant 1}$ is monotonically decreasing. Consequently, for every $k \geqslant 1$ we have

$$\begin{aligned} \mathcal{F}_1 & \geqslant \mathcal{F}_{k+1}+\frac{1}{2}\left\|x^{k+1}-x^k\right\|^2_{\mathbf{M}_3^k-C_0\mathbf{Id}}+\frac{1}{2}\left\|z^{k+1}-z^k\right\|^2_{\mathbf{M}_2^k} \\ & = h\left(x^{k+1}\right)+g\left(z^{k+1}\right)-\frac{1}{2r}\left\|y^{k+1}\right\|^2+\frac{r}{2}\left\|Ax^{k+1}-z^{k+1}+\frac{1}{r}y^{k+1}\right\|^2 \\ & \quad +T_0\left\|A^*\left(y^{k+1}-y^k\right)\right\|^2+\frac{1}{2}\left\|x^{k+1}-x^k\right\|^2_{\mathbf{M}_3^k-C_0\mathbf{Id}}+\frac{1}{2}\left\|z^{k+1}-z^k\right\|^2_{\mathbf{M}_2^k}+\frac{C_0}{2}\|x^{k+1}-x^k\|, \end{aligned}$$

which, thanks to (45), leads to

$$\begin{aligned} \mathcal{F}_1 \geqslant\ & h\left(x^{k+1}\right)+g\left(z^{k+1}\right)-\frac{T_1}{r}\left\|\nabla h\left(x^{k+1}\right)\right\|^2+\frac{r}{2}\left\|Ax^{k+1}-z^{k+1}+\frac{1}{r}y^{k+1}\right\|^2 \\ & +\frac{T_0}{2}\left\|A^*\left(y^{k+1}-y^k\right)\right\|^2+\frac{1}{2}\left\|x^{k+1}-x^k\right\|^2_{\mathbf{M}_3^k-C_0\mathbf{Id}}+\frac{1}{2}\left\|z^{k+1}-z^k\right\|^2_{\mathbf{M}_2^k}+\frac{C_0}{4}\|x^{k+1}-x^k\|^2. \end{aligned} \tag{56}$$

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 $r \geqslant (2+\gamma) T_1 L > 2T_1 L > 0$, there exists $\sigma > 0$ such that

$$\frac{1}{\sigma} - \frac{L}{2\sigma^2} = \frac{T_1}{r}.$$

From Proposition 1 and (56) we see that for every $k \geqslant 1$

$$\begin{aligned} & g\left(z^{k+1}\right) + \frac{r}{2}\left\|Ax^{k+1} - z^{k+1} + \frac{1}{r}y^{k+1}\right\|^2 + \frac{C_0}{4}\|x^{k+1} - x^k\|^2 \\ \leqslant\ & \mathcal{F}_1 - \inf_{x\in\mathbb{R}^n}\left\{h(x) - \frac{T_1}{r}\|\nabla h(x)\|^2\right\} < +\infty. \end{aligned}$$

Since g is coercive, it follows that the sequences $\left\{z^k\right\}_{k\geqslant 0}$, $\left\{Ax^k - z^k + r^{-1}y^k\right\}_{k\geqslant 0}$ and $\left\{x^{k+1} - x^k\right\}_{k\geqslant 0}$ are bounded. This implies that $\left\{A(x^{k+1} - x^k) - (z^{k+1} - z^k)\right\}_{k\geqslant 0}$ is bounded, from which we obtain the boundedness of $\left\{r^{-1}(y^{k+1} - y^k)\right\}_{k\geqslant 0}$. According to the third update in the iterative scheme, we obtain that $\left\{Ax^k - z^k\right\}_{k\geqslant 0}$ and thus $\left\{y^k\right\}_{k\geqslant 0}$ are 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 (18) there exists $\sigma > 0$ such that

$$\frac{1}{\sigma} - \frac{L}{2\sigma^2} = (1+\gamma)\frac{T_1}{r}.$$

We assume first that $\rho \neq 1$ or, equivalently, $T_0 \neq 0$. From Proposition 1 and (56) we see that for every $k \geqslant 1$

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

Since h is coercive, 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

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

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, $T_0 = 0$, we have that for every $k \geqslant 1$

$$\begin{aligned} & \left(1 - \frac{1}{\gamma}\right) h\left(x^{k+1}\right) + \frac{T_1}{2\gamma r}\left\|\nabla h\left(x^{k+1}\right)\right\|^2 + \frac{r}{2}\left\|Ax^{k+1} - z^{k+1} + \frac{1}{r}y^{k+1}\right\|^2 \\ \leqslant\ & \mathcal{F}_1 - g\left(z^{k+1}\right) - \frac{1}{\gamma}\inf_{x\in\mathbb{R}^n}\left\{h(x) - (1+\gamma)\frac{T_1}{2r}\|\nabla h(x)\|^2\right\} < +\infty, \end{aligned}$$

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) 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. □

We state now the first convergence result of this paper.

Theorem 7. *Let Assumption 1 be satisfied and $\left\{\left(x^k, z^k, y^k\right)\right\}_{k\geqslant 0}$ 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

$$\mathcal{F}_{k+1} + \frac{C_0}{4}\left\|x^{k+1} - x^k\right\|^2 + \frac{1}{2}\left\|z^{k+1} - z^k\right\|^2_{\mathbf{M}_2^k} \leqslant \mathcal{F}_k. \tag{57}$$

(ii) The sequence $\{\mathcal{F}_k\}_{k\geqslant 0}$ is bounded from below and convergent. Moreover,

$$x^{k+1} - x^k \to 0,\ z^{k+1} - z^k \to 0 \text{ and } y^{k+1} - y^k \to 0 \text{ as } k \to +\infty. \tag{58}$$

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

Proof. (i) According to (19) we have that $\mathbf{M}_3^k - C_0\mathbf{Id} \in \mathcal{P}^n_{\frac{C_0}{2}}$ and thus (55) implies (57).

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

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

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 (31), (33), (41) and (43) that for every $k \geqslant 1$

$$\begin{aligned}\frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2 \leqslant{}& \frac{C_1 - L}{2}\left\|x^{k+1} - x^k\right\|^2 + \frac{C_0}{2}\left\|x^k - x^{k-1}\right\|^2\\ &+ T_0\left\|A^*(y^k - y^{k-1})\right\|^2 - T_0\left\|A^*(y^{k+1} - y^k)\right\|^2.\end{aligned}$$

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

$$\begin{aligned}\frac{1}{\rho r}\sum_{k=1}^{K}\left\|y^{k+1} - y^k\right\|^2 \leqslant{}& \frac{C_1 - L}{2}\sum_{k=1}^{K}\left\|x^{k+1} - x^k\right\|^2 + \frac{C_0}{2}\sum_{k=1}^{K}\left\|x^k - x^{k-1}\right\|^2\\ &+ T_0\left\|A^*(y^1 - y^0)\right\|^2 - T_0\left\|A^*(y^{K+1} - y^K)\right\|^2\\ \leqslant{}& \frac{C_1 - L}{2}\sum_{k=1}^{K}\left\|x^{k+1} - x^k\right\|^2 + \frac{C_0}{2}\sum_{k=1}^{K}\left\|x^k - x^{k-1}\right\|^2 + T_0\left\|A^*(y^1 - y^0)\right\|^2.\end{aligned}$$

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

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

□

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. *Let Assumption 1 be satisfied and* $\{(x^k, z^k, y^k)\}_{k\geqslant 0}$ *be a sequence generated by Algorithm 1 or Algorithm 2. For every* $k \geqslant 0$ *we have*

$$d^{k+1} := \left(d_x^{k+1}, d_z^{k+1}, d_y^{k+1}\right) \in \partial\mathcal{L}_r\left(x^{k+1}, z^{k+1}, y^{k+1}\right), \tag{59}$$

where

$$d_x^{k+1} := C_2\left(\nabla h(x^{k+1}) - \nabla h(x^k)\right) + A^*(y^{k+1} - y^k) + \mathbf{M}_1^k(x^k - x^{k+1}), \tag{60a}$$

$$d_z^{k+1} := y^k - y^{k+1} + rA(x^k - x^{k+1}) + \mathbf{M}_2^k(z^k - z^{k+1}), \tag{60b}$$

$$d_y^{k+1} := \frac{1}{\rho r}\left(y^{k+1} - y^k\right). \tag{60c}$$

and

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

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

$$|||d^{k+1}||| \leqslant C_5 \left\| x^{k+1} - x^k \right\| + C_6 \left\| z^{k+1} - z^k \right\| + C_7 \left\| y^{k+1} - y^k \right\|, \tag{61}$$

where

$$C_5 := C_2 L + \mu_1 + r \left\| A \right\|, \quad C_6 := \mu_2, \quad C_7 := 1 + \left\| A \right\| + \frac{1}{\rho r}. \tag{62}$$

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

$$\nabla_x \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1} \right) = \nabla h \left(x^{k+1} \right) + A^* y^{k+1} + r A^* \left(A x^{k+1} - z^{k+1} \right), \tag{63a}$$

$$\partial_z \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1} \right) = \partial g \left(z^{k+1} \right) - y^{k+1} - r \left(A x^{k+1} - z^{k+1} \right), \tag{63b}$$

$$\nabla_y \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1} \right) = A x^{k+1} - z^{k+1}. \tag{63c}$$

Then (60c) follows directly from (63c) and (16c), respectively, (17c), while (60b) 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 (16a) and (17a), respectively. In order to derive (60a), let us notice that for Algorithm 1 we have (see (24))

$$-A^* y^k + \mathbf{M}_1^k \left(x^k - x^{k+1} \right) = \nabla h \left(x^{k+1} \right) + r A^* \left(A x^{k+1} - z^{k+1} \right), \tag{64}$$

while for Algorithm 2 we have (see (34))

$$-\nabla h \left(x^k \right) - A^* y^k + \mathbf{M}_1^k \left(x^k - x^{k+1} \right) = r A^* \left(A x^{k+1} - z^{k+1} \right). \tag{65}$$

By using (63a) we get the desired statement.

Relation (61) follows by combining the inequalities

$$\begin{aligned} \left\| d_x^{k+1} \right\| &\leqslant \left(C_2 L + \mu_1 \right) \left\| x^{k+1} - x^k \right\| + \left\| A \right\| \cdot \left\| y^{k+1} - y^k \right\|, \\ \left\| d_z^{k+1} \right\| &\leqslant \left\| y^k - y^{k+1} \right\| + r \left\| A \right\| \cdot \left\| x^{k+1} - x^k \right\| + \mu_2 \left\| z^{k+1} - z^k \right\|. \end{aligned}$$

with (5). □

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

$$D^{k+1} := \left(D_x^{k+1}, D_z^{k+1}, D_y^{k+1}, D_{x'}^{k+1}, D_{y'}^{k+1} \right) \in \partial \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k \right) \tag{66}$$

where

$$\begin{aligned} D_x^{k+1} &:= d_x^{k+1} + C_0 \left(x^{k+1} - x^k \right), \quad D_z^{k+1} := d_z^{k+1}, \quad D_y^{k+1} := d_y^{k+1} + 2T_0 AA^* \left(y^{k+1} - y^k \right), \\ D_{x'}^{k+1} &:= -C_0 \left(x^{k+1} - x^k \right), \quad D_{y'}^{k+1} := -2T_0 AA^* \left(y^{k+1} - y^k \right). \end{aligned} \tag{67}$$

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

$$|||D^{k+1}||| \leqslant C_8 \left\| x^{k+1} - x^k \right\| + C_9 \left\| z^{k+1} - z^k \right\| + C_{10} \left\| y^{k+1} - y^k \right\|, \tag{68}$$

where

$$C_8 := C_5 + 2C_0, \quad C_9 := C_6, \quad C_{10} := C_7 + 4T \left\| A \right\|^2. \tag{69}$$

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

$$\nabla_x \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k \right) := \nabla_x \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1} \right) + C_0 \left(x^{k+1} - x^k \right), \tag{70a}$$

$$\partial_z \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k \right) := \partial_z \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1} \right) \tag{70b}$$

$$\nabla_y \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k \right) := \nabla_y \mathcal{L}_r \left(x^{k+1}, z^{k+1}, y^{k+1} \right) + 2T_0 AA^* \left(y^{k+1} - y^k \right), \tag{70c}$$

$$\nabla_{x'} \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k \right) := -C_0 \left(x^{k+1} - x^k \right), \tag{70d}$$

$$\nabla_{y'} \mathcal{F}_r \left(x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k \right) := -2T_0 AA^* \left(y^{k+1} - y^k \right), \tag{70e}$$

Then (66) follows directly from the above relations and (59). Inequality (68) follows by combining

$$\begin{aligned}\left\|D_x^{k+1}\right\| &\leqslant \left\|d_x^{k+1}\right\| + C_0\left\|x^{k+1}-x^k\right\|,\\ \left\|D_y^{k+1}\right\| &\leqslant \left\|d_y^{k+1}\right\| + 2T\left\|A\right\|^2\cdot\left\|y^{k+1}-y^k\right\|.\end{aligned}$$

with (5). □

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

Corollary 10. *Let Assumption 1 be satisfied and $\left\{\left(x^k, z^k, y^k\right)\right\}_{k\geqslant 0}$ 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*

$$\begin{aligned}|||D^{k+1}||| \leqslant\ & C_{11}\left(\left\|x^{k+1}-x^k\right\| + \left\|x^k-x^{k-1}\right\| + \left\|x^{k-1}-x^{k-2}\right\|\right)\\ & + C_{12}\left(\left\|A^*\left(y^k-y^{k-1}\right)\right\| - \left\|A^*\left(y^{k+1}-y^k\right)\right\|\right)\\ & + C_{13}\left(\left\|A^*\left(y^{k-1}-y^{k-2}\right)\right\| - \left\|A^*\left(y^k-y^{k-1}\right)\right\|\right), \end{aligned}\tag{71}$$

where

$$\begin{aligned}C_{11} &:= \max\left\{C_8 + C_9\left\|A\right\| + C_3C_{10} + \frac{C_3C_9}{\rho r}, C_4C_{10} + \frac{C_3C_9}{\rho r}, \frac{C_4C_9}{\rho r}\right\},\\ C_{12} &:= \left(C_{10} + \frac{C_9}{\rho r}\right)T_2, C_{13} := \frac{C_9T_2}{\rho r}.\end{aligned}\tag{72}$$

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

Lemma 11. *Let Assumption 1 be satisfied and $\left\{\left(x^k, z^k, y^k\right)\right\}_{k\geqslant 0}$ 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 $q\to+\infty$, then

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

(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 $\lim_{k\to+\infty}\operatorname{dist}\left[\left(x^k, z^k, y^k\right), \omega\left(\left\{\left(x^k, z^k, y^k\right)\right\}_{k\geqslant 0}\right)\right] = 0$;

(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 $\mathcal{F}_* = \lim_{k\to+\infty}\mathcal{L}_r\left(x^k, z^k, y^k\right)$, as the objective function $g\circ A + h$ does on $\mathbf{Pr}_{\mathbb{R}^n}\left[\omega\left(\left\{\left(x^k, z^k, y^k\right)\right\}_{k\geqslant 0}\right)\right]$.

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 $q\to+\infty$.

(i) From either (16a) or (17a) we obtain for all $q\geqslant 1$

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

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

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

(ii) For the sequence $\left\{d^k\right\}_{k\geqslant 0}$ defined in (60a) - (60c), 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 $\left(x^{k_q}, z^{k_q}, y^{k_q}\right) \to \left(\widehat{x}, \widehat{z}, \widehat{y}\right)$ 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, $\left(\widehat{x}, \widehat{z}, \widehat{y}\right) \in \operatorname{crit}\left(\mathcal{L}_r\right)$. Choosing now an element $\left(\widehat{x}, \widehat{z}, \widehat{y}\right) \in \operatorname{crit}\left(\mathcal{L}_r\right)$, 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 [9], also by taking into consideration [9, 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 (58).

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

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

$$0 = A^*\partial g\left(A\widehat{x}\right) + \nabla h\left(\widehat{x}\right). \tag{73}$$

When A is injective this is further equivalent to

$$0 \in \partial(g \circ A)(\widehat{x}) + \nabla h\left(\widehat{x}\right) = \partial\left(g \circ A + h\right)\left(\widehat{x}\right), \tag{74}$$

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

On the other hand, when the functions g and h are convex, then (73) and (74) 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. *Let Assumption 1 be satisfied and* $\left\{\left(x^k, z^k, y^k\right)\right\}_{k\geqslant 0}$ *be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. Denote by* $\Omega := \omega\left(\left\{\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right\}_{k\geqslant 1}\right)$. *The following statements are true:*

(i) it holds

$$\Omega \subseteq \left\{\left(\widehat{x}, \widehat{z}, \widehat{y}, \widehat{x}, \widehat{y}\right) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m :\ \left(\widehat{x}, \widehat{z}, \widehat{y}\right) \in \operatorname{crit}\left(\mathcal{L}_r\right)\right\};$$

(ii) we have

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

(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* $\mathbf{Pr}_{\mathbb{R}^n}\Omega$.

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 *KŁ property.* The origins of this notion go back to the pioneering work of Kurdyka who introduced in [28] a general form of the Łojasiewicz inequality ([33]). A further extension to the nonsmooth setting has been proposed and studied in [6, 7, 8].

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

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 $\widehat{u} \in \operatorname{dom}\partial\Psi := \left\{u \in \mathbb{R}^N : \partial\Psi(u) \neq \varnothing\right\}$, if there exists $\eta \in (0, +\infty]$, a neighborhood U of $\widehat{u}$ and a function $\varphi \in \Phi_\eta$ such that for every
$$u \in U \cap \left[\Psi(\widehat{u}) < \Psi(u) < \Psi(\widehat{u}) + \eta\right]$$
the following inequality holds
$$\varphi'\left(\Psi(u) - \Psi(\widehat{u})\right) \cdot \operatorname{dist}\left(\mathbf{0}, \partial\Psi(u)\right) \geqslant 1.$$
2. If Ψ satisfies the KŁ property at each point of $\operatorname{dom}\partial\Psi$, 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, 6, 7, 8, 9] and to the references therein for more properties of KŁ functions and illustrating examples.

The following result, taken from [9, 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*
$$\left\{u \in \mathbb{R}^N : \operatorname{dist}(u, \Omega) < \varepsilon\right\} \cap \left[\Psi(\widehat{u}) < \Psi(u) < \Psi(\widehat{u}) + \eta\right]$$
it holds
$$\varphi'\left(\Psi(u) - \Psi(\widehat{u})\right) \cdot \operatorname{dist}\left(\mathbf{0}, \partial\Psi(u)\right) \geqslant 1.$$

Working in the hypotheses of Lemma 12, we define for every $k \geqslant 1$
$$\mathcal{E}_k := \mathcal{F}\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right) - \mathcal{F}_* = \mathcal{F}_k - \mathcal{F}_* \geqslant 0, \tag{75}$$
where $\mathcal{F}_*$ is the limit of $\left\{\mathcal{F}_k\right\}_{k\geqslant 1}$ as $k \to +\infty$. The sequence $\left\{\mathcal{E}_k\right\}_{k\geqslant 1}$ is monotonically decreasing and it converges to 0 as $k \to +\infty$.

The next result shows that when 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. *Let Assumption 1 be satisfied and $\left\{\left(x^k, z^k, y^k\right)\right\}_{k\geqslant 0}$ 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,
$$\sum_{k\geqslant 0}\left\|x^{k+1} - x^k\right\| < +\infty, \qquad \sum_{k\geqslant 0}\left\|z^{k+1} - z^k\right\| < +\infty, \qquad \sum_{k\geqslant 0}\left\|y^{k+1} - y^k\right\| < +\infty; \tag{76}$$

(ii) the sequence $\{(x^k, z^k, y^k)\}_{k \geqslant 0}$ *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 $\mathcal{F}_r(\widehat{x}, \widehat{z}, \widehat{y}, \widehat{x}, \widehat{y}) = \mathcal{F}_*$. 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.

First we assume 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 inequality (57) 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 $\sum_{k \geqslant 0} \|y^{k+1} - y^k\| < +\infty$. Finally, by using Lemma 5 (i), we get that $\sum_{k \geqslant 0} \|z^{k+1} - z^k\| < +\infty$.

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

$$\begin{aligned}&\{u \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m : \operatorname{dist}(u, \Omega) < \varepsilon\} \cap \\ &\{u \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m : \mathcal{F}_* < \mathcal{F}_r(u) < \mathcal{F}_* + \eta\}\end{aligned} \tag{77}$$

it holds

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

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

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

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

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

Thus, $\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)$ belongs to the intersection in (77) for all $k \geqslant k_0 := \max\{k_1, k_2, 3\}$, which further implies

$$\varphi'(\mathcal{F}_k - \mathcal{F}_*) \cdot \operatorname{dist}\left(\mathbf{0}, \partial \mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right) = \varphi'(\mathcal{E}_k) \cdot \operatorname{dist}\left(\mathbf{0}, \partial \mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right) \geqslant 1. \tag{78}$$

Define for two arbitrary nonnegative integers p and q

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

Then for all $K \geqslant k_0 \geqslant 1$ it holds

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

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 all $k \geqslant 1$

$$\Delta_{k,k+1} = \varphi(\mathcal{E}_k) - \varphi(\mathcal{E}_{k+1}) \geqslant \varphi'(\mathcal{E}_k)[\mathcal{E}_k - \mathcal{E}_{k+1}] = \varphi'(\mathcal{E}_k)[\mathcal{F}_k - \mathcal{F}_{k+1}] \geqslant \varphi'(\mathcal{E}_k) \frac{C_0}{4} \|x^{k+1} - x^k\|^2. \tag{79}$$

The last relation combined with (78) imply for all $k \geqslant k_0$

$$\begin{aligned}\|x^{k+1} - x^k\|^2 &\leqslant \varphi'(\mathcal{E}_k) \cdot \operatorname{dist}\left(\mathbf{0}, \partial \mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right) \|x^{k+1} - x^k\|^2 \\ &\leqslant \frac{4}{C_0} \Delta_{k,k+1} \cdot \operatorname{dist}\left(\mathbf{0}, \partial \mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right).\end{aligned}$$

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

$$\begin{aligned}
\left\|x^{k+1} - x^k\right\| &\leqslant \sqrt{\frac{4}{C_0}\Delta_{k,k+1} \cdot \operatorname{dist}\left(\mathbf{0}, \partial\mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right)} \\
&\leqslant \frac{\beta}{C_0}\Delta_{k,k+1} + \frac{1}{\beta}\operatorname{dist}\left(\mathbf{0}, \partial\mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right) \\
&\leqslant \frac{\beta}{C_0}\Delta_{k,k+1} + \frac{C_{11}}{\beta}\left(\left\|x^k - x^{k-1}\right\| + \left\|x^{k-1} - x^{k-2}\right\| + \left\|x^{k-2} - x^{k-3}\right\|\right) \\
&\quad + \frac{C_{12}}{\beta}\left(\left\|A^*\left(y^{k-1} - y^{k-2}\right)\right\| - \left\|A^*\left(y^k - y^{k-1}\right)\right\|\right) \\
&\quad + \frac{C_{13}}{\beta}\left(\left\|A^*\left(y^{k-2} - y^{k-3}\right)\right\| - \left\|A^*\left(y^{k-1} - y^{k-2}\right)\right\|\right).
\end{aligned} \tag{80}$$

We denote for every $k \geqslant 3$

$$\begin{aligned}
a^k &:= \left\|x^k - x^{k-1}\right\| \geqslant 0, \\
\delta_k &:= \frac{\beta}{C_0}\Delta_{k,k+1} + \frac{C_{12}}{\beta}\left(\left\|A^*\left(y^{k-1} - y^{k-2}\right)\right\| - \left\|A^*\left(y^k - y^{k-1}\right)\right\|\right) \\
&\quad + \frac{C_{13}}{\beta}\left(\left\|A^*\left(y^{k-2} - y^{k-3}\right)\right\| - \left\|A^*\left(y^{k-1} - y^{k-2}\right)\right\|\right).
\end{aligned}$$

The inequality (80) is nothing than (13) with $c_0 = c_1 = c_2 := \dfrac{C_{11}}{\beta}$. Observe that for every $K \geqslant k_0$ we have

$$\sum_{k=k_0}^{K} \delta_k \leqslant \frac{\beta}{C_0}\varphi\left(\mathcal{E}_{k_0}\right) + \frac{C_{12}}{\beta}\left\|A^*\left(y^{k_0-1} - y^{k_0-2}\right)\right\| + \frac{C_{13}}{\beta}\left\|A^*\left(y^{k_0-2} - y^{k_0-3}\right)\right\|$$

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 (76) 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 $(\widehat{x}, \widehat{z}, \widehat{y})$ which is, according to Lemmas 11, a KKT point of the optimization problem (1). □

Remark 4. The function $\mathcal{F}_r$ is a KŁ 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, [33]).

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

$$\left|\Psi\left(u\right) - \Psi\left(\widehat{u}\right)\right|^{\theta} \leqslant C_L \cdot \operatorname{dist}(0, \partial\Psi(u)) \ \forall u \in \operatorname{Ball}\left(\widehat{u}, \varepsilon\right), \tag{81}$$

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

Providing that the Assumption 1 is fulfilled and $\{(x^k, z^k, y^k)\}_{k\geqslant 0}$ is the sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded, we have seen in Lemma 12 that the set of cluster points $\Omega = \omega\left(\{(x^k, z^k, y^k, x^{k-1}, y^{k-1})\}_{k\geqslant 0}\right)$ is nonempty, compact and connected and $\mathcal{F}_r$ takes on Ω the value $\mathcal{F}_*$; moreover for any $(\widehat{x}, \widehat{z}, \widehat{y}, \widehat{x}, \widehat{y}) \in \Omega$, $(\widehat{x}, \widehat{z}, \widehat{y})$ 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 any

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

it holds

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

Obviously, $\mathcal{F}_r$ is a KŁ function with desingularization function $\varphi : [0, +\infty) \to [0, +\infty)$ such that $\varphi(s) := \frac{1}{1-\theta} C_L s^{1-\theta}$, 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 $\{(x^k, z^k, y^k, x^{k-1}, y^{k-1})\}_{k\geqslant 0}$ 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

$$\left|\mathcal{F}_r(x, z, y, x', y') - \mathcal{F}_*\right|^\theta \leqslant C_L \cdot \operatorname{dist}\left(0, \partial\mathcal{F}_r(x, z, y, x', y')\right) \ \forall (x, z, y, x', y') \in \operatorname{Ball}\left((\widehat{x}, \widehat{z}, \widehat{y}, \widehat{x}, \widehat{y}), \varepsilon\right). \tag{82}$$

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

The following lemma will provides 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 $k \geqslant k_0$*

$$e_{k-l_0} - e_k \geqslant C_e e_k^{2\theta}, \tag{83}$$

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 $k \geqslant k_0 + l_0$

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

Proof. Fix an integer $k \geqslant k_0$. Since $k_0 \geqslant l_0 \geqslant 0$, the recurrence inequality (83) 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 (83) we get

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

for every $k \geqslant k_0$, which actually leads to contradiction to 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 converges 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 $e_k \leqslant e_0$, thus $e_0^{2\theta-1} e_k \leqslant e_k^{2\theta}$, which leads to

$$e_{k-l_0} - e_k \geqslant C_e e_k^{2\theta} \geqslant C_e e_0^{2\theta-1} e_k \ \forall k \geqslant k_0.$$

Therefore

$$\begin{aligned} e_k &\leqslant \frac{1}{C_e e_0^{2\theta-1} + 1} e_{k-l} \leqslant \cdots \leqslant \left(\frac{1}{C_e e_0^{2\theta-1} + 1}\right)^{\left\lfloor \frac{k-k_0}{l_0} \right\rfloor} \max\{e_{k_0+j} \colon j = 0, \ldots, l_0\} \\ &\leqslant \left(\frac{1}{C_e e_0^{2\theta-1} + 1}\right)^{\frac{k}{l_0} - \frac{k_0}{l_0} - 1} e_0 = e_0\left(C_e e_0^{2\theta-1} + 1\right)^{\frac{k_0}{l_0} + 1} \left(\frac{1}{\sqrt[l_0]{C_e e_0^{2\theta-1} + 1}}\right)^k, \end{aligned}$$

where $\lfloor p \rfloor$ denotes the greatest integer that is less than or equal to the real number p. This provides the linear convergence rate, as $\frac{1}{\sqrt[l_0]{C_e e_0^{2\theta-1}+1}} \in [0,1)$.

(iii) The case when $\theta \in (1/2, 1)$. From (83) we get

$$C_e \leqslant (e_{k-l_0} - e_k)\, e_k^{-2\theta}. \tag{84}$$

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 (84) gives

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

  or, equivalently,

$$e_k^{1-2\theta} - e_{k-l_0}^{1-2\theta} \geqslant C_1', \quad \text{where } C_1' := \frac{(2\theta-1)\,C_e}{2} > 0. \tag{85}$$

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

$$\nu e_{k-l_0} \geqslant e_k \Leftrightarrow \nu^{1-2\theta} e_{k-l_0}^{1-2\theta} \leqslant e_k^{1-2\theta} \Leftrightarrow \left(\nu^{1-2\theta} - 1\right) e_{k-l_0}^{1-2\theta} \leqslant e_k^{1-2\theta} - e_{k-l_0}^{1-2\theta}.$$

  Recall that $\nu^{1-2\theta} - 1 > 0$, since $1-2\theta < 0$, and $e_0^{1-2\theta} \leqslant e_{k-l_0}^{1-2\theta}$, since $\{e_k\}_{k\geqslant 0}$ is monotonically decreasing, and thus

$$e_k^{1-2\theta} - e_{k-l_0}^{1-2\theta} \geqslant \left(\nu^{1-2\theta} - 1\right) e_{k-l_0}^{1-2\theta} \geqslant C_2', \quad \text{where } C_2' := \left(\nu^{1-2\theta} - 1\right) e_0^{2\theta-1} > 0. \tag{86}$$

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

$$e_i^{1-2\theta} - e_{i-l_0}^{1-2\theta} \geqslant C' := \min\left\{C_1', C_2'\right\} > 0, \tag{87}$$

where C_1' and C_2' are defined as in (85) and (86), respectively. For every $k \geqslant k_0 + 2l_0$, by summing up the inequalities (87) for $i = k_0 + l_0, \cdots, k$, we get

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

Using the fact that $1 - 2\theta < 0$ and the monotonicity of $\{e_i\}_{i\geqslant 0}$, it yields

$$e_{k_0+l_0} \leqslant \cdots \leqslant e_{k_0} \Leftrightarrow e_{k_0+l_0}^{1-2\theta} \geqslant \cdots \geqslant e_{k_0}^{1-2\theta} \Leftrightarrow -e_{k_0}^{1-2\theta} \geqslant \cdots \geqslant -e_{k_0+l_0}^{1-2\theta}$$

and thus

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

which gives

$$e_k^{1-2\theta} \geqslant e_{k_0}^{1-2\theta} + \frac{k - k_0 - l_0 + 1}{l_0} C'. \tag{88}$$

Moreover, we obtain from (87) that

$$e_{k_0}^{1-2\theta} \geqslant \left\lfloor \frac{k_0 + l_0}{l_0} \right\rfloor C' \geqslant \left(\frac{k_0 + l_0}{l_0} - 1\right) C' = \frac{k_0}{l_0} C'. \tag{89}$$

By plugging (89) into (88) we obtain

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

which implies

$$e_k \leqslant \left(\frac{C'}{l_0}\right)^{-\frac{1}{2\theta-1}} (k-l_0+1)^{-\frac{1}{2\theta-1}}. \tag{90}$$

This concludes the proof. □

Remark 5. The inequality in Lemma 15 (iii) can be writen in term of k instead of $k - l_0 + 1$ when k large enough. For instance, when $k \geqslant \frac{\gamma'}{\gamma'-1}(l_0+1)$ for some $\gamma' > 1$ then we have that $k - l_0 + 1 \geqslant \frac{1}{\gamma'} k$ and thus from (90) we get

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

3.2 Convergence rates

In this subsection we will study the convergence rates of Algorithm 1 and 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* (19) *which is replaced by*

$$\mathbf{M}_3^k := 2\mathbf{M}_1^k + rA^*A + (L - 2C_1)\,\mathbf{Id} \geqslant \frac{5}{2} C_0 \mathbf{Id} \quad \forall k \geqslant 0, \tag{91}$$

Notice that (91) can be written as

$$2\mathbf{M}_1^k + rA^*A - \left(L + r^{-1}C'_{\mathbf{M}}\right)\mathbf{Id} \geqslant 0 \;\forall k \geqslant 0, \text{ where } C'_{\mathbf{M}} := \begin{cases} \left(10\mu_1^2 + 8\left(L+\mu_1\right)^2\right) T_1, & \text{for Algorithm 1,} \\ \left(8\mu_1^2 + 10\left(L+\mu_1\right)^2\right) T_1, & \text{for Algorithm 2.} \end{cases} \tag{92}$$

Therefore (92) is nothing else than (20) after replacing $C_{\mathbf{M}}$ by the bigger constant $C'_{\mathbf{M}}$. So, all the examples in Remark 2 can be adapted to the new setting and provide frameworks which guarantee Assumption 2. The scenarios which ensure Assumption 2 evidently satisfy Assumption 1, 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 can be obtained thanks to Assumptions 2 by using similar techniques.

Firstly, by the same arguments as in Lemma 4, we have that for every $k \geqslant 1$ (see (22))

$$\begin{aligned} &\mathcal{L}_r\left(x^{k+1}, z^{k+1}, y^{k+1}\right) + \frac{1}{2}\left\|x^{k+1} - x^k\right\|^2_{2\mathbf{M}_1^k + rA^*A} - \frac{L}{2}\left\|x^{k+1} - x^k\right\|^2 + \frac{1}{2}\left\|z^{k+1} - z^k\right\|^2_{\mathbf{M}_2^k} \\ &\leqslant \mathcal{L}_r\left(x^k, z^k, y^k\right) + \frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2 \end{aligned} \tag{93}$$

and (see (31), (33), (41) and (43))

$$\begin{aligned} \frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2 \leqslant & \frac{C_1 - L}{2}\left\|x^{k+1} - x^k\right\|^2 + \frac{C_0}{2}\left\|x^k - x^{k-1}\right\|^2 + \\ & T_0\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 - T_0\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2. \end{aligned} \tag{94}$$

By multiplying (94) by 2 and by adding the resulting inequality to (93), we obtain for every $k \geqslant 1$

$$\begin{aligned} &\mathcal{L}_r\left(x^{k+1}, z^{k+1}, y^{k+1}\right) + \frac{1}{2}\left\|x^{k+1} - x^k\right\|^2_{\mathbf{M}_3^k} + \frac{1}{2}\left\|z^{k+1} - z^k\right\|^2_{\mathbf{M}_2^k} + \\ &2T_0\left\|A^*\left(y^{k+1} - y^k\right)\right\|^2 + \frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2 \\ &\leqslant \mathcal{L}_r\left(x^k, z^k, y^k\right) + 2T_0\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 + C_0\left\|x^k - x^{k-1}\right\|^2. \end{aligned} \tag{95}$$

We replace T_0 with $2T_0$ in the definition of the regularized augmented Lagrangian $\mathcal{F}_r$, which means that the sequence $\{\mathcal{F}_k\}_{k\geqslant 1}$ in (54) becomes now

$$\mathcal{F}_k := \mathcal{L}_r\left(x^k, z^k, y^k\right) + 2T_0\left\|A^*\left(y^k - y^{k-1}\right)\right\|^2 + C_0\left\|x^k - x^{k-1}\right\|^2 \quad \forall k \geqslant 1.$$

In this new context, the inequality (95) gives us for every $k \geqslant 1$

$$\mathcal{F}_{k+1} + \frac{C_0}{4}\left\|x^{k+1} - x^k\right\|^2 + \frac{1}{2}\left\|z^{k+1} - z^k\right\|^2_{\mathbf{M}_2^k} + \frac{1}{\rho r}\left\|y^{k+1} - y^k\right\|^2 \leqslant \mathcal{F}_k \tag{96}$$

and provides an inequality which is tighter than relation (57) 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 (67) (again by replacing T_0 by $2T_0$) we obtain for every $k \geqslant 2$ the following estimate, which is simpler than the estimate (71) in Corollary 10,

$$|||D^{k+1}||| \leqslant C_{14}\left\|x^{k+1} - x^k\right\| + C_{15}\left\|y^{k+1} - y^k\right\| + C_{16}\left\|y^k - y^{k-1}\right\|, \tag{97}$$

where

$$C_{14} := C_8 + C_9\|A\|, \quad C_{15} := C_{10} + \frac{C_9}{\rho r}, \quad C_{16} := \frac{C_9}{\rho r}. \tag{98}$$

This improvement provides instead of inequality (79) in the proof of Theorem 14 the following very useful estimate

$$\begin{aligned}\Delta_{k,k+1} &= \varphi(\mathcal{E}_k) - \varphi(\mathcal{E}_{k+1}) \geqslant \varphi'(\mathcal{E}_k)\left[\mathcal{E}_k - \mathcal{E}_{k+1}\right] \geqslant \min\left\{\frac{C_0}{4}, \frac{1}{\rho r}\right\}\left(\left\|x^{k+1} - x^k\right\|^2 + \left\|y^{k+1} - y^k\right\|^2\right)\\ &\geqslant C_{17}\left(\left\|x^{k+1} - x^k\right\| + \left\|y^{k+1} - y^k\right\|\right)^2,\end{aligned}$$

where

$$C_{17} := \frac{1}{2}\min\left\{\frac{C_0}{4}, \frac{1}{\rho r}\right\}. \tag{99}$$

The last relation together with (78) imply for all $k \geqslant k_0$

$$\begin{aligned}&\left(\left\|x^{k+1} - x^k\right\| + \left\|y^{k+1} - y^k\right\|\right)^2\\ &\leqslant \varphi'(\mathcal{E}_k)\cdot \operatorname{dist}\left(\mathbf{0}, \partial\mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right)\left(\left\|x^{k+1} - x^k\right\| + \left\|y^{k+1} - y^k\right\|\right)^2\\ &\leqslant \frac{\Delta_{k,k+1}}{C_{17}}\cdot \operatorname{dist}\left(\mathbf{0}, \partial\mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right).\end{aligned}$$

By the arithmetic mean-geometric mean inequality and (97) we have that for every $k \geqslant k_0$ and every $\beta > 0$

$$\begin{aligned}&\left\|x^{k+1} - x^k\right\| + \left\|y^{k+1} - y^k\right\|\\ &\leqslant \sqrt{\frac{\Delta_{k,k+1}}{C_{17}}\cdot \operatorname{dist}\left(\mathbf{0}, \partial\mathcal{F}_r\left(x^k, z^k, y^k, x^{k-1}, y^{k-1}\right)\right)} \leqslant \frac{\beta\Delta_{k,k+1}}{4C_{17}} + \frac{1}{\beta}\operatorname{dist}\left(\mathbf{0}, \partial\mathcal{F}_k\right)\\ &\leqslant \frac{\beta\Delta_{k,k+1}}{4C_{17}} + \frac{\max\left\{C_{14}, C_{15}\right\}}{\beta}\left(\left\|x^k - x^{k-1}\right\| + \left\|y^k - y^{k-1}\right\| + \left\|y^{k-1} - y^{k-2}\right\|\right).\end{aligned} \tag{100}$$

By denoting

$$a^k := \left(\left\|x^k - x^{k-1}\right\|, \left\|y^k - y^{k-1}\right\|\right) \in \mathbb{R}^2_+ \quad \text{and} \quad \delta_k := \frac{\beta\Delta_{k,k+1}}{4C_{17}}, \tag{101}$$

inequality (100) can be rewritten for every $k \geqslant k_0$ as

$$\left\langle \mathbb{1}, a^{k+1}\right\rangle \leqslant \left\langle c_0, a^k\right\rangle + \left\langle c_1, a^{k-1}\right\rangle + \delta_k, \tag{102}$$

where

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

Taking $\beta > 2\max\{C_{14}, C_{15}\}$, the conclusion in (76) follows by applying first Lemma 3 and after that Lemma 5.

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

Lemma 16. *Let Assumption 2 be satisfied and $\{(x^k, z^k, y^k)\}_{k\geqslant 0}$ 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 all $k \geqslant k_0$*

$$\mathcal{E}_{k-1} - \mathcal{E}_{k+1} \geqslant C_{19}\mathcal{E}_{k+1}^{2\theta}, \quad \text{where} \quad C_{19} := \frac{\min\left\{\frac{C_0}{4}, \frac{1}{\rho r}\right\}}{3C_L^2 \max\{C_{14}, C_{15}\}^2}. \tag{103}$$

Proof. For every $k \geqslant 2$ we obtain from (96)

$$\begin{aligned}
\mathcal{E}_{k-1} - \mathcal{E}_{k+1} &= \mathcal{F}_{k-1} - \mathcal{F}_k + \mathcal{F}_k - \mathcal{F}_{k+1} \\
&\geqslant \min\left\{\frac{C_0}{4}, \frac{1}{\rho r}\right\} \left(\|x^{k+1} - x^k\|^2 + \|y^{k+1} - y^k\|^2 + \|y^k - y^{k-1}\|^2\right) \\
&\geqslant \frac{1}{3}\min\left\{\frac{C_0}{4}, \frac{1}{\rho r}\right\} \left(\|x^{k+1} - x^k\| + \|y^{k+1} - y^k\| + \|y^k - y^{k-1}\|\right)^2 \\
&\geqslant C_{19}C_L^2 |||D^{k+1}|||^2.
\end{aligned}$$

Let $\varepsilon > 0$ be such that (82) is fulfilled and choose $k_0 \geqslant 1$ such that $(x^{k+1}, z^{k+1}, y^{k+1})$ belongs to $\operatorname{Ball}((\widehat{x}, \widehat{z}, \widehat{y}), \varepsilon)$ for every $k \geqslant k_0$. Then (82) implies (103) for every $k \geqslant k_0$. □

The following convergence rates follow by combining Lemma 15 with Lemma 16.

Theorem 17. *Let Assumption 2 be satisfied and $\{(x^k, z^k, y^k)\}_{k\geqslant 0}$ 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 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 $k \geqslant k_0$

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

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 $\{(x^k, z^k, y^k)\}_{k\geqslant 0}$.

Lemma 18. *Let Assumption 2 be satisfied and $\{(x^k, z^k, y^k)\}_{k\geqslant 0}$ 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 $C_L > 0$, Łojasiewicz exponent $\theta \in [0, 1)$ and desingularization function $\varphi : [0, +\infty) \to [0, +\infty), \varphi(s) := \frac{1}{1-\theta}C_L s^{1-\theta}$. Let $(\widehat{x}, \widehat{z}, \widehat{y})$ be the KKT point of the optimization problem (1) to which $\{(x^k, z^k, y^k)\}_{k\geqslant 0}$ converges as $k \to +\infty$. Then there exists $k_0 \geqslant 2$ such that the following estimates hold for every $k \geqslant k_0$*

$$\|x^k - \widehat{x}\| \leqslant C_{20} \max\left\{\sqrt{\mathcal{E}_k}, \varphi(\mathcal{E}_k)\right\}, \quad \text{where} \quad C_{20} := \frac{7}{\sqrt{C_{17}}} + \frac{1}{C_{17}}, \tag{104a}$$

$$\|y^k - \widehat{y}\| \leqslant C_{21} \max\left\{\sqrt{\mathcal{E}_k}, \varphi(\mathcal{E}_k)\right\}, \quad \text{where} \quad C_{21} := \frac{7}{2\sqrt{C_{17}}} + \frac{1}{2C_{17}}, \tag{104b}$$

$$\|z^k - \widehat{z}\| \leqslant C_{22} \max\left\{\sqrt{\mathcal{E}_{k-1}}, \varphi(\mathcal{E}_{k-1})\right\}, \quad \text{where} \quad C_{22} := C_{20}\|A\| + \frac{2C_{21}}{\rho r}. \tag{104c}$$

Proof. We assume that $\mathcal{E}_k > 0$ for every $k \geqslant 0$. Otherwise, the sequence $\{(x^k, z^k, y^k)\}_{k\geqslant 0}$ becomes identical to $(\widehat{x}, \widehat{z}, \widehat{y})$ beginning with a given index and the conclusion follows automatically as can be seen in the proof of Theorem 14.

Let $\varepsilon > 0$ be such that (82) is fulfilled and $k_0 \geqslant 2$ be such that $\left(x^{k+1}, z^{k+1}, y^{k+1}\right)$ belongs to $\operatorname{Ball}\left((\widehat{x}, \widehat{z}, \widehat{y}), \varepsilon\right)$ for all $k \geqslant k_0$. We fix $k \geqslant k_0$. One can easily notice that

$$\left\|x^k - \widehat{x}\right\| \leqslant \left\|x^{k+1} - x^k\right\| + \left\|x^{k+1} - \widehat{x}\right\| \leqslant \cdots \leqslant \sum_{l \geqslant k} \left\|x^{l+1} - x^l\right\|. \tag{105a}$$

Similarly, we derive

$$\left\|z^k - \widehat{z}\right\| \leqslant \sum_{l \geqslant k} \left\|z^{l+1} - z^l\right\| \text{ and } \left\|y^k - \widehat{y}\right\| \leqslant \sum_{l \geqslant k} \left\|y^{l+1} - y^l\right\|. \tag{105b}$$

Recall that by the notation in (101), the inequality (100) can be written as (102). For $\beta := 3\max\{C_{14}, C_{15}\} > 2\max\{C_{14}, C_{15}\}$, thanks to Lemma 3 and the estimate (96) we have that

$$\begin{aligned}
&\sum_{l \geqslant k} \left\|x^{l+1} - x^l\right\| = \sum_{l \geqslant k} a_1^{l+1} = \sum_{l \geqslant k+1} a_1^l \\
&\leqslant \left\|x^{k+1} - x^k\right\| + 2\left\|x^{k+2} - x^{k+1}\right\| + 3\left\|x^{k+3} - x^{k+2}\right\| + 2\left\|y^{k+1} - y^k\right\| \\
&\quad + 2\left\|y^{k+2} - y^{k+1}\right\| + 3\left\|y^{k+3} - y^{k+2}\right\| + \frac{\varphi\left(\mathcal{E}_k\right)}{C_{17}} \\
&\leqslant \frac{2}{\sqrt{C_{17}}}\sqrt{\mathcal{F}_k - \mathcal{F}_{k+1}} + \frac{2}{\sqrt{C_{17}}}\sqrt{\mathcal{F}_{k+1} - \mathcal{F}_{k+2}} + \frac{3}{\sqrt{C_{17}}}\sqrt{\mathcal{F}_{k+2} - \mathcal{F}_{k+3}} + \frac{\varphi\left(\mathcal{E}_k\right)}{C_{17}} \\
&\leqslant \frac{2}{\sqrt{C_{17}}}\sqrt{\mathcal{E}_k} + \frac{2}{\sqrt{C_{17}}}\sqrt{\mathcal{E}_{k+1}} + \frac{3}{\sqrt{C_{17}}}\sqrt{\mathcal{E}_{k+2}} + \frac{\varphi\left(\mathcal{E}_k\right)}{C_{17}}
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{l \geqslant k} \left\|y^{l+1} - y^l\right\| = \sum_{l \geqslant k} a_2^{l+1} = \sum_{l \geqslant k+1} a_2^l \\
&\leqslant \frac{1}{2}\left\|x^{k+1} - x^k\right\| + \left\|x^{k+2} - x^{k+1}\right\| + \frac{3}{2}\left\|x^{k+3} - x^{k+2}\right\| + \left\|y^{k+1} - y^k\right\| \\
&\quad + \left\|y^{k+2} - y^{k+1}\right\| + \frac{3}{2}\left\|y^{k+3} - y^{k+2}\right\| + \frac{\varphi\left(\mathcal{E}_k\right)}{2C_{17}} \\
&\leqslant \frac{1}{\sqrt{C_{17}}}\sqrt{\mathcal{F}_k - \mathcal{F}_{k+1}} + \frac{1}{\sqrt{C_{17}}}\sqrt{\mathcal{F}_{k+1} - \mathcal{F}_{k+2}} + \frac{3}{2\sqrt{C_{17}}}\sqrt{\mathcal{F}_{k+2} - \mathcal{F}_{k+3}} + \frac{\varphi\left(\mathcal{E}_k\right)}{2C_{17}} \\
&\leqslant \frac{1}{\sqrt{C_{17}}}\sqrt{\mathcal{E}_k} + \frac{1}{\sqrt{C_{17}}}\sqrt{\mathcal{E}_{k+1}} + \frac{3}{2\sqrt{C_{17}}}\sqrt{\mathcal{E}_{k+2}} + \frac{\varphi\left(\mathcal{E}_k\right)}{2C_{17}}.
\end{aligned}$$

By taking into account the relations above, (105a)-(105b) as well as

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

the estimates (104a) and (104b) follow. Statement (104c) follows from Lemma 5 and by considering (105b). □

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

Theorem 19. *Let Assumption 2 be satisfied and $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. Suppose further that $\mathcal{F}_r$ satisfies the 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 $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \geqslant 0}$ converges as $k \to +\infty$. Then the following statements are true:*

(i) if $\theta = 0$, then the algorithm converges 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 $k_0' \geqslant 2$ be such that for every $k \geqslant k_0'$ the inequalities (104a)-(104c) in Lemma 18 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 (96), the sequences $\{(x^k)\}_{k \geqslant 0}$ and $\{(y^k)\}_{k \geqslant 0}$ converge also in finite time. Further, by Lemma 5, it follows that $\{(z^k)\}_{k \geqslant 0}$ converges in finite time, too. In other words, the sequence $\{(x^k, z^k, y^k)\}_{k \geqslant 0}$ becomes identical to $(\widehat{x}, \widehat{z}, \widehat{y})$ starting from a given index and the conclusion follows.

(ii) If $\theta \in (0, 1/2]$, then $2\theta - 1 \leqslant 0$ and thus for all $k \geqslant k_0'$

$$\frac{1}{1-\theta} C_L \mathcal{E}_k^{1-\theta} \leqslant \mathcal{E}_k^{\frac{1}{2}},$$

which implies that

$$\max \left\{ \sqrt{\mathcal{E}_k}, \varphi(\mathcal{E}_k) \right\} = \sqrt{\mathcal{E}_k}.$$

By Theorem 17, 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 18 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 $2\theta - 1 > 0$ and thus for every $k \geqslant k_0'$

$$\mathcal{E}_k^{\frac{1}{2}} \leqslant \frac{1}{1-\theta} C_L \mathcal{E}_k^{1-\theta},$$

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 17, 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 18. □

Remark 6. In the case when $\rho = 1$ the same convergence rates can be obtained under the original Assumption 1. In fact, when $\rho = 1$ we have that $T_0 = 1$ and, as a consequence, the sequence $\{\mathcal{F}_k\}_{k \geqslant 1}$ in (54) becomes

$$\mathcal{F}_k = \mathcal{L}_r(x^k, z^k, y^k) + C_0 \left\| x^k - x^{k-1} \right\|^2 \ \forall k \geqslant 1.$$

In addition, the inequality (46) 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 nothing else than 0. Combining this inequality with (44) and, by taking into account Lemma 9, gives us (instead of (71))

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

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

$$\begin{aligned}
\mathcal{E}_{k-2} - \mathcal{E}_{k+1} &= \mathcal{F}_{k-2} - \mathcal{F}_{k-1} + \mathcal{F}_{k-1} - \mathcal{F}_k + \mathcal{F}_k - \mathcal{F}_{k+1} \\
&\geqslant \frac{C_0}{4} \left(\left\| x^{k-1} - x^{k-2} \right\|^2 + \left\| x^k - x^{k-1} \right\|^2 + \left\| x^{k+1} - x^k \right\|^2 \right) \\
&\geqslant \frac{C_0}{12} \left(\left\| x^{k-1} - x^{k-2} \right\| + \left\| x^k - x^{k-1} \right\| + \left\| x^{k+1} - x^k \right\| \right)^2 \\
&\geqslant \frac{C_0}{12 C_{11}^2} |||D^{k+1}|||^2.
\end{aligned}$$

Let $\varepsilon > 0$ be such that (82) is fulfilled and $k_0 \geqslant 3$ be such that $\left(x^{k+1}, z^{k+1}, y^{k+1}\right)$ belongs to $\text{Ball}\left(\left(\widehat{x}, \widehat{z}, \widehat{y}\right), \varepsilon\right)$ for all $k \geqslant k_0$. Then (82) implies for every $k \geqslant k_0$

$$\mathcal{E}_{k-2} - \mathcal{E}_{k+1} \geqslant C_{23} \mathcal{E}_{k+1}, \quad \text{where} \quad C_{23} := \frac{C_0}{12 C_L^2 C_{11}^2}, \tag{106}$$

which is the key inequality for deriving the rates of convergence, as we have seen when working in the hypotheses of Assumption 2.

Acknowledgements. The authors are thankful to Ernö Robert Csetnek (University of Vienna) for his valuable comments which improved the quality of the paper.

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–Lojasiewicz 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] **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)

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

[8] **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)

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

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

[11] **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)

[12] **R. I. Boţ, E. R. Csetnek**. *ADMM for monotone operators: convergence analysis and rates.* arXiv:1705.01913 (2017)

[13] **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)

[14] **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)

[15] **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), pp.1–122 (2010)

[16] **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)

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

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

[19] **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)

[20] **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)

[21] **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)

[22] **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)

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

[24] **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)

[25] **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)

[26] **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)

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

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

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

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

[31] **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)

[32] **Q. Liu, X. Shen, Y. Gu**. *Linearized ADMM for Non-convex Non-smooth Optimization with Convergence Analysis.* `arXiv:1705.02502` (2017)

[33] **S. Łojasiewicz** . *Une propriété topologique des sous-ensembles analytiques réels, Les Équations aux Dérivées Partielles.* Éditions du Centre National de la Recherche Scientifique, Paris, 8–89 (1963)

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

[35] **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)

[36] **Y. Ouyang, Y. Chen, G. Lan and E. Pasiliao, Jr.**. *An Accelerated Linearized Alternating Direction Method of Multipliers.* SIAM Journal on Imaging Sciences 8(1), 644–681 (2015).

[37] **X. Ren, Z. Lin**. *Linearized Alternating Direction 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).

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

[39] **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)

[40] **D. Sun, K.-C. Toh, L. Yang**. *A Convergent 3-Block SemiProximal Alternating Direction Method of Multipliers for Conic Programming with 4-Type Constraints* SIAM Journal on Optimization 25(2), 882–915 (2015)

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

[42] **Y. Wang, W. Yin, and J. Zeng**. *Global convergence of ADMM in nonconvex nonsmooth optimization.* UCLA CAM Report 15-62, UCLA (2015)

[43] **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)

[44] **L. Yang, T. K. Pong and 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)

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