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

Radu Ioan Bot
Dang-Khoa Nguyen
March 22, 2019

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

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

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

1 Introduction

1.1 Problem formulation and motivation

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

\[ \min_{x \in \mathbb{R}^n} \{ g(x) + h(x) \}, \]

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

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

\[ \min_{x \in \mathbb{R}^n} \{ f(x) + g(Ax) + h(x) \}, \]

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

\[ \min_{(x,z) \in \mathbb{R}^n \times \mathbb{R}^m} \{ f(x) + g(z) + h(x) \}. \]
For a fixed real number $r > 0$, the augmented Lagrangian associated with problem (3) reads

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

Given a starting vector $(x^0, z^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$ and $(\{M_1^k\}_{k \geq 0} \subseteq \mathbb{R}^{n \times n}, \{M_2^k\}_{k \geq 0} \subseteq \mathbb{R}^{m \times m}$, two sequences of symmetric and positive semidefinite matrices, the following proximal ADMM algorithm formulated in the presence of a smooth function and involving variable metrics has been proposed and investigated in [5]:

generate the sequence $\{(x^k, z^k, y^k)\}_{k \geq 0}$ for every $k \geq 0$ as

$$x^{k+1} = \arg \min_{x \in \mathbb{R}^n} \left\{ f(x) + \langle x - x^k, \nabla h(x^k) \rangle + \frac{r}{2} \|Ax - z^k + \frac{1}{r}y^k\|^2 + \frac{1}{2} \left\|x - x^k\right\|^2_{M_1^k} \right\}, \quad (4a)$$

$$z^{k+1} = \arg \min_{z \in \mathbb{R}^m} \left\{ g(z) + \frac{r}{2} \|Ax^{k+1} - z + \frac{1}{r}y^k\|^2 + \frac{1}{2} \left\|z - z^k\right\|^2_{M_2^k} \right\}, \quad (4b)$$

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

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

In case $h = 0$, the above iterative scheme encompasses as special cases different numerical algorithms considered in the literature. If $M_1^k = M_2^k = 0$ for all $k \geq 0$, then (4a)-(4c) becomes the classical ADMM algorithm (17, 21, 23, 27), which lately gained a huge popularity in the optimization community, despite its poor implementation properties caused by the fact that, in general, the calculation of the sequence of primal variables $\{x^k\}_{k \geq 0}$ does not correspond to a proximal step. For an inertial version of the classical ADMM algorithm we refer the reader to [12]. On the other hand, if $M_1^k = M_1$ and $M_2^k = M_2$ for all $k \geq 0$, then (4a)-(4c) recovers the proximal ADMM algorithm investigated by Shefi and Teboulle in [43] (see also [22, 23]). It has been pointed out in [43] that, for suitable choices of the matrices $M_1$ and $M_2$, the proximal ADMM algorithm becomes a primal-dual splitting algorithm in the sense of those considered in [15, 18, 21, 45], and which, due to its full splitting character, overcomes the drawbacks of the classical ADMM algorithm. Recently, in [44] it has been shown that, if $f$ is strongly convex, then suitable choices of the non-constant sequences $\{M_1^k\}_{k \geq 0}$ and $\{M_2^k\}_{k \geq 0}$ lead to a rate of convergence of $O(1/k)$ for the sequence of primal iterates.

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

In the following we will comment on previous works addressing the ADMM algorithm in the nonconvex setting. None of the papers which have addressed nonconvex optimization problems involving compositions with linear operators propose and investigate iterative schemes designed in the spirit of full splitting algorithms. In [34], the convergence of the ADMM algorithm for solving the problem (1) is studied under the assumption that $h$ is twice continuously differentiable with bounded Hessian. In [35], the algorithm is used to minimize the sum of finitely many smooth nonconvex functions and a nonsmooth convex function, by rewriting it as an general consensus problem. No linear operator occurs in the formulation of the optimization problem under investigation. In [1], the ADMM algorithm is used to solve a DC optimization problem over the unit ball which occurs in the penalized zero-variance linear discriminant analysis. In [25], a nonconvex ADMM algorithm involving proximal terms induced via Bregman distances is introduced and investigated, however, without addressing the question of the boundedness of the generated iterates. On the other hand, in [28], in order to guarantee boundedness of the iterates a strong assumption on $g$ is made, which is proved to hold for the normed-square function.
In [47], a lot of efforts are made to guarantee boundedness for the generated iterates of the nonconvex ADMM algorithm, which is an essential component of the convergence analysis, however, this is done by assuming that the objective function is continuous and coercive over the feasible set, while its non-smooth part is either restricted prox-regular or piecewise linear. Similar ingredients are used in [59] in the convergence analysis of a nonconvex linearized ADMM algorithm.

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

The major strengths of our paper are:

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

2. We prove convergence for relaxed variants of the nonconvex ADMM algorithms, which allow to chose in the update of the dual sequence $p \in (0, 2)$. We notice that $p = 1$ is the standard choice in the literature ([11, 14, 34, 35, 47]). Gabay and Mercier proved in [27] in the convex setting that $p$ may be chosen in $(0, 2)$, however, the majority of the extensions of the convex relaxed ADMM algorithm assume that $p \in (0, \frac{1+\sqrt{2}}{2})$ (see [22, 23, 26, 44, 48, 49]) or ask for a particular choice of $p$, which is interpreted as a step size (see [31]). In [49], an alternating minimization algorithm for the minimization of the sum of a simple nonsmooth function and a smooth function in the nonconvex setting, which allows for a parameter $p$ different from 1, has been proposed.

3. By appropriate choices of the matrix sequences, we derive from the proposed iterative schemes full splitting algorithms for solving the nonconvex complexly structured optimization problem (1). More precisely, (P-ADMM) gives rise to an iterative scheme formulated only in terms of proximal steps for the functions $g$ and $h$ and of forward evaluations of the matrix $A$, while (PL-ADMM) gives rise to an iterative scheme in which the function $h$ is performed via a gradient step. Exact formulas for proximal operators are available not only for large classes of convex functions ([6, 20]), but also of nonconvex functions ([3, 29, 33]). The fruitful idea to linearize the step involving the smooth term for the functions $g$ and $h$ is interpreted as a step size (see [31]). In [49], an alternating minimization algorithm for the minimization of the sum of a simple nonsmooth function and a smooth function in the nonconvex setting, which allows for a parameter $p$ different from 1, has been proposed.

1.2 Notations and preliminaries

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

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

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

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

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

$$\frac{1}{\sqrt{p}} \sum_{i=1}^{p} \|u_i\| \leq \|u\| = \sqrt{\sum_{i=1}^{p} \|u_i\|^2} \leq \sum_{i=1}^{p} \|u_i\|. \quad (5)$$

3
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
\[
\|x\|^2_M := \langle Mx, x \rangle \quad \forall x \in \mathbb{R}^N.
\]
The Loewner partial ordering on $\mathbb{S}_+^n$ is defined for $M, M' \in \mathbb{S}_+^n$ as
\[
M \succeq M' \iff \|x\|^2_M \geq \|x\|^2_{M'} \quad \forall x \in \mathbb{R}^N.
\]
Thus $M \in \mathbb{S}_+^n$ is nothing else than $M \succeq 0$. For $\alpha > 0$ we set
\[
\mathcal{P}_\alpha := \{ M \in \mathbb{S}_+^n : M \geq \alpha \text{Id} \},
\]
where $\text{Id}$ denotes the identity matrix in $\mathbb{R}^{N \times N}$. If $M \in \mathcal{P}_\alpha$, then the semi-norm $\| \cdot \|_M$ becomes a norm.

The linear operator $A$ is surjective if and only if its associated matrix has full row rank. This assumption is further equivalent to the fact that the matrix associated to $AA^*$, where $A^*$ denotes the adjoint operator of $A$, is positively definite. Since
\[
\lambda_{\min}(AA^*) \|y\|^2 \leq \|y\|^2_{AA^*} = \langle AA^*y, y \rangle = \|A^*y\|^2 \quad \forall y \in \mathbb{R}^n,
\]
is further equivalent to $\lambda_{\min}(AA^*) > 0$ (and $AA^* \in \mathcal{P}_{\min(\lambda_{\min}(AA^*))}$), 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 $A^*A \in \mathcal{P}_{\min(\lambda_{\min}(A^*A))}$).

**Proposition 1.** Let $\Psi : \mathbb{R}^N \to \mathbb{R}$ 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
\[
\Psi(y) \leq \Psi(x) + \langle \nabla \Psi(z), y - x \rangle + \frac{L}{2} \|y - x\|^2; \tag{6}
\]

2. If $\Psi$ is bounded from below, then for every $\sigma > 0$ it holds
\[
\inf_{x \in \mathbb{R}^N} \left\{ \Psi(x) - \left( \frac{1}{\sigma} - \frac{L}{2\sigma^2} \right) \|\nabla \Psi(x)\|^2 \right\} > -\infty.
\]

**Proof.**

1. Let be $x, y \in \mathbb{R}^N$ and $z := (1-t)x + ty$ for $t \in [0, 1]$. By the fundamental theorem for line integrals we have
\[
\Psi(y) - \Psi(x) = \int_0^1 \langle \nabla \Psi((1-s)x + sy), y - x \rangle ds
\]
\[
= \int_0^1 \langle \nabla \Psi((1-s)x + sy) - \nabla \Psi(z), y - x \rangle ds + \langle \nabla \Psi(z), y - x \rangle. \tag{7}
\]

Since
\[
\left| \int_0^1 \langle \nabla \Psi((1-s)x + sy) - \nabla \Psi(z), y - x \rangle ds \right|
\]
\[
\leq \int_0^1 \|\nabla \Psi((1-s)x + sy) - \nabla \Psi(z)\| \cdot \|y - x\| ds \leq L \|y - x\|^2 \int_0^1 |s - t| ds
\]
\[
= L \|y - x\|^2 \left( \int_0^t (-s + t) ds + \int_t^1 (s - t) ds \right) = L \left( \frac{1}{2} - t (1 - t) \right) \|x - y\|^2. \tag{8}
\]
The inequality in (6) follows by combining (7) and (8) and by using that $0 \leq t \leq 1$.

2. The inequality in (6) gives for every $x \in \mathbb{R}^N$
\[
-\infty < \inf_{y \in \mathbb{R}^N} \Psi(y) \leq \Psi(x) - \frac{1}{\sigma} \|\nabla \Psi(x)\|
\]
\[
\leq \Psi(x) + \left( \frac{1}{\sigma} - \frac{L}{2\sigma^2} \right) \|\nabla \Psi(x)\|^2,
\]
which leads to the desired conclusion. \qed

---

4
Remark 1. The so-called Descent Lemma, which says that for a Fréchet differentiable function $Ψ: \mathbb{R}^N \to \mathbb{R}$ having Lipschitz continuous gradient with constant $L > 0$ it holds

$$Ψ(y) \leq Ψ(x) + \langle ∇Ψ(x), y - x \rangle + \frac{L}{2} \|y - x\|^2 \quad \forall x, y \in \mathbb{R}^N,$$

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

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

$$Ψ(x) \geq Ψ(y) + \langle ∇Ψ(y), x - y \rangle - \frac{L}{2} \|x - y\|^2 \quad \forall x, y \in \mathbb{R}^N,$$

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

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

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

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

$$\hat{∂}Ψ(x) := \{d \in \mathbb{R}^N: \text{exist sequences } x^k \to x \text{ and } d^k \to d \text{ as } k \to +\infty \text{ such that } Ψ(x^k) \to Ψ(x) \text{ as } k \to +\infty \text{ and } d^k \in \hat{∂}Ψ(x^k) \text{ for all } k \geq 0\}.$$ 

For $x \notin \text{dom}(Ψ)$, we set $\hat{∂}Ψ(x) = \partial Ψ(x) := \emptyset$.

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

$$\hat{Ψ}(x) = \partial Ψ(x) = \{d \in \mathbb{R}^N: Ψ(y) \geq Ψ(x) + \langle d, y - x \rangle \quad \forall y \in \mathbb{R}^N\} \text{ for all } x \in \text{dom}Ψ.$$

If $x \in \mathbb{R}^N$ is a local minimum of $Ψ$, then $0 \in \partial Ψ(x)$. We denote by $\text{crit}(Ψ) = \{x \in \mathbb{R}^N: 0 \in \partial Ψ(x)\}$ the set of critical points of $Ψ$. The limiting subdifferential fulfills the closedness criterion: if $\{x^k\}_{k \geq 0}$ and $\{d^k\}_{k \geq 0}$ are sequence in $\mathbb{R}^N$ such that $d^k \in \hat{∂}Ψ(x^k)$ for all $k \geq 0$, and $(x^k, d^k) \to (x, d)$ and $Ψ(x^k) \to Ψ(x)$ as $k \to +\infty$, then $d \in \hat{∂}Ψ(x)$.

We have the following subdifferential sum rule holds ([33] Proposition 1.107), ([42] Exercise 8.8): if $Φ: \mathbb{R}^N \to \mathbb{R}$ is a continuously differentiable function, then $\hat{∂}(Φ + Ψ)(x) = \hat{∂}Φ(x) + Ψ(x)$ for all $x \in \mathbb{R}^N$; and the following subdifferential rule for the composition with a linear operator $A: \mathbb{R}^N \to \mathbb{R}^N$ ([33] Proposition 1.112), ([42] Exercise 10.7)): if $x \in \text{dom}Ψ$ and $A$ is injective, then $\hat{∂}(Ψ ∘ A)(x) = A^* Ψ(Ax)$.

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

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

$$b_{k+1} + ξ_k \leq b_k.$$

Then the following statements hold:

1. the sequence $\{ξ_k\}_{k \geq 0}$ is summable, namely $\sum_{k \geq 0} ξ_k < +\infty$;

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

The following lemma, which is an extension of [13] Lemma 2.3 (see, also [16] Lemma 3]), is of interest by its own.
Lemma 3. Let \( \{a^k := (a_{1}^{k}, a_{2}^{k}, \ldots, a_{N}^{k})\}_{k \geq 0} \) be a sequence in \( \mathbb{R}_{+}^{N} \) and \( \{\delta_{k}\}_{k \geq 0} \) a sequence in \( \mathbb{R} \) such that

\[
\langle 1, a^{k+1} \rangle \leq \langle c_{0}, a^{k} \rangle + \langle c_{1}, a^{k-1} \rangle + \langle c_{2}, a^{k-2} \rangle + \delta_{k} \quad \forall k \geq 2,
\]

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 \( c_{0} + c_{1} + c_{2} < 1 \). Assume further that there exists \( \delta \geq 0 \) such that for every \( \overline{K} \geq K \geq 2 \)

\[
\sum_{k=K}^{\overline{K}} \delta_{k} \leq \delta.
\]

Then, for every \( i = 1, \ldots, N \), it holds

\[
\sum_{k=0}^{\infty} a_{i}^{k} < +\infty.
\]

In particular, for every \( i = 1, \ldots, N \) and every \( \overline{K} \geq K \geq 2 \), it holds

\[
\sum_{k=K}^{\overline{K}} a_{i}^{k} \leq \frac{\sum_{j=1}^{N} \left[ (1 - c_{0,j} - c_{1,j}) a_{j}^{K} + (1 - c_{0,j}) a_{j}^{K+1} + a_{j}^{K+2} \right] + \delta}{1 - c_{0,i} - c_{1,i} - c_{2,i}}.
\]

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

\[
\langle 1, \sum_{k=K+2}^{\overline{K}} a^{k+1} \rangle \leq \langle 1, \sum_{k=K+2}^{\overline{K}} a^{k} \rangle + \langle 1, \sum_{k=K+2}^{\overline{K}} a^{k-1} \rangle + \langle 1, \sum_{k=K+2}^{\overline{K}} a^{k-2} \rangle + \sum_{k=K+2}^{\overline{K}} \delta_{k}.
\]

Since

\[
\begin{align*}
\sum_{k=K+2}^{\overline{K}} a^{k+1} &= \sum_{k=K+3}^{\overline{K}+1} a^{k} = \sum_{k=K}^{\overline{K}} a^{k} + a^{K+1} - a^{K+2}, \\
\sum_{k=K+2}^{\overline{K}} a^{k} &= \sum_{k=K}^{\overline{K}} a^{k} - (a^{K} + a^{K+1}), \\
\sum_{k=K+2}^{\overline{K}} a^{k-1} &= \sum_{k=K+1}^{\overline{K}+1} a^{k} = \sum_{k=K}^{\overline{K}} a^{k} - (a^{K} + a^{K+1}), \\
\sum_{k=K+2}^{\overline{K}} a^{k-2} &= \sum_{k=K}^{\overline{K}} a^{k} - (a^{K} + \overline{a}^{K+1}),
\end{align*}
\]

the inequality in (11) can be rewritten as

\[
\langle 1, \sum_{k=K}^{\overline{K}} a^{k} \rangle + \langle 1, a^{K+1} - a^{K} - a^{K+1} - a^{K+2} \rangle \leq \langle c_{0}, \sum_{k=K}^{\overline{K}} a^{k} \rangle - \langle c_{0}, a^{K} + a^{K+1} \rangle \\
+ \langle c_{1}, \sum_{k=K}^{\overline{K}} a^{k} \rangle - \langle c_{1}, a^{K} + a^{K+1} \rangle + \langle c_{2}, \sum_{k=K}^{\overline{K}} a^{k} \rangle - \langle c_{2}, a^{K+1} + a^{K+2} \rangle + \sum_{k=K+2}^{\overline{K}} \delta_{k},
\]

which further implies

\[
\sum_{j=1}^{N} \left[ (1 - c_{0,j} - c_{1,j} - c_{2,j}) \sum_{k=K}^{\overline{K}} a_{j}^{k} \right] = \langle 1, c_{0} - c_{1} - c_{2}, \sum_{k=K}^{\overline{K}} a^{k} \rangle \\
\leq \langle 1, c_{0} - c_{1}, a^{K} \rangle + \langle 1, c_{0}, a^{K+1} \rangle + \langle 1, a^{K+2} \rangle + \sum_{k=K+2}^{\overline{K}} \delta_{k} \\
= \sum_{j=1}^{N} \left[ (1 - c_{0,j} - c_{1,j}) a_{j}^{K} + (1 - c_{0,j}) a_{j}^{K+1} + a_{j}^{K+2} \right] + \sum_{k=K+2}^{\overline{K}} \delta_{k}.
\]
Hence, for every $i = 1, \ldots, N$, it holds

$$(1 - c_{0,i} - c_{1,i} - c_{2,i}) \sum_{k=1}^{N} a_k^i \leq \sum_{j=1}^{N} \left[ (1 - c_{0,j} - c_{1,j}) a_j K + (1 - c_{0,j}) a_j K^2 + a_j K^3 + a_j K^4 \right] + \delta$$

and the conclusion follows by taking into consideration that $c_0 + c_1 + c_2 < 1$.

\[ \square \]

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

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

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

2.1 General formulations and full proximal splitting algorithms as particular instances

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

$$z^{k+1} \in \arg \min_{z \in \mathbb{R}^m} \left\{ \mathcal{L}_r (x^k, z, y^k) + \frac{1}{2} \| z - z^k \|^2_{M^k_r} \right\}$$

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

$$x^{k+1} \in \arg \min_{x \in \mathbb{R}^n} \left\{ \mathcal{L}_r (x, z^{k+1}, y^k) + \frac{1}{2} \| x - x^k \|^2_{M^k_r} \right\}$$

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

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

(12c)

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

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

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

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

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

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

The above particular instance of Algorithm 1 is an iterative scheme formulated in the spirit of full splitting numerical methods; in other words, the functions $g$ and $h$ are evaluated by their proximal operators, while the linear operator $A$ and its adjoint operator are evaluated by simple forward steps. Exact formulas for the proximal operator are available not only for large classes of convex functions \([6, 20]\), but also for many nonconvex functions occurring in applications \([8, 20, 33]\).

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

7
Algorithm 2. Let be the matrix sequences \( \{M^k_1\}_{k \geq 0} \in \mathbb{S}^n_+ \), \( \{M^k_2\}_{k \geq 0} \in \mathbb{S}^n_+ \), \( r > 0 \) and \( 0 < \rho < 2 \). For a given starting vector \( (x^0, z^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \), generate the sequence \( \{(x^k, z^k, y^k)\}_{k \geq 0} \) for every \( k \geq 0 \) as:

\[
\begin{align*}
  z^{k+1} & \in \arg \min_{z \in \mathbb{R}^m} \left\{ g(z) + \frac{r}{2} \|Ax^k - z\|^2 + \frac{1}{2} \|z - z^k\|^2_{M^k_2} \right\}, \\
  x^{k+1} & \in \arg \min_{x \in \mathbb{R}^n} \left\{ \langle x - x^k, \nabla h(x^k) \rangle + \langle y^k, Ax - z^{k+1} \rangle + \frac{r}{2} \|Ax - z^{k+1}\|^2 + \frac{1}{2} \|x - x^k\|^2_{M^k_1} \right\}, \\
  y^{k+1} & := y^k + \rho r (Ax^{k+1} - z^{k+1}).
\end{align*}
\]

Due to the presence of the variable metric inducing matrix sequences, Algorithm 2 represents a unifying scheme for several linearized ADMM algorithms from the literature (see \cite{35, 36, 40, 41, 48, 50}). By choosing as above \( M^k_1 := t_k \text{Id} - rA^*A \), where \( t_k \) is positive such that \( t_k \geq r \|A\|^2 \), and \( M^k_2 := 0 \), for every \( k \geq 0 \), Algorithm 2 translates for every \( k \geq 0 \) into:

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

In this iterative scheme the smooth term is evaluated via a gradient step, which is an improvement with respect to other nonconvex ADMM algorithms, such as \cite{47, 49}, where the smooth function is involved in a subproblem, which may be difficult to solve, unless it can be reformulated as a proximal step (see \cite{54}).

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

Assumption 1. We assume that

1. \( g \) and \( h \) are bounded from below;

2. \( A \) is surjective and thus the constant

\[
T_0 := \begin{cases} \\
\frac{1}{\lambda_{\min}(AA^*)} & \text{if } 0 < \rho \leq 1, \\
\frac{1}{\lambda_{\min}(AA^*) (2 - \rho^2)} & \text{if } 1 < \rho < 2,
\end{cases}
\]

is well-defined;

3. \( \mu_1 := \sup_{k \geq 0} \|M^k_1\| < +\infty \) and \( \mu_2 := \sup_{k \geq 0} \|M^k_2\| < +\infty \);

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

\[
r \geq 4T_0 L > 0
\]

and

\[
2M^k_1 + rA^*A \succeq \left( L + \frac{C_M}{r} \right) \text{Id} \quad \forall k \geq 0,
\]

where

\[
C_M := \begin{cases} \\
\left( 6\mu_1^2 + 4(L + \mu_1)^2 \right) T_0, & \text{for Algorithm 1} \\
\left( 4\mu_1^2 + 6(L + \mu_1)^2 \right) T_0, & \text{for Algorithm 2}
\end{cases}
\]

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

2. In the following we discuss possible choices of the matrix sequence \( \{M^k_1\}_{k \geq 0} \) which fulfill Assumption 1.
(a) If \( \sup_{k \geq 0} \| M_k^t \| = \mu_1 > \frac{L}{2} \), then, for every

\[
\begin{align*}
r & \geq \max \left\{ 4T_0L, \frac{C_M}{2\mu_1 - L} \right\} > 0,
\end{align*}
\]

there exists \( \alpha_1 > 0 \) such that

\[
\mu_1 \geq \alpha_1 \geq \frac{1}{2} \left( L + \frac{C_M}{r} \right) > 0.
\]

The inequality in (15) is ensured for \( M_k^t \) chosen such that

\[
\mu_1 \operatorname{Id} \geq M_k^t \geq \alpha_1 \operatorname{Id} \quad \forall k \geq 0.
\]

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

\[
r \geq \max \left\{ 4T_0L, \frac{L + \sqrt{L^2 + 4\lambda_{\min} (A^* A) C_M}}{2\lambda_{\min} (A^* A)} \right\} > 0,
\]

it follows that \( r^2 \lambda_{\min} (A^* A) - rL - C_M \geq 0 \). Thus,

\[
rA^* A - (L + r^{-1}C_M) \operatorname{Id} \geq 0,
\]

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

(c) For \( t > 0 \), we take \( M_k^t := t \operatorname{Id} - rA^* A \) for every \( k \geq 0 \) (see also Section 6.3 in \[11\]). Then

\[
\mu_1 = \| t \operatorname{Id} - rA^* A \| = \lambda_{\max} (t \operatorname{Id} - rA^* A) = t - r \lambda_{\min} (A^* A).
\]

Condition (15) is equivalent to

\[
2t - r \| A \|^2 - \left( L + \frac{C_M}{r} \right) \geq 0
\]

and is guaranteed for both algorithms when

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

or, equivalently,

\[
10T_0\mu_1^2 - 2 (r - 6T_0L) \mu_1 + 6T_0L^2 + r^2 \left( \| A \|^2 - 2\lambda_{\min} (A^* A) \right) - Lr \leq 0.
\]

This quadratic inequality in \( \mu_1 \geq 0 \) has nonnegative solutions if, for instance, \( r \geq 6T_0L \) (thus (14) holds) and the reduced discriminant

\[
\Delta := (r - 6T_0L)^2 - 60T_0^2L^2 - 10T_0r^2 \left( \| A \|^2 - 2\lambda_{\min} (A^* A) \right) + 10T_0Lr
\]

\[
= \left[ 1 + 10T_0 \left( 2\lambda_{\min} (A^* A) - \| A \|^2 \right) \right] r^2 - 2T_0Lr - 24T_0^2L^2
\]

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

\[
\kappa(A^* A) := \frac{\lambda_{\max} (A^* A)}{\lambda_{\min} (A^* A)} = \frac{\| A \|^2}{\lambda_{\min} (A^* A)} \leq 2.
\]

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

\[
r \geq 6T_0L
\]

and \( t \) such that

\[
r\lambda_{\min} (A^* A) \leq t \leq r\lambda_{\min} (A^* A) + \frac{1}{10T_0} \left( r - 6T_0L + \sqrt{\Delta} \right).
\]

For a similar choice for the
Lemma 4. Suppose that Assumption 1 holds true and let $k$ be fixed. In both cases the proof builds on showing that the following inequality

$$L_r \left( x^{k+1}, z^{k+1}, y^{k+1} \right) + T_1 \left\| A^* \left( y^{k+1} - y^k \right) \right\|^2 + \frac{1}{2} \left\| z^{k+1} - x^k \right\|^2_M + \frac{1}{2} \left\| z^{k+1} - z^k \right\|^2_M \leq L_r \left( x^k, z^k, y^k \right) + T_1 \left\| A^* \left( y^k - y^{k-1} \right) \right\|^2 + \frac{C_1}{2} \left\| x^k - x^{k-1} \right\|^2.$$

(16)

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

$$L_r \left( x^{k+1}, z^{k+1}, y^{k+1} \right) + \frac{1}{2} \left\| z^{k+1} - z^k \right\|^2_M + r \left\| A^* \left( y^{k+1} - y^k \right) \right\|^2 + \frac{L}{2} \left\| x^{k+1} - x^k \right\|^2 + \frac{1}{2} \left\| z^{k+1} - z^k \right\|^2_M \leq L_r \left( x^k, z^k, y^k \right) + \frac{1}{pr} \left\| y^{k+1} - y^k \right\|^2.$$

(17)

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

1. For Algorithm 2 From (12a) we have

$$g \left( z^{k+1} \right) + \left\langle y^k, Ax^k - z^{k+1} \right\rangle + r \left\| A^* \left( y^k - y^{k-1} \right) \right\|^2 + \frac{L}{2} \left\| x^{k+1} - x^k \right\|^2 + \frac{1}{2} \left\| z^{k+1} - z^k \right\|^2_M.$$

(18)

The optimality criterion of (12a) is

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

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

$$h \left( x^{k+1} \right) \leq h \left( x^k \right) + \left\langle y^k, Ax^k - Ax^{k+1} \right\rangle + r \left\langle Ax^{k+1} - z^{k+1}, Ax^k - Ax^{k+1} \right\rangle$$

$$- \left\| x^{k+1} - x^k \right\|^2_M + \frac{L}{2} \left\| x^{k+1} - x^k \right\|^2.$$

(20)

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

By using the notation $u_l^1 := -\nabla h \left( x^l \right) + M_l^{-1} \left( x^{l-1} - x^l \right) \forall l \geq 1$

(21)

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

$$A^* y^{l+1} = \rho u_l^{l+1} + (1 - \rho) A^* y^l \forall l \geq 0.$$

(22)
• The case $0 < \rho \leq 1$. We have
\[ A^* (y^{k+1} - y^k) = \rho \,(u_1^{k+1} - u_1^k) + (1 - \rho) \,A^* \,(y^k - y^{k-1}). \]
Since $0 < \rho \leq 1$, the convexity of $\|\cdot\|^2$ gives
\[ \|A^* (y^{k+1} - y^k)\|^2 \leq \rho \|u_1^{k+1} - u_1^k\|^2 + (1 - \rho) \|A^* (y^k - y^{k-1})\|^2 \]
and from here we get
\[ \lambda_{\min}(AA^*)\rho \|y^{k+1} - y^k\|^2 \leq \rho \|A^* (y^{k+1} - y^k)\|^2 \]
\[ \leq \rho \|u_1^{k+1} - u_1^k\|^2 + (1 - \rho) \|A^* (y^k - y^{k-1})\|^2 - (1 - \rho) \|A^* (y^{k+1} - y^k)\|^2. \tag{23} \]
By using the Lipschitz continuity of $\nabla h$ we have
\[ \|u_1^{k+1} - u_1^k\| \leq (L + \mu_1) \|x^{k+1} - x^k\| + \mu_1 \|x^k - x^{k-1}\|, \tag{24} \]
thus
\[ \|u_1^{k+1} - u_1^k\|^2 \leq 2 (L + \mu_1)^2 \|x^{k+1} - x^k\|^2 + 2\mu_1^2 \|x^k - x^{k-1}\|^2. \tag{25} \]
After plugging (25) into (23) we get
\[ \frac{1}{\rho r} \|y^{k+1} - y^k\|^2 \leq \frac{2(L + \mu_1)^2}{\lambda_{\min}(AA^*)r} \|x^{k+1} - x^k\|^2 + \frac{2\mu_1^2}{\lambda_{\min}(AA^*)r^2} \|x^k - x^{k-1}\|^2 \]
\[ + \frac{(1 - \rho)}{\lambda_{\min}(AA^*)r^2} \|A^* (y^k - y^{k-1})\|^2 - \frac{(1 - \rho)}{\lambda_{\min}(AA^*)r^2} \|A^* (y^{k+1} - y^k)\|^2, \tag{26} \]
which, combined with (17), provides (16).

• The case $1 < \rho < 2$. This time we have from (22) that
\[ A^* (y^{k+1} - y^k) = (2 - \rho) \frac{\rho}{2 - \rho} \,(u_1^{k+1} - u_1^k) + (\rho - 1) \,A^* \,(y^{k-1} - y^k). \]
As $1 < \rho < 2$, the convexity of $\|\cdot\|^2$ gives
\[ \|A^* (y^{k+1} - y^k)\|^2 \leq \frac{\rho^2}{2 - \rho} \|u_1^{k+1} - u_1^k\|^2 + (\rho - 1) \|A^* (y^k - y^{k-1})\|^2 \]
and from here it follows
\[ \lambda_{\min}(AA^*) \,(2 - \rho) \|y^{k+1} - y^k\|^2 \leq (2 - \rho) \|A^* (y^{k+1} - y^k)\|^2 \]
\[ \leq \frac{\rho^2}{2 - \rho} \|u_1^{k+1} - u_1^k\|^2 + (\rho - 1) \|A^* (y^k - y^{k-1})\|^2 - (\rho - 1) \|A^* (y^{k+1} - y^k)\|^2. \tag{27} \]
After plugging (25) into (27) we get
\[ \frac{1}{\rho r} \|y^{k+1} - y^k\|^2 \leq \frac{2\rho(L + \mu_1)^2}{\lambda_{\min}(AA^*)(2 - \rho)^2 r} \|x^{k+1} - x^k\|^2 + \frac{2\rho\mu_1^2}{\lambda_{\min}(AA^*)(2 - \rho)^2 r} \|x^k - x^{k-1}\|^2 \]
\[ + \frac{(\rho - 1)}{\lambda_{\min}(AA^*)(2 - \rho)^2 r} \|A^* (y^k - y^{k-1})\|^2 \]
\[ - \frac{(\rho - 1)}{\lambda_{\min}(AA^*)(2 - \rho)^2 r} \|A^* (y^{k+1} - y^k)\|^2, \tag{28} \]
which, combined with (17), provides (16).

2. For Algorithm 3 the optimality criterion of (13) is
\[ \nabla h (x^k) = - A^* y^k - r A^* (Ax^{k+1} - z^{k+1}) + M_k^e (x^k - x^{k+1}). \tag{29} \]
From (36) (applied for $z := x^k$) we get
\[
 h(x^{k+1}) \leq h(x^k) + \langle y^k, Ax^k - Ax^{k+1} \rangle + rRk + 1, Ax^k - Ax^{k+1} \rangle \\
- \|x^{k+1} - x^k\|^2 M_1 + L \|x^{k+1} - x^k\|^2.
\] (30)

Since the definition of $z^{k+1}$ in (13a) leads also to (15), by combining this inequality with (30) and (13c), after some rearrangements, (17) follows. By using this time the notation
\[
u_2 := -\nabla h(x^{l-1}) + M_1^{-1} (x^{l-1} - x^l) \quad \forall l \geq 1
\] (31)
and by taking into consideration (13c), we can rewrite (29) as
\[
A^* y^{l+1} = \rho \nu_2^{l+1} + (1 - \rho) A^* y^l \quad \forall l \geq 0.
\] (32)

- **The case $0 < \rho \leq 1$.** As in (23) we obtain
\[
\lambda_{\min}(AA^*) \rho \|y^{k+1} - y^k\|^2 \leq \rho \|A^* (y^{k+1} - y^k)\|^2 \\
\leq \rho \|u_2^{k+1} - u_2^k\|^2 + \|A^* (y^k - y^{k-1})\|^2 - (1 - \rho) \|A^* (y^{k+1} - y^k)\|^2.
\] (33)

By using the Lipschitz continuity of $\nabla h$ we have
\[
\|u_2^{k+1} - u_2^k\| \leq \mu_1 \|x^{k+1} - x^k\| + (L + \mu_1) \|x^k - x^{k-1}\|,
\] (34)

thus
\[
\|u_2^{k+1} - u_2^k\|^2 \leq 2 \mu_1^2 \|x^{k+1} - x^k\|^2 + 2 (L + \mu_1)^2 \|x^k - x^{k-1}\|^2.
\] (35)

After plugging (35) into (33) it follows
\[
\frac{1}{\rho^2} \|y^{k+1} - y^k\|^2 \leq \frac{2 \mu_1^2}{\lambda_{\min}(AA^*) \rho \mu_1} \|x^{k+1} - x^k\|^2 + \frac{2 (L + \mu_1)^2}{\lambda_{\min}(AA^*) \rho^2} \|x^k - x^{k-1}\|^2 \\
+ \frac{(1 - \rho)}{\lambda_{\min}(AA^*) \rho \mu_1} \|A^* (y^k - y^{k-1})\|^2 - \frac{(1 - \rho)}{\lambda_{\min}(AA^*) \rho^2} \|A^* (y^{k+1} - y^k)\|^2,
\] (36)

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

- **The case $1 < \rho < 2$.** As in (27) we obtain
\[
\lambda_{\min}(AA^*) (2 - \rho) \|y^{k+1} - y^k\|^2 \leq (2 - \rho) \|A^* (y^{k+1} - y^k)\|^2 \\
\leq \frac{\rho^2}{2} \|u_2^{k+1} - u_2^k\|^2 + (\rho - 1) \|A^* (y^k - y^{k-1})\|^2 - (\rho - 1) \|A^* (y^{k+1} - y^k)\|^2.
\] (37)

After plugging (35) into (37) it follows
\[
\frac{1}{\rho^2} \|y^{k+1} - y^k\|^2 \leq \frac{2 \mu_1^2}{\lambda_{\min}(AA^*) (2 - \rho)^2} \|x^{k+1} - x^k\|^2 + \frac{2 (L + \mu_1)^2}{\lambda_{\min}(AA^*) (2 - \rho)^2} \|x^k - x^{k-1}\|^2 \\
+ \frac{(\rho - 1)}{\lambda_{\min}(AA^*) (2 - \rho) \mu_1} \|A^* (y^k - y^{k-1})\|^2 \\
- \frac{(\rho - 1)}{\lambda_{\min}(AA^*) (2 - \rho) \mu_1} \|A^* (y^{k+1} - y^k)\|^2.
\] (38)

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

This concludes the proof.

The following three estimates will be useful in the sequel.

**Lemma 5.** Suppose that Assumption holds true and let $\{x^k, z^k, y^k\}_{k \geq 0}$ be a sequence generated by Algorithm or Algorithm. Then the following statements are true:
(i) for every \( k \geq 1 \)
\[
\|z^{k+1} - z^k\| \leq \|A\| \cdot \|x^{k+1} - x^k\| + \|Ax^{k+1} - z^{k+1}\| + \|Ax^k - z^k\|
= \|A\| \cdot \|x^{k+1} - x^k\| + \frac{1}{pr} \|y^{k+1} - y^k\| + \frac{1}{pr} \|y^k - y^{k-1}\|; \quad (39)
\]

(ii) for every \( k \geq 0 \)
\[
\frac{1}{2r} \|y^{k+1}\|^2 \leq \frac{T_1}{2} \|A^* (y^{k+1} - y^k)\|^2 + \frac{T_0}{r} \|\nabla h (x^{k+1})\|^2 + \frac{C_1}{4} \|x^{k+1} - x^k\|^2; \quad (40)
\]

(iii) for every \( k \geq 1 \)
\[
\|y^{k+1} - y^k\| \leq C_3 \|x^{k+1} - x^k\| + C_4 \|x^k - x^{k-1}\| + T_2 (\|A^* (y^k - y^{k-1})\| - \|A^* (y^{k+1} - y^k)\|), \quad (41)
\]

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

Proof. The statement in \( (39) \) is straightforward.

From \( (22) \) and \( (32) \) we have for every \( k \geq 0 \)
\[
A^* y^{k+1} = \rho u^{k+1} + (1 - \rho) A^* y^k
\]
or, equivalently,
\[
\rho A^* y^{k+1} = \rho u^{k+1} + (1 - \rho) A^* (y^k - y^{k+1}),
\]
where \( u^{k+1} \) is defined as being equal to \( u_1^{k+1} \) in \( (21) \), for Algorithm 1 and, respectively, to \( u_2^{k+1} \) in \( (31) \), for Algorithm 2.

For \( 0 < \rho \leq 1 \) we have
\[
\lambda_{\min}(AA^*) \rho^2 \|y^{k+1}\|^2 \leq \rho^2 \|A^* y^{k+1}\|^2 \leq \rho \|u^{k+1}\|^2 + (1 - \rho) \|A^* (y^{k+1} - y^k)\|^2, \quad (42)
\]
while, for \( 1 < \rho < 2 \), we have
\[
\lambda_{\min}(AA^*) \rho^2 \|y^{k+1}\|^2 \leq \rho^2 \|A^* y^{k+1}\|^2 \leq \frac{\rho^2}{2 - \rho} \|u^{k+1}\|^2 + (\rho - 1) \|A^* (y^{k+1} - y^k)\|^2. \quad (43)
\]

Notice further that for \( 1 < \rho < 2 \) we have \( \frac{1}{\rho} < 1 \) and \( 1 < \frac{\rho}{2 - \rho} \).

In case \( u^{k+1} \) is defined as in \( (21) \) it holds
\[
\|u^{k+1}\|^2 = \|u_1^{k+1}\|^2 \leq 2 \|\nabla h (x^{k+1})\|^2 + 2 \mu_1^2 \|x^{k+1} - x^k\|^2 \forall k \geq 0, \quad (44)
\]
while, in case \( u_2^{k+1} \) is defined as in \( (31) \), it holds
\[
\|u^{k+1}\|^2 = \|u_2^{k+1}\|^2 \leq 2 \|\nabla h (x^{k+1})\|^2 + 2 (L + \mu_1)^2 \|x^{k+1} - x^k\|^2 \forall k \geq 0. \quad (45)
\]

13
We divide (42) and (43) by $2\lambda_{\min}(AA^*)\rho^2 r > 0$ and plug (14) and, respectively, (45) into the resulting inequalities. This gives us (40). Finally, in order to prove (41), we notice that for every $k \geq 1$ it holds
\[
\left\| A^* (y^{k+1} - y^k) \right\| \leq \rho \left\| u^{k+1} - u^k \right\| + |1 - \rho| \left\| A^* (y^k - y^{k-1}) \right\|,
\]
so,
\[
\sqrt{\lambda_{\min}(AA^*)} (1 - |1 - \rho|) \left\| y^{k+1} - y^k \right\| \leq (1 - |1 - \rho|) \left\| A^* (y^{k+1} - y^k) \right\| - |1 - \rho| \left\| A^* (y^k - y^{k-1}) \right\|.
\]
(46)

We plug into (46) the estimates for $\left\| u^{k+1} - u^k \right\|$ derived in (24) and, respectively, (34) and divide the resulting inequality by $\sqrt{\lambda_{\min}(AA^*)} (1 - |1 - \rho|) > 0$. 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
\[
F_r : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\},
\]
\[
F_r(x, z, y, x', y') = L_r(x, z, y) + T_1 \left\| A^* (y - y') \right\|^2 + \frac{C_1}{2} \left\| x - x' \right\|^2,
\]
where $T_1$ and $C_1$ are defined in Assumption [3]. For every $k \geq 1$ we denote
\[
F_k := F_r(x^k, z^k, y^k, x^{k-1}, y^{k-1}) = L_r(x^k, z^k, y^k) + T_1 \left\| A^* (y^k - y^{k-1}) \right\|^2 + \frac{C_1}{2} \left\| x^k - x^{k-1} \right\|^2. (47)
\]

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

**Theorem 6.** Suppose that Assumption [7] holds true and let $\{(x^k, z^k, y^k)\}_{k \geq 0}$ be a sequence generated by Algorithm [3] or Algorithm [2]. Suppose that one of the following conditions holds:

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

(B-II) $h$ is coercive.

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

**Proof.** From Lemma [3] we have that for every $k \geq 1$
\[
F_{k+1} + \frac{1}{2} \left\| x^{k+1} - x^k \right\|^2 M_3^{-1} - C_1 I_d + \frac{1}{2} \left\| z^{k+1} - z^k \right\|^2 M_2^{-1} \leq F_k
\]
(48)
which shows, according to (15), that $\{F_k\}_{k \geq 1}$ is monotonically decreasing. Consequently, for every $k \geq 1$ we have
\[
F_1 \geq F_{k+1} + \frac{1}{2} \left\| x^{k+1} - x^k \right\|^2 M_3^{-1} - C_1 I_d + \frac{1}{2} \left\| z^{k+1} - z^k \right\|^2 M_2^{-1}
\]
\[
= h(x^{k+1}) + g(z^{k+1}) - \frac{1}{2r} \left\| y^{k+1} \right\|^2 + \frac{r}{2} \left\| A x^{k+1} - z^{k+1} + \frac{1}{r} y^{k+1} \right\|^2
\]
\[
+ T_1 \left\| A^* (y^{k+1} - y^k) \right\|^2 + \frac{1}{2} \left\| x^{k+1} - x^k \right\|^2 M_3^{-1} - C_1 I_d + \frac{1}{2} \left\| z^{k+1} - z^k \right\|^2 M_2^{-1} + C_1 \left\| x^{k+1} - x^k \right\|^2,
\]
which, thanks to (40), leads to
\[
F_1 \geq h(x^{k+1}) + g(z^{k+1}) - \frac{T_0}{r} \left\| \nabla h(x^{k+1}) \right\|^2 + \frac{r}{2} \left\| A x^{k+1} - z^{k+1} + \frac{1}{r} y^{k+1} \right\|^2
\]
\[
+ \frac{T_1}{2} \left\| A^* (y^{k+1} - y^k) \right\|^2 + \frac{1}{2} \left\| x^{k+1} - x^k \right\|^2 M_3^{-1} - C_1 I_d + \frac{1}{2} \left\| z^{k+1} - z^k \right\|^2 M_2^{-1} + \frac{C_1}{4} \left\| x^{k+1} - x^k \right\|^2. (49)
\]
Next we will prove the boundedness of $\{(x^k, z^k, y^k)\}_{k \geq 0}$ under each of the two scenarios.
(B-I) Since \( r \geq 4T_0L \), there exists \( \sigma > 0 \) such that
\[
\frac{1}{\sigma} - \frac{L}{2\sigma^2} = \frac{T_0}{r}.
\]
From Proposition 1 and the relation (49), we see that for every \( k \geq 1 \)
\[
g(z^{k+1}) + \frac{r}{2} \left\| Ax^{k+1} - z^{k+1} + \frac{1}{r}y^{k+1} \right\|^2 + \frac{C_1}{4} \left\| x^{k+1} - x^k \right\|^2
\leq \mathcal{F}_1 - \inf_{z \in \mathbb{R}^n} \left\{ h(x) - \frac{T_0}{r} \left\| \nabla h(x) \right\|^2 \right\} < +\infty.
\]
Since \( g \) is coercive, it follows that the sequence \( \{z^k\}_{k \geq 0} \) is bounded. On the other hand, since \( g \) is bounded from below, it follows that the sequences \( \{Ax^k - z^k + r^{-1}y^k\}_{k \geq 0} \) and \( \{x^{k+1} - x^k\}_{k \geq 0} \) are bounded as well. In addition, since for every \( k \geq 0 \) it holds
\[
\left\| A(x^{k+1} - x^k) - (z^{k+1} - z^k) \right\| \leq \|A\| \cdot \left\| x^{k+1} - x^k \right\| + \left\| y^{k+1} \right\| + \left\| z^k \right\|
\]
it follows that \( \{A(x^{k+1} - x^k) - (z^{k+1} - z^k)\}_{k \geq 0} \) is bounded, thus \( \{r^{-1}(y^{k+1} - y^k)\}_{k \geq 0} \) is bounded. According to the third update in the iterative scheme we obtain that \( \{Ax^k - z^k\}_{k \geq 0} \) is bounded and from here that \( \{y^k\}_{k \geq 0} \) is also bounded. This implies the boundedness of \( \{Ax^k\}_{k \geq 0} \) and, finally, since \( A \) is invertible, the boundedness of \( \{x^k\}_{k \geq 0} \).

(B-II) Again thanks to (14) there exists \( \sigma > 0 \) such that
\[
\frac{1}{\sigma} - \frac{L}{2\sigma^2} = \frac{3T_0}{2r}.
\]
We assume first that \( \rho \neq 1 \) or, equivalently, \( T_1 \neq 0 \). From Proposition 1 and (49) we see that for every \( k \geq 1 \)
\[
\frac{1}{2} h(z^{k+1}) + \frac{T_0}{4r} \left\| \nabla h(x^{k+1}) \right\|^2 + \frac{r}{2} \left\| Ax^{k+1} - z^{k+1} + \frac{1}{r}y^{k+1} \right\|^2 + \frac{T_1}{2} \left\| A^*(y^{k+1} - y^k) \right\|^2
\leq \mathcal{F}_1 - \inf_{z \in \mathbb{R}^n} g(z) - \frac{1}{2} \inf_{x \in \mathbb{R}^n} \left\{ h(x) - \frac{3T_0}{2r} \left\| \nabla h(x) \right\|^2 \right\} < +\infty.
\]
Since \( h \) is coercive and bounded from below, we obtain that \( \{x^k\}_{k \geq 0}, \{Ax^k - z^k + r^{-1}y^k\}_{k \geq 0} \) and \( \{A^*(y^{k+1} - y^k)\}_{k \geq 0} \) are bounded. For every \( k \geq 0 \) we have that
\[
\lambda_{\min}(A^*A) \rho^2 r^2 \left\| Ax^{k+1} - z^{k+1} \right\|^2 = \lambda_{\min}(A^*A) \left\| y^{k+1} - y^k \right\|^2 \leq \left\| A^*(y^{k+1} - y^k) \right\|^2,
\]
thus \( \{Ax^k - z^k\}_{k \geq 0} \) is bounded. Consequently, \( \{y^k\}_{k \geq 0} \) and \( \{z^k\}_{k \geq 0} \) are bounded.

In case \( \rho = 1 \) or, equivalently, \( T_1 = 0 \), we have that for every \( k \geq 1 \)
\[
\frac{1}{2} h(z^{k+1}) + \frac{T_0}{4r} \left\| \nabla h(x^{k+1}) \right\|^2 + \frac{r}{2} \left\| Ax^{k+1} - z^{k+1} + \frac{1}{r}y^{k+1} \right\|^2
\leq \mathcal{F}_1 - \inf_{z \in \mathbb{R}^n} g(z) - \frac{1}{2} \inf_{x \in \mathbb{R}^n} \left\{ h(x) - \frac{3T_0}{2r} \left\| \nabla h(x) \right\|^2 \right\} < +\infty
\]
from which we deduce that \( \{x^k\}_{k \geq 0} \) and \( \{Ax^k - z^k + r^{-1}y^k\}_{k \geq 0} \) are bounded. From Lemma 5 (iii), which now reads
\[
\left\| y^{k+1} - y^k \right\| \leq C_3 \left\| z^{k+1} - x^k \right\| + C_4 \left\| x^k - x^{k-1} \right\| \quad \forall k \geq 1,
\]
it yields that \( \{y^{k+1} - y^k\}_{k \geq 0} \) is bounded, thus, \( \{Ax^k - z^k\}_{k \geq 0} \) is bounded. Consequently, \( \{y^k\}_{k \geq 0} \) and \( \{z^k\}_{k \geq 0} \) are bounded.

Both considered scenarios lead to the conclusion that the sequence \( \{(x^k, z^k, y^k)\}_{k \geq 0} \) is bounded. □

We state now the first convergence result of this paper.

**Theorem 7.** Suppose that Assumption 1 holds true and let \( \{(x^k, z^k, y^k)\}_{k \geq 0} \) be a sequence generated by Algorithm 2 or Algorithm 3, which is assumed to be bounded. The following statements are true:
(i) for every \( k \geq 1 \) it holds
\[
F_{k+1} + \frac{C_1}{4} \|x^{k+1} - x^k\|^2 + \frac{1}{2} \|z^{k+1} - z^k\|^2_{M_z^k} \leq F_k;
\]
(ii) the sequence \( \{F_k\}_{k \geq 0} \) is bounded from below and convergent. In addition,
\[
x^{k+1} - x^k \to 0, \quad z^{k+1} - z^k \to 0 \quad \text{and} \quad y^{k+1} - y^k \to 0 \quad \text{as} \quad k \to +\infty;
\]
(iii) the sequences \( \{F_k\}_{k \geq 0} \), \( \{\mathcal{L}_r(x^k, z^k, y^k)\}_{k \geq 0} \) and \( \{h(x^k) + g(z^k)\}_{k \geq 0} \) have the same limit, which we denote by \( \mathcal{F}_\ast \in \mathbb{R} \).

Proof. (i) According to \((15)\) we have that \( M_z^k - C_1 \text{Id} \in \mathcal{P}_\varnothing_{\varnothing} \) and thus \((48)\) implies \((50)\).

(ii) We will show that \( \{\mathcal{L}_r(x^k, z^k, y^k)\}_{k \geq 0} \) is bounded from below, which will imply that \( \{F_k\}_{k \geq 0} \) is bounded from below as well. Assuming the contrary, as \( \{ (x^k, z^k, y^k) \}_{k \geq 0} \) is bounded, there exists a subsequence \( \{ (x^{k_n}, z^{k_n}, y^{k_n}) \}_{n \geq 0} \) converging to an element \( (\bar{x}, \bar{z}, \bar{y}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \) such that \( \{\mathcal{L}_r(x^{k_n}, z^{k_n}, y^{k_n})\}_{n \geq 0} \) converges to \( -\infty \) as \( n \to +\infty \). However, using the lower semicontinuity of \( g \) and the continuity of \( h \), we obtain
\[
\liminf_{q \to +\infty} \mathcal{L}_r(x^q, z^q, y^q) \geq h(\bar{x}) + g(\bar{z}) + \langle \bar{y}, A\bar{x} - \bar{z} \rangle + \frac{r}{2} \|A\bar{x} - \bar{z}\|^2,
\]
which leads to a contradiction. From Lemma \( 2 \) we conclude that \( \{F_k\}_{k \geq 1} \) is convergent and
\[
\sum_{k \geq 0} \|x^{k+1} - x^k\|^2 < +\infty,
\]
thus \( x^{k+1} - x^k \to 0 \) as \( k \to +\infty \).

We proved in \((26), (28), (36)\) and \((38)\) that for every \( k \geq 1 \)
\[
\frac{1}{pr} \|y^{k+1} - y^k\|^2 \leq \frac{C_0 - L}{2} \|x^{k+1} - x^k\|^2 + \frac{C_1}{2} \|x^k - x^{k-1}\|^2
+ T_1 \|A^* (y^k - y^{k-1})\|^2 - T_1 \|A^* (y^k - y^{k-1})\|^2.
\]
Summing up the above inequality for \( k = 1, \ldots, K \), for \( K > 1 \), we get
\[
\frac{1}{pr} \sum_{k=1}^K \|y^{k+1} - y^k\|^2 \leq \frac{C_0 - L}{2} \sum_{k=1}^K \|x^{k+1} - x^k\|^2 + \frac{C_1}{2} \sum_{k=1}^K \|x^k - x^{k-1}\|^2
+ T_1 \|A^* (y^1 - y^0)\|^2 - T_1 \|A^* (y^1 - y^0)\|^2
\leq \frac{C_0 - L}{2} \sum_{k=1}^K \|x^{k+1} - x^k\|^2 + \frac{C_1}{2} \sum_{k=1}^K \|x^k - x^{k-1}\|^2 + T_1 \|A^* (y^1 - y^0)\|^2.
\]
We let \( K \) converge to \( +\infty \) and conclude
\[
pr \sum_{k \geq 0} \|Ax^{k+1} - z^{k+1}\|^2 = \frac{1}{pr} \sum_{k \geq 0} \|y^{k+1} - y^k\|^2 < +\infty,
\]
thus \( Ax^{k+1} - z^{k+1} \to 0 \) and \( y^{k+1} - y^k \to 0 \) as \( k \to +\infty \). Since \( x^{k+1} - x^k \to 0 \) as \( k \to +\infty \), it follows that \( z^{k+1} - z^k \to 0 \) as \( k \to +\infty \).

(iii) By using \((51)\) and the fact that \( \{y^k\}_{k \geq 0} \) is bounded, it follows
\[
F_\ast = \lim_{k \to +\infty} F_k = \lim_{k \to +\infty} \mathcal{L}_r(x^k, z^k, y^k) = \lim_{k \to +\infty} \{h(x^k) + g(z^k)\},
\]
which is the desired statement.

The following lemmas provides upper estimates in terms of the iterates for limiting subgradients of the augmented Lagrangian and the regularized augmented Lagrangian \( \mathcal{F}_r \), respectively.
Lemma 8. Suppose that Assumption \[1\] holds true and let \(\{(x^k, z^k, y^k)\}_{k \geq 0}\) be a sequence generated by Algorithm \[1\] or Algorithm \[2\]. For every \(k \geq 0\) we have
\[
d^{k+1} := (d_x^{k+1}, d_z^{k+1}, d_y^{k+1}) \in \partial \mathcal{L}_r (x^{k+1}, z^{k+1}, y^{k+1}),
\]
where
\[
d_x^{k+1} := C_2 (\nabla h (x^{k+1}) - \nabla h (x^k)) + A^* (y^{k+1} - y^k) + M^k_1 (x^k - x^{k+1}), \tag{53a}
d_z^{k+1} := y^k - y^{k+1} + r A (x^k - x^{k+1}) + M^k_2 (z^k - z^{k+1}), \tag{53b}
d_y^{k+1} := \frac{1}{\rho r} (y^{k+1} - y^k). \tag{53c}
\]
and
\[
C_2 := \begin{cases} 
0, & \text{for Algorithm } 1 \vspace{1mm} \\
1, & \text{for Algorithm } 2
\end{cases}
\]
Moreover, for every \(k \geq 0\) it holds
\[
\|d^{k+1}\| \leq C_5 \|x^{k+1} - x^k\| + C_6 \|z^{k+1} - z^k\| + C_7 \|y^{k+1} - y^k\|, \tag{54}
\]
where
\[
C_5 := C_2 L + \mu_1 + r \|A\|, \quad C_6 := \mu_2, \quad C_7 := 1 + \|A\| + \frac{1}{\rho r}.
\]
Proof. Let \(k \geq 0\) be fixed. Applying the calculus rules of the limiting subdifferential, we obtain
\[
\nabla \mathcal{L}_r (x^{k+1}, z^{k+1}, y^{k+1}) = \nabla h (x^{k+1}) + A^* (y^{k+1} - y^k) + r A^* (Ax^{k+1} - z^{k+1}), \tag{55a}
\]
\[
\partial \mathcal{L}_r (x^{k+1}, z^{k+1}, y^{k+1}) = \partial g (z^{k+1}) - y^{k+1} - r (Ax^{k+1} - z^{k+1}), \tag{55b}
\]
\[
\nabla \mathcal{L}_r (x^{k+1}, z^{k+1}, y^{k+1}) = Ax^{k+1} - z^{k+1}. \tag{55c}
\]
Then (53c) follows directly from (55c) and (12a), respectively, while (53b) follows from
\[
y^k + r (Ax^k - z^{k+1}) + M^k_2 (z^k - z^{k+1}) \in \partial g (z^{k+1}),
\]
which is a consequence of the optimality criterion of (12a) and (13a), respectively. In order to derive (53a), let us notice that for Algorithm 1 we have (see (19))
\[
-A^* y^k + M^k_1 (x^k - x^{k+1}) = \nabla h (x^{k+1}) + r A^* (Ax^{k+1} - z^{k+1}), \tag{56}
\]
while for Algorithm 2 we have (see (29))
\[
-\nabla h (x^k) - A^* y^k + M^k_1 (x^k - x^{k+1}) = r A^* (Ax^{k+1} - z^{k+1}). \tag{57}
\]
By using (55a) we get the desired statement.

Relation (54) follows by combining the inequalities
\[
\|d_x^{k+1}\| \leq (C_2 L + \mu_1) \|x^{k+1} - x^k\| + \|A\| \cdot \|y^{k+1} - y^k\|,
\]
\[
\|d_z^{k+1}\| \leq \|y^k - y^{k+1}\| + r \|A\| \cdot \|x^{k+1} - x^k\| + \mu_2 \|z^{k+1} - z^k\|
\]
and (5). \qed

Lemma 9. Suppose that Assumption \[1\] holds true and let \(\{(x^k, z^k, y^k)\}_{k \geq 0}\) be a sequence generated by Algorithm \[1\] or Algorithm \[2\]. For every \(k \geq 0\) we have
\[
D^{k+1} := (D_x^{k+1}, D_z^{k+1}, D_y^{k+1}, D_x^{k+1}, D_y^{k+1}) \in \partial \mathcal{F}_r (x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k)
\]
where
\[
D_x^{k+1} := d_x^{k+1} + C_1 (x^{k+1} - x^k), \quad D_z^{k+1} := d_z^{k+1} + C_2 (y^{k+1} - y^k), \quad D_y^{k+1} := -C_3 (x^{k+1} - x^k), \quad D_y^{k+1} := -C_4 (y^{k+1} - y^k).
\]
Moreover, for every \(k \geq 0\) it holds
\[
\|D^{k+1}\| \leq C_8 \|x^{k+1} - x^k\| + C_9 \|z^{k+1} - z^k\| + C_{10} \|y^{k+1} - y^k\|,
\]
where
\[
C_8 := 2C_1 + C_5, \quad C_9 := C_6, \quad C_{10} := C_7 + 4T_1 \|A\|^2.
\]
Proof. Let \( k \geq 0 \) be fixed. Applying the calculus rules of the limiting subdifferential it follows

\[
\begin{align*}
\nabla_x \mathcal{F}_r (x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k) &:= \nabla_x \mathcal{L}_r (x^{k+1}, z^{k+1}, y^{k+1}) + C_1 (x^{k+1} - x^k), \\
\nabla_z \mathcal{F}_r (x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k) &:= \nabla_z \mathcal{L}_r (x^{k+1}, z^{k+1}, y^{k+1}), \\
\nabla_y \mathcal{F}_r (x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k) &:= -C_1 (x^{k+1} - x^k), \\
\nabla_y \mathcal{F}_r (x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k) &:= -2T_1 A A^* (y^{k+1} - y^k).
\end{align*}
\]

Then \((68)\) follows directly from the above relations and \((52)\). Inequality \((60)\) follows by combining

\[
\begin{align*}
\|D^k_{x+1}\| &\leq \|d^k_{x+1}\| + C_1 \|x^{k+1} - x^k\|, \\
\|D^k_{y+1}\| &\leq \|d^k_{y+1}\| + 2T_1 \|A\|^2 \|y^{k+1} - y^k\|.
\end{align*}
\]

and \((5)\). \( \square \)

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

**Corollary 10.** Suppose that Assumption \((7)\) holds true and let \(\{(x^k, z^k, y^k)\}_{k \geq 0}\) be a sequence generated by Algorithm \((2)\) or Algorithm \((3)\). Then the norm of the element \(D^k_{x+1} \in \partial \mathcal{F}_r (x^{k+1}, z^{k+1}, y^{k+1}, x^k, y^k)\) defined in the previous lemma verifies for every \(k \geq 2\) the following estimate

\[
\|D^k_{x+1}\| \leq C_{11} \left( \|x^{k+1} - x^k\| + \|x^k - x^{k-1}\| + \|x^{k-1} - x^{k-2}\| \right) + C_{12} \left( \|A^* (y^k - y^{k-1})\| - \|A^* (y^{k+1} - y^k)\| \right) + C_{13} \left( \|A^* (y^{k-1} - y^{k-2})\| - \|A^* (y^{k} - y^{k-1})\| \right),
\]

where

\[
C_{11} := \max \left\{ C_8 + C_9 \|A\| + C_3 C_{10} + \frac{C_3 C_9}{pr}, C_4 C_{10} + \frac{C_3 C_9}{pr}, \frac{C_4 C_9}{pr} \right\},
\]

\[
C_{12} := \left( C_{10} + \frac{C_9}{pr} \right) T_2, \quad C_{13} := \frac{C_6 T_2}{pr}.
\]

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

**Lemma 11.** Suppose that Assumption \((7)\) holds true and let \(\{(x^k, z^k, y^k)\}_{k \geq 0}\) be a sequence generated by Algorithm \((2)\) or Algorithm \((3)\) which is assumed to be bounded. The following statements are true:

(i) if \(\{(x^k, z^k, y^k)\}_{q \geq 0}\) is a subsequence of \(\{(x^k, z^k, y^k)\}_{k \geq 0}\) which converges to \((\hat{x}, \hat{z}, \hat{y})\) as \(q \to +\infty\), then

\[
\lim_{q \to +\infty} \mathcal{L}_r (x^k, z^k, y^k) = \mathcal{L}_r (\hat{x}, \hat{z}, \hat{y});
\]

(ii) it holds

\[
\omega \left(\{(x^k, z^k, y^k)\}_{k \geq 0}\right) \subseteq \text{crit} \left(\mathcal{L}_r\right)
\]

\[
\subseteq \{ (\hat{x}, \hat{z}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m : -A^* \hat{y} = \nabla h (\hat{x}), \hat{y} \in \partial g (\hat{z}), \hat{z} = A \hat{x} \};
\]

(iii) we have \(\lim_{k \to +\infty} \text{dist} \left(\{(x^k, z^k, y^k)\}_{k \geq 0}\}, \omega \left(\{(x^k, z^k, y^k)\}_{k \geq 0}\right)\right) = 0\);

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

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

**Proof.** Let \((\hat{x}, \hat{z}, \hat{y}) \in \omega \left(\{(x^k, z^k, y^k)_{k \geq 0}\}\right)\) and \(\{(x^k, z^k, y^k)\}_{q \geq 0}\) be a subsequence of \(\{(x^k, z^k, y^k)_{k \geq 0}\}\) converging to \((\hat{x}, \hat{z}, \hat{y})\) as \(q \to +\infty\).
(i) From either (12a) or (13a) we obtain for every \( q \geq 1 \)

\[
\begin{align*}
g(z^k) + \langle y^{k-1}, Ax^{k-1} - z^k \rangle + \frac{r}{2} \| Ax^{k-1} - z^k \|^2 + \frac{1}{2} \| z^{k-1} - z^k \|^2 M^2_{k-1} \leq g(\bar{z}) + \langle y^{k-1}, Ax^{k-1} - \bar{z} \rangle + \frac{r}{2} \| Ax^{k-1} - \bar{z} \|^2 + \frac{1}{2} \| \bar{z} - z^{k-1} \|^2 M^2_{k-1}.
\end{align*}
\]

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

\[
\limsup_{q \to +\infty} g(z^k) \leq g(\bar{z}),
\]

which, combined with the lower semicontinuity of \( g \), leads to

\[
\lim_{q \to +\infty} g(z^k) = g(\bar{z}).
\]

Since \( h \) is continuous, we further obtain

\[
\lim_{q \to +\infty} L_r(x^{k_q}, z^{k_q}, y^{k_q}) = \lim_{q \to +\infty} \left[ g(z^k) + h(x^k) + \langle y^k, Ax^k - z^k \rangle + \frac{r}{2} \| Ax^k - z^k \|^2 \right] = g(\bar{z}) + \langle \bar{y}, \bar{x} - \bar{z} \rangle + \frac{r}{2} \| \bar{x} - \bar{z} \|^2 = L_r(\bar{x}, \bar{z}, \bar{y}).
\]

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

\[
\begin{align*}
0 &= \nabla h(\bar{x}) + A^* \bar{y} + r A^* (A \bar{x} - \bar{z}) \\
0 &\in \partial g(\bar{z}) - \bar{y} - r (A \bar{x} - \bar{z}) \\
0 &= A \bar{x} - \bar{z},
\end{align*}
\]

which is further equivalent to

\[
-A^* \bar{y} = \nabla h(\bar{x}), \quad \bar{y} \in \partial g(\bar{z}), \quad \bar{z} = A \bar{x}.
\]

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

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

\[\square\]

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

\[
-A^* \bar{y} = \nabla h(\bar{x}), \quad \bar{y} \in \partial g(\bar{z}), \quad \bar{z} = A \bar{x}
\]

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

\[
0 = A^* \partial g(A \bar{x}) + \nabla h(\bar{x}). \tag{63}
\]

When \( A \) is injective this is further equivalent to

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

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

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

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

**Lemma 12.** Suppose that Assumption 7 holds true and let \( \{(x^k, z^k, y^k)\}_{k \geq 0} \) be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. Let \( \Omega := \omega \left\{ (x^k, z^k, y^k, x^{k-1}, y^{k-1}) \right\}_{k \geq 1} \). The following statements are true:
Lemma 13. and to the references therein for more properties of K L functions and illustrating examples.

functions and convex functions satisfying a growth condition. We refer the reader to \[2, 3, 4, 7, 8, 9, 10\]

to the class of K L functions belong semialgebraic, real subanalytic, uniformly convex

K L property reveals the possibility to reparameterize the values of \(\Psi\) in order to avoid flatness around

it holds

in the intersection 

\(u\)

\(\Omega\)

\(\Omega \subseteq \{(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^m : (\hat{x}, \hat{y}) \in \text{crit}(L)\};

(ii) we have

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

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

(iv) the regularized augmented Lagrangian \(F_r\) takes on \(\Omega\) the value \(F_\ast = \lim_{k \to +\infty} F_k\), as the objective

function \(g \circ A + h\) does on the projection of the set \(\Omega\) onto the space \(\mathbb{R}^n\) corresponding to the first component.

2.3 Convergence analysis under Kurdyka-Lojasiewicz assumptions

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

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

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

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

2. \(\varphi\) is \(C^1\) on \((0, \eta)\) and continuous at 0;

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

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

1. The function \(\Psi\) is said to have the Kurdyka-Łojasiewicz (KL) property at a point \(\hat{u} \in \text{dom} \partial \Psi := \{u \in \mathbb{R}^N : \partial \Psi(u) \neq \emptyset\}\), if there exists \(\eta \in (0, +\infty]\), a neighborhood \(U\) of \(\hat{u}\) and a function \(\varphi \in \Phi_\eta\) such that for every

\[
u \in U \cap [\Psi(\hat{u}) < \Psi(u) < \Psi(\hat{u}) + \eta]
\]

it holds

\[
\varphi'(\Psi(u) - \Psi(\hat{u})) \cdot \text{dist}(0, \partial \Psi(u)) \geq 1.
\]

2. If \(\Psi\) satisfies the KL property at each point of \(\text{dom} \partial \Psi\), then \(\Psi\) is called KL function.

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

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

Lemma 13. (Uniformized KL property) Let \(\Omega\) be a compact set and \(\Psi : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}\) be a proper and lower semicontinuous function. Assume that \(\Psi\) is constant on \(\Omega\) and satisfies the KL property at each point of \(\Omega\). Then there exist \(\varepsilon > 0, \eta > 0\) and \(\varphi \in \Phi_\eta\) such that for every \(\hat{u} \in \Omega\) and every element \(u\) in the intersection

\[
\{u \in \mathbb{R}^N : \text{dist}(u, \Omega) < \varepsilon\} \cap [\Psi(\hat{u}) < \Psi(u) < \Psi(\hat{u}) + \eta]
\]

it holds

\[
\varphi' (\Psi(u) - \Psi(\hat{u})) \cdot \text{dist}(0, \partial \Psi(u)) \geq 1.
\]
Thus, and telescoping sum arguments, it yields for every \( x \in \mathbb{R}^n \) that the inequality in (50) with Lemma 5, it yields that
\[
\phi(x) - \phi^* \to 0 \quad \text{as} \quad k \to +\infty.
\]
The next result shows that, if the regularization of the augmented Lagrangian \( F_r \) is a KL function, then the sequence \( \{ (x^k, z^k, y^k) \} \) converges to a KKT point of the optimization problem 1.

**Theorem 14.** Suppose that Assumption \( \square \) holds true and let \( \{ (x^k, z^k, y^k) \} \) be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. If \( F_r \) is a KL function, then the following statements are true:

(i) the sequence \( \{ (x^k, z^k, y^k) \} \) has finite length, namely,
\[
\sum_{k \geq 0} \| x^{k+1} - x^k \| < +\infty, \quad \sum_{k \geq 0} \| z^{k+1} - z^k \| < +\infty, \quad \sum_{k \geq 0} \| y^{k+1} - y^k \| < +\infty;
\]

(ii) the sequence \( \{ (x^k, z^k, y^k) \} \) converges to a KKT point of the optimization problem 1.

**Proof.** As in Lemma 12 we denote by \( \Omega := \phi \left( \{ (x^k, z^k, y^k, x^{k-1}, y^{k-1}) \} \right) \), which is a nonempty set. Let be \( (\hat{x}, \hat{z}, \hat{y}) \in \Omega \), thus \( F_r(\hat{x}, \hat{z}, \hat{y}) = \phi^* \). We have seen that \( \phi_k = F_k - F^* \) converges to 0 as \( k \to +\infty \) and will consider, consequently, two cases.

We assume first that there exists an integer \( k' \geq 0 \) such that \( \phi_{k'} = 0 \) or, equivalently, \( F_{k'} = \phi^* \). Due to the monotonicity of \( \phi_k \), it follows that \( \phi_k = 0 \) or, equivalently, \( F_k = \phi^* \) for all \( k \geq k' \). Combining the inequality in (60) with Lemma 5, it yields that \( x^{k+1} - x^k = 0 \) for all \( k \geq k' + 1 \). Using Lemma 5 (iii) and telescoping sum arguments, it yields \( \sum_{k \geq 0} \| y^{k+1} - y^k \| < +\infty \). Finally, by using Lemma 5 (i), we obtain that \( \sum_{k \geq 0} \| x^{k+1} - x^k \| < +\infty \).

Consider now the case when \( \phi_k > 0 \) or, equivalently, \( F_k > \phi^* \) for every \( k \geq 1 \). According to Lemma 13 there exist \( \epsilon > 0 \), \( \eta > 0 \) and a desingularization function \( \phi \) such that for every element \( u \) in the intersection
\[
\{ u \in \mathbb{R}^n \times \mathbb{R}^m : \text{dist}(u, \Omega) < \epsilon \} \cap \{ u \in \mathbb{R}^n \times \mathbb{R}^m : F^* < F_r(u) < F^* + \eta \}
\]
it holds
\[
\phi'(F_r(u) - \phi^*) \cdot \text{dist}(0, \partial F_r(u)) \geq 1.
\]
Let be \( k_1 \geq 1 \) such that for every \( k \geq k_1 \)
\[
F_* < F_k < F^* + \eta.
\]
Since \( \lim_{k \to +\infty} \text{dist}((x^k, z^k, y^k, x^{k-1}, y^{k-1}), \Omega) = 0 \), see Lemma 12 (ii), there exists \( k_2 \geq 1 \) such that for every \( k \geq k_2 \)
\[
\text{dist}((x^k, z^k, y^k, x^{k-1}, y^{k-1}), \Omega) < \epsilon.
\]
Thus, \( (x^k, z^k, y^k, x^{k-1}, y^{k-1}) \) belongs to the intersection in (66) for every \( k \geq k_0 := \max \{ k_1, k_2, 3 \} \), which further implies
\[
\phi'(F_k - \phi^*) \cdot \text{dist}(0, \partial F_r(x^k, z^k, y^k, x^{k-1}, y^{k-1})) = \phi'(\phi_k) \cdot \text{dist}(0, \partial F_r(x^k, z^k, y^k, x^{k-1}, y^{k-1})) \geq 1.
\]

Define for two arbitrary nonnegative integers \( p, q \)
\[
\Delta_{p,q} := \phi(F_p - \phi^*) - \phi(F_q - \phi^*) = \phi(\phi_p) - \phi(\phi_q).
\]
For every \( K \geq k_0 \geq 1 \) it holds
\[
\Delta_{k,k+1} = \Delta_{k_0,K+1} = \phi(\phi_k) - \phi(\phi_{K+1}) \leq \phi(\phi_k),
\]

21
from which we get \( \sum_{k \geq 1} \Delta_{k,k+1} < +\infty \).

By combining Theorem \([7]\) (i) with the concavity of \( \varphi \) we obtain for every \( k \geq 1 \)

\[
\Delta_{k,k+1} = \varphi(\mathbf{E}_k) - \varphi(\mathbf{E}_{k+1}) \geq \varphi'(\mathbf{E}_k) [\mathbf{E}_k - \mathbf{E}_{k+1}] = \varphi'(\mathbf{E}_k) [\mathcal{F}_k - \mathcal{F}_{k+1}] \geq \varphi'(\mathbf{E}_k) \frac{C_1}{4} \| x^{k+1} - x^k \|^2.
\]

The last relation combined with \([67]\) imply

\[
\| x^{k+1} - x^k \|^2 \leq \varphi'(\mathbf{E}_k) \cdot \text{dist} \left( 0, \partial \mathcal{F}_r \left( x^k, z^k, y^k, x^{k-1}, y^{k-1} \right) \right) \| x^{k+1} - x^k \|^2 \\
\leq \frac{4}{C_1} \Delta_{k,k+1} \cdot \text{dist} \left( 0, \partial \mathcal{F}_r \left( x^k, z^k, y^k, x^{k-1}, y^{k-1} \right) \right) \forall k \geq k_0.
\]

By the arithmetic mean-geometric mean inequality and Corollary \([10]\) we have that for every \( k \geq k_0 \) and every \( \beta > 0 \)

\[
\| x^{k+1} - x^k \| \leq \sqrt{\frac{4}{C_1} \Delta_{k,k+1} \cdot \text{dist} \left( 0, \partial \mathcal{F}_r \left( x^k, z^k, y^k, x^{k-1}, y^{k-1} \right) \right)} \\
\leq \frac{\beta}{C_1} \Delta_{k,k+1} + \frac{1}{\beta} \text{dist} \left( 0, \partial \mathcal{F}_r \left( x^k, z^k, y^k, x^{k-1}, y^{k-1} \right) \right) \\
\leq \frac{\beta}{C_1} \Delta_{k,k+1} + \frac{C_{11}}{\beta} \left( \| x^k - x^{k-1} \| + \| x^{k-1} - x^{k-2} \| + \| x^{k-2} - x^{k-3} \| \right) \\
+ \frac{C_{12}}{\beta} \left( \| A^* (y^{k-1} - y^{k-2}) \| - \| A^* (y^k - y^{k-1}) \| \right) \\
+ \frac{C_{13}}{\beta} \left( \| A^* (y^{k-2} - y^{k-3}) \| - \| A^* (y^{k-1} - y^{k-2}) \| \right).
\]

We denote for every \( k \geq 3 \)

\[
a^k := \| x^k - x^{k-1} \| \geq 0, \\
\delta_k := \frac{\beta}{C_1} \Delta_{k,k+1} + \frac{C_{12}}{\beta} \left( \| A^* (y^{k-1} - y^{k-2}) \| - \| A^* (y^k - y^{k-1}) \| \right) \\
+ \frac{C_{13}}{\beta} \left( \| A^* (y^{k-2} - y^{k-3}) \| - \| A^* (y^{k-1} - y^{k-2}) \| \right).
\]

The inequality \([69]\) is nothing than \([9]\) with \( c_0 = c_1 = c_2 := \frac{C_{11}}{\beta} \). Observe that for every \( K \geq k_0 = 0 \), we have

\[
\sum_{k=k_0}^K \delta_k \leq \frac{\beta}{C_1} \varphi(\mathbf{E}_{k_0}) + \frac{C_{12}}{\beta} \| A^* (y^{k_0-1} - y^{k_0-2}) \| + \frac{C_{13}}{\beta} \| A^* (y^{k_0-2} - y^{k_0-3}) \|.
\]

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

\[
\sum_{k \geq 0} \| x^{k+1} - x^k \| < +\infty.
\]

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

\[\square\]

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

### 3 Convergence rates under Łojasiewicz assumptions

In this section we derive convergence rates for the sequence \( \{ (x^k, z^k, y^k) \}_{k \geq 0} \) 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 Lojasiewicz property and a technical lemma

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

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

$$|\Psi(u) - \Psi(\hat{u})|^\theta \leq C_L \cdot \text{dist}(0, \partial \Psi(u)) \forall u \in \text{Ball}(\hat{u}, \varepsilon),$$

where $\text{Ball}(\hat{u}, \varepsilon)$ denotes the open ball with centre $\hat{u}$ and radius $\varepsilon$.

If Assumption 1 is fulfilled and $\{(x^k, z^k, y^k)\}_{k \geq 0}$ is the sequence generated by Algorithm 1 or Algorithm 2, assumed to be bounded, then, as seen in Lemma 12, the set of cluster points $\Omega$ where

$$\text{Ball}(\hat{u}, \varepsilon) = \omega(\{(x^k, z^k, y^k, x^{k-1}, y^{k-1})\}_{k \geq 0})$$

is nonempty, compact and connected and $\mathcal{F}_r$ takes on $\Omega$ the value $\mathcal{F}_a$; in addition, for every $(\hat{x}, \hat{z}, \hat{y}, \hat{x}, \hat{y}) \in \Omega$, $(\hat{x}, \hat{z}, \hat{y})$ belongs to crit($\mathcal{L}_r$). According to [2] Lemma 1, if $\mathcal{F}_r$ has the Lojasiewicz property, then there exist $C_L > 0$, $\theta \in [0, 1]$ and $\varepsilon > 0$ such that for every

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

it holds

$$|\mathcal{F}_r(x, z, y, x', y') - \mathcal{F}_a|^\theta \leq C_L \cdot \text{dist}(0, \partial \mathcal{F}_r(x, z, y, x', y')).$$

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

$$|\mathcal{F}_r(x, z, y, x', y') - \mathcal{F}_a|^\theta \leq C_L \cdot \text{dist}(0, \partial \mathcal{F}_r(x, z, y, x', y')) \forall (x, z, y, x', y') \in \text{Ball}(\hat{x}, \hat{z}, \hat{y}, \varepsilon).$$

(70)

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

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

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

$$e_{k-l_0} - e_k \geq C_e e_k^\theta,$$

(71)

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

(i) if $\theta = 0$, then $\{e_k\}_{k \geq 0}$ 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 $k \geq k_0$

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

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

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

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

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

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

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

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

23
(ii) The case when $\theta \in (0, 1/2]$. We have $e_k \leq e_0$, which leads to $e_{k-l_0} - e_k \geq C_{e_0} e_k^{2\theta} \geq C_{e_0} e_0^{2\theta-1} e_k$ for every $k \geq k_0$. Therefore,

$$e_k \leq \left( \frac{1}{C_{e_0} e_0^{2\theta-1} + 1} \right)^{k_0 - l_0 - 1} e_0 = e_0 \left( \frac{1}{C_{e_0} e_0^{2\theta-1} + 1} \right)^{l_0 + 1} \left( \frac{1}{Q C_{e_0} e_0^{2\theta-1} + 1} \right)^k.
$$

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

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

Define $\zeta$: $(0, +\infty) \to \mathbb{R}, \zeta(s) = s^{-2\theta}$. 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, $\zeta(e_{k-l_0}) \leq \zeta(s)$ for all $s \in [e_k, e_{k-l_0}]$.

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

$$C_e \leq 2 \zeta(e_{k-l_0}) \int_{e_k}^{e_{k-l_0}} 1 ds \leq 2 \int_{e_k}^{e_{k-l_0}} \zeta(s) ds = \frac{2}{2\theta - 1} \left( e_1^{1-2\theta} - e_1^{1-2\theta} \right)
$$

or, equivalently,

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

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

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

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

$$e_i^{1-2\theta} - e_{i-l_0}^{1-2\theta} \geq C' := \min \{ C_1', C_2' \} > 0,
$$

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

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

Since

$$l_0 \left( e_1^{1-2\theta} - e_{k_0}^{1-2\theta} \right) \geq \sum_{j=0}^{l_0 - 1} \left( e_{k-j}^{1-2\theta} - e_{k_0 + j}^{1-2\theta} \right) \geq (k - k_0 - l_0 + 1) C',
$$

we have

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

We obtain from (75) that

$$e_k^{1-2\theta} \geq \left[ \frac{k_0 + l_0}{l_0} \right] C' \geq \left( \frac{k_0 + l_0}{l_0} - 1 \right) C' = k_0 l_0 C',
$$

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

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

which implies

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

This concludes the proof.
From (26), (28), (36) and (38) it follows that

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

### 3.2 Convergence rates

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

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

\[ 2M_1^2 + rA^*A \geq \left( L + \frac{C'M}{r} \right) \text{Id} \quad \forall k \geq 0, \tag{79} \]

where

\[ C'M := \begin{cases} 
10\mu^2_1 + 8 (L + \mu_1)^2 & \text{for Algorithm 1} \\
8\mu^2_1 + 10 (L + \mu_1)^2 & \text{for Algorithm 2} 
\end{cases} \]

The condition (79) is nothing else than (15) after replacing \( C_M \) by the bigger constant \( C'M \).

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

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

\[ \mathcal{L}_r (x^{k+1}, z^{k+1}, y^{k+1}) + 2T_1 \| A^* (y^{k+1} - y^k) \|^2 + \frac{1}{2} \| x^{k+1} - x^k \|^2_{M_3} + \frac{1}{2} \| z^{k+1} - z^k \|^2_{M_2} \]

\[ + \frac{1}{\rho r} \| y^{k+1} - y^k \|^2 \]

\[ \leq \mathcal{L}_r (x^k, z^k, y^k) + 2T_1 \| A^* (y^k - y^{k-1}) \|^2 + C_1 \| y^k - y^{k-1} \|^2. \tag{80} \]

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

\[ \mathcal{L}_r (x^{k+1}, z^{k+1}, y^{k+1}) + \frac{1}{2} \| x^{k+1} - x^k \|^2_{2M_1^2 + rA^*A} - \frac{L}{2} \| x^{k+1} - x^k \|^2 + \frac{1}{2} \| z^{k+1} - z^k \|^2_{M_2} \]

\[ \leq \mathcal{L}_r (x^k, z^k, y^k) + \frac{1}{\rho r} \| y^{k+1} - y^k \|^2. \tag{81} \]

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

\[ \frac{1}{\rho r} \| y^{k+1} - y^k \|^2 \leq \frac{C_0 - L}{2} \| x^{k+1} - x^k \|^2 + \frac{C_1}{2} \| x^k - x^{k-1} \|^2 + T_1 \| A^* (y^k - y^{k-1}) \|^2 - T_1 \| A^* (y^{k+1} - y^k) \|^2. \tag{82} \]

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

We replace \( T_1 \) with \( 2T_1 \) in the definition of the regularized augmented Lagrangian \( \mathcal{F}_r \), thus, the sequence \( \{ \mathcal{F}_k \} \) in (17) becomes

\[ \mathcal{F}_k := \mathcal{L}_r (x^k, z^k, y^k) + 2T_1 \| A^* (y^k - y^{k-1}) \|^2 + C_1 \| y^k - y^{k-1} \|^2 \forall k \geq 1. \]

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

\[ \mathcal{F}_{k+1} + \frac{C_1}{4} \| x^{k+1} - x^k \|^2 + \frac{1}{2} \| z^{k+1} - z^k \|^2_{M_2} + \frac{1}{2} \| y^{k+1} - y^k \|^2 \leq \mathcal{F}_k \tag{83} \]
and provides an inequality which is tighter than relation (50) in Theorem 7. Furthermore, for a subgradient $D^{k+1}$ of $F_r$ at $(x^{k+1}, z^{k+1}, y^{k+1}, x^k, z^k)$ defined as in (59) (again by replacing $T_1$ by $2T_1$) we obtain for every $k \geq 2$ the following estimate, which is simpler than (62) in Corollary 10:

$$
\|D^{k+1}\| \leq C_{14} \|x^{k+1} - x^k\| + C_{15} \|y^{k+1} - y^k\| + C_{16} \|y^k - y^{k-1}\|,
$$

where

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

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

$$
\Delta_{k,k+1} = \varphi(E_k) - \varphi(E_{k+1}) \geq \varphi'(E_k) \min \left\{ \frac{C_1}{4}, \frac{1}{pr} \right\} \left( \|x^{k+1} - x^k\|^2 + \|y^{k+1} - y^k\|^2 \right),
$$

where

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

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

$$
(\|x^{k+1} - x^k\| + \|y^{k+1} - y^k\|)^2 \leq \frac{\Delta_{k,k+1}}{C_{17}} \cdot \text{dist} (0, \partial F_r (x^k, z^k, y^k, x^{k-1}, y^{k-1}))
$$

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

$$
\|x^{k+1} - x^k\| + \|y^{k+1} - y^k\| \leq \frac{\beta \Delta_{k,k+1}}{4C_{17}} + \max \left\{ \frac{C_{14} + C_{15}}{\beta} \right\} \left( \|x^k - x^{k-1}\| + \|y^k - y^{k-1}\| + \|y^k - y^{k-2}\| \right).
$$

By denoting

$$
a^k := \left( \|x^k - x^{k-1}\|, \|y^k - y^{k-1}\| \right) \in \mathbb{R}^2
$$

and $\delta_k := \frac{\beta \Delta_{k,k+1}}{4C_{17}}$

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

$$
\langle 1, a^{k+1} \rangle \leq \langle c_0, a^k \rangle + \langle c_1, a^{k-1} \rangle + \delta_k,
$$

where

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

Choosing $\beta > 2 \max \{C_{14}, C_{15}\}$, Lemma 8 and Lemma 9 imply that $\{ (x^k, z^k, y^k) \}_{k \geq 0}$ has finite length (see (65)).

Next we prove a recurrence inequality for the sequence $\{E_k\}_{k \geq 0}$.

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

$$
E_{k-1} - E_{k+1} \geq C_{19} E_{k+1}^{2\theta}, \quad \text{where} \quad C_{19} := \frac{\min \left\{ \frac{C_1}{4}, \frac{1}{pr} \right\}}{3C_L^2 \max \{C_{14} + C_{15}\}^{1/\theta}}.
$$

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

$$
E_{k-1} - E_{k+1} = F_{k-1} - F_k + F_k - F_{k+1}
$$

$$
\geq \min \left\{ \frac{C_1}{4}, \frac{1}{pr} \right\} \left( \|x^{k+1} - x^k\|^2 + \|y^{k+1} - y^k\|^2 + \|y^k - y^{k-1}\|^2 \right)
$$

$$
\geq \frac{1}{3} \min \left\{ \frac{C_1}{4}, \frac{1}{pr} \right\} \left( \|x^{k+1} - x^k\|^2 + \|y^{k+1} - y^k\|^2 + \|y^k - y^{k-1}\|^2 \right)^2
$$

$$
\geq C_{19} C_L^2 \|D^{k+1}\|^2.
$$

Let $\varepsilon > 0$ be such that (70) is fulfilled and choose $k_0 \geq 1$ such that $(x^{k+1}, z^{k+1}, y^{k+1})$ belongs to Ball $((\bar{x}, \bar{z}, \bar{y}), \varepsilon)$ for every $k \geq k_0$. Then (70) implies (86) for every $k \geq k_0$. \qed
The following convergence rates follow by combining Lemma 15 with Lemma 17.

Theorem 18. Suppose that Assumption 3 holds true and let \( \{(x^k, z^k, y^k)\}_{k \geq 0} \) be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. If \( F_r \) satisfies the Lojasiewicz property with Lojasiewicz constant \( C_L > 0 \) and Lojasiewicz exponent \( \theta \in [0, 1) \), then the following statements are true:

(i) if \( \theta = 0 \), then \( \{F_k\}_{k \geq 1} \) converges in finite time;

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

\[
0 \leq F_k - F_* \leq \tilde{C}_0 Q^k;
\]

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

\[
0 \leq F_k - F_* \leq \tilde{C}_1 (k - 1)^{-\frac{1}{\theta - 1}}.
\]

The next lemma will play an important role when transferring the convergence rates for \( \{F_k\}_{k \geq 0} \) to the sequence of iterates \( \{(x^k, z^k, y^k)\}_{k \geq 0} \) (see [23] for a similar statement).

Lemma 19. Suppose that Assumption 3 holds true and let \( \{(x^k, z^k, y^k)\}_{k \geq 0} \) be a sequence generated by Algorithm 1 or Algorithm 2, which is assumed to be bounded. Suppose further that \( F_r \) satisfies the Lojasiewicz property with Lojasiewicz constant \( C_L > 0 \), Lojasiewicz exponent \( \theta \in [0, 1) \) and desingularization function \( \varphi : [0, +\infty) \to [0, +\infty), \varphi(s) := \frac{1}{1 - \theta} C_L s^{1-\theta} \). Let \( (\bar{x}, \bar{z}, \bar{y}) \) be the KKT point of the optimization problem (1) to which \( \{(x^k, z^k, y^k)\}_{k \geq 0} \) converges as \( k \to +\infty \). Then there exists \( k_0 \geq 2 \) such that the following estimates hold for every \( k \geq k_0 \)

\[
\|x^k - \bar{x}\| \leq C_{20} \max \left\{ \sqrt{E_k}, \varphi(E_k) \right\}, \quad \text{where} \quad C_{20} := \frac{7}{\sqrt{C_{17}}} + \frac{1}{C_{17}}, \tag{87a}
\]

\[
\|y^k - \bar{y}\| \leq C_{21} \max \left\{ \sqrt{E_k}, \varphi(E_k) \right\}, \quad \text{where} \quad C_{21} := \frac{7}{2\sqrt{C_{17}}} + \frac{1}{2C_{17}}, \tag{87b}
\]

\[
\|z^k - \bar{z}\| \leq C_{22} \max \left\{ \sqrt{E_{k-1}}, \varphi(E_{k-1}) \right\}, \quad \text{where} \quad C_{22} := C_{20} \|A\| + \frac{2C_{21}}{\beta}. \tag{87c}
\]

Proof. We assume that \( E_k > 0 \) for every \( k \geq 0 \). Otherwise, beginning with a given index, the sequence \( \{(x^k, z^k, y^k)\}_{k \geq 0} \) becomes identical to \( (\bar{x}, \bar{z}, \bar{y}) \) and the conclusion follows as in the proof of Theorem 14.

Let \( \varepsilon > 0 \) be such that (70) is fulfilled and \( k_0 \geq 2 \) such that \( (x^{k+1}, z^{k+1}, y^{k+1}) \) belongs to \( \text{Ball}((\bar{x}, \bar{z}, \bar{y}), \varepsilon) \) for every \( k \geq k_0 \). We fix \( k \geq k_0 \). One can easily notice that

\[
\|x^k - \bar{x}\| \leq \|x^{k+1} - x^k\| + \|x^k - \bar{x}\| \leq \cdots \leq \sum_{l \geq k} \|x^{l+1} - x^l\| \tag{88a}
\]

and, similarly,

\[
\|z^k - \bar{z}\| \leq \sum_{l \geq k} \|z^{l+1} - z^l\| \quad \text{and} \quad \|y^k - \bar{y}\| \leq \sum_{l \geq k} \|y^{l+1} - y^l\|. \tag{88b}
\]

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

\[
\sum_{l \geq k} \|x^{l+1} - x^l\| = \sum_{l \geq k} \|a_{l+1}^1 - a_l^1\| \leq \|x^{k+1} - x^k\| + 2 \|x^k - y^k\| + \|y^{k+1} - y^k\| + \|y^{k+1} - y^k\| \frac{\varphi(E_k)}{C_{17}}
\]

\[
\leq \frac{2}{\sqrt{C_{17}}} \sqrt{F_k - F_{k+1}} + \frac{2}{\sqrt{C_{17}}} \sqrt{F_{k+1} - F_{k+2}} + \frac{3}{\sqrt{C_{17}}} \sqrt{F_{k+2} - F_{k+3}} + \frac{\varphi(E_k)}{C_{17}}.
\]

\[
\leq \frac{2}{\sqrt{C_{17}}} \sqrt{E_k} + \frac{2}{\sqrt{C_{17}}} \sqrt{E_{k+1}} + \frac{3}{\sqrt{C_{17}}} \sqrt{E_{k+2}} + \frac{\varphi(E_k)}{C_{17}}.
\]

\[
\leq 27.
and, similarly,
\[ \sum_{i=2}^{k} \| y_i^{+1} - y_i^+ \| \leq \frac{1}{\sqrt{C_{17}}} \sqrt{\mathcal{E}_k} + \frac{1}{\sqrt{C_{17}}} \sqrt{\mathcal{E}_{k+1}} + \frac{3}{2\sqrt{C_{17}}} \sqrt{\mathcal{E}_{k+2}} + \frac{\varphi(\mathcal{E}_k)}{2C_{17}}. \]

By taking into account the relations above, (88a)-(88b) as well as
\[ \sqrt{\mathcal{E}_{k+2}} \leq \sqrt{\mathcal{E}_{k+1}} \leq \sqrt{\mathcal{E}_k} \quad \text{and} \quad \varphi(\mathcal{E}_{k+1}) \leq \varphi(\mathcal{E}_k) \quad \forall k \geq 1, \]
the estimates (87a) and (87b) follow. Statement (87c) follows from Lemma 3 and by considering (88b).

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

**Theorem 20.** Suppose that Assumption 2 holds true and let \( \{(x^k, z^k, y^k)\}_{k \geq 0} \) be a sequence generated by Algorithm 2 or Algorithm 3, which is assumed to be bounded. Suppose further that \( F_r \) satisfies the Lojasiewicz property with Lojasiewicz constant \( C_L > 0 \) and Lojasiewicz exponent \( \theta \in [0, 1) \). Let \( (\hat{x}, \hat{z}, \hat{y}) \) be the KKT point of the optimization problem (1) to which \( \{(x^k, z^k, y^k)\}_{k \geq 0} \) converges as \( k \to +\infty \). Then the following statements are true:

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

(ii) if \( \theta \in (0, 1/2] \), then there exist \( k_0 \geq 1, \hat{C}_{0,1}, \hat{C}_{0,2}, \hat{C}_{0,3} > 0 \) and \( \hat{Q} \in [0, 1) \) such that for every \( k \geq k_0 \)
\[ \|x^k - \hat{x}\| \leq \hat{C}_{0,1} \hat{Q}^k, \quad \|y^k - \hat{y}\| \leq \hat{C}_{0,2} \hat{Q}^k, \quad \|z^k - \hat{z}\| \leq \hat{C}_{0,3} \hat{Q}^k; \]

(iii) if \( \theta \in (1/2, 1) \), then there exist \( k_0 \geq 3 \) and \( \hat{C}_{1,1}, \hat{C}_{1,2}, \hat{C}_{1,3} > 0 \) such that for every \( k \geq k_0 \)
\[ \|x^k - \hat{x}\| \leq \hat{C}_{1,1} (k - 1)^{-\frac{1-\theta}{2\bar{p}}}, \quad \|y^k - \hat{y}\| \leq \hat{C}_{1,2} (k - 1)^{-\frac{1-\theta}{2\bar{p}}}, \quad \|z^k - \hat{z}\| \leq \hat{C}_{1,3} (k - 2)^{-\frac{1-\theta}{2\bar{p}}}. \]

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 \( k_0' \geq 2 \) such that for every \( k \geq k_0' \) the inequalities (87a)-(87c) in Lemma 19 and \( \mathcal{E}_k \leq \left( \frac{1}{1-\theta} C_L \right)^{\frac{1}{2\bar{p}}} \) hold.

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

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

The conclusion follows from Lemma 19 for \( k_0 := \max\{k_0', k_0''\} \), by noticing that
\[ \sqrt{\mathcal{E}_{k-1}} \leq \sqrt{\hat{C}_0 \hat{Q}^{k-1}} = \frac{\sqrt{\hat{C}_0 \hat{Q}^{k-1}}}{\sqrt{\hat{Q}}} \hat{Q}^k \quad \text{and} \quad \sqrt{\mathcal{E}_{k-2}} \leq \sqrt{\hat{C}_0 \hat{Q}^{k-2}} = \frac{\sqrt{\hat{C}_0 \hat{Q}^{k-2}}}{\sqrt{\hat{Q}^2}} \hat{Q}^k \forall k \geq k_0. \]

(iii) If \( \theta \in (1/2, 1) \), then \( \mathcal{E}_k^{\frac{1}{2}} \leq \frac{1}{1-\theta} C_L \mathcal{E}_k^{1-\theta} \), for every \( k \geq k_0' \), which implies that max \( \{\sqrt{\mathcal{E}_k}, \varphi(\mathcal{E}_k)\} \) = \( \varphi(\mathcal{E}_k) = \frac{1}{1-\theta} C_L \mathcal{E}_k^{1-\theta} \). By Theorem 18 there exist \( k_0' \geq 3 \) and \( \hat{C}_1 > 0 \) such that for all \( k \geq k_0' \)
\[ \frac{1}{1-\theta} C_L \mathcal{E}_k^{1-\theta} \leq \frac{1}{1-\theta} C_L \hat{C}_1^{1-\theta} (k - 2)^{-\frac{1}{2\bar{p}}}. \]

The conclusion follows again for \( k_0 := \max\{k_0', k_0''\} \) from Lemma 19. \( \square \)
Remark 6. For $\rho = 1$ the same convergence rates can be obtained under the original Assumption. Indeed, when $\rho = 1$ we have that $T_1 = 0$ and, as a consequence, the sequence $\{\mathcal{F}_k\}_{k \geq 1}$ defined in (41) becomes

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

In addition, the inequality (41) simplifies to

$$\|y^{k+1} - y^k\| \leq C_3 \|x^{k+1} - x^k\| + C_4 \|x^k - x^{k-1}\| \quad \forall k \geq 1,$$

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

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

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

$$\mathcal{E}_{k-2} - \mathcal{E}_{k+1} = \mathcal{F}_{k-2} - \mathcal{F}_{k-1} - \mathcal{F}_{k-1} + \mathcal{F}_k - \mathcal{F}_{k+1} \geq C_1 \left( \|x^{k+1} - x^{k-2}\|^2 + \|x^k - x^{k-1}\|^2 + \|x^{k+1} - x^k\|^2 \right),$$

$$\geq \frac{C_1}{12} \left( \|x^{k+1} - x^{k-2}\|^2 + \|x^k - x^{k-1}\|^2 + \|x^{k-1} - x^k\|^2 \right),$$

$$\geq \frac{C_1}{12C_{11}^2} \|D^{k+1}\|^2.$$ 

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

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

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

Acknowledgements. The authors are thankful to Ernő Robert Csetnek (University of Vienna) and to two anonymous reviewers valuable comments which improved the quality of the paper.

References


References


