

A proximal minimization algorithm for structured nonconvex and nonsmooth problems

Radu Ioan Bot^{*}

Ernö Robert Csetnek[†]

Dang-Khoa Nguyen[‡]

May 28, 2018

Abstract. We propose a proximal algorithm for minimizing objective functions consisting of three summands: the composition of a nonsmooth function with a linear operator, another nonsmooth function, each of the nonsmooth summands depending on an independent block variable, and a smooth function which couples the two block variables. The algorithm is a full splitting method, which means that the nonsmooth functions are processed via their proximal operators, the smooth function via gradient steps, and the linear operator via matrix times vector multiplication. We provide sufficient conditions for the boundedness of the generated sequence and prove that any cluster point of the latter is a KKT point of the minimization problem. In the setting of the Kurdyka-Łojasiewicz property we show global convergence, and derive convergence rates for the iterates in terms of the Łojasiewicz exponent.

Key Words. structured nonconvex and nonsmooth optimization, proximal algorithm, full splitting scheme, Kurdyka-Łojasiewicz property, limiting subdifferential

AMS subject classification. 65K10, 90C26, 90C30

1 Introduction

1.1 Problem formulation and motivation

In this paper we propose a full splitting algorithm for solving nonconvex and nonsmooth problems of the form

$$\min_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} \{F(Ax) + G(y) + H(x, y)\}, \quad (1.1)$$

where $F: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$ and $G: \mathbb{R}^q \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper and lower semicontinuous functions, $H: \mathbb{R}^m \times \mathbb{R}^q \rightarrow \mathbb{R}$ is a Fréchet differentiable function with Lipschitz continuous gradient, and $A: \mathbb{R}^m \rightarrow \mathbb{R}^p$ is a linear operator. It is noticeable that neither for the nonsmooth nor for the smooth functions convexity is assumed.

In case $m = p$ and A is the identity operator, Bolte, Sabach and Teboulle formulated in [9], also in the nonconvex setting, a proximal alternating linearization method (PALM) for solving (1.1). PALM is a proximally regularized variant of the Gauss-Seidel alternating minimization scheme and basically consists of two proximal-gradient steps. It had a significant impact in the optimization community, as it can be used to solve a large variety of nonconvex and nonsmooth problems arising in applications such as: matrix factorization, image deblurring and denoising, the feasibility problem, compressed sensing, etc. An inertial version of PALM has been proposed by Pock and Sabach in [23].

A naive approach of PALM for solving (1.1) would require the calculation of the proximal operator of the function $F \circ A$, for which, in general, even in the convex case, a closed formula is not available. In the last decade, an impressive progress can be noticed in the field of primal-dual/proximal ADMM algorithms, designed to solve convex optimization problems involving compositions with linear operators in the spirit of the full splitting paradigm. One of the pillars of this development is the conjugate duality theory which is available for convex optimization problems.

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

[†]Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria, e-mail: ernoe.robert.csetnek@univie.ac.at. Research supported by FWF (Austrian Science Fund), project P 29809-N32.

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

The algorithm which we propose in this paper for solving the nonconvex and nonsmooth problem (1.1) is a full splitting scheme, too; the nonsmooth functions are processed via their proximal operators, the smooth function via gradient steps, and the linear operator via matrix times vector multiplication. In case $G(y) = 0$ and $H(x, y) = H(x)$ for any $(x, y) \in \mathbb{R}^m \times \mathbb{R}^q$, where $H : \mathbb{R}^m \rightarrow \mathbb{R}$ is a Fréchet differentiable function with Lipschitz continuous gradient, it furnishes a full splitting iterative scheme for solving the nonsmooth and nonconvex optimization problem

$$\min_{x \in \mathbb{R}^m} \{F(Ax) + H(x)\}. \quad (1.2)$$

Splitting algorithms for solving problems of the form (1.2) have been considered in [19], under the assumption that H is twice continuously differentiable with bounded Hessian, in [25], under the assumption that one of the summands is convex and continuous on its effective domain, and in [13], as a particular case of a general nonconvex proximal ADMM algorithm. We would like to mention in this context also [10] for the case when A is nonlinear.

The convergence analysis we will carry out in this paper relies on a descent inequality, which we prove for a regularization of the augmented Lagrangian $L_\beta : \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\}$

$$L_\beta(x, y, z, u) = F(z) + G(y) + H(x, y) + \langle u, Ax - z \rangle + \frac{\beta}{2} \|Ax - z\|^2, \beta > 0,$$

associated with problem (1.1). This is obtained by an appropriate tuning of the parameters involved in the description of the algorithm. In addition, we provide sufficient conditions in terms of the input functions F, G and H for the boundedness of the generated sequence of iterates. We also show that any cluster point of this sequence is a KKT point of the optimization problem (1.1). By assuming that the above-mentioned regularization of the augmented Lagrangian satisfies the Kurdyka-Łojasiewicz property, we prove global convergence. If this function satisfies the Łojasiewicz property, then we can even derive convergence rates for the sequence of iterates formulated in terms of the Łojasiewicz exponent. For similar approaches relying on the use of the Kurdyka-Łojasiewicz property in the proof of the global convergence of nonconvex optimization algorithms we refer to the papers of Attouch and Bolte [1], Attouch, Bolte and Svaiter [3], and Bolte, Sabach and Teboulle [9].

1.2 Notations and preliminaries

Every space \mathbb{R}^d , where d is a positive integer, is assumed to be equipped with the Euclidean inner product $\langle \cdot, \cdot \rangle$ and associated norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$. The Cartesian product $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \dots \times \mathbb{R}^{d_k}$ of the Euclidean spaces $\mathbb{R}^{d_i}, i = 1, \dots, k$, will be endowed with inner product and associated norm defined for $x := (x_1, \dots, x_k), y := (y_1, \dots, y_k) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \dots \times \mathbb{R}^{d_k}$ by

$$\langle x, y \rangle = \sum_{i=1}^k \langle x_i, y_i \rangle \quad \text{and} \quad \|x\| = \sqrt{\sum_{i=1}^k \|x_i\|^2},$$

respectively. For every $x := (x_1, \dots, x_k) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \dots \times \mathbb{R}^{d_k}$ we have

$$\frac{1}{\sqrt{k}} \sum_{i=1}^k \|x_i\| \leq \|x\| = \sqrt{\sum_{i=1}^k \|x_i\|^2} \leq \sum_{i=1}^k \|x_i\|. \quad (1.3)$$

Let $\psi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function and x an element of its effective domain $\text{dom}\psi := \{y \in \mathbb{R}^d : \psi(y) < +\infty\}$. The Fréchet (viscosity) subdifferential of ψ at x is

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

and the limiting (Mordukhovich) subdifferential of ψ at x is

$$\begin{aligned} \partial\psi(x) := \{d \in \mathbb{R}^d : \text{exist sequences } x_n \rightarrow x \text{ and } d_n \rightarrow d \text{ as } n \rightarrow +\infty \\ \text{such that } \psi(x_n) \rightarrow \psi(x) \text{ as } n \rightarrow +\infty \text{ and } d_n \in \hat{\partial}\psi(x_n) \text{ for any } n \geq 0\}. \end{aligned}$$

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

The inclusion $\widehat{\partial}\psi(x) \subseteq \partial\psi(x)$ holds for each $x \in \mathbb{R}^d$. If ψ is convex, then the two subdifferentials coincide with the convex subdifferential of ψ , thus

$$\widehat{\partial}\psi(x) = \partial\psi(x) = \{d \in \mathbb{R}^d : \psi(y) \geq \psi(x) + \langle d, y - x \rangle \quad \forall y \in \mathbb{R}^d\} \text{ for any } x \in \mathbb{R}^d.$$

If $x \in \mathbb{R}^d$ is a local minimum of ψ , then $0 \in \partial\psi(x)$. We denote by $\text{crit}(\psi) := \{x \in \mathbb{R}^d : 0 \in \partial\psi(x)\}$ the set of critical points of ψ . The limiting subdifferential fulfils the following closedness criterion: if $\{x_n\}_{n \geq 0}$ and $\{d_n\}_{n \geq 0}$ are sequence in \mathbb{R}^d such that $d_n \in \partial\psi(x_n)$ for any $n \geq 0$ and $(x_n, d_n) \rightarrow (x, d)$ and $\psi(x_n) \rightarrow \psi(x)$ as $n \rightarrow +\infty$, then $d \in \partial\psi(x)$. We also have the following subdifferential sum formula (see [21, Proposition 1.107], [24, Exercise 8.8]): if $\Phi: \mathbb{R}^d \rightarrow \mathbb{R}$ is a continuously differentiable function, then $\partial(\psi + \phi)(x) = \partial\psi(x) + \nabla\phi(x)$ for any $x \in \mathbb{R}^d$; and a formula for the subdifferential of the composition of ψ with a linear operator $A: \mathbb{R}^k \rightarrow \mathbb{R}^d$ (see [21, Proposition 1.112], [24, Exercise 10.7]): if A is injective, then $\partial(\psi \circ A)(x) = A^T \partial\psi(Ax)$ for any $x \in \mathbb{R}^k$.

The following proposition collects some important properties of a (not necessarily convex) Fréchet differentiable function with Lipschitz continuous gradient. For the proof of this result we refer to [13, Proposition 1].

Proposition 1. *Let $\psi: \mathbb{R}^d \rightarrow \mathbb{R}$ be Fréchet differentiable such that its gradient is Lipschitz continuous with constant $\ell > 0$. Then the following statements are true:*

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

$$\psi(y) \leq \psi(x) + \langle \nabla\psi(z), y - x \rangle + \frac{\ell}{2} \|y - x\|^2; \quad (1.4)$$

(ii) *For any $\gamma \in \mathbb{R} \setminus \{0\}$ it holds*

$$\inf_{x \in \mathbb{R}^d} \left\{ \psi(x) - \left(\frac{1}{\gamma} - \frac{\ell}{2\gamma^2} \right) \|\nabla\psi(x)\|^2 \right\} \geq \inf_{x \in \mathbb{R}^d} \psi(x). \quad (1.5)$$

The Descent Lemma, which says that for a Fréchet differentiable function $\psi: \mathbb{R}^d \rightarrow \mathbb{R}$ having a Lipschitz continuous gradient with constant $\ell > 0$ it holds

$$\psi(y) \leq \psi(x) + \langle \nabla\psi(x), y - x \rangle + \frac{\ell}{2} \|y - x\|^2 \quad \forall x, y \in \mathbb{R}^d,$$

follows from (1.4) for $z := x$.

In addition, by taking in (1.4) $z := y$ we obtain

$$\psi(x) \geq \psi(y) + \langle \nabla\psi(y), x - y \rangle - \frac{\ell}{2} \|x - y\|^2 \quad \forall x, y \in \mathbb{R}^d.$$

This is equivalent to the fact that $\psi + \frac{\ell}{2} \|\cdot\|^2$ is a convex function, which is the same with ψ is ℓ -semiconvex ([8]). In other words, a consequence of Proposition (1) is, that a Fréchet differentiable function with ℓ -Lipschitz continuous gradient is ℓ -semiconvex.

We close this introductory section by presenting two convergence results for real sequences that will be used in the sequel in the convergence analysis. The following lemma is useful when proving convergence of numerical algorithms relying on Fejér monotonicity techniques (see, for instance, [11, Lemma 2.2], [12, Lemma 2]).

Lemma 2. *Let $\{\xi_n\}_{n \geq 0}$ be a sequence of real numbers and $\{\omega_n\}_{n \geq 0}$ a sequence of real nonnegative numbers. Assume that $\{\xi_n\}_{n \geq 0}$ is bounded from below and that for any $n \geq 0$*

$$\xi_{n+1} + \omega_n \leq \xi_n.$$

Then the following statements hold:

(i) *the sequence $\{\omega_n\}_{n \geq 0}$ is summable, namely $\sum_{n \geq 0} \omega_n < +\infty$;*

(ii) *the sequence $\{\xi_n\}_{n \geq 0}$ is monotonically decreasing and convergent.*

The following lemma can be found in [11, Lemma 2.3] (see, also [12, Lemma 3]).

Lemma 3. *Let $\{a_n\}_{n \geq 0}$ and $\{b_n\}_{n \geq 1}$ be sequences of real nonnegative numbers such that for any $n \geq 1$*

$$a_{n+1} \leq \chi_0 a_n + \chi_1 a_{n-1} + b_n, \quad (1.6)$$

where $\chi_0 \in \mathbb{R}$ and $\chi_1 \geq 0$ fulfill $\chi_0 + \chi_1 < 1$, and $\sum_{n \geq 1} b_n < +\infty$. Then $\sum_{n \geq 0} a_n < +\infty$.

2 The algorithm

The numerical algorithm we propose for solving (1.1) has the following formulation.

Algorithm 1. Let $\mu, \beta, \tau > 0$ and $0 < \sigma \leq 1$. For a given starting point $(x_0, y_0, z_0, u_0) \in \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p$ generate the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ for any $n \geq 0$ as follows

$$y_{n+1} \in \arg \min_{y \in \mathbb{R}^q} \left\{ G(y) + \langle \nabla_y H(x_n, y_n), y \rangle + \frac{\mu}{2} \|y - y_n\|^2 \right\} \quad (2.1a)$$

$$z_{n+1} \in \arg \min_{z \in \mathbb{R}^p} \left\{ F(z) + \langle u_n, Ax_n - z \rangle + \frac{\beta}{2} \|Ax_n - z\|^2 \right\} \quad (2.1b)$$

$$x_{n+1} := x_n - \tau^{-1} (\nabla_x H(x_n, y_{n+1}) + A^T u_n + \beta A^T (Ax_n - z_{n+1})) \quad (2.1c)$$

$$u_{n+1} := u_n + \sigma \beta (Ax_{n+1} - z_{n+1}). \quad (2.1d)$$

The proximal point operator with parameter $\gamma > 0$ (see [22]) of a proper and lower semicontinuous function $\psi: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ is the set-valued operator defined as

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

Exact formulas for the proximal operator are available not only for large classes of convex functions ([4, 5, 14]), but also for various nonconvex functions ([2, 15, 18]). In view of the above definition, the iterative scheme (2.1a) - (2.1d) reads for every $n \geq 0$

$$\begin{aligned} y_{n+1} &\in \text{prox}_{\mu^{-1}G}(y_n - \mu^{-1}\nabla_y H(x_n, y_n)) \\ z_{n+1} &\in \text{prox}_{\beta^{-1}F}(Ax_n + \beta^{-1}u_n) \\ x_{n+1} &:= x_n - \tau^{-1} (\nabla_x H(x_n, y_{n+1}) + A^T u_n + \beta A^T (Ax_n - z_{n+1})) \\ u_{n+1} &:= u_n + \sigma \beta (Ax_{n+1} - z_{n+1}). \end{aligned}$$

One can notice the full splitting character of Algorithm 1 and also that the first two steps can be performed in parallel.

Remark 1. (i) In case $G(y) = 0$ and $H(x, y) = H(x)$ for any $(x, y) \in \mathbb{R}^m \times \mathbb{R}^q$, where $H: \mathbb{R}^m \rightarrow \mathbb{R}$ is a Fréchet differentiable function with Lipschitz continuous gradient, Algorithm 1 gives rise to an iterative scheme which has been proposed in [13] for solving the optimization problem (1.2). This reads for any $n \geq 0$

$$\begin{aligned} z_{n+1} &\in \text{prox}_{\beta^{-1}F}(Ax_n + \beta^{-1}u_n) \\ x_{n+1} &:= x_n - \tau^{-1} (\nabla H(x_n) + A^T u_n + \beta A^T (Ax_n - z_{n+1})) \\ u_{n+1} &:= u_n + \sigma \beta (Ax_{n+1} - z_{n+1}). \end{aligned}$$

(ii) In case $m = p$ and $A = \text{Id}$ is the identity operator on \mathbb{R}^m , Algorithm 1 gives rise to an iterative scheme for solving

$$\min_{(x, y) \in \mathbb{R}^m \times \mathbb{R}^q} \{F(x) + G(y) + H(x, y)\}, \quad (2.2)$$

which reads for any $n \geq 0$

$$\begin{aligned} y_{n+1} &\in \text{prox}_{\mu^{-1}G}(y_n - \mu^{-1}\nabla_y H(x_n, y_n)) \\ z_{n+1} &\in \text{prox}_{\beta^{-1}F}(x_n + \beta^{-1}u_n) \\ x_{n+1} &:= x_n - \tau^{-1} (\nabla_x H(x_n, y_{n+1}) + u_n + \beta(x_n - z_{n+1})) \\ u_{n+1} &:= u_n + \sigma \beta (x_{n+1} - z_{n+1}). \end{aligned}$$

This algorithm provides an alternative to PALM ([9]) for solving optimization problems of the form (2.2).

(iii) In case $m = p$, $A = \text{Id}$, $F(x) = 0$ and $H(x, y) = H(y)$ for any $(x, y) \in \mathbb{R}^m \times \mathbb{R}^q$, where $H: \mathbb{R}^q \rightarrow \mathbb{R}$ is a Fréchet differentiable function with Lipschitz continuous gradient, Algorithm 1 gives rise to an iterative scheme for solving

$$\min_{y \in \mathbb{R}^q} \{G(y) + H(y)\}, \quad (2.3)$$

which reads for any $n \geq 0$

$$y_{n+1} \in \text{prox}_{\mu^{-1}G}(y_n - \mu^{-1}\nabla H(y_n)),$$

and is nothing else than the proximal-gradient method. An inertial version of the proximal-gradient method for solving (2.3) in the fully nonconvex setting has been considered in [12].

2.1 A descent inequality

We will start with the convergence analysis of Algorithm (1) by proving a descent inequality, which will play a fundamental role in our investigations. We will analyse Algorithm (1) under the following assumptions, which we will be later even weakened.

Assumption 1. (i) *the functions F, G and H are bounded from below;*

(ii) *the linear operator A is surjective;*

(iii) *for any fixed $y \in \mathbb{R}^q$ there exists $\ell_1(y) \geq 0$ such that*

$$\|\nabla_x H(x, y) - \nabla_x H(x', y)\| \leq \ell_1(y) \|x - x'\| \quad \forall x, x' \in \mathbb{R}^m, \quad (2.4a)$$

and for any fixed $x \in \mathbb{R}^m$ there exist $\ell_2(x), \ell_3(x) \geq 0$ such that

$$\|\nabla_y H(x, y) - \nabla_y H(x, y')\| \leq \ell_2(x) \|y - y'\| \quad \forall y, y' \in \mathbb{R}^q, \quad (2.4b)$$

$$\|\nabla_x H(x, y) - \nabla_x H(x, y')\| \leq \ell_3(x) \|y - y'\| \quad \forall y, y' \in \mathbb{R}^q; \quad (2.4c)$$

(iv) *there exist $\ell_{i,+} > 0, i = 1, 2, 3$, such that*

$$\sup_{n \geq 0} \ell_1(y_n) \leq \ell_{1,+}, \quad \sup_{n \geq 0} \ell_2(x_n) \leq \ell_{2,+}, \quad \sup_{n \geq 0} \ell_3(x_n) \leq \ell_{3,+}. \quad (2.5)$$

Remark 2. Some comments on Assumption 1 are in order.

(i) Assumption (i) ensures that the sequence generated by Algorithm 1 is well-defined. It has also as consequence that

$$\underline{\Psi} := \inf_{(x,y,z) \in \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p} \{F(z) + G(y) + H(x, y)\} > -\infty. \quad (2.6)$$

(ii) Comparing the assumptions in (iii) and (iv) to the ones in [9], one can notice the presence of the additional condition (2.4c), which is essential in particular when proving the boundedness of the sequence of generated iterates. Notice that in iterative schemes of gradient type, proximal-gradient type or forward-backward-forward type (see [9, 11, 12]) the boundedness of the iterates follow by combining a descent inequality expressed in terms of the objective function with coercivity assumptions on the later. In our setting this undertaken is less simple, since the descent inequality which we obtain below is in terms of the augmented Lagrangian associated with problem (1.1).

(iii) The linear operator A is surjective if and only if its associated matrix has full row rank, which is the same with the fact that the matrix associated to AA^T is positively definite. Since

$$\lambda_{\min}(AA^T) \|z\|^2 \leq \langle AA^T z, z \rangle = \|A^T z\|^2 \quad \forall z \in \mathbb{R}^p,$$

this is further equivalent to $\lambda_{\min}(AA^T) > 0$, where $\lambda_{\min}(M)$ denotes the minimal eigenvalue of a square matrix M . We also denote by $\kappa(M)$ the condition number, namely the ratio between the maximal eigenvalue $\lambda_{\max}(M)$ and the minimal eigenvalue of the square matrix M ,

$$\kappa(M) := \frac{\lambda_{\max}(M)}{\lambda_{\min}(M)} = \frac{\|M\|^2}{\lambda_{\min}(M)} \geq 1.$$

The convergence analysis will make use of the following regularized augmented Lagrangian function

$$\Psi: \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R} \cup \{+\infty\},$$

defined as

$$(x, y, z, u, x', u') \mapsto F(z) + G(y) + H(x, y) + \langle u, Ax - z \rangle + \frac{\beta}{2} \|Ax - z\|^2 \\ + C_0 \|A^T(u - u') + \sigma B(x - x')\|^2 + C_1 \|x - x'\|^2,$$

where

$$B := \tau \text{Id} - \beta A^T A, \quad C_0 := \frac{4(1 - \sigma)}{\sigma^2 \beta \lambda_{\min}(AA^T)} \geq 0 \quad \text{and} \quad C_1 := \frac{8(\sigma\tau + \ell_{1,+})^2}{\sigma \beta \lambda_{\min}(AA^T)} > 0.$$

Notice that

$$\|B\| \leq \tau,$$

whenever $2\tau \geq \beta \|A\|^2$. Indeed, this is a consequence of the relation

$$\|Bx\|^2 = \tau^2 \|x\|^2 - 2\tau\beta \|Ax\|^2 + \beta^2 \|A^T Ax\|^2 \leq \tau^2 \|x\|^2 + \beta(\beta \|A\|^2 - 2\tau) \|Ax\|^2 \quad \forall x \in \mathbb{R}^m.$$

For simplification, we introduce the following notations

$$\mathbf{R} := \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R}^p \\ \mathbf{X} := (x, y, z, u, x', u') \\ \mathbf{X}_n := (x_n, y_n, z_n, u_n, x_{n-1}, u_{n-1}) \quad \forall n \geq 1 \\ \Psi_n := \Psi(\mathbf{X}_n) \quad \forall n \geq 1.$$

The next result provides the announced descent inequality.

Lemma 4. *Let Assumption 1 be satisfied, $2\tau \geq \beta \|A\|^2$ and $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ be a sequence generated by Algorithm 1. Then for any $n \geq 1$ it holds*

$$\Psi_{n+1} + C_2 \|x_{n+1} - x_n\|^2 + C_3 \|y_{n+1} - y_n\|^2 + C_4 \|u_{n+1} - u_n\|^2 \leq \Psi_n, \quad (2.7)$$

where

$$C_2 := \tau - \frac{\ell_{1,+} + \beta \|A\|^2}{2} - \frac{4\sigma\tau^2}{\beta \lambda_{\min}(AA^T)} - \frac{8(\sigma\tau + \ell_{1,+})^2}{\sigma \beta \lambda_{\min}(AA^T)}, \quad (2.8a)$$

$$C_3 := \frac{\mu - \ell_{2,+}}{2} - \frac{8\ell_{3,+}^2}{\sigma \beta \lambda_{\min}(AA^T)}, \quad (2.8b)$$

$$C_4 := \frac{1}{\sigma\beta}. \quad (2.8c)$$

Proof. Let $n \geq 1$ be fixed. We will show first that

$$F(z_{n+1}) + G(y_{n+1}) + H(x_{n+1}, y_{n+1}) + \langle u_{n+1}, Ax_{n+1} - z_{n+1} \rangle + \frac{\beta}{2} \|Ax_{n+1} - z_{n+1}\|^2 \\ + \left(\tau - \frac{\ell_{1,+} + \beta \|A\|^2}{2} \right) \|x_{n+1} - x_n\|^2 + \frac{\mu - \ell_{2,+}}{2} \|y_{n+1} - y_n\|^2 + \frac{1}{\sigma\beta} \|u_{n+1} - u_n\|^2 \\ \leq F(z_n) + G(y_n) + H(x_n, y_n) + \langle u_n, Ax_n - z_n \rangle + \frac{\beta}{2} \|Ax_n - z_n\|^2 + \frac{2}{\sigma\beta} \|u_{n+1} - u_n\|^2 \quad (2.9)$$

and provide afterwards an upper estimate for the term $\|u_{n+1} - u_n\|^2$ on the right-hand side of (2.9).

From (2.1a) and (2.1b) we obtain

$$G(y_{n+1}) + \langle \nabla_y H(x_n, y_n), y_{n+1} - y_n \rangle + \frac{\mu}{2} \|y_{n+1} - y_n\|^2 \leq G(y_n) \quad (2.10)$$

and

$$F(z_{n+1}) + \langle u_n, Ax_n - z_{n+1} \rangle + \frac{\beta}{2} \|Ax_n - z_{n+1}\|^2 \leq F(z_n) + \langle u_n, Ax_n - z_n \rangle + \frac{\beta}{2} \|Ax_n - z_n\|^2 \quad (2.11)$$

respectively. On the other hand, according to the Descent Lemma we have

$$\begin{aligned} H(x_n, y_{n+1}) &\leq H(x_n, y_n) + \langle \nabla_y H(x_n, y_n), y_{n+1} - y_n \rangle + \frac{\ell_2(x_n)}{2} \|y_{n+1} - y_n\|^2 \\ &\leq H(x_n, y_n) + \langle \nabla_y H(x_n, y_n), y_{n+1} - y_n \rangle + \frac{\ell_{2,+}}{2} \|y_{n+1} - y_n\|^2 \end{aligned}$$

and, further, by taking into consideration (2.1c),

$$\begin{aligned} H(x_{n+1}, y_{n+1}) &\leq H(x_n, y_{n+1}) + \langle \nabla_x H(x_n, y_{n+1}), x_{n+1} - x_n \rangle + \frac{\ell_1(y_{n+1})}{2} \|x_{n+1} - x_n\|^2 \\ &= H(x_n, y_{n+1}) - \langle u_n, Ax_{n+1} - Ax_n \rangle - \beta \langle Ax_n - z_{n+1}, Ax_{n+1} - Ax_n \rangle \\ &\quad - \left(\tau - \frac{\ell_1(y_{n+1})}{2} \right) \|x_{n+1} - x_n\|^2 \\ &\leq H(x_n, y_{n+1}) - \langle u_n, Ax_{n+1} - Ax_n \rangle + \frac{\beta}{2} \|Ax_n - z_{n+1}\|^2 - \frac{\beta}{2} \|Ax_{n+1} - z_{n+1}\|^2 \\ &\quad - \left(\tau - \frac{\ell_{1,+} + \beta \|A\|^2}{2} \right) \|x_{n+1} - x_n\|^2. \end{aligned}$$

Summing this inequality with (2.10) and (2.11) gives (2.9).

Next we will focus on estimating $\|u_{n+1} - u_n\|^2$. Combining (2.1c) and (2.1d), we obtain

$$A^T u_{n+1} + \sigma B(x_{n+1} - x_n) = (1 - \sigma) A^T u_n - \sigma \nabla_x H(x_n, y_{n+1})$$

and

$$A^T u_n + \sigma B(x_n - x_{n-1}) = (1 - \sigma) A^T u_{n-1} - \sigma \nabla_x H(x_{n-1}, y_n).$$

Subtracting these relations and making use of the notations

$$\begin{aligned} w_n &:= A^T(u_n - u_{n-1}) + \sigma B(x_n - x_{n-1}) \\ v_n &:= \sigma B(x_n - x_{n-1}) + \nabla_x H(x_{n-1}, y_n) - \nabla_x H(x_n, y_{n+1}), \end{aligned}$$

it yields

$$w_{n+1} = (1 - \sigma) w_n + \sigma v_n.$$

The convexity of $\|\cdot\|^2$ guarantees that (notice that $0 < \sigma \leq 1$)

$$\|w_{n+1}\|^2 \leq (1 - \sigma) \|w_n\|^2 + \sigma \|v_n\|^2. \quad (2.12)$$

In addition, from the definitions of w_n and v_n , we obtain

$$\|A^T(u_{n+1} - u_n)\| \leq \|w_{n+1}\| + \sigma \|B\| \|x_{n+1} - x_n\| \leq \|w_{n+1}\| + \sigma \tau \|x_{n+1} - x_n\| \quad (2.13)$$

and

$$\begin{aligned} \|v_n\| &\leq \sigma \|B\| \|x_n - x_{n-1}\| + \|\nabla_x H(x_{n-1}, y_n) - \nabla_x H(x_n, y_{n+1})\| \\ &\leq \sigma \tau \|x_n - x_{n-1}\| + \|\nabla_x H(x_{n-1}, y_n) - \nabla_x H(x_n, y_n)\| + \|\nabla_x H(x_n, y_n) - \nabla_x H(x_n, y_{n+1})\| \\ &\leq (\sigma \tau + \ell_{1,+}) \|x_n - x_{n-1}\| + \ell_{3,+} \|y_{n+1} - y_n\| \end{aligned} \quad (2.14)$$

respectively. Using the Cauchy-Schwarz inequality, (2.13) yields

$$\frac{\lambda_{\min}(AA^T)}{2} \|u_{n+1} - u_n\|^2 \leq \frac{1}{2} \|A^T(u_{n+1} - u_n)\|^2 \leq \|w_{n+1}\|^2 + \sigma^2 \tau^2 \|x_{n+1} - x_n\|^2$$

and (2.14) yields

$$\|v_n\|^2 \leq 2(\sigma \tau + \ell_{1,+})^2 \|x_n - x_{n-1}\|^2 + 2\ell_{3,+}^2 \|y_{n+1} - y_n\|^2.$$

After combining these two inequalities with (2.12), we get

$$\begin{aligned} &\frac{\sigma \lambda_{\min}(AA^T)}{2} \|u_{n+1} - u_n\|^2 + (1 - \sigma) \|w_{n+1}\|^2 \\ &\leq (1 - \sigma) \|w_n\|^2 + \sigma^3 \tau^2 \|x_{n+1} - x_n\|^2 + 2\sigma(\sigma \tau + \ell_{1,+})^2 \|x_n - x_{n-1}\|^2 + 2\sigma \ell_{3,+}^2 \|y_{n+1} - y_n\|^2. \end{aligned}$$

The desired statement follows after we multiply the above relation by $\frac{4}{\sigma^2 \beta \lambda_{\min}(AA^T)} > 0$ and combine the resulting inequality with (2.9). \square

The following result provides one possibility to choose the parameters in Algorithm 1, such that all three constants C_2, C_3 and C_4 that appear in (2.7) are positive.

Lemma 5. *Let*

$$0 < \sigma < \frac{1}{24\kappa(AA^T)} \quad (2.15a)$$

$$\beta > \frac{4\nu}{1 - 24\sigma\kappa(AA^T)} \left(4 + 3\sigma + \sqrt{24 + 24\sigma + 9\sigma^2 - 192\sigma\kappa(AA^T)} \right) > 0 \quad (2.15b)$$

$$\max \left\{ \frac{\beta \|A\|^2}{2}, \frac{\beta \lambda_{\min}(AA^T)}{24\sigma} \left(1 - \frac{16\nu}{\beta} - \sqrt{\Delta'_\tau} \right) \right\} < \tau < \frac{\beta \lambda_{\min}(AA^T)}{24\sigma} \left(1 - \frac{16\nu}{\beta} + \sqrt{\Delta'_\tau} \right) \quad (2.15c)$$

$$\mu > \ell_{2,+} + \frac{16\ell_{3,+}^2}{\sigma\beta\lambda_{\min}(AA^T)} > 0, \quad (2.15d)$$

where

$$\nu := \frac{\ell_{1,+}}{\lambda_{\min}(AA^T)} > 0 \text{ and } \Delta'_\tau := 1 - \frac{32\nu}{\beta} - \frac{128\nu^2}{\beta^2} - \frac{24\nu\sigma}{\beta} - 24\sigma\kappa(AA^T) > 0. \quad (2.15e)$$

Then we have

$$\min \{C_2, C_3, C_4\} > 0.$$

Furthermore, there exist $\gamma_1, \gamma_2 \in \mathbb{R} \setminus \{0\}$ such that

$$\frac{1}{\gamma_1} - \frac{\ell_{1,+}}{2\gamma_1^2} = \frac{1}{\beta\lambda_{\min}(AA^T)} \quad \text{and} \quad \frac{1}{\gamma_2} - \frac{\ell_{1,+}}{2\gamma_2^2} = \frac{2}{\beta\lambda_{\min}(AA^T)}. \quad (2.16)$$

Proof. We will prove first that

$$C_2 > 0 \Leftrightarrow \frac{24\sigma\tau^2}{\beta\lambda_{\min}(AA^T)} - 2 \left(1 - \frac{16\ell_{1,+}}{\beta\lambda_{\min}(AA^T)} \right) \tau + \frac{16\ell_{1,+}^2}{\sigma\beta\lambda_{\min}(AA^T)} + \ell_{1,+} + \beta \|A\|^2 < 0. \quad (2.17)$$

The reduced discriminant of the quadratic function in τ in the above relation reads

$$\begin{aligned} \Delta'_\tau &:= \left(1 - \frac{16\ell_{1,+}}{\beta\lambda_{\min}(AA^T)} \right)^2 - \frac{384\ell_{1,+}^2}{\beta^2\lambda_{\min}^2(AA^T)} - \frac{24\ell_{1,+}\sigma}{\beta\lambda_{\min}(AA^T)} - 24\sigma\kappa(AA^T) \\ &= \left(1 - \frac{16\nu}{\beta} \right)^2 - \frac{384\nu^2}{\beta^2} - \frac{24\nu\sigma}{\beta} - 24\sigma\kappa(AA^T) \\ &= 1 - \frac{32\nu}{\beta} - \frac{128\nu^2}{\beta^2} - \frac{24\nu\sigma}{\beta} - 24\sigma\kappa(AA^T) > 0, \end{aligned} \quad (2.18)$$

if σ and β are being chosen as in (2.15a) and (2.15b), respectively. Therefore, for

$$\frac{\beta\lambda_{\min}(AA^T)}{24\sigma} \left(1 - \frac{16\nu}{\beta} - \sqrt{\Delta'_\tau} \right) < \tau < \frac{\beta\lambda_{\min}(AA^T)}{24\sigma} \left(1 - \frac{16\nu}{\beta} + \sqrt{\Delta'_\tau} \right),$$

(2.17) is satisfied. It remains to verify the feasibility of τ in (2.15c), in other words, to prove that

$$\frac{\beta \|A\|^2}{2} < \frac{\beta\lambda_{\min}(AA^T)}{24\sigma} \left(1 - \frac{16\nu}{\beta} + \sqrt{\Delta'_\tau} \right).$$

This is easy to see, as, according to (2.18), we have

$$\frac{\beta \|A\|^2}{2} < \frac{\beta\lambda_{\min}(AA^T)}{24\sigma} \left(1 - \frac{16\nu}{\beta} \right) \Leftrightarrow 1 - \frac{16\nu}{\beta} - 12\sigma\kappa(AA^T) > 0.$$

The positivity of C_3 follows from the choice of μ in (2.15d), while, obviously, $C_4 > 0$.

Finally, as

$$\beta > \frac{4\nu}{1 - 24\sigma\kappa(AA^T)} > 4\nu,$$

it follows that each of the two quadratic equations in (2.16) (in γ_1 and, respectively, γ_2) has a nonzero real solution. \square

Remark 3. Hong and Luo proved recently in [16] linear convergence for the iterates generated by a Lagrangian-based algorithm in the convex setting, without any strong convexity assumption. To this end a certain error bound condition must hold true and the step size of the dual update, which is also assumed to depend on the error bound constants, must be taken small. The authors also mention that this choice of the dual step size may be too conservative and cumbersome to compute unless the objective function is strongly convex. As shown in previous lemma, the step size of the dual update in our algorithm can be computed without assuming strong convexity and indeed it depends only on the linear operator A .

Theorem 6. *Let Assumption 1 be satisfied and the parameters in Algorithm 1 be such that $2\tau \geq \beta \|A\|^2$ and the constants defined in Lemma 4 fulfil $\min\{C_2, C_3, C_4\} > 0$. If $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ is a sequence generated by Algorithm 1, then the following statements are true:*

(i) *the sequence $\{\Psi_n\}_{n \geq 1}$ is bounded from below and convergent;*

(ii)

$$x_{n+1} - x_n \rightarrow 0, \quad y_{n+1} - y_n \rightarrow 0, \quad z_{n+1} - z_n \rightarrow 0 \quad \text{and} \quad u_{n+1} - u_n \rightarrow 0 \quad \text{as} \quad n \rightarrow +\infty. \quad (2.19)$$

Proof. First, we show that $\underline{\Psi}$ defined in (2.6) is a lower bound of $\{\Psi_n\}_{n \geq 2}$. Suppose the contrary, namely that there exists $n_0 \geq 2$ such that $\Psi_{n_0} - \underline{\Psi} < 0$. According to Lemma 4, $\{\Psi_n\}_{n \geq 1}$ is a nonincreasing sequence and thus for any $N \geq n_0$

$$\sum_{n=1}^N (\Psi_n - \underline{\Psi}) \leq \sum_{n=1}^{n_0-1} (\Psi_n - \underline{\Psi}) + (N - n_0 + 1) (\Psi_{n_0} - \underline{\Psi}),$$

which implies that

$$\lim_{N \rightarrow +\infty} \sum_{n=1}^N (\Psi_n - \underline{\Psi}) = -\infty.$$

On the other hand, for any $n \geq 1$ it holds

$$\begin{aligned} \Psi_n - \underline{\Psi} &\geq F(z_n) + G(y_n) + H(x_n, y_n) + \langle u_n, Ax_n - z_n \rangle - \underline{\Psi} \\ &\geq \langle u_n, Ax_n - z_n \rangle = \frac{1}{\sigma\beta} \langle u_n, u_n - u_{n-1} \rangle = \frac{1}{2\sigma\beta} \|u_n\|^2 + \frac{1}{2\sigma\beta} \|u_n - u_{n-1}\|^2 - \frac{1}{2\sigma\beta} \|u_{n-1}\|^2. \end{aligned}$$

Therefore, for any $N \geq 1$, we have

$$\sum_{n=1}^N (\Psi_n - \underline{\Psi}) \geq \frac{1}{2\sigma\beta} \sum_{n=1}^N \|u_n - u_{n-1}\|^2 + \frac{1}{2\sigma\beta} \|u_N\|^2 - \frac{1}{2\sigma\beta} \|u_0\|^2 \geq -\frac{1}{2\sigma\beta} \|u_0\|^2,$$

which leads to a contradiction. As $\{\Psi_n\}_{n \geq 1}$ is bounded from below, we obtain from Lemma 2 statement (i) and also that

$$x_{n+1} - x_n \rightarrow 0, \quad y_{n+1} - y_n \rightarrow 0 \quad \text{and} \quad u_{n+1} - u_n \rightarrow 0 \quad \text{as} \quad n \rightarrow +\infty.$$

Since for any $n \geq 1$ it holds

$$\begin{aligned} \|z_{n+1} - z_n\| &\leq \|A\| \|x_{n+1} - x_n\| + \|Ax_{n+1} - z_{n+1}\| + \|Ax_n - z_n\| \\ &= \|A\| \|x_{n+1} - x_n\| + \frac{1}{\sigma\beta} \|u_{n+1} - u_n\| + \frac{1}{\sigma\beta} \|u_n - u_{n-1}\|, \end{aligned} \quad (2.20)$$

it follows that $z_{n+1} - z_n \rightarrow 0$ as $n \rightarrow +\infty$. □

Usually, for nonconvex algorithms, the fact that the sequences of differences of consecutive iterates converge to zero is shown by assuming that the generated sequences are bounded (see [13, 19, 25]). In our analysis the only ingredients for obtaining statement (ii) in Theorem 6 are the descent property and Lemma 2.

2.2 General conditions for the boundedness of $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$

In the following we will formulate general conditions in terms of the input data of the optimization problem (1.1) which guarantee the boundedness of the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$. Working in the setting of Theorem 6, thanks to (2.19), we have that the sequences $\{x_{n+1} - x_n\}_{n \geq 0}$, $\{y_{n+1} - y_n\}_{n \geq 0}$, $\{z_{n+1} - z_n\}_{n \geq 0}$ and $\{u_{n+1} - u_n\}_{n \geq 0}$ are bounded. Denote

$$s_* := \sup_{n \geq 0} \{\|x_{n+1} - x_n\|, \|y_{n+1} - y_n\|, \|z_{n+1} - z_n\|, \|u_{n+1} - u_n\|\} < +\infty.$$

Even though this observation does not imply immediately that $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ is bounded, this will follow under standard coercivity assumptions. Recall that a function $\psi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ is called coercive, if $\lim_{\|x\| \rightarrow +\infty} \psi(x) = +\infty$.

Theorem 7. *Let Assumption 1 be satisfied and the parameters in Algorithm 1 be such that $2\tau \geq \beta \|A\|^2$, the constants defined in Lemma 4 fulfil $\min\{C_2, C_3, C_4\} > 0$ and there exist $\gamma_1, \gamma_2 \in \mathbb{R} \setminus \{0\}$ such that (2.16) holds. Suppose that one of the following conditions hold:*

- (i) *the function H is coercive;*
- (ii) *the operator A is invertible, and F and G are coercive.*

Then every sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ generated by Algorithm 1 is bounded.

Proof. Let $n \geq 1$ be fixed. According to Lemma 4 we have that

$$\begin{aligned} \Psi_1 &\geq \dots \geq \Psi_n \geq \Psi_{n+1} \\ &\geq F(z_{n+1}) + G(y_{n+1}) + H(x_{n+1}, y_{n+1}) - \frac{1}{2\beta} \|u_{n+1}\|^2 + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2. \end{aligned} \quad (2.21)$$

Combine (2.1c) and (2.1d) we get

$$\begin{aligned} A^T u_{n+1} &= \left(1 - \frac{1}{\sigma}\right) A^T (u_{n+1} - u_n) + B(x_n - x_{n+1}) \\ &\quad + \nabla_x H(x_{n+1}, y_{n+1}) - \nabla_x H(x_n, y_{n+1}) - \nabla_x H(x_{n+1}, y_n), \end{aligned} \quad (2.22)$$

which implies

$$\begin{aligned} \|A^T u_{n+1}\| &\leq \left(\frac{1}{\sigma} - 1\right) \|A\| \|u_{n+1} - u_n\| + (\tau + \ell_{1,+}) \|x_{n+1} - x_n\| + \|\nabla_x H(x_{n+1}, y_{n+1})\| \\ &\leq \left(\left(\frac{1}{\sigma} - 1\right) \|A\| + \tau + \ell_{1,+}\right) s_* + \|\nabla_x H(x_{n+1}, y_{n+1})\|. \end{aligned}$$

By using the Cauchy-Schwarz inequality we further obtain

$$\lambda_{\min}(AA^T) \|u_{n+1}\|^2 \leq \|A^T u_{n+1}\|^2 \leq 2 \left(\left(\frac{1}{\sigma} - 1\right) \|A\| + \tau + \ell_{1,+}\right)^2 s_*^2 + 2 \|\nabla_x H(x_{n+1}, y_{n+1})\|^2.$$

Multiplying the above relation by $\frac{1}{2\beta\lambda_{\min}(AA^T)}$ and combining it with (2.21), we get

$$\begin{aligned} \Psi_1 &\geq F(z_{n+1}) + G(y_{n+1}) + H(x_{n+1}, y_{n+1}) - \frac{1}{\beta\lambda_{\min}(AA^T)} \|\nabla_x H(x_{n+1}, y_{n+1})\|^2 \\ &\quad - \frac{1}{\beta\lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1\right) \|A\| + \tau + \ell_{1,+}\right)^2 s_*^2 + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2. \end{aligned} \quad (2.23)$$

We will prove the boundedness of $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ in each of the two scenarios.

(i) According to (2.23) and Proposition 1, we have that for any $n \geq 1$

$$\begin{aligned}
& \frac{1}{2}H(x_{n+1}, y_{n+1}) + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta}u_{n+1} \right\|^2 \\
& \leq \Psi_1 + \frac{1}{\beta\lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 - \inf_{z \in \mathbb{R}^p} F(z) - \inf_{y \in \mathbb{R}^m} G(y) \\
& \quad - \frac{1}{2} \inf_{n \geq 1} \left\{ H(x_{n+1}, y_{n+1}) - \left(\frac{1}{\gamma_1} - \frac{\ell_{1,+}}{2\gamma_1^2} \right) \|\nabla_x H(x_{n+1}, y_{n+1})\|^2 \right\} \\
& \leq \Psi_1 + \frac{1}{\beta\lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 - \inf_{z \in \mathbb{R}^p} F(z) - \inf_{y \in \mathbb{R}^q} G(y) - \inf_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} H(x, y) \\
& < +\infty.
\end{aligned}$$

Since H is coercive and bounded from below, it follows that $\{(x_n, y_n)\}_{n \geq 0}$ and $\left\{ Ax_n - z_n + \frac{1}{\beta}u_n \right\}_{n \geq 0}$ are bounded. As, according to (2.1d), $\{Ax_n - z_n\}_{n \geq 0}$ is bounded, it follows that $\{u_n\}_{n \geq 0}$ and $\{z_n\}_{n \geq 0}$ are also bounded.

(ii) According to (2.23) and Proposition 1, we have this time that for any $n \geq 1$

$$\begin{aligned}
& F(z_{n+1}) + G(y_{n+1}) + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta}u_{n+1} \right\|^2 \\
& \leq \Psi_1 + \frac{1}{\beta\lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 \\
& \quad - \inf_{n \geq 1} \left\{ H(x_{n+1}, y_{n+1}) - \left(\frac{1}{\gamma_2} - \frac{\ell_{1,+}}{2\gamma_2^2} \right) \|\nabla_x H(x_{n+1}, y_{n+1})\|^2 \right\} \\
& \leq \Psi_1 + \frac{1}{\beta\lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 - \inf_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} H(x, y) < +\infty.
\end{aligned}$$

Since F and G are coercive and bounded from below, it follows that the sequences $\{(y_n, z_n)\}_{n \geq 0}$ and $\left\{ Ax_n - z_n + \frac{1}{\beta}u_n \right\}_{n \geq 0}$ are bounded. As, according to (2.1d), $\{Ax_n - z_n\}_{n \geq 0}$ is bounded, it follows that $\{u_n\}_{n \geq 0}$ and $\{Ax_n\}_{n \geq 0}$ are bounded. The fact that A is invertible implies that $\{x_n\}_{n \geq 0}$ is bounded. \square

2.3 The cluster points of $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ are KKT points

We will close this section dedicated to the convergence analysis of the sequence generated by Algorithm 1 in a general framework by proving that any cluster point of $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ is a KKT point of the optimization problem (1.1). We provided above general conditions which guarantee both the descent inequality (2.7), with positive constants C_2, C_3 and C_4 , and the boundedness of the generated iterates. Lemma 5 and Theorem 7 provide one possible setting that ensures these two fundamental properties of the convergence analysis. We do not want to restrict ourselves to this particular setting and, therefore, we will work, from now on, under the following assumptions.

Assumption 2. (i) *the functions F, G and H are bounded from below;*

(ii) *the linear operator A is surjective;*

(iii) *every sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ generated by the Algorithm 1 is bounded;*

(iv) *∇H is Lipschitz continuous with constant $L > 0$ on a convex bounded subset $B_1 \times B_2 \subseteq \mathbb{R}^m \times \mathbb{R}^q$ containing $\{(x_n, y_n)\}_{n \geq 0}$. In other words, for any $(x, y), (x', y') \in B_1 \times B_2$ it holds*

$$\|(\nabla_x H(x, y) - \nabla_x H(x', y'), \nabla_y H(x, y) - \nabla_y H(x', y'))\| \leq L \|(x, y) - (x', y')\|; \quad (2.24)$$

(v) the parameters $\mu, \beta, \tau > 0$ and $0 < \sigma \leq 1$ are such that $2\tau \geq \beta \|A\|^2$ and $\min\{C_2, C_3, C_4\} > 0$, where

$$C_2 := \tau - \frac{L\sqrt{2} + \beta \|A\|^2}{2} - \frac{4\sigma\tau^2}{\beta\lambda_{\min}(AA^T)} - \frac{8(\sigma\tau + L\sqrt{2})^2}{\sigma\beta\lambda_{\min}(AA^T)}, \quad (2.25a)$$

$$C_3 := \frac{\mu - L\sqrt{2}}{2} - \frac{16L^2}{\sigma\beta\lambda_{\min}(AA^T)}, \quad (2.25b)$$

$$C_4 := \frac{1}{\sigma\beta}. \quad (2.25c)$$

Remark 4. Being facilitated by the boundedness of the generated sequence, Assumption 2 (iv) not only guarantee the fulfilment of Assumption 1 (iii) and (iv) on a convex bounded set, but it also arises in a more natural way (see also [9]). Assumption 2 (iv) holds, for instance, if H is twice continuously differentiable. In addition, as (2.24) implies for any $(x, y), (x', y') \in B_1 \times B_2$ that

$$\|\nabla_x H(x, y) - \nabla_x H(x', y')\| + \|\nabla_y H(x, y) - \nabla_y H(x', y')\| \leq L\sqrt{2} (\|x - x'\| + \|y - y'\|),$$

we can take

$$\ell_{1,+} = \ell_{2,+} = \ell_{3,+} := L\sqrt{2}. \quad (2.26)$$

As (2.4a) - (2.4c) are valid also on a convex bounded set, the descent inequality

$$\Psi_{n+1} + C_2 \|x_{n+1} - x_n\|^2 + C_3 \|y_{n+1} - y_n\|^2 + C_4 \|u_{n+1} - u_n\|^2 \leq \Psi_n \quad \forall n \geq 1 \quad (2.27)$$

remains true, where the constants on the left-hand sided are given in (2.25) and follow from (2.8) under the consideration of (2.26). A possible choice of the parameters of the algorithm such that $\min\{C_2, C_3, C_4\} > 0$ can be obtained also from Lemma 5.

The next result provide upper estimates for the limiting subgradients of the regularized function Ψ at (x_n, y_n, z_n, u_n) for every $n \geq 1$.

Lemma 8. *Let Assumption 2 be satisfied and $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ be a sequence generated by Algorithm 1. Then for any $n \geq 1$ it holds*

$$D_n := (d_x^n, d_y^n, d_z^n, d_u^n, d_{x'}^n, d_{u'}^n) \in \partial\Psi(\mathbf{X}_n), \quad (2.28)$$

where

$$d_x^n := \nabla_x H(x_n, y_n) + A^T u_n + \beta A^T (Ax_n - z_n) + 2C_1 (x_n - x_{n-1}) + 2\sigma C_0 B^T (A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})), \quad (2.29a)$$

$$d_y^n := \nabla_y H(x_n, y_n) - \nabla_y H(x_{n-1}, y_{n-1}) + \mu (y_{n-1} - y_n), \quad (2.29b)$$

$$d_z^n := u_{n-1} - u_n + \beta A (x_{n-1} - x_n), \quad (2.29c)$$

$$d_u^n := Ax_n - z_n + 2C_0 A (A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})), \quad (2.29d)$$

$$d_{x'}^n := -2\sigma C_0 B^T (A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})) - 2C_1 (x_n - x_{n-1}), \quad (2.29e)$$

$$d_{u'}^n := -2C_0 A (A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})). \quad (2.29f)$$

In addition, for any $n \geq 1$ it holds

$$\|D_n\| \leq C_5 \|x_n - x_{n-1}\| + C_6 \|y_n - y_{n-1}\| + C_7 \|u_n - u_{n-1}\|, \quad (2.30)$$

where

$$C_5 := 2\sqrt{2} \cdot L + \tau + \beta \|A\| + 4(\sigma\tau + \|A\|) \sigma\tau C_0 + 4C_1, \quad (2.31a)$$

$$C_6 := L\sqrt{2} + \mu, \quad (2.31b)$$

$$C_7 := 1 + \frac{1}{\sigma\beta} + \left(\frac{2}{\sigma} - 1\right) \|A\| + 4(\sigma\tau + \|A\|) C_0 \|A\|. \quad (2.31c)$$

Proof. Let $n \geq 1$ be fixed. Applying the calculus rules of the limiting subdifferential we get

$$\begin{aligned} \nabla_x \Psi(\mathbf{X}_n) &= \nabla_x H(x_n, y_n) + A^T u_n + \beta A^T (Ax_n - z_n) + 2C_1(x_n - x_{n-1}) \\ &\quad + 2\sigma C_0 B^T (A^T (u_n - u_{n-1}) + \sigma B(x_n - x_{n-1})), \end{aligned} \quad (2.32a)$$

$$\partial_y \Psi(\mathbf{X}_n) = \partial G(y_n) + \nabla_y H(x_n, y_n), \quad (2.32b)$$

$$\partial_z \Psi(\mathbf{X}_n) = \partial F(z_n) - u_n - \beta (Ax_n - z_n), \quad (2.32c)$$

$$\nabla_u \Psi(\mathbf{X}_n) = Ax_n - z_n + 2C_0 A (A^T (u_n - u_{n-1}) + \sigma B(x_n - x_{n-1})), \quad (2.32d)$$

$$\nabla_{x'} \Psi(\mathbf{X}_n) = -2\sigma C_0 B^T (A^T (u_n - u_{n-1}) + \sigma B(x_n - x_{n-1})) - 2C_1(x_n - x_{n-1}), \quad (2.32e)$$

$$\nabla_{u'} \Psi(\mathbf{X}_n) = -2C_0 A (A^T (u_n - u_{n-1}) + \sigma B(x_n - x_{n-1})). \quad (2.32f)$$

Then (2.29a) and (2.29d) - (2.29f) follow directly from (2.32a) and (2.32d) - (2.32f), respectively. By combining (2.32b) with the optimality criterion for (2.1a)

$$0 \in G(y_n) + \nabla_y H(x_{n-1}, y_{n-1}) + \mu(y_n - y_{n-1}),$$

we obtain (2.29b). Similarly, by combining (2.32c) with the optimality criterion for (2.1b)

$$0 \in F(z_n) - u_{n-1} - \beta (Ax_{n-1} - z_n),$$

we get (2.29c).

In the following we will derive the upper estimates for the components of the limiting subgradient. From (2.22) it follows

$$\begin{aligned} \|d_x^n\| &\leq \|\nabla_x H(x_n, y_n) + A^T u_n\| + \beta \|A\| \|Ax_n - z_n\| + 2(C_1 + \sigma^2 \tau^2 C_0) \|x_n - x_{n-1}\| \\ &\quad + 2\sigma \tau C_0 \|A\| \|u_n - u_{n-1}\| \\ &\leq (L\sqrt{2} + \tau + 2C_1 + 2\sigma^2 \tau^2 C_0) \|x_n - x_{n-1}\| + \left(\frac{2}{\sigma} - 1 + 2\sigma \tau C_0\right) \|A\| \|u_n - u_{n-1}\|. \end{aligned}$$

In addition, we have

$$\begin{aligned} \|d_y^n\| &\leq L\sqrt{2} \|x_n - x_{n-1}\| + (L\sqrt{2} + \mu) \|y_n - y_{n-1}\|, \\ \|d_z^n\| &\leq \beta \|A\| \|x_n - x_{n-1}\| + \|u_n - u_{n-1}\|, \\ \|d_u^n\| &\leq 2\sigma \tau C_0 \|A\| \|x_n - x_{n-1}\| + \left(\frac{1}{\sigma \beta} + 2C_0 \|A\|^2\right) \|u_n - u_{n-1}\|, \\ \|d_{x'}^n\| &\leq 2(\sigma^2 \tau^2 C_0 + C_1) \|x_n - x_{n-1}\| + 2\sigma \tau C_0 \|A\| \|u_n - u_{n-1}\|, \\ \|d_{u'}^n\| &\leq 2\sigma \tau C_0 \|A\| \|x_n - x_{n-1}\| + 2C_0 \|A\|^2 \|u_n - u_{n-1}\|. \end{aligned}$$

The inequality (2.30) follows by combining the above relations with (1.3). \square

We denote by $\Omega := \Omega(\{\mathbf{X}_n\}_{n \geq 1})$ the set of cluster points of the sequence $\{\mathbf{X}_n\}_{n \geq 1} \subseteq \mathbf{R}$, which is nonempty thanks to the boundedness of $\{\mathbf{X}_n\}_{n \geq 1}$. The distance function of the set Ω is defined for any $\mathbf{X} \in \mathbf{R}$ by $\text{dist}(\mathbf{X}, \Omega) := \inf\{\|\mathbf{X} - \mathbf{Y}\| : \mathbf{Y} \in \Omega\}$. The main result of this section follows.

Theorem 9. *Let Assumption 2 be satisfied and $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ be a sequence generated by Algorithm 1. The following statements are true:*

- (i) *if $\{(x_{n_k}, y_{n_k}, z_{n_k}, u_{n_k})\}_{k \geq 0}$ is a subsequence of $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ which converges to (x_*, y_*, z_*, u_*) as $k \rightarrow +\infty$, then*

$$\lim_{k \rightarrow +\infty} \Psi_{n_k} = \Psi(x_*, y_*, z_*, u_*, x_*, u_*);$$

- (ii) *it holds*

$$\begin{aligned} \Omega &\subseteq \text{crit}(\Psi) \\ &\subseteq \{\mathbf{X}_* \in \mathbf{R} : -A^T u_* = \nabla_x H(x_*, y_*), 0 \in \partial G(y_*) + \nabla_y H(x_*, y_*), u_* \in \partial F(z_*), z_* = Ax_*\}, \end{aligned} \quad (2.33)$$

where $\mathbf{X}_* := (x_*, y_*, z_*, u_*, x_*, u_*)$;

(iii) it holds $\lim_{n \rightarrow +\infty} \text{dist}(\mathbf{X}_n, \Omega) = 0$;

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

(v) the function Ψ takes on Ω the value $\Psi_* = \lim_{n \rightarrow +\infty} \Psi_n = \lim_{n \rightarrow +\infty} \{F(z_n) + G(y_n) + H(x_n, y_n)\}$.

Proof. Let $(x_*, y_*, z_*, u_*) \in \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p$ be such that the subsequence

$$\{\mathbf{X}_{n_k} := (x_{n_k}, y_{n_k}, z_{n_k}, u_{n_k}, x_{n_k-1}, u_{n_k-1})\}_{k \geq 1}$$

of $\{\mathbf{X}_n\}_{n \geq 1}$ converges to $\mathbf{X}_* := (x_*, y_*, z_*, u_*, x_*, u_*)$.

(i) From (2.1a) and (2.1b) we have for any $k \geq 1$

$$\begin{aligned} & G(y_{n_k}) + \langle \nabla_y H(x_{n_k-1}, y_{n_k-1}), y_{n_k} - y_{n_k-1} \rangle + \frac{\mu}{2} \|y_{n_k} - y_{n_k-1}\|^2 \\ & \leq G(y_*) + \langle \nabla_y H(x_{n_k-1}, y_{n_k-1}), y_* - y_{n_k-1} \rangle + \frac{\mu}{2} \|y_* - y_{n_k-1}\|^2 \end{aligned}$$

and

$$\begin{aligned} & F(z_{n_k}) + \langle u_{n_k-1}, Ax_{n_k-1} - z_{n_k} \rangle + \frac{\beta}{2} \|Ax_{n_k-1} - z_{n_k}\|^2 \\ & \leq F(z_*) + \langle u_{n_k-1}, Ax_{n_k-1} - z_* \rangle + \frac{\beta}{2} \|Ax_{n_k-1} - z_*\|^2, \end{aligned}$$

respectively. From (2.1d) and Theorem 6 follows $Ax_* = z_*$. Taking the limit superior as $k \rightarrow +\infty$ on both sides of the above inequalities, we get

$$\limsup_{k \rightarrow +\infty} F(z_{n_k}) \leq F(z_*) \quad \text{and} \quad \limsup_{k \rightarrow +\infty} G(y_{n_k}) \leq G(y_*)$$

which, combined with the lower semicontinuity of F and G , lead to

$$\lim_{k \rightarrow +\infty} F(z_{n_k}) = F(z_*) \quad \text{and} \quad \lim_{k \rightarrow +\infty} G(y_{n_k}) = G(y_*).$$

The desired statement follows thanks to the continuity of H .

(ii) For the sequence $\{D_n\}_{n \geq 0}$ defined in (2.28) - (2.29), we have that $D_{n_k} \in \partial\Psi(\mathbf{X}_{n_k})$ for any $k \geq 1$ and $D_{n_k} \rightarrow 0$ as $k \rightarrow +\infty$, while $\mathbf{X}_{n_k} \rightarrow \mathbf{X}_*$ and $\Psi_{n_k} \rightarrow \Psi(\mathbf{X}_*)$ as $k \rightarrow +\infty$. The closedness criterion of the limiting subdifferential guarantees that $0 \in \partial\Psi(\mathbf{X}_*)$ or, in other words, $\mathbf{X}_* \in \text{crit}(\Psi)$.

Choosing now an element $\mathbf{X}_* \in \text{crit}(\Psi)$, it holds

$$\begin{cases} 0 & = \nabla_x H(x_*, y_*) + A^T u_* + \beta A^T (Ax_* - z_*), \\ 0 & \in \partial G(y_*) + \nabla_y H(x_*, y_*), \\ 0 & \in \partial F(z_*) - u_* - \beta (Ax_* - z_*), \\ 0 & = Ax_* - z_*, \end{cases}$$

which is further equivalent to (2.33).

(iii)-(iv) The proof follows in the lines of the proof of Theorem 5 (ii)-(iii) in [9], also by taking into consideration [9, Remark 5], according to which the properties in (iii) and (iv) are generic for sequences satisfying $\mathbf{X}_n - \mathbf{X}_{n-1} \rightarrow 0$ as $n \rightarrow +\infty$, which is indeed the case due to (2.19).

(v) Due to (2.19) and the fact that $\{u_n\}_{n \geq 0}$ is bounded, the sequences $\{F(z_n) + G(y_n) + H(x_n, y_n)\}_{n \geq 0}$ and $\{\Psi_n\}_{n \geq 0}$ have the same limit

$$\Psi_* = \lim_{n \rightarrow +\infty} \Psi_n = \lim_{n \rightarrow +\infty} \{F(z_n) + G(y_n) + H(x_n, y_n)\}.$$

The conclusion follows by taking into consideration the first two statements of this theorem. \square

Remark 5. An element (x_*, y_*, z_*, u_*) fulfilling (2.33) is a so-called KKT point of the optimization problem (1.1). Such a KKT point obviously fulfils

$$0 \in A^T \partial F(Ax_*) + \nabla_x H(x_*, y_*), \quad 0 \in \partial G(y_*) + \nabla_y H(x_*, y_*). \quad (2.34)$$

If A is injective, then this system of inclusions is further equivalent to

$$\begin{aligned} 0 &\in \partial(F \circ A)(x_*) + \nabla_x H(x_*, y_*) = \partial_x(F \circ A + H), \\ 0 &\in \partial G(y_*) + \nabla_y H(x_*, y_*) = \partial_y(G + H), \end{aligned} \tag{2.35}$$

in other words, (x_*, y_*) is a critical point of the optimization problem (1.1). On the other hand, if the functions F, G and H are convex, then, even without asking A to be injective, (2.34) and (2.35) are equivalent, which means that (x_*, y_*) is a global minimum of the optimization problem (1.1).

3 Global convergence and rates

In this section we will prove global convergence for the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ generated by Algorithm 1 in the context of the Kurdyka-Łojasiewicz property and provide convergence rates for it in the context of the Łojasiewicz property.

3.1 Global convergence under Kurdyka-Łojasiewicz assumptions

The origins of this notion go back to the pioneering work of Kurdyka who introduced in [17] a general form of the Łojasiewicz inequality [20]. An extension to the nonsmooth setting has been proposed and studied in [6, 7, 8].

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

- (i) $\varphi(0) = 0$;
- (ii) φ is \mathcal{C}^1 on $(0, \eta)$ and continuous at 0;
- (iii) for any $s \in (0, \eta) : \varphi'(s) > 0$.

Definition 2. Let $\Psi: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper and lower semicontinuous.

- (i) The function Ψ is said to have the Kurdyka-Łojasiewicz (KL) property at a point $\hat{v} \in \text{dom} \partial \Psi := \{v \in \mathbb{R}^d : \partial \Psi(v) \neq \emptyset\}$, if there exists $\eta \in (0, +\infty]$, a neighborhood V of \hat{v} and a function $\varphi \in \Phi_\eta$ such that for any

$$v \in V \cap [\Psi(\hat{v}) < \Psi(v) < \Psi(\hat{v}) + \eta]$$

the following inequality holds

$$\varphi'(\Psi(v) - \Psi(\hat{v})) \cdot \text{dist}(\mathbf{0}, \partial \Psi(v)) \geq 1.$$

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

The functions φ belonging to the set Φ_η for $\eta \in (0, +\infty]$ are called desingularization functions. The KL property reveals the possibility to reparametrize the values of Ψ 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 to [1, 2, 3, 6, 7, 8, 9] for more properties of KL functions and illustrating examples.

The following result, the proof of which can be found in [9, Lemma 6], will play an essential role in our convergence analysis.

Lemma 10. (Uniformized KL property) *Let Ω be a compact set and $\Psi: \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. Assume that Ψ is constant on Ω and satisfies the KL property at each point of Ω . Then there exist $\varepsilon > 0, \eta > 0$ and $\varphi \in \Phi_\eta$ such that for any $\hat{v} \in \Omega$ and every element u in the intersection*

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

it holds

$$\varphi'(\Psi(v) - \Psi(\hat{v})) \cdot \text{dist}(\mathbf{0}, \partial \Psi(v)) \geq 1.$$

From now on we will use the following notations

$$C_8 := \frac{1}{\min\{C_2, C_3, C_4\}}, \quad C_9 := \max\{C_5, C_6, C_7\} \quad \text{and} \quad \mathcal{E}_n := \Psi_n - \Psi_* \quad \forall n \geq 1,$$

where $\Psi_* = \lim_{n \rightarrow +\infty} \Psi_n$.

The next result shows that if Ψ is a KL function, then the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ converges to a KKT point of the optimization problem (1.1). This hypothesis is fulfilled if, for instance, F, G and H are semi-algebraic functions.

Theorem 11. *Let Assumption 2 be satisfied and $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ be a sequence generated by Algorithm 1. If Ψ is a KL function, then the following statements are true:*

(i) *the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ has finite length, namely,*

$$\sum_{n \geq 0} \|x_{n+1} - x_n\| < +\infty, \quad \sum_{n \geq 0} \|y_{n+1} - y_n\| < +\infty, \quad \sum_{n \geq 0} \|z_{n+1} - z_n\| < +\infty, \quad \sum_{n \geq 0} \|u_{n+1} - u_n\| < +\infty; \quad (3.1)$$

(ii) *the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ converges to a KKT point of the optimization problem (1.1).*

Proof. Let be $\mathbf{X}_* \in \Omega$, thus $\Psi(\mathbf{X}_*) = \Psi_*$. Recall that $\{\mathcal{E}_n\}_{n \geq 1}$ is monotonically decreasing and converges to 0 as $n \rightarrow +\infty$. We consider two cases.

Case 1. Assume that there exists an integer $n' \geq 1$ such that $\mathcal{E}_{n'} = 0$ or, equivalently, $\Psi_{n'} = \Psi_*$. Due to the monotonicity of $\{\mathcal{E}_n\}_{n \geq 1}$, it follows that $\mathcal{E}_n = 0$ or, equivalently, $\Psi_n = \Psi_*$ for any $n \geq n'$. The inequality (2.27) yields for any $n \geq n' + 1$

$$x_{n+1} - x_n = 0, \quad y_{n+1} - y_n = 0 \quad \text{and} \quad u_{n+1} - u_n = 0.$$

The inequality (2.20) gives us further $z_{n+1} - z_n = 0$ for any $n \geq n' + 2$. This proves (3.1).

Case 2. Consider now the case when $\mathcal{E}_n > 0$ or, equivalently, $\Psi_n > \Psi_*$ for any $n \geq 1$. According to Lemma 10, there exist $\varepsilon > 0, \eta > 0$ and a desingularization function φ such that for any element \mathbf{X} in the intersection

$$\{\mathbf{Z} \in \mathbf{R}: \text{dist}(\mathbf{Z}, \Omega) < \varepsilon\} \cap \{\mathbf{Z} \in \mathbf{R}: \Psi_* < \Psi(\mathbf{Z}) < \Psi_* + \eta\} \quad (3.2)$$

it holds

$$\varphi'(\Psi(\mathbf{X}) - \Psi_*) \cdot \text{dist}(\mathbf{0}, \partial\Psi(\mathbf{X})) \geq 1.$$

Let be $n_1 \geq 1$ such that for any $n \geq n_1$

$$\Psi_* < \Psi_n < \Psi_* + \eta.$$

Since $\lim_{n \rightarrow +\infty} \text{dist}(\mathbf{X}_n, \Omega) = 0$ (see Lemma 9 (iii)), there exists $n_2 \geq 1$ such that for any $n \geq n_2$

$$\text{dist}(\mathbf{X}_n, \Omega) < \varepsilon.$$

Consequently, $\mathbf{X}_n = (x_n, y_n, z_n, u_n, x_{n-1}, u_{n-1})$ belongs to the intersection in (3.2) for any $n \geq n_0 := \max\{n_1, n_2\}$, which further implies

$$\varphi'(\Psi_n - \Psi_*) \cdot \text{dist}(\mathbf{0}, \partial\Psi(\mathbf{X}_n)) = \varphi'(\mathcal{E}_n) \cdot \text{dist}(\mathbf{0}, \partial\Psi(\mathbf{X}_n)) \geq 1. \quad (3.3)$$

Define for two arbitrary nonnegative integers i and j

$$\Delta_{i,j} := \varphi(\Psi_i - \Psi_*) - \varphi(\Psi_j - \Psi_*) = \varphi(\mathcal{E}_i) - \varphi(\mathcal{E}_j).$$

The monotonicity of the sequence $\{\Psi_n\}_{n \geq 0}$ and of the function φ implies that $\Delta_{i,j} \geq 0$ for any $1 \leq i \leq j$. In addition, for any $N \geq n_0 \geq 1$ it holds

$$\sum_{n=n_0}^N \Delta_{n,n+1} = \Delta_{n_0, N+1} = \varphi(\mathcal{E}_{n_0}) - \varphi(\mathcal{E}_{N+1}) \leq \varphi(\mathcal{E}_{n_0}),$$

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

By combining Lemma 4 with the concavity of φ we obtain for any $n \geq 1$

$$\begin{aligned}\Delta_{n,n+1} &= \varphi(\mathcal{E}_n) - \varphi(\mathcal{E}_{n+1}) \geq \varphi'(\mathcal{E}_n)(\mathcal{E}_n - \mathcal{E}_{n+1}) = \varphi'(\mathcal{E}_n)(\Psi_n - \Psi_{n+1}) \\ &\geq \min\{C_2, C_3, C_4\} \varphi'(\mathcal{E}_n) \left(\|x_{n+1} - x_n\|^2 + \|y_{n+1} - y_n\|^2 + \|u_{n+1} - u_n\|^2 \right).\end{aligned}$$

Thus, (3.3) implies for any $n \geq n_0$

$$\begin{aligned}& \|x_{n+1} - x_n\|^2 + \|y_{n+1} - y_n\|^2 + \|u_{n+1} - u_n\|^2 \\ & \leq \text{dist}(\mathbf{0}, \partial\Psi(\mathbf{X}_n)) \cdot \varphi'(\mathcal{E}_n) \left(\|x_{n+1} - x_n\|^2 + \|y_{n+1} - y_n\|^2 + \|u_{n+1} - u_n\|^2 \right) \\ & \leq C_8 \cdot \text{dist}(\mathbf{0}, \partial\Psi(\mathbf{X}_n)) \cdot \Delta_{n,n+1}.\end{aligned}$$

By the Cauchy-Schwarz inequality, the arithmetic mean-geometric mean inequality and Lemma 8, we have that for any $n \geq n_0$ and every $\alpha > 0$

$$\begin{aligned}& \|x_{n+1} - x_n\| + \|y_{n+1} - y_n\| + \|u_{n+1} - u_n\| \\ & \leq \sqrt{3} \cdot \sqrt{\|x_{n+1} - x_n\|^2 + \|y_{n+1} - y_n\|^2 + \|u_{n+1} - u_n\|^2} \\ & \leq \sqrt{3C_8} \cdot \sqrt{\text{dist}(\mathbf{0}, \partial\Psi(\mathbf{X}_n)) \cdot \Delta_{n,n+1}} \\ & \leq \alpha \cdot \text{dist}(\mathbf{0}, \partial\Psi(\mathbf{X}_n)) + \frac{3C_8}{4\alpha} \Delta_{n,n+1} \\ & \leq \alpha C_9 (\|x_n - x_{n-1}\| + \|y_n - y_{n-1}\| + \|u_n - u_{n-1}\|) + \frac{3C_8}{4\alpha} \Delta_{n,n+1}.\end{aligned}\tag{3.4}$$

If we denote for any $n \geq 0$

$$a_n := \|x_n - x_{n-1}\| + \|y_n - y_{n-1}\| + \|u_n - u_{n-1}\| \quad \text{and} \quad b_n := \frac{3C_8}{4\alpha} \Delta_{n,n+1},\tag{3.5}$$

then the above inequality is nothing else than (1.6) with

$$\chi_0 := \alpha C_9 \quad \text{and} \quad \chi_1 := 0.$$

Since $\sum_{n \geq 1} b_n < +\infty$, by choosing $\alpha < 1/C_9$, we can apply Lemma 3 to conclude that

$$\sum_{n \geq 0} \left(\|x_{n+1} - x_n\| + \|y_{n+1} - y_n\| + \|u_{n+1} - u_n\| \right) < +\infty.$$

The proof of (3.1) is completed by taking into account once again (2.20).

From (i) it follows that the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ is Cauchy, thus it converges to an element (x_*, y_*, z_*, u_*) which is, according to Lemmas 9, a KKT point of the optimization problem (1.1). \square

3.2 Convergence rates

In this section we derive convergence rates for the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ generated by Algorithm 1 as well as for $\{\Psi_n\}_{n \geq 0}$, if the regularized augmented Lagrangian Ψ satisfies the Łojasiewicz property. The following definition is from [1] (see also [20]).

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

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

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

If Assumption 2 is fulfilled and $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ is the sequence generated by Algorithm 1, then, according to Theorem 9, the set of cluster points Ω is nonempty, compact and connected and Ψ takes on Ω the value Ψ_* ; in addition, $\Omega \subseteq \text{crit}(\Psi)$.

According to [1, Lemma 1], if Ψ has the Łojasiewicz property, then there exist $C_L > 0$, $\theta \in [0, 1)$ and $\varepsilon > 0$ such that for any

$$\mathbf{X} \in \{\mathbf{Z} \in \mathbf{R}: \text{dist}(\mathbf{Z}, \Omega) < \varepsilon\},$$

it holds

$$|\Psi(\mathbf{X}) - \Psi_*|^\theta \leq C_L \cdot \text{dist}(0, \partial\Psi(\mathbf{X})).$$

Obviously, Ψ is a KL function with desingularization function

$$\varphi : [0, +\infty) \rightarrow [0, +\infty), \quad \varphi(s) := \frac{1}{1-\theta} C_L s^{1-\theta},$$

which, according to Theorem 11, means that Ω contains a single element \mathbf{X}_* , which is the limit of $\{\mathbf{X}_n\}_{n \geq 1}$ as $n \rightarrow +\infty$. In other words, if Ψ has the Lojasiewicz property, then there exist $C_L > 0$, $\theta \in [0, 1)$ and $\varepsilon > 0$ such that for any $\mathbf{X} \in \text{Ball}(\mathbf{X}_*, \varepsilon)$

$$|\Psi(\mathbf{X}) - \Psi_*|^\theta \leq C_L \cdot \text{dist}(0, \partial\Psi(\mathbf{X})). \quad (3.6)$$

In this case, Ψ is said to satisfy the Lojasiewicz property with Lojasiewicz constant $C_L > 0$ and Lojasiewicz exponent $\theta \in [0, 1)$.

The following lemma will provide convergence rates for a particular class of monotonically decreasing real sequences converging to 0. Its proof can be found in [13, Lemma 15].

Lemma 12. *Let $\{e_n\}_{n \geq 0}$ be a monotonically decreasing sequence of nonnegative numbers converging to 0. Assume further that there exists natural numbers $n_0 \geq 1$ such that for any $n \geq n_0$*

$$e_{n-1} - e_n \geq C_e e_n^{2\theta},$$

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

(i) if $\theta = 0$, then $\{e_n\}_{n \geq 0}$ converges in finite time;

(ii) if $\theta \in (0, 1/2]$, then there exist $C_{e,0} > 0$ and $Q \in [0, 1)$ such that for any $n \geq n_0$

$$0 \leq e_n \leq C_{e,0} Q^n;$$

(iii) if $\theta \in (1/2, 1)$, then there exists $C_{e,1} > 0$ such that for any $n \geq n_0 + 1$

$$0 \leq e_n \leq C_{e,1} n^{-\frac{1}{2\theta-1}}.$$

We prove a recurrence inequality for the sequence $\{\mathcal{E}_n\}_{n \geq 0}$.

Lemma 13. *Let Assumption 2 be satisfied and $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ be a sequence generated by Algorithm 1. If Ψ satisfies the Lojasiewicz property with Lojasiewicz constant $C_L > 0$ and Lojasiewicz exponent $\theta \in [0, 1)$, then there exists $n_0 \geq 1$ such that the following estimate holds for any $n \geq n_0$*

$$\mathcal{E}_{n-1} - \mathcal{E}_n \geq C_{10} \mathcal{E}_n^{2\theta}, \quad \text{where} \quad C_{10} := \frac{C_8}{3(C_L \cdot C_9)^2}. \quad (3.7)$$

Proof. For every $n \geq 2$ we obtain from Lemma 4

$$\begin{aligned} \mathcal{E}_{n-1} - \mathcal{E}_n &= \Psi_{n-1} - \Psi_n \\ &\geq C_8 \left(\|x_n - x_{n-1}\|^2 + \|y_n - y_{n-1}\|^2 + \|u_n - u_{n-1}\|^2 \right) \\ &\geq \frac{1}{3} C_8 \left(\|x_n - x_{n-1}\| + \|y_n - y_{n-1}\| + \|u_n - u_{n-1}\| \right)^2 \\ &\geq C_{10} C_L^2 \|D_n\|^2, \end{aligned}$$

where $D_n \in \partial\Psi(\mathbf{X}_n)$. Let $\varepsilon > 0$ be such that (3.6) is fulfilled and choose $n_0 \geq 1$ with the property that for any $n \geq n_0$, \mathbf{X}_n belongs to $\text{Ball}(\mathbf{X}_*, \varepsilon)$. Relation (3.6) implies (3.7) for any $n \geq n_0$. \square

The following result follows by combining Lemma 12 with Lemma 13.

Theorem 14. *Let Assumption 2 be satisfied and $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ be a sequence generated by Algorithm 1. If Ψ 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 $\{\Psi_n\}_{n \geq 1}$ converges in finite time;

(ii) if $\theta \in (0, 1/2]$, then there exist $n_0 \geq 1$, $\hat{C}_0 > 0$ and $Q \in [0, 1)$ such that for any $n \geq n_0$

$$0 \leq \Psi_n - \Psi_* \leq \hat{C}_0 Q^n;$$

(iii) if $\theta \in (1/2, 1)$, then there exist $n_0 \geq 1$ and $\hat{C}_1 > 0$ such that for any $n \geq n_0 + 1$

$$0 \leq \Psi_n - \Psi_* \leq \hat{C}_1 n^{-\frac{1}{2\theta-1}}.$$

The next lemma will play an important role when transferring the convergence rates for $\{\Psi_n\}_{n \geq 0}$ to the sequence of iterates $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$.

Lemma 15. *Let Assumption 2 be satisfied and $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ be a sequence generated by Algorithm 1. Let (x_*, y_*, z_*, u_*) be the KKT point of the optimization problem (1.1) to which $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ converges as $n \rightarrow +\infty$. Then there exists $n_0 \geq 1$ such that the following estimates hold for any $n \geq n_0$*

$$\begin{aligned} \|x_n - x_*\| &\leq C_{11} \max \left\{ \sqrt{\mathcal{E}_n}, \varphi(\mathcal{E}_n) \right\}, & \|y_n - y_*\| &\leq C_{11} \max \left\{ \sqrt{\mathcal{E}_n}, \varphi(\mathcal{E}_n) \right\}, \\ \|z_n - z_*\| &\leq C_{12} \max \left\{ \sqrt{\mathcal{E}_n}, \varphi(\mathcal{E}_n) \right\}, & \|u_n - u_*\| &\leq C_{11} \max \left\{ \sqrt{\mathcal{E}_n}, \varphi(\mathcal{E}_n) \right\}, \end{aligned} \quad (3.8)$$

where

$$C_{11} := 2\sqrt{3C_8} + 3C_8C_9 \quad \text{and} \quad C_{12} := \left(\|A\| + \frac{2}{\sigma\beta} \right) C_{11}.$$

Proof. We assume that $\mathcal{E}_n > 0$ for any $n \geq 0$. Otherwise, the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ becomes identical to (x_*, y_*, z_*, u_*) beginning with a given index and the conclusion follows automatically (see the proof of Theorem 11).

Let $\varepsilon > 0$ be such that (3.6) is fulfilled and $n_0 \geq 2$ be such that X_n belongs to $\text{Ball}(\mathbf{X}_*, \varepsilon)$ for any $n \geq n_0$.

We fix $n \geq n_0$ now. One can easily notice that

$$\|x_n - x_*\| \leq \|x_{n+1} - x_n\| + \|x_{n+1} - x_*\| \leq \dots \leq \sum_{k \geq n} \|x_{k+1} - x_k\|.$$

Similarly, we derive

$$\|y_n - y_*\| \leq \sum_{k \geq n} \|y_{k+1} - y_k\|, \quad \|z_n - z_*\| \leq \sum_{k \geq n} \|z_{k+1} - z_k\|, \quad \|u_n - u_*\| \leq \sum_{k \geq n} \|u_{k+1} - u_k\|.$$

On the other hand, in view of (3.5) and by taking $\alpha := \frac{1}{2C_9}$ the inequality (3.4) can be written as

$$a_{n+1} \leq \frac{1}{2}a_n + b_n \quad \forall n \geq n_0.$$

Let us fix now an integer $N \geq n$. Summing up the above inequality for $k = n, \dots, N$, we have

$$\begin{aligned} \sum_{k=n}^N a_{k+1} &\leq \frac{1}{2} \sum_{k=n}^N a_k + \sum_{k=n}^N b_k = \frac{1}{2} \sum_{k=n}^N a_{k+1} + a_n - a_{N+1} + \sum_{k=n}^N b_k \\ &\leq \frac{1}{2} \sum_{k=n}^N a_{k+1} + a_n + \frac{3C_8C_9}{2} \varphi(\mathcal{E}_n). \end{aligned}$$

By passing $N \rightarrow +\infty$, we obtain

$$\begin{aligned} \sum_{k \geq n} a_{k+1} &= \sum_{k \geq n} (\|x_{k+1} - x_k\| + \|y_{k+1} - y_k\| + \|u_{k+1} - u_k\|) \\ &\leq 2(\|x_{n+1} - x_n\| + \|y_{n+1} - y_n\| + \|u_{n+1} - u_n\|) + 3C_8C_9\varphi(\mathcal{E}_n) \\ &\leq 2\sqrt{3} \cdot \sqrt{\|x_{n+1} - x_n\|^2 + \|y_{n+1} - y_n\|^2 + \|u_{n+1} - u_n\|^2} + 3C_8C_9\varphi(\mathcal{E}_n) \\ &\leq 2\sqrt{3C_8} \cdot \sqrt{\mathcal{E}_n - \mathcal{E}_{n+1}} + 3C_8C_9\varphi(\mathcal{E}_n), \end{aligned}$$

which gives the desired statement. \square

We can now formulate convergence rates for the sequence of generated iterates.

Theorem 16. *Let Assumption 2 be satisfied and $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ be a sequence generated by Algorithm 1. Suppose further that Ψ satisfies the Lojasiewicz property with Lojasiewicz constant $C_L > 0$ and Lojasiewicz exponent $\theta \in [0, 1)$. Let (x_*, y_*, z_*, u_*) be the KKT point of the optimization problem (1.1) to which $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ converges as $n \rightarrow +\infty$. Then the following statements are true:*

- (i) if $\theta = 0$, then the algorithm converges in finite time;
- (ii) if $\theta \in (0, 1/2]$, then there exist $n_0 \geq 1$, $\hat{C}_{0,1}, \hat{C}_{0,2}, \hat{C}_{0,3}, \hat{C}_{0,4} > 0$ and $\hat{Q} \in [0, 1)$ such that for any $n \geq n_0$

$$\|x_n - x_*\| \leq \hat{C}_{0,1} \hat{Q}^k, \quad \|y_n - y_*\| \leq \hat{C}_{0,2} \hat{Q}^k, \quad \|z_n - z_*\| \leq \hat{C}_{0,3} \hat{Q}^k, \quad \|u_n - u_*\| \leq \hat{C}_{0,4} \hat{Q}^k;$$

- (iii) if $\theta \in (1/2, 1)$, then there exist $n_0 \geq 1$ and $\hat{C}_{1,1}, \hat{C}_{1,2}, \hat{C}_{1,3}, \hat{C}_{1,4} > 0$ such that for any $n \geq n_0 + 1$

$$\begin{aligned} \|x_n - x_*\| &\leq \hat{C}_{1,1} n^{-\frac{1-\theta}{2\theta-1}}, & \|y_n - y_*\| &\leq \hat{C}_{1,2} n^{-\frac{1-\theta}{2\theta-1}}, \\ \|z_n - z_*\| &\leq \hat{C}_{1,3} n^{-\frac{1-\theta}{2\theta-1}}, & \|u_n - u_*\| &\leq \hat{C}_{1,4} n^{-\frac{1-\theta}{2\theta-1}}. \end{aligned}$$

Proof. Let

$$\varphi : [0, +\infty) \rightarrow [0, +\infty), \quad s \mapsto \frac{1}{1-\theta} C_L s^{1-\theta},$$

be the desingularization function.

- (i) If $\theta = 0$, then $\{\Psi_n\}_{n \geq 1}$ converges in finite time. As seen in the proof of Theorem 11, the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ becomes identical to (x_*, y_*, z_*, u_*) starting from a given index. In other words, the sequence $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ converges also in finite time and the conclusion follows.

Let be $\theta \neq \frac{1}{2}$ and $n'_0 \geq 1$ such that for any $n \geq n'_0$ the inequalities (3.8) in Lemma 15 and

$$\mathcal{E}_n \leq \left(\frac{1}{1-\theta} C_L \right)^{\frac{2}{2\theta-1}}$$

hold.

- (ii) If $\theta \in (0, 1/2)$, then $2\theta - 1 < 0$ and thus for any $n \geq n'_0$

$$\frac{1}{1-\theta} C_L \mathcal{E}_n^{1-\theta} \leq \sqrt{\mathcal{E}_n},$$

which implies that

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

If $\theta = 1/2$, then

$$\varphi(\mathcal{E}_n) = 2C_L \sqrt{\mathcal{E}_n},$$

thus

$$\max \left\{ \sqrt{\mathcal{E}_n}, \varphi(\mathcal{E}_n) \right\} = \max \{1, 2C_L\} \cdot \sqrt{\mathcal{E}_n} \quad \forall n \geq 1.$$

In both cases we have

$$\max \left\{ \sqrt{\mathcal{E}_n}, \varphi(\mathcal{E}_n) \right\} \leq \max \{1, 2C_L\} \cdot \sqrt{\mathcal{E}_n} \quad \forall n \geq n'_0.$$

By Theorem 14, there exist $n''_0 \geq 1$, $\hat{C}_0 > 0$ and $Q \in [0, 1)$ such that for $\hat{Q} := \sqrt{Q}$ and every $n \geq n''_0$ it holds

$$\sqrt{\mathcal{E}_n} \leq \sqrt{\hat{C}_0} Q^{n/2} = \sqrt{\hat{C}_0} \hat{Q}^n.$$

The conclusion follows from Lemma 15 for $n_0 := \max \{n'_0, n''_0\}$.

(iii) If $\theta \in (1/2, 1)$, then $2\theta - 1 > 0$ and thus for any $n \geq n'_0$

$$\sqrt{\mathcal{E}_n} \leq \frac{1}{1-\theta} C_L \mathcal{E}_n^{1-\theta},$$

which implies that

$$\max \left\{ \sqrt{\mathcal{E}_n}, \varphi(\mathcal{E}_n) \right\} = \varphi(\mathcal{E}_n) = \frac{1}{1-\theta} C_L \mathcal{E}_n^{1-\theta}.$$

By Theorem 14, there exist $n''_0 \geq 1$ and $\hat{C}_1 > 0$ such that for any $n \geq n''_0$

$$\frac{1}{1-\theta} C_L \mathcal{E}_n^{1-\theta} \leq \frac{1}{1-\theta} C_L \hat{C}_1^{1-\theta} (n-2)^{-\frac{1-\theta}{2\theta-1}}.$$

The conclusion follows again for $n_0 := \max \{n'_0, n''_0\}$ from Lemma 15. \square

References

- [1] **H. Attouch, J. Bolte.** *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features.* Mathematical Programming 116(1), 5–16 (2009)
- [2] **H. Attouch, J. Bolte, P. Redont, A. Soubeyran.** *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Lojasiewicz inequality.* Mathematics of Operations Research 35(2), 438–457 (2010)
- [3] **H. Attouch, J. Bolte, B. F. Svaiter.** *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods.* Mathematical Programming 137(1–2), 91–129 (2013)
- [4] **H.H. Bauschke, P.L. Combettes.** *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* CMS Books in Mathematics. Springer, New York (2011)
- [5] **A. Beck.** *First-Order Methods in Optimization.* MOS-SIAM Series on Optimization. SIAM, Philadelphia (2017)
- [6] **J. Bolte, A. Daniilidis, A. Lewis.** *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems.* SIAM Journal on Optimization 17(4), 1205–1223 (2006)
- [7] **J. Bolte, A. Daniilidis, A. Lewis, M. Shiota.** *Clarke subgradients of stratifiable functions.* SIAM Journal on Optimization 18(2), 556–572 (2007)
- [8] **J. Bolte, A. Daniilidis, O. Ley, L. Mazet.** *Characterizations of Lojasiewicz inequalities: subgradient flows, talweg, convexity.* Transactions of the American Mathematical Society 362(6), 3319–3363 (2010)
- [9] **J. Bolte, S. Sabach, M. Teboulle.** *Proximal alternating linearized minimization for nonconvex and nonsmooth problems.* Mathematical Programming, 146(1): 459–494 (2014)
- [10] **J. Bolte, S. Sabach, M. Teboulle.** *Nonconvex Lagrangian-based optimization: monitoring schemes and global convergence.* Mathematics of Operations Research, to appear. <https://doi.org/10.1287/moor.2017.0900>
- [11] **R. I. Boţ, E. R. Csetnek.** *An inertial Tseng’s type proximal algorithm for nonsmooth and nonconvex optimization problems.* Journal of Optimization Theory and Applications 171(2), 600–616 (2016)
- [12] **R. I. Boţ, E. R. Csetnek, S. C. László.** *An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions.* EURO Journal on Computational Optimization 4(1), 3–25 (2016)
- [13] **R. I. Boţ, D.-K. Nguyen.** *The proximal alternating direction method of multipliers in the nonconvex setting: convergence analysis and rates.* arXiv:1801.01994

- [14] **P. L. Combettes, V. R. Wajs.** *Signal recovery by proximal forward-backward splitting.* Multiscale Modeling and Simulation 4(4), 1168–1200 (2005)
- [15] **W. Hare, C. Sagastizábal.** *Computing proximal points of nonconvex functions.* Mathematical Programming 116(1-2), 221–258 (2009)
- [16] **M. Hong, Z.-Q. Luo.** *On the linear convergence of the alternating direction method of multipliers.* Mathematica Programming 162, 165–199 (2017)
- [17] **K. Kurdyka.** *On gradients of functions definable in o-minimal structures.* Annales de l’Institut Fourier 48, 769–783 (1998)
- [18] **A. Lewis, J. Malick.** *Alternating projection on manifolds.* Mathematics of Operations Research 33(1), 216–234 (2008)
- [19] **G. Li, T. K. Pong.** *Global convergence of splitting methods for nonconvex composite optimization.* SIAM Journal on Optimization 25(4), 2434–2460 (2015)
- [20] **S. Łojasiewicz .** *Une propriété topologique des sous-ensembles analytiques réels, Les Équations aux Dérivées Partielles.* Éditions du Centre National de la Recherche Scientifique, Paris, 8–89 (1963)
- [21] **B. Mordukhovich.** *Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications.* Springer, Berlin (2006)
- [22] **J. Moreau.** *Fonctions convexes duales et points proximaux dans un espace hilbertien.* Comptes Rendus de l’Académie des Sciences (Paris), Série A 255, 2897–2899 (1962)
- [23] **T. Pock, S. Sabach.** *Inertial proximal alternating linearized minimization (iPALM) for nonconvex and nonsmooth problems,* SIAM Journal Imaging Sciences 9(4), 1756–1787 (2016)
- [24] **R. T. Rockafellar, R. J.-B. Wets.** *Variational Analysis.* Fundamental Principles of Mathematical Sciences 317. Springer, Berlin (1998)
- [25] **L. Yang, T. K. Pong and X. Chen.** *Alternating Direction Method of Multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction.* SIAM Journal on Imaging Sciences, 10(1), 74–110 (2017)