SIAM J. OPTIM.
Vol. 29, No. 2, pp. 1300–1328

A PROXIMAL MINIMIZATION ALGORITHM FOR STRUCTURED NONCONVEX AND NONSMOOTH PROBLEMS*

RADU IOAN BOŢ†, ERNÖ ROBERT CSETNEK†, AND DANG-KHOA NGUYEN†

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 (with 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 classifications. 65K10, 90C26, 90C30

DOI. 10.1137/18M1190689

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

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

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

In the case in which $m = p$ and A is the identity operator, Bolte, Sabach, and Teboulle formulated in [12], 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 it 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, and compressed sensing, among others. An inertial version of PALM was proposed by Pock and Sabach in [26].

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,

*Received by the editors May 29, 2018; accepted for publication (in revised form) February 8, 2019; published electronically May 14, 2019.
http://www.siam.org/journals/siopt/29-2/M119068.html
Funding: The first author's research was partially supported by the FWF (Austrian Science Fund), project I 2419-N32. The second author's research was supported by the FWF, project P 29809-N32. The research of the third author was supported by the Doctoral Programme Vienna Graduate School on Computational Optimization (VGSCO), which is funded by the FWF, project W1260-N35.
†Faculty of Mathematics, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria (radu.bot@univie.ac.at, ernoe.robert.csetnek@univie.ac.at, dang-khoa.nguyen@univie.ac.at).

a closed formula is not available. In the last decade, impressive progress has been made 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. In addition, several fundamental algorithms, like the proximal method, the forward-backward splitting method, the regularized Gauss–Seidel method, the proximal alternating method, the forward-backward-forward method, and some of their inertial variants, have been exported from the convex to the nonconvex setting and proved to converge globally in the setting of the Kurdyka–Łojasiewicz property (see, for instance, [1, 2, 3, 12, 6, 7]). However, a similar undertaking for structured optimization problems, such as those which involve compositions with linear operators and require primal-dual methods with a full-splitting character, was not very successful. The main reason for that is the absence in the nonconvex setting of an analogue of convex conjugate duality theory.

Despite these premises we succeed in providing in this paper a full splitting algorithm for solving the nonconvex and nonsmooth problem (1.1); more precisely, the nonsmooth functions are processed via their proximal operators, the smooth function via gradient steps, and the linear operator via matrix times vector multiplication. The convergence analysis is based on a descent inequality, which we prove for a regularization of the augmented Lagrangian $L_\beta : \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$,

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

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 based on the use of the Kurdyka–Łojasiewicz property in the proof of the global convergence of nonconvex optimization algorithms we refer the reader to the papers of Attouch and Bolte [1], Attouch, Bolte, and Svaiter [3], and Bolte, Sabach, and Teboulle [12].

One of the benefits which comes with the new algorithm is that it furnishes a full splitting iterative scheme for the nonsmooth and nonconvex optimization problem

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

which follows as a particular case of (1.1) for $G(y) = 0$ and $H(x, y) = H(x)$ for any $(x, y) \in \mathbb{R}^m \times \mathbb{R}^q$, where $H : \mathbb{R}^m \to \mathbb{R}$ is a Fréchet differentiable function with Lipschitz continuous gradient.

In the last years, several articles have been devoted to the design and convergence analysis of algorithms for solving structured optimization problems in the nonconvex and nonsmooth setting. They all focus on algorithms relying on the alternating direction method of multipliers (ADMM), which is well known not to be a full splitting algorithm. Nonconvex ADMM algorithms for (1.2) have been proposed in [22], under the assumption that H is twice continuously differentiable with bounded Hessian, and in [30], under the assumption that one of the summands is convex and continuous on its effective domain. In [29], a general nonconvex optimization problem involving compositions with linear operators and smooth coupling functions is considered and the importance of providing sufficient conditions for the boundedness of the iterates generated by the proposed nonconvex ADMM algorithm is recognized. This is achieved by assuming that the objective function is continuous and coercive over the feasible set, while its nonsmooth part is either restricted prox-regular or piecewise linear. Similar ingredients are used in [23] in the convergence analysis of a nonconvex linearized ADMM algorithm. In [17], the ADMM technique is used to minimize the sum of finitely many smooth nonconvex functions and a nonsmooth convex function by reformulating it as a general consensus problem. In [28], a multi-block Bregman ADMM algorithm is proposed and analyzed in a setting based on restrictive strong convexity assumptions. On the other hand, in [18], two proximal variants of the ADMM algorithm are introduced and the analysis is focused on providing iteration complexity bounds to reach an ε-KKT solutions.

We would also like to mention in this context the recent publication [13] for the case when A is replaced by a nonlinear continuously differentiable operator.

1.2. Notation 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 $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \cdots \times \mathbb{R}^{d_k}$ of the Euclidean spaces $\mathbb{R}^{d_i}$, $i = 1, \ldots, k$, will be endowed with inner product and associated norm defined for $x := (x_1, \ldots, x_k)$, $y := (y_1, \ldots, y_k) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \cdots \times \mathbb{R}^{d_k}$ by

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

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

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

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

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

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

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

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

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

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

If $x \in \mathbb{R}^d$ is a local minimum of ψ, then $0 \in \partial\psi(x)$. We denote by $\operatorname{crit}(\psi) := \{x \in \mathbb{R}^d : 0 \in \partial\psi(x)\}$ 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 sequences in $\mathbb{R}^d$ such that $d_n \in \partial\psi(x_n)$ for any $n \geq 0$ and $(x_n, d_n) \to (x, d)$ and $\psi(x_n) \to \psi(x)$ as $n \to +\infty$, then $d \in \partial\psi(x)$. We also have the following subdifferential sum formula (see [24, Proposition 1.107] and [27, Exercise 8.8]): if $\Phi\colon \mathbb{R}^d \to \mathbb{R}$ 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 $A\colon \mathbb{R}^k \to \mathbb{R}^d$ (see [24, Proposition 1.112] and [27, 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 the reader to [8, Proposition 1].

Proposition 1.1. *Let $\psi\colon \mathbb{R}^d \to \mathbb{R}$ 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 $z \in [x, y] = \{(1-t)x + ty\colon t \in [0,1]\}$ it holds that*

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

(ii) *for any $\gamma \in \mathbb{R}\backslash\{0\}$ it holds that*

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

The descent lemma, which says that for a Fréchet differentiable function $\psi\colon \mathbb{R}^d \to \mathbb{R}$ having a Lipschitz continuous gradient with constant $\ell > 0$ it holds that

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

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 when ψ is ℓ-semiconvex [11]. In other words, a consequence of Proposition (1.1) 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 what follows in the convergence analysis. The following lemma is useful when proving convergence of numerical algorithms relying on Fejér monotonicity techniques (see, for instance, [6, Lemma 2.2] and [7, Lemma 2]).

Lemma 1.2. *Let $\{\xi_n\}_{n\geq 0}$ be a sequence of real numbers and $\{\omega_n\}_{n\geq 0}$ be 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 [6, Lemma 2.3] (see also [7, Lemma 3]).

LEMMA 1.3. *Let* $\{a_n\}_{n\geq 0}$ *and* $\{b_n\}_{n\geq 1}$ *be sequences of real nonnegative numbers such that, for any* $n \geq 1$,

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

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 2.1. *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:*

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

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

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

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

The proximal point operator with parameter $\gamma > 0$ (see [25]) of a proper and lower semicontinuous function $\psi\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is the set-valued operator defined as

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

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, 21]. In view of the above definition, the iterative scheme (2.1a)–(2.1d) reads, for every $n \geq 0$,

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

One can notice the full splitting character of Algorithm 2.1 and also that the first two steps can be performed in parallel.

Remark 2.2.
(i) In the case in which $G(y) = 0$ and $H(x, y) = H(x)$ for any $(x, y) \in \mathbb{R}^m \times \mathbb{R}^q$, where $H : \mathbb{R}^m \to \mathbb{R}$ is a Fréchet differentiable function with Lipschitz continuous gradient, Algorithm 2.1 gives rise to an iterative scheme for solving (1.2) (see also [8]) that reads, for any $n \geq 0$,

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

(ii) In the case in which $m = p$ and $A = \mathrm{Id}$ is the identity operator on $\mathbb{R}^m$, Algorithm 2.1 gives rise to an iterative scheme for solving

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

which reads, for any $n \geq 0$,

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

We notice that, similar to PALM [12], which is also designed to solve optimization problems of the form (2.2), the algorithm evaluates F and G by proximal steps, while H is evaluated by gradient steps for each of the two blocks.

(iii) In the case in which $m = p$, $A = \mathrm{Id}$, $F(x) = 0$, and $H(x,y) = H(y)$ for any $(x,y) \in \mathbb{R}^m \times \mathbb{R}^q$, where $H : \mathbb{R}^q \to \mathbb{R}$ is a Fréchet differentiable function with Lipschitz continuous gradient, Algorithm 2.1 gives rise to an iterative scheme for solving

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

which reads, for any $n \geq 0$,

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

and is none other 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 [7].

2.1. A descent inequality. We will start with the convergence analysis of Algorithm 2.1 by proving a descent inequality which will play a fundamental role in our investigations. We will analyse Algorithm 2.1 under the following assumptions, which we later weaken even further.

Assumption 2.3.
(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

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

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

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

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

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

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

Remark 2.4. Some comments on Assumption 2.3 are in order.

(i) Assumption (i) ensures that the sequence generated by Algorithm 2.1 is well defined. It also has as a consequence that

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

(ii) Comparing the assumptions in (iii) and (iv) to the ones in [12], 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 [12, 6, 7]) the boundedness of the iterates follows by combining a descent inequality expressed in terms of the objective function with coercivity assumptions on the latter. In our setting this undertaking 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 as the fact that the matrix associated with AA^T is positive definite. Since

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

this is furthermore 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 of M, namely the ratio between the maximal eigenvalue $\lambda_{\max}(M)$ and the minimal eigenvalue of the square matrix M,

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

Here, $\|M\|$ denotes the operator norm of M induced by the Euclidean vector norm.

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

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

defined as

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

where

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

Notice that

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

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

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

For simplification, we introduce the following notation:

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

The next result provides the announced descent inequality.

Lemma 2.5. *Let Assumption 2.3 be satisfied, let $2\tau \geq \beta\|A\|^2$, and let the sequence $\{(x_n, y_n, z_n, u_n)\}_{n\geq 0}$ be generated by Algorithm 2.1. Then for any $n \geq 1$ it holds that*

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

where

$$\begin{aligned}C_2 &:= \tau - \frac{\ell_{1,+} + \beta\|A\|^2}{2} - \frac{4\sigma\tau^2}{\beta\lambda_{\min}(AA^T)} - \frac{8(\sigma\tau + \ell_{1,+})^2}{\sigma\beta\lambda_{\min}(AA^T)}, \\ C_3 &:= \frac{\mu - \ell_{2,+}}{2} - \frac{8\ell_{3,+}^2}{\sigma\beta\lambda_{\min}(AA^T)}, \\ C_4 &:= \frac{1}{\sigma\beta}.\end{aligned}$$

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

$$\begin{aligned}&F(z_{n+1}) + G(y_{n+1}) + H(x_{n+1}, y_{n+1}) + \langle u_{n+1}, Ax_{n+1} - z_{n+1}\rangle \\ &\quad + \frac{\beta}{2}\|Ax_{n+1} - z_{n+1}\|^2 + \left(\tau - \frac{\ell_{1,+} + \beta\|A\|^2}{2}\right)\|x_{n+1} - x_n\|^2 \\ &\quad + \frac{\mu - \ell_{2,+}}{2}\|y_{n+1} - y_n\|^2 + \frac{1}{\sigma\beta}\|u_{n+1} - u_n\|^2 \\ &\qquad \leq F(z_n) + G(y_n) + H(x_n, y_n) + \langle u_n, Ax_n - z_n\rangle \\ &\qquad\quad + \frac{\beta}{2}\|Ax_n - z_n\|^2 + \frac{2}{\sigma\beta}\|u_{n+1} - u_n\|^2\end{aligned} \tag{2.8}$$

and provide afterwards an upper estimate for the term $\|u_{n+1} - u_n\|^2$ on the right-hand side of (2.8).

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

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

and

$$\begin{aligned}&F(z_{n+1}) + \langle u_n, Ax_n - z_{n+1}\rangle + \frac{\beta}{2}\|Ax_n - z_{n+1}\|^2 \\ &\qquad \leq F(z_n) + \langle u_n, Ax_n - z_n\rangle + \frac{\beta}{2}\|Ax_n - z_n\|^2,\end{aligned}$$

respectively. Adding these two inequalities yields

$$\begin{aligned} & F\left(z_{n+1}\right)+G\left(y_{n+1}\right)+\left\langle u_n, A x_n-z_{n+1}\right\rangle+\frac{\beta}{2}\left\|A x_n-z_{n+1}\right\|^2 \\ & \quad+\left\langle\nabla_y H\left(x_n, y_n\right), y_{n+1}-y_n\right\rangle+\frac{\mu}{2}\left\|y_{n+1}-y_n\right\|^2 \\ & \leq F\left(z_n\right)+G\left(y_n\right)+\left\langle u_n, A x_n-z_n\right\rangle+\frac{\beta}{2}\left\|A x_n-z_n\right\|^2 . \end{aligned} \tag{2.9}$$

On the other hand, according to the descent lemma we have

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

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

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

Combining the two above estimates, we get

$$\begin{aligned} & H\left(x_{n+1}, y_{n+1}\right)+\left\langle u_n, A x_{n+1}-A x_n\right\rangle-\frac{\beta}{2}\left\|A x_n-z_{n+1}\right\|^2 \\ & \quad+\frac{\beta}{2}\left\|A x_{n+1}-z_{n+1}\right\|^2-\frac{\ell_{2,+}}{2}\left\|y_{n+1}-y_n\right\|^2 \\ & \quad+\left(\tau-\frac{\ell_{1,+}+\beta\|A\|^2}{2}\right)\left\|x_{n+1}-x_n\right\|^2 \\ & \leq H\left(x_n, y_n\right)+\left\langle\nabla_y H\left(x_n, y_n\right), y_{n+1}-y_n\right\rangle . \end{aligned} \tag{2.10}$$

We obtain (2.8) after we sum up (2.9) and (2.10), use (2.1d), and add $\frac{2}{\sigma\beta}\|u_{n+1}-u_n\|^2$ to both sides of the resulting inequality.

Next we will focus on estimating $\|u_{n+1}-u_n\|^2$. We can rewrite (2.1c) as

$$\begin{aligned} \tau\left(x_n-x_{n+1}\right) & =\nabla_x H\left(x_n, y_{n+1}\right)+A^T u_n+\beta A^T\left(A x_{n+1}-z_{n+1}\right) \\ & \qquad +\beta A^T A\left(x_n-x_{n+1}\right) \\ & =\nabla_x H\left(x_n, y_{n+1}\right)+A^T u_n+\frac{1}{\sigma} A^T\left(u_{n+1}-u_n\right)+\beta A^T A\left(x_n-x_{n+1}\right), \end{aligned}$$

where the last equation is due to (2.1d). After multiplying both sides by σ and rearranging the terms, we get

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

Since n has been arbitrarily chosen, we also have

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

Subtracting these relations and making use of the notation

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

yields

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

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

$$\left\|w_{n+1}\right\|^2 \leq (1-\sigma)\left\|w_n\right\|^2 + \sigma\left\|v_n\right\|^2. \tag{2.11}$$

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

$$\left\|A^T\left(u_{n+1} - u_n\right)\right\| \leq \left\|w_{n+1}\right\| + \sigma\|B\|\left\|x_{n+1} - x_n\right\| \leq \left\|w_{n+1}\right\| + \sigma\tau\left\|x_{n+1} - x_n\right\| \tag{2.12}$$

and

$$\begin{aligned} \left\|v_n\right\| &\leq \sigma\|B\|\left\|x_n - x_{n-1}\right\| + \left\|\nabla_x H\left(x_{n-1}, y_n\right) - \nabla_x H\left(x_n, y_{n+1}\right)\right\| \\ &\leq \sigma\tau\left\|x_n - x_{n-1}\right\| + \left\|\nabla_x H\left(x_{n-1}, y_n\right) - \nabla_x H\left(x_n, y_n\right)\right\| \\ &\quad + \left\|\nabla_x H\left(x_n, y_n\right) - \nabla_x H\left(x_n, y_{n+1}\right)\right\| \\ &\leq \left(\sigma\tau + \ell_{1,+}\right)\left\|x_n - x_{n-1}\right\| + \ell_{3,+}\left\|y_{n+1} - y_n\right\|, \end{aligned} \tag{2.13}$$

respectively. Using the Cauchy–Schwarz inequality, (2.12) yields

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

and (2.13) yields

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

After combining these two inequalities with (2.11), we get

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

Multiplying the above relation by $\frac{4}{\sigma^2\beta\lambda_{\min}(AA^T)} > 0$ and adding the resulting inequality to (2.8) yields

$$\begin{aligned} &F\left(z_{n+1}\right) + G\left(y_{n+1}\right) + H\left(x_{n+1}, y_{n+1}\right) + \left\langle u_{n+1}, Ax_{n+1} - z_{n+1}\right\rangle + \frac{\beta}{2}\left\|Ax_{n+1} - z_{n+1}\right\|^2 \\ &\quad + \frac{4(1-\sigma)}{\sigma^2\beta\lambda_{\min}(AA^T)}\|A^T(u_{n+1} - u_n) + \sigma B(x_{n+1} - x_n)\|^2 \end{aligned}$$

$$
\begin{aligned}
&+ \frac{8\left(\sigma\tau+\ell_{1,+}\right)^2}{\sigma\beta\lambda_{\min}\left(AA^T\right)}\|x_{n+1}-x_n\|^2 \\
&+ \left(\tau - \frac{\ell_{1,+}+\beta\|A\|^2}{2} - \sigma^3\tau^2 - \frac{8\left(\sigma\tau+\ell_{1,+}\right)^2}{\sigma\beta\lambda_{\min}\left(AA^T\right)}\right)\|x_{n+1}-x_n\|^2 \\
&+ \left(\frac{\mu-\ell_{2,+}}{2} - \frac{8\ell_{3,+}^2}{\sigma\beta\lambda_{\min}\left(AA^T\right)}\right)\|y_{n+1}-y_n\|^2 + \frac{1}{\sigma\beta}\|u_{n+1}-u_n\|^2 \\
&\quad \leq F(z_n) + G(y_n) + H(x_n,y_n) + \langle u_n, Ax_n - z_n\rangle + \frac{\beta}{2}\|Ax_n - z_n\|^2 \\
&\qquad + \frac{4(1-\sigma)}{\sigma^2\beta\lambda_{\min}(AA^T)}\|A^T(u_n-u_{n-1}) + \sigma B(x_n - x_{n-1})\|^2 \\
&\qquad + \frac{8\left(\sigma\tau+\ell_{1,+}\right)^2}{\sigma\beta\lambda_{\min}\left(AA^T\right)}\|x_n - x_{n-1}\|^2,
\end{aligned}
$$

which is none other than (2.7). ∎

The following result provides one possibility for choosing the parameters in Algorithm 2.1 such that all three constants C_2, C_3, and C_4 that appear in (2.7) are positive.

LEMMA 2.6. *Let*

$$0 < \sigma < \frac{1}{24\kappa\left(AA^T\right)}, \tag{2.14a}$$

$$\beta > \frac{\nu}{1-24\sigma\kappa\left(AA^T\right)}\left(4+3\sigma+\sqrt{24+24\sigma+9\sigma^2-192\sigma\kappa\left(AA^T\right)}\right) > 0, \tag{2.14b}$$

$$\max\left\{\frac{\beta\|A\|^2}{2}, \frac{\beta\lambda_{\min}\left(AA^T\right)}{24\sigma}\left(1-\frac{4\nu}{\beta}-\sqrt{\Delta'_\tau}\right)\right\} < \tau < \frac{\beta\lambda_{\min}\left(AA^T\right)}{24\sigma}\left(1-\frac{4\nu}{\beta}+\sqrt{\Delta'_\tau}\right), \tag{2.14c}$$

$$\mu > \ell_{2,+} + \frac{16\ell_{3,+}^2}{\sigma\beta\lambda_{\min}\left(AA^T\right)} > 0, \tag{2.14d}$$

where

$$\nu := \frac{4\ell_{1,+}}{\lambda_{\min}\left(AA^T\right)} > 0 \qquad \textit{and} \qquad \Delta'_\tau := 1 - \frac{8\nu}{\beta} - \frac{8\nu^2}{\beta^2} - \frac{6\nu\sigma}{\beta} - 24\sigma\kappa\left(AA^T\right) > 0.$$

Then we have

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

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

$$\frac{1}{\gamma_1} - \frac{\ell_{1,+}}{2\gamma_1^2} = \frac{1}{\beta\lambda_{\min}\left(AA^T\right)} \qquad \textit{and} \qquad \frac{1}{\gamma_2} - \frac{\ell_{1,+}}{2\gamma_2^2} = \frac{2}{\beta\lambda_{\min}\left(AA^T\right)}. \tag{2.15}$$

Proof. We will prove first that $C_2 > 0$ or, equivalently,

$$-2C_2 = \frac{24\sigma\tau^2}{\beta\lambda_{\min}\left(AA^T\right)} - 2\left(1-\frac{16\ell_{1,+}}{\beta\lambda_{\min}\left(AA^T\right)}\right)\tau + \frac{16\ell_{1,+}^2}{\sigma\beta\lambda_{\min}\left(AA^T\right)} + \ell_{1,+} + \beta\|A\|^2 < 0. \tag{2.16}$$

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

$$\begin{aligned}
\Delta'_\tau &:= \left(1 - \frac{16\ell_{1,+}}{\beta\lambda_{\min}\left(AA^T\right)}\right)^2 - \frac{384\ell_{1,+}^2}{\beta^2\lambda_{\min}^2\left(AA^T\right)} - \frac{24\ell_{1,+}\sigma}{\beta\lambda_{\min}\left(AA^T\right)} - 24\sigma\kappa\left(AA^T\right) \\
&= \left(1 - \frac{4\nu}{\beta}\right)^2 - \frac{24\nu^2}{\beta^2} - \frac{6\nu\sigma}{\beta} - 24\sigma\kappa\left(AA^T\right) \\
&= 1 - \frac{8\nu}{\beta} - \frac{8\nu^2}{\beta^2} - \frac{6\nu\sigma}{\beta} - 24\sigma\kappa\left(AA^T\right) > 0 \qquad (2.17)
\end{aligned}$$

if σ and β are chosen as in (2.14a) and (2.14b), respectively. Indeed, the inequality (2.17) is equivalent to

$$\left(1 - 24\sigma\kappa\left(AA^T\right)\right)\beta^2 - 2\left(4 + 3\sigma\right)\nu\beta - 8\nu^2 > 0.$$

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

$$\begin{aligned}
\Delta_\beta &:= \left[(4 + 3\sigma)^2 + 8\left(1 - 24\sigma\kappa\left(AA^T\right)\right)\right]\nu^2 \\
&= \left[24 + 24\sigma + 9\sigma^2 - 192\sigma\kappa\left(AA^T\right)\right]\nu^2 > 0
\end{aligned}$$

as $24 - 192\sigma\kappa(AA^T) = 16 + 8(1 - 24\sigma\kappa(AA^T)) > 0$ for every σ that satisfies (2.14a). Hence, for every σ satisfying (2.14a) and every β satisfying (2.14b) we have that (2.17) holds. Therefore, (2.16) is satisfied for every

$$\frac{\beta\lambda_{\min}\left(AA^T\right)}{24\sigma}\left(1 - \frac{4\nu}{\beta} - \sqrt{\Delta'_\tau}\right) < \tau < \frac{\beta\lambda_{\min}\left(AA^T\right)}{24\sigma}\left(1 - \frac{4\nu}{\beta} + \sqrt{\Delta'_\tau}\right).$$

It remains to verify the feasibility of τ in (2.14c), in other words, to prove that

$$\frac{\beta\left\|A\right\|^2}{2} < \frac{\beta\lambda_{\min}\left(AA^T\right)}{24\sigma}\left(1 - \frac{4\nu}{\beta} + \sqrt{\Delta'_\tau}\right).$$

This is easy to see, as, according to (2.17), we have

$$\frac{\beta\left\|A\right\|^2}{2} < \frac{\beta\lambda_{\min}\left(AA^T\right)}{24\sigma}\left(1 - \frac{4\nu}{\beta}\right) \Leftrightarrow 1 - \frac{4\nu}{\beta} - 12\sigma\kappa\left(AA^T\right) > 0.$$

The positivity of C_3 follows from the choice of μ in (2.14d), while, obviously, $C_4 > 0$.

Finally, we notice that the reduced discriminants of the two quadratic equations in (2.15) (in γ_1 and γ_2) are

$$\Delta_{\gamma_1} := 1 - \frac{2\ell_{1,+}}{\beta\lambda_{\min}\left(AA^T\right)} = 1 - \frac{\nu}{2\beta} \qquad \text{and} \qquad \Delta_{\gamma_2} := 1 - \frac{\ell_{1,+}}{\beta\lambda_{\min}\left(AA^T\right)} = 1 - \frac{\nu}{4\beta}.$$

Since

$$\beta > \frac{\nu}{1 - 24\sigma\kappa\left(AA^T\right)} > \frac{\nu}{2},$$

it follows that $\Delta_{\gamma_1}, \Delta_{\gamma_2} > 0$ and hence each of the two equations has a nonzero real solution. ∎

Remark 2.7. 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. Huong and Luo also mentioned that the dual step size may be cumbersome to compute unless the objective function is strongly convex. As one can see in (2.14a) and (2.14b), the step size of the dual update in our algorithm can be chosen only in dependence of the condition number of AA^T.

THEOREM 2.8. *Let Assumption* 2.3 *be satisfied and let the parameters in Algorithm* 2.1 *be such that* $2\tau \geq \beta\|A\|^2$ *and the constants defined in Lemma* 2.5 *satisfy* $\min\{C_2, C_3, C_4\} > 0$. *If* $\{(x_n, y_n, z_n, u_n)\}_{n\geq 0}$ *is a sequence generated by Algorithm* 2.1*, then the following statements are true:*

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

(ii) *we have*

$$x_{n+1} - x_n \to 0, \; y_{n+1} - y_n \to 0, \; z_{n+1} - z_n \to 0, \quad \text{and } u_{n+1} - u_n \to 0 \text{ as } n \to +\infty. \tag{2.18}$$

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 2.5, $\{\Psi_n\}_{n\geq 1}$ 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

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

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

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

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 1.2 statement (i) and also that

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

Since for any $n \geq 1$ it holds that

$$\begin{aligned} \|z_{n+1} - z_n\| &\leq \|A\| \, \|x_{n+1} - x_n\| + \|Ax_{n+1} - z_{n+1}\| + \|Ax_n - z_n\| \\ &= \|A\| \, \|x_{n+1} - x_n\| + \frac{1}{\sigma\beta} \|u_{n+1} - u_n\| + \frac{1}{\sigma\beta} \|u_n - u_{n-1}\| , \end{aligned} \tag{2.19}$$

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

Remark 2.9. 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 [8, 22, 30]). In our analysis the only ingredients for obtaining statement (ii) in Theorem 2.8 are the descent property and Lemma 1.2.

As one can notice, the assumption that $\min\{C_2, C_3, C_4\} > 0$ plays an essential role in our analysis. In Lemma 2.6 we provide possible choices of the algorithm parameters, which lead to the fulfillment of this assumption. However, these choices depend on $\ell_{+,1}$, which, in turn, is defined as being a finite upper bound for the sequence of Lipschitz constants $(\ell_1(y_n))_{n\geq 0}$ (see (2.5)). This condition is definitely fulfilled when ℓ_1 is globally bounded. This is, for instance, the case when H depends only on x and has a Lipschitz continuous gradient (see Remark 2.2(i)), but also when H depends only on y.

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 2.8, 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 $\{u_{n+1}-u_n\}_{n\geq 0}$, thanks to (2.18), are bounded. Define

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

Even though this observation does not immediately imply 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 $\psi : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is called coercive if $\lim_{\|x\|\to +\infty} \psi(x) = +\infty$.

THEOREM 2.10. *Let Assumption* 2.3 *be satisfied and let the parameters in Algorithm* 2.1 *be such that* $2\tau \geq \beta\|A\|^2$, *the constants defined in Lemma* 2.5 *fulfil* $\min\{C_2, C_3, C_4\} > 0$, *and there exist* $\gamma_1, \gamma_2 \in \mathbb{R} \setminus \{0\}$ *such that* (2.15) *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* 2.1 *is bounded.*

Proof. Let $n \geq 1$ be fixed. According to Lemma 2.5 we have that

$$\begin{aligned} \Psi_1 \geq \cdots \geq \Psi_n \geq \Psi_{n+1} & \\ &\geq F(z_{n+1}) + G(y_{n+1}) + H(x_{n+1}, y_{n+1}) - \frac{1}{2\beta} \|u_{n+1}\|^2 \\ &\quad + \frac{\beta}{2} \left\| Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1} \right\|^2 . \end{aligned} \tag{2.20}$$

After multiplying (2.1c) by $-\tau$ and using (2.1d), we have

$$\begin{aligned}
A^T u_{n+1} &= A^T u_n + \sigma\beta A^T(Ax_{n+1} - z_{n+1}) \\
&= A^T u_n + (\sigma - 1)\beta A^T(Ax_{n+1} - z_{n+1}) + \beta A^T(Ax_{n+1} - z_{n+1}) \\
&= \left(1 - \frac{1}{\sigma}\right) A^T \left(u_{n+1} - u_n\right) + A^T u_n + \beta A^T(Ax_n - z_{n+1}) \\
&\quad + \beta A^T A(x_{n+1} - x_n) \\
&= \left(1 - \frac{1}{\sigma}\right) A^T \left(u_{n+1} - u_n\right) + (\tau \mathrm{Id} - \beta A^T A)(x_n - x_{n+1}) \\
&\quad - \nabla_x H\left(x_n, y_{n+1}\right) \\
&= \left(1 - \frac{1}{\sigma}\right) A^T \left(u_{n+1} - u_n\right) + B\left(x_n - x_{n+1}\right) \\
&\quad + \nabla_x H\left(x_{n+1}, y_{n+1}\right) - \nabla_x H\left(x_n, y_{n+1}\right) - \nabla_x H\left(x_{n+1}, y_{n+1}\right). \qquad (2.21)
\end{aligned}$$

This implies

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

By using the Cauchy–Schwarz inequality we further obtain

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

Multiplying the above relation by $\frac{1}{2\beta\lambda_{\min}(AA^T)}$ and combining it with (2.20), we get

$$\begin{aligned}
\Psi_1 &\geq F\left(z_{n+1}\right) + G\left(y_{n+1}\right) + H\left(x_{n+1}, y_{n+1}\right) \\
&\quad - \frac{1}{\beta\lambda_{\min}\left(AA^T\right)} \left\|\nabla_x H\left(x_{n+1}, y_{n+1}\right)\right\|^2 \\
&\quad - \frac{1}{\beta\lambda_{\min}\left(AA^T\right)} \left(\left(\frac{1}{\sigma} - 1\right) \|A\| + \tau + \ell_{1,+}\right)^2 s_*^2 \\
&\quad + \frac{\beta}{2} \left\|Ax_{n+1} - z_{n+1} + \frac{1}{\beta} u_{n+1}\right\|^2. \qquad (2.22)
\end{aligned}$$

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 (2.22) and Proposition 1.1, we have that, for any $n \geq 1$,

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

$$\leq \Psi_1 + \frac{1}{\beta \lambda_{\min}(AA^T)} \left(\left(\frac{1}{\sigma} - 1 \right) \|A\| + \tau + \ell_{1,+} \right)^2 s_*^2 - \inf_{z \in \mathbb{R}^p} F(z) - \inf_{y \in \mathbb{R}^q} G(y) - \inf_{(x,y) \in \mathbb{R}^m \times \mathbb{R}^q} H(x,y) < +\infty.$$

Since H is coercive and bounded from below, we have that the sequences

$$\{(x_n, y_n)\}_{n \geq 0} \quad \text{and} \quad \left\{ Ax_n - z_n + \frac{1}{\beta} u_n \right\}_{n \geq 0}$$

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 (2.22) and Proposition 1.1, we have that, for any $n \geq 1$,

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

Since F and G are coercive and bounded from below, it follows that the sequences $\{(y_n, z_n)\}_{n \geq 0}$ and $\{Ax_n - z_n + \frac{1}{\beta} u_n\}_{n \geq 0}$ are bounded. Since, 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. ∎

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 2.1 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 2.6 and Theorem 2.10 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 2.11.

(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 Algorithm 2.1 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 that

$$\begin{aligned} &|||(\nabla_x H(x,y) - \nabla_x H(x',y'), \nabla_y H(x,y) - \nabla_y H(x',y'))||| \\ &\qquad \leq L|||(x,y) - (x',y')|||. \end{aligned} \tag{2.23}$$

(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

$$C_2 := \tau - \frac{L\sqrt{2} + \beta \|A\|^2}{2} - \frac{4\sigma\tau^2}{\beta\lambda_{\min}(AA^T)} - \frac{8\left(\sigma\tau + L\sqrt{2}\right)^2}{\sigma\beta\lambda_{\min}(AA^T)},$$
$$C_3 := \frac{\mu - L\sqrt{2}}{2} - \frac{16L^2}{\sigma\beta\lambda_{\min}(AA^T)},$$
$$C_4 := \frac{1}{\sigma\beta}.$$

Remark 2.12. Being facilitated by the boundedness of the generated sequence, Assumption 2.11(iv) not only guarantees the fulfilment of Assumption 2.3(iii) and (iv) on a convex bounded set, but it also arises in a more natural way (see also [12]). Assumption 2.11(iv) holds, for instance, if H is twice continuously differentiable. In addition, as (2.23) implies for any $(x, y), (x', y') \in B_1 \times B_2$ that

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

we can take

$$\ell_{1,+} = \ell_{2,+} = \ell_{3,+} := L\sqrt{2}. \tag{2.24}$$

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

$$\Psi_{n+1} + C_2 \|x_{n+1} - x_n\|^2 + C_3 \|y_{n+1} - y_n\|^2 + C_4 \|u_{n+1} - u_n\|^2 \leq \Psi_n \ \forall n \geq 1 \tag{2.25}$$

remains true for constants $C_2,\ C_3,\ C_4$ taken as in Lemma 2.5 and by taking into consideration (2.24). A possible choice of the parameters of the algorithm such that $\min\{C_2, C_3, C_4\} > 0$ can also be obtained from Lemma 2.6.

The next result provides 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 2.13. *Let Assumption* 2.11 *be satisfied and* $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ *be a sequence generated by Algorithm* 2.1. *Then for any* $n \geq 1$ *it holds that*

$$D_n := \left(d_x^n, d_y^n, d_z^n, d_u^n, d_{x'}^n, d_{u'}^n\right) \in \partial\Psi(\mathbf{X}_n), \tag{2.26}$$

where

$$\begin{aligned} d_x^n &:= \nabla_x H(x_n, y_n) + A^T u_n + \beta A^T (A x_n - z_n) + 2C_1 (x_n - x_{n-1}) \\ &\quad + 2\sigma C_0 B^T \left(A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})\right), \end{aligned} \tag{2.27a}$$
$$d_y^n := \nabla_y H(x_n, y_n) - \nabla_y H(x_{n-1}, y_{n-1}) + \mu (y_{n-1} - y_n), \tag{2.27b}$$
$$d_z^n := u_{n-1} - u_n + \beta A (x_{n-1} - x_n), \tag{2.27c}$$
$$d_u^n := A x_n - z_n + 2C_0 A \left(A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})\right), \tag{2.27d}$$
$$d_{x'}^n := -2\sigma C_0 B^T \left(A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})\right) - 2C_1 (x_n - x_{n-1}), \tag{2.27e}$$
$$d_{u'}^n := -2C_0 A \left(A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})\right). \tag{2.27f}$$

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

$$|||D_n||| \leq C_5 \|x_n - x_{n-1}\| + C_6 \|y_n - y_{n-1}\| + C_7 \|u_n - u_{n-1}\|, \tag{2.28}$$

where

$$C_5 := 2\sqrt{2} \cdot L + \tau + \beta \|A\| + 4 (\sigma\tau + \|A\|) \sigma\tau C_0 + 4C_1,$$
$$C_6 := L\sqrt{2} + \mu,$$
$$C_7 := 1 + \frac{1}{\sigma\beta} + \left(\frac{2}{\sigma} - 1\right) \|A\| + 4 (\sigma\tau + \|A\|) C_0 \|A\| .$$

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

$$\begin{aligned} \nabla_x \Psi (\mathbf{X}_n) = {} & \nabla_x H (x_n, y_n) + A^T u_n + \beta A^T (Ax_n - z_n) + 2C_1 (x_n - x_{n-1}) \\ & + 2\sigma C_0 B^T \left(A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})\right), \end{aligned} \tag{2.29a}$$

$$\partial_y \Psi (\mathbf{X}_n) = \partial G (y_n) + \nabla_y H (x_n, y_n), \tag{2.29b}$$

$$\partial_z \Psi (\mathbf{X}_n) = \partial F (z_n) - u_n - \beta (Ax_n - z_n), \tag{2.29c}$$

$$\nabla_u \Psi (\mathbf{X}_n) = Ax_n - z_n + 2C_0 A \left(A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})\right), \tag{2.29d}$$

$$\begin{aligned} \nabla_{x'} \Psi (\mathbf{X}_n) = {} & -2\sigma C_0 B^T \left(A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})\right) \\ & - 2C_1 (x_n - x_{n-1}), \end{aligned} \tag{2.29e}$$

$$\nabla_{u'} \Psi (\mathbf{X}_n) = -2C_0 A \left(A^T (u_n - u_{n-1}) + \sigma B (x_n - x_{n-1})\right). \tag{2.29f}$$

Then (2.27a) and (2.27d)–(2.27f) follow directly from (2.29a) and (2.29d)–(2.29f), respectively. By combining (2.29b) with the optimality criterion for (2.1a),

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

we obtain (2.27b). Similarly, by combining (2.29c) with the optimality criterion for (2.1b),

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

we get (2.27c).

In what follows we will derive the upper estimates for the components of the limiting subgradient. From (2.21) it follows that

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

In addition, we have

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

Inequality (2.28) follows by combining the above relations with (1.3). ∎

We denote by $\Omega := \Omega(\{\mathbf{X}_n\}_{n\geq 1})$ the set of cluster points of $\{\mathbf{X}_n\}_{n\geq 1} \subseteq \mathbf{R}$, 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 $\operatorname{dist}(\mathbf{X}, \Omega) := \inf\{|||\mathbf{X} - \mathbf{Y}|||\colon \mathbf{Y} \in \Omega\}$. The main result of this section is as follows.

THEOREM 2.14. *Let Assumption* 2.11 *be satisfied and* $\{(x_n, y_n, z_n, u_n)\}_{n\geq 0}$ *be a sequence generated by Algorithm* 2.1. *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* $\{(x_n, y_n, z_n, u_n)\}_{n\geq 0}$ *that converges to* (x_*, y_*, z_*, u_*) *as* $k \to +\infty$*, then*

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

(ii) *it holds that*

$$\begin{aligned}\Omega \subseteq \operatorname{crit}(\Psi) \subseteq \{\mathbf{X}_* \in \mathbf{R} : -A^T u_* = \nabla_x H\left(x_*, y_*\right),\\ 0 \in \partial G\left(y_*\right) + \nabla_y H\left(x_*, y_*\right),\ u_* \in \partial F\left(z_*\right),\ z_* = Ax_*\},\end{aligned} \tag{2.30}$$

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

(iii) *it holds that* $\lim_{n\to+\infty} \operatorname{dist}(\mathbf{X}_n, \Omega) = 0$*;*

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

(v) *the function* Ψ *takes on* Ω *the value*

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

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 $k \geq 1$,

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

and

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

respectively. From (2.1d) and Theorem 2.8, we have $Ax^* = z^*$. Taking the limit superior as $k \to +\infty$ on both sides of the above inequalities, we get

$$\limsup_{k\to+\infty} F\left(z_{n_k}\right) \leq F\left(z_*\right) \qquad \text{and} \qquad \limsup_{k\to+\infty} G\left(y_{n_k}\right) \leq G\left(y_*\right),$$

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

$$\lim_{k\to+\infty} F\left(z_{n_k}\right) = F\left(z_*\right) \qquad \text{and} \qquad \lim_{k\to+\infty} G\left(y_{n_k}\right) = G\left(y_*\right).$$

The desired statement follows thanks to the continuity of H.

(ii) For the sequence $\{D_n\}_{n\geq 0}$ defined in (2.26)–(2.27), we have that $D_{n_k} \in \partial\Psi(\mathbf{X}_{n_k})$ for any $k \geq 1$ and $D_{n_k} \to 0$ as $k \to +\infty$, while $\mathbf{X}_{n_k} \to \mathbf{X}_*$ and $\Psi_{n_k} \to \Psi(\mathbf{X}_*)$ as $k \to +\infty$. The closedness criterion of the limiting subdifferential guarantees that $0 \in \partial\Psi(\mathbf{X}_*)$, or, in other words, $\mathbf{X}_* \in \operatorname{crit}(\Psi)$.

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

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

which is furthermore equivalent to (2.30).

(iii), (iv) The proof follows along the lines of the proof of Theorem 5(ii) and (iii) in [12] by also taking into consideration [12, Remark 5], according to which the properties in (iii) and (iv) are generic for sequences satisfying $\mathbf{X}_n - \mathbf{X}_{n-1} \to 0$ as $n \to +\infty$, which is indeed the case due to (2.18).

(v) Due to (2.18) and the fact that $\{u_n\}_{n\geq 0}$ is bounded, the sequences $\{\Psi_n\}_{n\geq 0}$ and $\{F(z_n) + G(y_n) + H(x_n, y_n)\}_{n\geq 0}$ have the same limit:

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

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

Remark 2.15. An element (x_*, y_*, z_*, u_*) fulfilling (2.30) is a so-called KKT point of the optimization problem (1.1). Such a KKT point obviously fulfils

$$0 \in A^T \partial F\left(Ax_*\right) + \nabla_x H\left(x_*, y_*\right), \qquad 0 \in \partial G\left(y_*\right) + \nabla_y H\left(x_*, y_*\right). \tag{2.31}$$

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

$$\begin{aligned} &0 \in \partial\left(F \circ A\right)\left(x_*\right) + \nabla_x H\left(x_*, y_*\right) = \partial_x\left(F \circ A + H\right), \\ &0 \in \partial G\left(y_*\right) + \nabla_y H\left(x_*, y_*\right) = \partial_y\left(G + H\right); \end{aligned} \tag{2.32}$$

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, (2.31) and (2.32) 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 2.1 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 [19] a general form of the Łojasiewicz inequality [20]. An extension to the nonsmooth setting has been proposed and studied in [9, 10, 11].

Definition 3.1. *Let $\eta \in (0, +\infty]$. We denote by Φ_η the set of all concave and continuous functions $\varphi\colon [0, \eta) \to [0, +\infty)$ which satisfy the following conditions:*

(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 3.2. *Let $\Psi\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be proper and lower semicontinuous.*

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

$$v \in V \cap \left\{v \in \mathbb{R}^d : \Psi(\widehat{v}) < \Psi(v) < \Psi(\widehat{v}) + \eta\right\}$$

the following inequality holds:

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

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

The functions φ belonging to the set Φ_η for $\eta \in (0, +\infty]$ are called desingularization functions. The KŁ property reveals the possibility of reparametrizing the values of Ψ in order to avoid flatness around the critical points. To the class of KŁ functions belong semialgebraic, real subanalytic, uniformly convex functions and convex functions satisfying a growth condition. We refer the reader to [1, 2, 3, 9, 10, 11, 12] for more properties of KŁ functions and illustrative examples.

The following result, the proof of which can be found in [12, Lemma 6], will play an essential role in our convergence analysis.

Lemma 3.3 (uniformized KŁ property). *Let Ω be a compact set and $\Psi\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. Assume that Ψ is constant on Ω and satisfies the KŁ property at each point of Ω. Then there exist $\varepsilon > 0$, $\eta > 0$, and $\varphi \in \Phi_\eta$ such that for any $\widehat{v} \in \Omega$ and every element u in the intersection*

$$\left\{v \in \mathbb{R}^d\colon \mathrm{dist}(v, \Omega) < \varepsilon\right\} \cap \left\{v \in \mathbb{R}^d : \Psi(\widehat{v}) < \Psi(v) < \Psi(\widehat{v}) + \eta\right\}$$

it holds that

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

From now on we will use the notation

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

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

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

Theorem 3.4. *Let Assumption* 2.11 *be satisfied and $\{(x_n, y_n, z_n, u_n)\}_{n\geq 0}$ be a sequence generated by Algorithm* 2.1*. If Ψ is a KŁ 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,*

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

(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 $\mathbf{X}_* \in \Omega$; thus $\Psi(\mathbf{X}_*) = \Psi_*$. Recall that $\{\mathcal{E}_n\}_{n\geq 1}$ is monotonically decreasing and converges to 0 as $n \to +\infty$. 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 $n \geq n'$. Inequality (2.25) yields, for any $n \geq n' + 1$,

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

Inequality (2.19) gives us, furthermore, that $z_{n+1} - z_n = 0$ 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 3.3, there exist $\varepsilon > 0$, $\eta > 0$, and a desingularization function φ such that for any element $\mathbf{X}$ in the intersection

$$(3.2) \qquad \{\mathbf{Z} \in \mathbf{R} \colon \operatorname{dist}(\mathbf{Z}, \Omega) < \varepsilon\} \cap \{\mathbf{Z} \in \mathbf{R} \colon \Psi_* < \Psi(\mathbf{Z}) < \Psi_* + \eta\}$$

it holds that

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

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

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

Since $\lim_{n\to+\infty} \operatorname{dist}(\mathbf{X}_n, \Omega) = 0$ (see Lemma 2.14 (iii)), there exists $n_2 \geq 1$ such that, for any $n \geq n_2$,

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

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

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

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 that

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

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

By combining Lemma 2.5 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 $n \geq n_0$,

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

By the Cauchy–Schwarz inequality, the arithmetic-mean–geometric-mean inequality and Lemma 2.13, we have that, for any $n \geq n_0$ and every $\alpha > 0$,

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

If we define, for any $n \geq 0$,

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

then the above inequality is none other than (1.6) with

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

Since $\sum_{n\geq 1} b_n < +\infty$, by choosing $\alpha < 1/C_9$, we can apply Lemma 1.3 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 once again taking into account (2.19).

From (i) it follows that the sequence $\{(x_n, y_n, z_n, u_n)\}_{n\geq 0}$ is Cauchy, and thus it converges to an element (x_*, y_*, z_*, u_*) which is, according to Lemma 2.14, a KKT point of the optimization problem (1.1). ∎

3.2. Convergence rates. In what follows we derive convergence rates for the sequence $\{(x_n, y_n, z_n, u_n)\}_{n\geq 0}$ generated by Algorithm 2.1 and for $\{\Psi_n\}_{n\geq 0}$, provided that the regularized augmented Lagrangian Ψ satisfies the Łojasiewicz property. The following definition is from [1] (see also [20]).

DEFINITION 3.5. *Let $\Psi\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be proper and lower semicontinuous. Then Ψ satisfies the Łojasiewicz property if for any critical point $\widehat{v}$ of Ψ there exists $C_L > 0$, $\theta \in [0, 1)$, and $\varepsilon > 0$ such that*

$$|\Psi(v) - \Psi(\widehat{v})|^\theta \leq C_L \cdot \operatorname{dist}(0, \partial\Psi(v)) \;\; \forall v \in \operatorname{Ball}(\widehat{v}, \varepsilon),$$

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

If Assumption 2.11 is fulfilled and $\{(x_n, y_n, z_n, u_n)\}_{n\geq 0}$ is the sequence generated by Algorithm 2.1, then, according to Theorem 2.14, 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

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

it holds that

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

Obviously, Ψ is a KŁ function with desingularization function

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

which, according to Theorem 3.4, means that Ω contains a single element $\mathbf{X}_*$, which is the limit of $\{\mathbf{X}_n\}_{n\geq 1}$ as $n \to +\infty$. In other words, if Ψ has the Łojasiewicz property, then there exist $C_L > 0$, $\theta \in [0, 1)$, and $\varepsilon > 0$ such that, for any $\mathbf{X} \in \text{Ball}(\mathbf{X}_*, \varepsilon)$,

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

In this case, Ψ is said to satisfy the Łojasiewicz property with Łojasiewicz constant $C_L > 0$ and Łojasiewicz 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 [8, Lemma 15].

LEMMA 3.6. *Let $\{e_n\}_{n\geq 0}$ be a monotonically decreasing sequence of nonnegative numbers converging to* 0*. Assume further that there exist 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 3.7. *Let Assumption* 2.11 *be satisfied and $\{(x_n, y_n, z_n, u_n)\}_{n\geq 0}$ be a sequence generated by Algorithm* 2.1*. If Ψ satisfies the Łojasiewicz property with Łojasiewicz constant $C_L > 0$ and Łojasiewicz exponent $\theta \in [0, 1)$, then there exists $n_0 \geq 1$ such that the following estimate holds for any $n \geq n_0$:*

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

Proof. For every $n \geq 2$ we obtain from Lemma 2.5

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

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 $n \geq n_0$, X_n belongs to $\operatorname{Ball}(\mathbf{X}_*, \varepsilon)$. Relation (3.6) implies (3.7) for any $n \geq n_0$. ∎

The following result follows by combining Lemma 3.6 with Lemma 3.7.

THEOREM 3.8. *Let Assumption* 2.11 *be satisfied and* $\{(x_n, y_n, z_n, u_n)\}_{n\geq 0}$ *be a sequence generated by Algorithm* 2.1. *If* Ψ *satisfies the Łojasiewicz property with Łojasiewicz constant* $C_L > 0$ *and Łojasiewicz exponent* $\theta \in [0, 1)$, *then the following statements are true:*

(i) *if* $\theta = 0$, *then* $\{\Psi_n\}_{n\geq 1}$ *converges in finite time;*

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

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

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

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

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 3.9. *Let Assumption* 2.11 *be satisfied and* $\{(x_n, y_n, z_n, u_n)\}_{n\geq 0}$ *be a sequence generated by Algorithm* 2.1. *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* $n \to +\infty$. *Then there exists* $n_0 \geq 1$ *such that the following estimates hold for any* $n \geq n_0$:

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

where

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

Proof. We assume that $\mathcal{E}_n > 0$ for any $n \geq 0$. Otherwise, $\{(x_n, y_n, z_n, u_n)\}_{n\geq 0}$ becomes identical to (x_*, y_*, z_*, u_*) beginning with a given index and the conclusion follows automatically (see the proof of Theorem 3.4).

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

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

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

Similarly, we derive

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

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 now fix an integer $N \geq n$. Summing up the above inequality for $k = n, \ldots, N$, 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 letting $N \to +\infty$, we obtain

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

which gives the desired statement. ∎

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

THEOREM 3.10. *Let Assumption* 2.11 *be satisfied and* $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ *be a sequence generated by Algorithm* 2.1. *Suppose further that* Ψ *satisfies the Łojasiewicz property with Łojasiewicz constant* $C_L > 0$ *and Łojasiewicz exponent* $\theta \in [0, 1)$. *Let* (x_*, y_*, z_*, u_*) *be the KKT point of the optimization problem* (1.1) *to which the sequence* $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ *converges as* $n \to +\infty$. *Then the following statements are true:*

(i) *if* $\theta = 0$, *then the algorithm converges in finite time;*

(ii) *if* $\theta \in (0, 1/2]$, *then there exist* $n_0 \geq 1$, $\widehat{C}_{0,1}, \widehat{C}_{0,2}, \widehat{C}_{0,3}, \widehat{C}_{0,4} > 0$, *and* $\widehat{Q} \in [0, 1)$ *such that, for any* $n \geq n_0$,

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

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

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

Proof. Let

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

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 3.4, 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 $\{(x_n, y_n, z_n, u_n)\}_{n \geq 0}$ also converges in finite time and the conclusion follows.

Let $\theta \neq \frac{1}{2}$ and $n_0' \geq 1$ be such that for any $n \geq n_0'$ the inequalities (3.8) in Lemma 3.9 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

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

and 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 3.8, there exist $n_0'' \geq 1$, $\widehat{C}_0 > 0$, and $Q \in [0, 1)$ such that for $\widehat{Q} := \sqrt{Q}$ and every $n \geq n_0''$ it holds that

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

The conclusion follows from Lemma 3.9 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

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

By Theorem 3.8, there exist $n_0'' \geq 1$ and $\widehat{C}_1 > 0$ such that, for any $n \geq n_0''$,

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

The conclusion follows again for $n_0 := \max\{n_0', n_0''\}$ from Lemma 3.9. ∎

Acknowledgment. The authors are thankful to two anonymous referees for their comments and recommendations, which improved the quality of the paper.

REFERENCES

[1] H. Attouch and J. Bolte, *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*, Math. Program., 116 (2009), pp. 5–16.
[2] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka–Łojasiewicz inequality*, Math. Oper. Res., 35 (2010), pp. 438–457.
[3] H. Attouch, J. Bolte, and B. F. Svaiter, *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss–Seidel methods*, Math. Program., 137 (2013), pp. 91–129.
[4] H. Bauschke and P. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 1st ed., CMS Books Math., Springer, New York, 2011.
[5] A. Beck, *First-Order Methods in Optimization*, 1st ed., MOS-SIAM Ser. Optim., SIAM, Philadelphia, PA, 2017.
[6] R. I. Boţ and E. R. Csetnek, *An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems*, J. Optim. Theory Appl., 171 (2016), pp. 600–616.
[7] R. I. Boţ, E. R. Csetnek, and S. C. László, *An inertial forward-backward algorithm for the minimization of the sum of two nonconvex functions*, EURO J. Comput. Optim., 4 (2016), pp. 3–25.
[8] R. I. Boţ and D. K. Nguyen, *The Proximal Alternating Direction Method of Multipliers in the Nonconvex Setting: Convergence Analysis and Rates*, preprint, https://arxiv.org/abs/1801.01994, 2018.
[9] J. Bolte, A. Daniilidis, and A. Lewis, *The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*, SIAM J. Optim., 17 (2006), pp. 1205–1223.
[10] J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota, *Clarke subgradients of stratifiable functions*, SIAM J. Optim., 18 (2007), pp. 556–572.
[11] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet, *Characterizations of Łojasiewicz inequalities: Subgradient flows, talweg, convexity*, Trans. Amer. Math. Soc., 362 (2010), pp. 3319–3363.
[12] J. Bolte, S. Sabach, and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program., 146 (2014), pp. 459–494.
[13] J. Bolte, S. Sabach, and M. Teboulle, *Nonconvex Lagrangian-based optimization: Monitoring schemes and global convergence*, Math. Oper. Res., 43 (2018), pp. 1210–1232.
[14] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Model. Simul., 4 (2005), pp. 1168–1200.
[15] W. Hare and C. Sagastizábal, *Computing proximal points of nonconvex functions*, Math. Program., 116 (2009), pp. 221–258.
[16] M. Hong and Z. Q. Luo, *On the linear convergence of the alternating direction method of multipliers*, Math. Program., 162 (2017), pp. 165–199.

[17] M. Hong, Z. Q. Luo, and M. Razaviyayn, *Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems*, SIAM J. Optim., 26 (2016), pp. 337–364.

[18] B. Jiang, T. Lin, S. Ma, and S. Zhang, *Structured nonconvex and nonsmooth optimization: Algorithms and iteration complexity analysis*, Comput. Optim. Appl., 72 (2019), pp. 115–157.

[19] K. Kurdyka, *On gradients of functions definable in o-minimal structures*, Ann. Inst. Fourier (Grenoble), 48 (1998), pp. 769–783.

[20] S. Łojasiewicz, *Une propriété topologique des sous-ensembles analytiques réels*, in Les Équations aux Dérivées Partielles, Éditions du Centre National de la Recherche Scientifique, Paris, 1963, pp. 87–89.

[21] A. Lewis and J. Malick, *Alternating projection on manifolds*, Math. Oper. Res., 33 (2008), pp. 216–234.

[22] G. Li and T. K. Pong, *Global convergence of splitting methods for nonconvex composite optimization*, SIAM J. Optim., 25 (2015), pp. 2434–2460.

[23] Q. Liu, X. Shen, and Y. Gu, *Linearized ADMM for Non-convex Non-smooth Optimization with Convergence Analysis*, preprint, https://arxiv.org/abs/1705.02502, 2017.

[24] B. Mordukhovich, *Variational Analysis and Generalized Differentiation,* I: *Basic Theory*, Grundlehren Math. Wiss. 330, Springer, Berlin, 2006.

[25] J. Moreau, *Fonctions convexes duales et points proximaux dans un espace hilbertien*, C. R. Acad. Sci. Sér. A, 255 (1962), pp. 2897–2899.

[26] T. Pock and S. Sabach, *Inertial proximal alternating linearized minimization (iPALM) for nonconvex and nonsmooth problems*, SIAM J. Imaging Sci., 9 (2016), pp. 1756–1787.

[27] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, Springer, Berlin, 1998.

[28] F. Wang, W. Cao, and Z. Xu, *Convergence of multi-block Bregman ADMM for nonconvex composite problems*, Sci. China Inf. Sci., 61 (2018), pp. 1–12.

[29] Y. Wang, W. Yin, and J. Zeng, *Global convergence of ADMM in nonconvex nonsmooth optimization*, J. Sci. Comput., 78 (2019), pp. 29–63.

[30] 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 J. Imaging Sci., 10 (2017), pp. 74–110.