Xue and Ma *Journal of Inequalities and Applications* (2024) 2024:124
https://doi.org/10.1186/s13660-024-03206-1

Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH

Open Access

A modified inertial proximal minimization algorithm for structured nonconvex and nonsmooth problem

Zhonghui Xue[1*] and Qianfeng Ma[1]

*Correspondence:
hnlgxzh@163.com
[1]Shanghai Publishing and Printing College, Shanghai, 200093, China

Abstract

We introduce an enhanced inertial proximal minimization algorithm tailored for a category of structured nonconvex and nonsmooth optimization problems. The objective function in question is an aggregation of a smooth function with an associated linear operator, a nonsmooth function dependent on an independent variable, and a mixed function involving two variables. Throughout the iterative procedure, parameters are selected employing a straightforward approach, and weak inertial terms are incorporated into two subproblems within the update sequence. Under a set of lenient conditions, we demonstrate that the sequence engendered by our algorithm is bounded. Furthermore, we establish the global and strong convergence of the algorithmic sequence, contingent upon the assumption that the principal function adheres to the Kurdyka–Łojasiewicz (KL) property. Ultimately, the numerical outcomes corroborate the algorithm's feasibility and efficacy.

Mathematics Subject Classification: 49J53; 49K99

Keywords: Weak inertial; Proximal minimization algorithm; Nonconvex-nonsmooth optimization; Kurdyka–Łojasiewicz property

1 Introduction

In this paper, we consider the following nonconvex and nonsmooth problem:

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

where the function $F:\mathbb{R}^p \to \mathbb{R}$ is continuously Lipschitz differentiable, $G:\mathbb{R}^q \to \mathbb{R}\cup\{+\infty\}$ is a proper lower semicontinuous function, $H:\mathbb{R}^m\times\mathbb{R}^q \to \mathbb{R}$ is a Fréchet differentiable function with Lipschitz continuous gradient, and $A:\mathbb{R}^m \to \mathbb{R}^p$ is a linear operator. Many application problems can be modeled as (1.1), e.g., compressed sensing [2, 14], matrix factorization [5], sparse approximations of signals and images [22, 27], and so on.

Obviously, when $m = p$ and A is the identity operator, (1.1) can be written as

$$\min_{(x,y)\in\mathbb{R}^m\times\mathbb{R}^q} \{F(x) + G(y) + H(x,y)\}. \tag{1.2}$$

Utilizing the two-block structure, a natural method to solve (1.2) is the alternating minimization method. For a given initial point $(x^0, y^0) \in \mathbb{R}^m \times \mathbb{R}^q$, it generates the iterative sequence $\{(x^k, y^k)\}$ by the following scheme:

$$x^{k+1} \in \arg\min \left\{ F(x) + G(y^k) + H(x, y^k) : x \in \mathbb{R}^m \right\},$$
$$y^{k+1} \in \arg\min \left\{ F(x^{k+1}) + G(y) + H(x^{k+1}, y) : y \in \mathbb{R}^q \right\}.$$

The method is also called the Gauss–Seidel method or the block coordinate descent method, and its convergence results can be found in [3, 23, 29]. However, the convergence of the above methods is in the setting of convex case. In the nonconvex setting, the situation becomes much harder. Bolte et al. [5] considered a proximal alternating linearized minimization (PALM) algorithm for solving problem (1.2) in the nonconvex and nonsmooth case, which has the following form:

$$x^{k+1} \in \arg\min \left\{ F(x) + \left\langle \nabla_x H\left(x^k, y^k\right), x - x^k \right\rangle + \frac{c_k}{2} \left\| x - x^k \right\|^2 : x \in \mathbb{R}^m \right\},$$
$$y^{k+1} \in \arg\min \left\{ G(y) + \left\langle \nabla_y H\left(x^{k+1}, y^k\right) y - y^k \right\rangle + \frac{d_k}{2} \left\| y - y^k \right\|^2 : y \in \mathbb{R}^q \right\},$$

where c_k and d_k are positive real numbers. They proved the global convergence under the assumption that the augmented Lagrangian function satisfies the Kurdyka–Łojasiewicz property. Driggs et al. [15] proposed a generic stochastic version of PALM algorithm for nonsmooth nonconvex optimization problem, where various variance-reduced gradient approximations were allowed. PALM can be considered as a blockwise application of the well-known proximal forward–backward algorithm [13, 20] in the nonconvex setting. In [9], Bot et al. chose a continuous forward–backward method and introduced a dynamical system consisting of partial gradients of smooth coupling functions and proximal point operators of two nonsmooth functions, and Attouch et al. [1] proposed an alternating proximal minimization algorithm for nonconvex structured problem (1.2).

When $G(y) = 0$ and $H(x, y) = H(x)$ for all $(x, y) \in \mathbb{R}^m \times \mathbb{R}^q$, problem (1.1) is translated into the following problem:

$$\min_{x \in \mathbb{R}^m} \{F(Ax) + H(x)\}, \tag{1.3}$$

where $H : \mathbb{R}^m \to \mathbb{R}$ is a Fréchet differentiable function with Lipschitz continuous gradient. Problem (1.3) can be writen as

$$\min_{x \in \mathbb{R}^m} \; F(z) + H(x)$$
$$s.t. \quad Ax = z.$$

In the convex case, the linearized ADMM was adopted to solve this problem in [24, 31, 32] in the following form:

$$z^{k+1} \in \arg\min_z \{F(z) + \langle u^k, Ax^k - z \rangle + \frac{\beta}{2} \left\| Ax^k - z \right\|^2\},$$
$$x^{k+1} = x^k - \frac{1}{\tau} \left(\nabla H\left(x^k\right) + A^T u^k + \beta A^T \left(Ax^k - z^{k+1}\right)\right),$$

$$u^{k+1} = u^k + \sigma\beta\left(Ax^{k+1} - z^{k+1}\right), \tag{1.4}$$

where u is the Lagrangian multiplier, β is the penalty parameter, and $\tau, \sigma > 0$ are step-sizes. Furthermore, the linearized ADMM was applied in the nonconvex case in [10, 21]. Motivated by [15], Bian et al. [4] extended the result to the case of ADMM, which combined ADMM with a class of stochastic gradient with variance reduction. Li et al. [19] examined two types of splitting methods for solving this nonconvex optimization problem (1.3), the alternating direction method of multipliers and the proximal gradient algorithm.

Inertial effect, as an accelerated technique, started from the so-called heavy-ball method of Polyak [26], was very efficient in improving numerical performance of the algorithm. Recently, the research of inertial-type algorithm has attracted more and more attention, such as inertial versions of the ADMM for maximal monotone operator inclusion problem [6], inertial forward–backward–forward method [7] based on Tseng's approach [28], and general inertial proximal point method for the mixed variational inequality problem [12]. Notably, Guo, Zhao, and Dong [17] have proposed a stochastic two-step inertial Bregman proximal alternating linearized minimization algorithm, which presents a significant advancement in the field. Also, Zhao, Dong, Rassias, and Wang [34] have contributed to this area with their two-step inertial Bregman alternating minimization algorithm, enhancing the understanding of nonconvex and nonsmooth optimization problems. Additionally, Zhang and He [33] have introduced an inertial proximal alternating minimization method, which further broadens the scope of application for inertial techniques. Specially, for problem (1.2) in the nonconvex setting, Pock and Sabach [25] proposed an inertial version of PALM (IPALM), and Gao et al. [16] proposed a Gauss–Seidel-type inertial proximal alternating linearized minimization (GIPALM) algorithm. For problem (1.3), Chao et al. [11] combined an inertial technique with ADMM and employed the KL assumption to obtain the global convergence in the nonconvex setting.

For problem (1.1), Hong et al. [18] analyzed the behavior of the alternating direction method of multipliers (ADMM) in the nonconvex case. Wang et al. [30] studied the convergence of the alternating direction method of multipliers (ADMM) for problem (1.1) in the nonconvex and possibly nonsmooth case. Bolt et al. [8] transformed problem (1.1) into a three-block nonseparable problem by introducing a new variable, which has the following form

$$\begin{aligned} \min_{(x,y,z)\in\mathbb{R}^m\times\mathbb{R}^q\times\mathbb{R}^p} \quad & F(z) + G(y) + H(x,y) \\ s.t. \quad & Ax = z. \end{aligned} \tag{1.5}$$

Then the augmented Lagrangian function $L_\beta : \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R} \cup \{+\infty\}$ of problem (1.5) was defined as

$$L_\beta(x,y,z,u) = F(z) + G(y) + H(x,y) + \langle u, Ax - z\rangle + \frac{\beta}{2}\|Ax - z\|^2, \quad \beta > 0, \tag{1.6}$$

where u is the Lagrangian multiplier, and β is the penalty parameter. Bolt [8] gave the following proximal minimization algorithm (PMA) to solve it:

$$y^{k+1} \in \arg\min_{y\in\mathbb{R}^q}\left\{G(y) + \langle\nabla_y H\left(x^k, y^k\right), y\rangle + \frac{\mu}{2}\left\|y - y^k\right\|^2\right\},$$

$$z^{k+1} \in \arg\min_{z\in\mathbb{R}^p} \left\{ F(z) + \langle u^k, Ax^k - z\rangle + \frac{\beta}{2}\left\| Ax^k - z\right\|^2 \right\},$$
$$x^{k+1} := x^k - \tau^{-1}\left(\nabla_x H\left(x^k, y^{k+1}\right) + A^T u^k + \beta A^T\left(Ax^k - z^{k+1}\right)\right),$$
$$u^{k+1} := u^k + \sigma\beta\left(Ax^{k+1} - z^{k+1}\right).$$

Moreover, they provided sufficient conditions for the boundedness of the generated sequence and proved that any cluster point of the latter is a KKT point of the minimization problem. They also showed the global convergence under the Kurdyka–Łojasiewicz property.

Inspired by the above algorithms, in this paper, we propose a weak inertial proximal minimization algorithm for the nonconvex and nonsmooth problem (1.1). The main contributions of the paper are as follows.

• Comparing with [8], inertial effect can effectively improve the convergence. Under the action of the inertial effect, we also show that the sequence generated by the proposed method is bounded and the algorithm is global and strongly convergent under the Kurdyka–Łojasiewicz assumption.

• Comparing with [11], problem (1.1) is different from problem (2) in [11]. Here our problem owns a nonsepable term $H(x,y)$, which leads to big difficulty for showing the convergence.

The paper is organized as follows. In Sect. 2, some useful definitions and results are collected for the convergence analysis of the proposed algorithm. In Sect. 3, we propose a modified inertial proximal minimization algorithm and analyze its convergence. Section 4 tests a numerical experiment to conclude the effectiveness of our algorithm. Finally, some conclusions are drawn in Sect. 5.

2 Notation and preliminaries

In this section, we summarize some basic notations and some conclusions, which will be used in the subsequent analysis.

In the following, $\mathbb{R}^n$ stands for the n-dimensional Euclidean space with

$$\langle x, y\rangle = x^T y = \sum_{i=1}^n x_i y_i, \quad \|x\| = \sqrt{\langle x, x\rangle},$$

where T stands for the transpose operation. For a set $S \subset \mathbb{R}^n$ and a point $x \in \mathbb{R}^n$, let $d(x,S) = \inf_{y\in S}\left\|y - x\right\|^2$. If $S = \emptyset$, then we set $d(x,S) = +\infty$ for all $x \in \mathbb{R}^n$.

Definition 2.1 (Lipschitz differentiability) A function $f(x)$ is said to be L_f Lipschitz differentiable if for all x, y, we have

$$\left\|\nabla f(x) - \nabla f(y)\right\|_2 \le L_f\left\|x - y\right\|_2.$$

Lemma 2.1 [21] (*Descending lemma*) *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be Fréchet differentiable such that its gradient is Lipschitz continuous with constant* $\ell > 0$. *Then the following statements are true*:

(*i*) *For all* $x, y \in \mathbb{R}^n$ *and* $z \in \left[x, y\right] = \left\{(1-t)x + ty : t \in [0,1]\right\}$, *we have*

$$f\left(y\right) \le f\left(x\right) + \left\langle \nabla f\left(z\right), y - x\right\rangle + \frac{\ell}{2}\left\|y - x\right\|^2;$$

(*ii*) *For all* $\gamma \in \mathbb{R}\backslash\{0\}$, *we have*

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

Remark 2.1 The descending lemma can be written as follows:

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

which follows from (i) by taking $z := x$. In addition, by taking $z := y$ in (i) we obtain

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

Lemma 2.2 [25] *Let* $\{a_n\}_{n\geq 0}$ *be a sequence of real numbers bounded from below, and let* $\{b_n\}_{n\geq 0}$ *be a sequence of real nonnegative numbers. Assume that for all* $n \geq 0$,

$$a_{n+1}+b_n\leq a_n.$$

Then the following statements hold:

(*i*) *The sequence* $\{a_n\}_{n\geq 0}$ *is monotonically decreasing and convergent*;

(*ii*) *The sequence* $\{b_n\}_{n\geq 0}$ *is summable, that is,* $\sum_{n\geq 0} b_n < \infty$.

Lemma 2.3 *Let* $\{a_n\}_{n\in N}$ *and* $\{b_n\}_{n\in N}$ *be nonnegative real sequences such that* $\sum_{n\in N} b_n < \infty$ *and* $a_{n+1} \leq a \cdot a_n + b \cdot a_{n-1} + b_n$ *for all* $n \geq 1$, *where* $a \in \mathbb{R}$, $b \geq 0$, *and* $a + b < 1$. *Then* $\sum_{n\in N} a_n < \infty$.

We now introduce a function satisfying the Kurdyka–Łojasiewicz property. This class of functions will play a crucial role in the convergence results of the proposed algorithm.

Definition 2.2 Let $\eta \in (0, +\infty]$. We denote by Φ_η the set of all concave continuous functions $\varphi : [0, \eta) \to [0, +\infty)$. A function φ belonging to the set Φ_η for $\eta \in (0, +\infty]$ is called a desingularization function if it satisfies the following conditions:

(i) $\varphi(0) = 0$;

(ii) φ is continuously differentiable on $(0, \eta)$ and continuous at 0;

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

The KŁ property reveals the possibility of reparameterizing the values of the function to avoid flatness around the critical points. To the class of KŁ functions there belong semi-algebraic, real subanalytic, uniformly convex functions, and convex functions satisfying a growth condition.

Definition 2.3 [5] (Kurdyka–Łojasiewicz property) Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. The function f is said to have the Kundyka–Łojasiewicz (KŁ) property at a point $\hat{v} \in \operatorname{dom}\partial f := \{v\in\mathbb{R}^n : \partial f(v)\neq\emptyset\}$ if there exist $\eta \in (0, +\infty]$, a neighborhood V of $\hat{v}$, and a function $\varphi \in f_\eta$ such that

$$\varphi'(f(v)-f(\hat{v}))\cdot \operatorname{dist}(\mathbf{0},\partial f(v))\geq 1$$

for all

$$v \in V \cap \left\{ v \in \mathbb{R}^n : f(\hat{v}) < f(v) < f(\hat{v}) + \eta \right\}.$$

If f satisfies the KŁ property at each point of $\mathrm{dom}\partial f$, then f is called a KŁ function. Next, we recall the following result, which is called the uniformized KŁ property.

Lemma 2.4 (*Uniformized KŁ property*) *Let Ω be a compact set, and let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. Assume that f is constant on Ω and satisfies the KŁ property at each point of Ω. Then there exist $\varepsilon > 0$, $\eta > 0$, and $\varphi \in f_\eta$ such that*

$$\varphi'(f(v) - f(\hat{v})) \cdot \mathrm{dist}(\mathbf{0}, \partial f(v)) \geq 1$$

for allllny $\hat{v} \in \Omega$ and every element v in the intersection

$$\left\{ v \in \mathbb{R}^n : \mathrm{dist}(v, \Omega) < \varepsilon \right\} \cap \left\{ v \in \mathbb{R}^d : f(\hat{v}) < f(v) < f(\hat{v}) + \eta \right\}.$$

Definition 2.4 (Coercivity) A function $\psi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called coercive if

$$\lim_{\|x\| \to +\infty} \psi(x) = +\infty.$$

Definition 2.5 [5] (Subdifferentials) Let $\sigma : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper lower semicontinuous function.

(i) For a given $x \in \mathrm{dom}\,\sigma$, the Fréchet subdifferential of σ at x, written $\widehat{\partial}\sigma(x)$, is the set of all vectors $u \in \mathbb{R}^n$ that satisfy

$$\liminf_{y \neq x} \frac{\sigma(y) - \sigma(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0.$$

When $x \notin \mathrm{dom}\,\sigma$, we set $\widehat{\partial}\sigma(x) = \emptyset$.

(ii) The limiting subdifferential, or simply the subdifferential, of σ at $x \in \mathbb{R}^n$, written $\partial\sigma(x)$, is defined through the following closure process $\partial\sigma(x) := \{u \in \mathbb{R}^n : \exists x^k \to x, \sigma\left(x^k\right) \to \sigma(x) \text{ and } u^k \in \widehat{\partial}\sigma\left(x^k\right) \to u \text{ as } k \to \infty\}$.

3 Algorithm and its convergence

In this section, we propose a weak inertial proximal minimization algorithm for solving the optimization problem (1.1) and study the convergence behavior of the algorithm.

Algorithm 3.1 Modified Inertial Proximal Minimization Algorithm (MIPMA)

Let β, $\tau > 0$, $0 < \theta_k < 1$, and $\mu > 0$. For the starting points $(x^0, y^0, z^0) \in \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^p$, $u^0 \in \mathbb{R}^p$, $(x^{-1}, y^{-1}) \in \mathbb{R}^m \times \mathbb{R}^q$, the sequence $\left\{\left(x^k, y^k, z^k, u^k\right)\right\}_{k\geq 0}$ for any $k \geq 0$ is generated by

$$y^{k+1} \in \arg\min_{y\in\mathbb{R}^q} \left\{ G(y) + \left\langle \nabla_y H\left(x^k, y^k\right), y \right\rangle + \frac{\mu}{2}\left\| y - y^k \right\|^2 + \theta\left\langle y, y^k - y^{k-1} \right\rangle \right\}, \tag{3.1a}$$

$$z^{k+1} \in \arg\min_{z\in\mathbb{R}^p} \left\{ F(z) + \left\langle u^k, Ax^k - z \right\rangle + \frac{\beta}{2}\left\| Ax^k - z \right\|^2 + \frac{\mu}{2}\left\| z - z^k \right\|^2 \right\}, \tag{3.1b}$$

$$x^{k+1}:=x^k-\tau^{-1}\left(\nabla_x H\left(x^k,y^{k+1}\right)+A^T u^k+\beta A^T\left(Ax^k-z^{k+1}\right)+\theta\left(x^k-x^{k-1}\right)\right), \tag{3.1c}$$

$$u^{k+1}:=u^k+\beta\left(Ax^{k+1}-z^{k+1}\right). \tag{3.1d}$$

Remark 3.1 Based on the algorithm in [8], inertial terms $\theta\langle y, y^k-y^{k-1}\rangle$ and $\theta\left(x^k-x^{k-1}\right)$ are added into the y-subproblem and the x-subproblem, respectively, and a regular term $\frac{\mu}{2}\|z-z^k\|^2$ is admitted to z-subproblem.

The following assumptions are important for the convergence analysis.

Assumption A (i) The functions F, G, and H are bounded from below.
(ii) F is Lipschitz differentiable, i.e.,

$$\left\|\nabla F(z)-\nabla F(z')\right\|^2\le l_F^2\left\|z-z'\right\|^2.$$

(iii) The Function H is L_h Lipschitz differentiable, and ∇H is $L_H(x,y)$ Lipschitz continuous, i.e.,

$$\left\|\nabla_y H\left(x,y\right)-\nabla_y H\left(x',y'\right)\right\|^2\le L_H^2(x,y)\left(\left\|x-x'\right\|^2+\left\|y-y'\right\|^2\right).$$

For any fixed $y\in\mathbb{R}^q$, there exists $\ell_1(y)\ge 0$ such that

$$\left\|\nabla_x H(x,y)-\nabla_x H\left(x',y\right)\right\|\le \ell_1(y)\left\|x-x'\right\|,\quad \forall x,x'\in\mathbb{R}^m,$$

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

$$\left\|\nabla_y H(x,y)-\nabla_y H\left(x,y'\right)\right\|\le \ell_2(x)\left\|y-y'\right\|\quad \forall y,y'\in\mathbb{R}^q.$$

Furthermore, there exist $\ell_{1,+}>0$, $\ell_{2,+}>0$, ℓ_h such that

$$\sup_{y\in\mathbb{R}^q}\ell_1(y)\le\ell_{1,+},\quad \sup_{x\in\mathbb{R}^m}\ell_2(x)\le\ell_{2,+},\quad \sup_{(x,y)\in\mathbb{R}^m\times\mathbb{R}^q}L_H(x,y)\le\ell_h.$$

(iv)We assume that

$$\mu>\ell_{2,+}+2\theta,\beta>\max\left\{\frac{10l_F^2+20\mu^2}{\mu},3l_F\right\},\tau>10\beta\|A\|^2+\frac{\beta\|A\|^2+\ell_{1,+}}{2}+\theta.$$

Remark 3.2 Assumption (i) ensures that the sequence generated by Algorithm 3.1 is well defined. It also has as the consequence that

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

Before the proof, let us present the variational characterization of scheme (3.1a)–(3.1d). By the optimality conditions for (3.1a) and (3.1b) we have

$$0\in\partial G(y^{k+1})+\nabla_y H(x^k,y^k)+\mu(y^{k+1}-y^k)+\theta\left(y^k-y^{k-1}\right), \tag{3.2a}$$

$$0=\nabla F(z^{k+1})-u^k-\beta(Ax^k-z^{k+1})+\mu(z^{k+1}-z^k). \tag{3.2b}$$

By substituting (3.1d) into (3.2b) and rearranging terms we obtain

$$u^{k+1} = \nabla F(z^{k+1}) + \beta(Ax^{k+1} - Ax^k) + \mu(z^{k+1} - z^k). \tag{3.3}$$

The convergence analysis is based on a descent inequality, which will play an important role in our research.

Lemma 3.1 *Suppose that Assumption A holds. Suppose $L_\beta\left(\omega^k\right)$ is defined as (1.6). Then we have*

$$\begin{aligned}
&L_\beta\left(x^{k+1},y^{k+1},z^{k+1},u^{k+1}\right) + \frac{5\mu^2}{\beta}\left\|z^{k+1}-z^k\right\|^2 + (5\beta\|A\|^2 + \frac{\theta}{2})\left\|x^{k+1}-x^k\right\|^2 \\
&+ \frac{\theta}{2}\left\|y^{k+1}-y^k\right\|^2 + C_1\left\|z^{k+1}-z^k\right\|^2 + C_2\left\|x^{k+1}-x^k\right\|^2 + C_3\left\|y^{k+1}-y^k\right\|^2 \\
&\le L_\beta\left(x^k,y^k,z^k,u^k\right) + \frac{5\mu^2}{\beta}\left\|z^k-z^{k-1}\right\|^2 + (5\beta\|A\|^2 + \frac{\theta}{2})\left\|x^k-x^{k-1}\right\|^2 \\
&+ \frac{\theta}{2}\left\|y^k-y^{k-1}\right\|^2,
\end{aligned}$$

where

$$\begin{aligned}
&C_1 = \frac{\mu}{2} - \frac{5l_F^2 + 10\mu^2}{\beta}, \\
&C_2 = \tau - 10\beta\|A\|^2 - \frac{\beta\|A\|^2 + \ell_{1,+}}{2} - \theta, \\
&C_3 = \frac{\mu}{2} - \frac{\ell_{2,+}}{2} - \theta.
\end{aligned}$$

Proof According to the descent lemma and (3.1c), we have

$$\begin{aligned}
&H\left(x^{k+1},y^{k+1}\right) \\
\le& H\left(x^k,y^{k+1}\right) + \left\langle \nabla_x H\left(x^k,y^{k+1}\right), x^{k+1}-x^k\right\rangle + \frac{\ell_1\left(y^{k+1}\right)}{2}\left\|x^{k+1}-x^k\right\|^2 \\
\le& H\left(x^k,y^{k+1}\right) + \frac{\ell_{1,+}}{2}\left\|x^{k+1}-x^k\right\|^2 \\
&+ \left\langle -\tau\left(x^{k+1}-x^k\right) - A^T u^k - \beta A^T\left(Ax^k - z^{k+1}\right) - \theta(x^k - x^{k-1}), x^{k+1}-x^k\right\rangle \\
=& H\left(x^k,y^{k+1}\right) - \left\langle u^k, Ax^{k+1}-Ax^k\right\rangle - \beta\left\langle Ax^k - z^{k+1}, Ax^{k+1}-Ax^k\right\rangle \\
&- \theta\left\langle x^k - x^{k-1}, x^{k+1}-x^k\right\rangle + \left(\frac{\ell_{1,+}}{2} - \tau\right)\left\|x^{k+1}-x^k\right\|^2 \\
\le& H\left(x^k,y^{k+1}\right) - \left\langle u^k, Ax^{k+1}-Ax^k\right\rangle + \frac{\beta}{2}\left\|Ax^k - z^{k+1}\right\|^2 - \frac{\beta}{2}\left\|Ax^{k+1}-z^{k+1}\right\|^2 \\
&+ \left(\frac{\beta\|A\|^2 + \ell_{1,+}}{2} - \tau + \frac{\theta}{2}\right)\left\|x^{k+1}-x^k\right\|^2 + \frac{\theta}{2}\left\|x^k - x^{k-1}\right\|^2,
\end{aligned}$$

which implies that

$$H\left(x^{k+1},y^{k+1}\right) + \left\langle u^k, Ax^{k+1}-z^{k+1}\right\rangle + \frac{\beta}{2}\left\|Ax^{k+1}-z^{k+1}\right\|^2$$

$$\le H\left(x^k,y^{k+1}\right)+\left\langle u^k,Ax^k-z^{k+1}\right\rangle+\frac{\beta}{2}\left\|Ax^k-z^{k+1}\right\|^2+\frac{\theta}{2}\left\|x^k-x^{k-1}\right\|^2$$
$$+\left(\frac{\beta\|A\|^2+\ell_{1,+}}{2}-\tau+\frac{\theta}{2}\right)\left\|x^{k+1}-x^k\right\|^2.$$

By the definition of L_β it can be rewritten as

$$L_\beta\left(x^{k+1},y^{k+1},z^{k+1},u^k\right)\le L_\beta\left(x^k,y^{k+1},z^{k+1},u^k\right)+\frac{\theta}{2}\left\|x^k-x^{k-1}\right\|^2$$
$$+\left(\frac{\beta\|A\|^2+\ell_{1,+}}{2}-\tau+\frac{\theta}{2}\right)\left\|x^{k+1}-x^k\right\|^2. \tag{3.4a}$$

According to the descent lemma, we easily get

$$H\left(x^k,y^{k+1}\right)\le H\left(x^k,y^k\right)+\left\langle\nabla_y H\left(x^k,y^k\right),y^{k+1}-y^k\right\rangle+\frac{\ell_2\left(x^k\right)}{2}\left\|y^{k+1}-y^k\right\|^2$$
$$\le H\left(x^k,y^k\right)+\left\langle\nabla_y H\left(x^k,y^k\right),y^{k+1}-y^k\right\rangle+\frac{\ell_{2,+}}{2}\left\|y^{k+1}-y^k\right\|^2. \tag{3.5}$$

From (3.1a) and (3.1b) we obtain

$$G\left(y^{k+1}\right)+\left\langle\nabla_y H\left(x^k,y^k\right),y^{k+1}-y^k\right\rangle$$
$$+\frac{\mu}{2}\left\|y^{k+1}-y^k\right\|^2+\theta\left\langle y^{k+1}-y^k,y^k-y^{k-1}\right\rangle\le G\left(y^k\right)$$

and

$$F(z^{k+1})+\left\langle u^k,Ax^k-z^{k+1}\right\rangle+\frac{\beta}{2}\left\|Ax^k-z^{k+1}\right\|^2+\frac{\mu}{2}\left\|z^{k+1}-z^k\right\|^2$$
$$\le F(z^k)+\left\langle u^k,Ax^k-z^k\right\rangle+\frac{\beta}{2}\left\|Ax^k-z^k\right\|^2,$$

respectively. Adding the above three inequalities yields

$$F(z^{k+1})+G\left(y^{k+1}\right)+H\left(x^k,y^{k+1}\right)+\left\langle u^k,Ax^k-z^{k+1}\right\rangle$$
$$+\frac{\beta}{2}\left\|Ax^k-z^{k+1}\right\|^2+\left(\frac{\mu}{2}-\frac{\ell_{2,+}}{2}\right)\left\|y^{k+1}-y^k\right\|^2+\frac{\mu}{2}\left\|z^{k+1}-z^k\right\|^2$$
$$\le F(z^k)+G\left(y^k\right)+H\left(x^k,y^k\right)+\left\langle u^k,Ax^k-z^k\right\rangle+\frac{\beta}{2}\left\|Ax^k-z^k\right\|^2$$
$$+\theta\left\langle y^k-y^{k+1},y^k-y^{k-1}\right\rangle.$$

By the definition of L_β we have

$$L_\beta\left(x^k,y^{k+1},z^{k+1},u^k\right)+\left(\frac{\mu}{2}-\frac{\ell_{2,+}}{2}\right)\left\|y^{k+1}-y^k\right\|^2+\frac{\mu}{2}\left\|z^{k+1}-z^k\right\|^2$$
$$\le L_\beta\left(x^k,y^k,z^k,u^k\right)+\theta\left\langle y^k-y^{k+1},y^k-y^{k-1}\right\rangle$$
$$\le L_\beta\left(x^k,y^k,z^k,u^k\right)+\theta\left\|y^k-y^{k+1}\right\|\left\|y^k-y^{k-1}\right\|$$
$$\le L_\beta\left(x^k,y^k,z^k,u^k\right)+\frac{\theta}{2}\left\|y^k-y^{k+1}\right\|^2+\frac{\theta}{2}\left\|y^k-y^{k-1}\right\|^2. \tag{3.4b}$$

Then we get

$$L_\beta\left(x^k,y^{k+1},z^{k+1},u^k\right)+\left(\frac{\mu}{2}-\frac{\ell_{2,+}}{2}-\frac{\theta}{2}\right)\left\|y^{k+1}-y^k\right\|^2+\frac{\mu}{2}\left\|z^{k+1}-z^k\right\|^2 \\ \le L_\beta\left(x^k,y^k,z^k,u^k\right)+\frac{\theta}{2}\left\|y^k-y^{k-1}\right\|^2. \tag{3.6}$$

Combining the definition of L_β and (3.1d), we have

$$\begin{aligned} & L_\beta\left(x^{k+1},y^{k+1},z^{k+1},u^{k+1}\right)-L_\beta\left(x^{k+1},y^{k+1},z^{k+1},u^k\right) \\ &=\left\langle u^{k+1},Ax^{k+1}-z^{k+1}\right\rangle-\left\langle u^k,Ax^{k+1}-z^{k+1}\right\rangle \\ &=\left\langle u^{k+1}-u^k,Ax^{k+1}-z^{k+1}\right\rangle \\ &=\frac{1}{\beta}\left\|u^{k+1}-u^k\right\|^2. \end{aligned}$$

From (3.3) it follows that

$$\begin{aligned} & \left\|u^{k+1}-u^k\right\|^2 \\ &=\left\|\nabla F(z^{k+1})-\nabla F(z^k)+\beta(Ax^{k+1}-Ax^k)-\beta(Ax^k-Ax^{k-1})\right. \\ &\quad \left.+\mu(z^{k+1}-z^k)-\mu(z^k-z^{k-1})\right\|^2 \\ &\le 5l_F^2\left\|z^{k+1}-z^k\right\|^2+5\beta^2\|A\|^2\left\|x^{k+1}-x^k\right\|^2+5\beta^2\|A\|^2\left\|x^k-x^{k-1}\right\|^2 \\ &\quad +5\mu^2\left\|z^{k+1}-z^k\right\|^2+5\mu^2\left\|z^k-z^{k-1}\right\|^2 \\ &=(5l_F^2+5\mu^2)\left\|z^{k+1}-z^k\right\|^2+5\mu^2\left\|z^k-z^{k-1}\right\|^2 \\ &\quad +5\beta^2\|A\|^2\left\|x^{k+1}-x^k\right\|^2+5\beta^2\|A\|^2\left\|x^k-x^{k-1}\right\|^2. \end{aligned} \tag{3.7}$$

Then we have

$$\begin{aligned} & L_\beta\left(x^{k+1},y^{k+1},z^{k+1},u^{k+1}\right)-L_\beta\left(x^{k+1},y^{k+1},z^{k+1},u^k\right) \\ &=\frac{1}{\beta}\left\|u^{k+1}-u^k\right\|^2 \\ &\le\frac{5l_F^2+5\mu^2}{\beta}\left\|z^{k+1}-z^k\right\|^2+\frac{5\mu^2}{\beta}\left\|z^k-z^{k-1}\right\|^2 \\ &+5\beta\|A\|^2\left\|x^{k+1}-x^k\right\|^2+5\beta\|A\|^2\left\|x^k-x^{k-1}\right\|^2. \end{aligned} \tag{3.4c}$$

Hence, combining (3.4a) and (3.4b) with (3.4c), we obtain

$$\begin{aligned} & L_\beta\left(x^{k+1},y^{k+1},z^{k+1},u^{k+1}\right) \\ &\le L_\beta\left(x^{k+1},y^{k+1},z^{k+1},u^k\right)+\frac{5l_F^2+5\mu^2}{\beta}\left\|z^{k+1}-z^k\right\|^2+\frac{5\mu^2}{\beta}\left\|z^k-z^{k-1}\right\|^2 \\ &+5\beta\|A\|^2\left\|x^{k+1}-x^k\right\|^2+5\beta\|A\|^2\left\|x^k-x^{k-1}\right\|^2 \\ &\le L_\beta\left(x^k,y^{k+1},z^{k+1},u^k\right)+\frac{5l_F^2+5\mu^2}{\beta}\left\|z^{k+1}-z^k\right\|^2+\frac{5\mu^2}{\beta}\left\|z^k-z^{k-1}\right\|^2 \end{aligned}$$

$$+(5\beta\|A\|^2+\frac{\beta\|A\|^2+\ell_{1,+}}{2}-\tau+\frac{\theta}{2})\|x^{k+1}-x^k\|^2+(5\beta\|A\|^2+\frac{\theta}{2})\|x^k-x^{k-1}\|^2$$
$$\le L_\beta\left(x^k,y^k,z^k,u^k\right)+(\frac{5l_F^2+5\mu^2}{\beta}-\frac{\mu}{2})\|z^{k+1}-z^k\|^2+\frac{5\mu^2}{\beta}\|z^k-z^{k-1}\|^2$$
$$+(5\beta\|A\|^2+\frac{\beta\|A\|^2+\ell_{1,+}}{2}-\tau+\frac{\theta}{2})\|x^{k+1}-x^k\|^2$$
$$+(5\beta\|A\|^2+\frac{\theta}{2})\|x^k-x^{k-1}\|^2+\frac{\theta}{2}\|y^k-y^{k-1}\|^2+\left(\frac{\theta}{2}+\frac{\ell_{2,+}}{2}-\frac{\mu}{2}\right)\|y^{k+1}-y^k\|^2.$$

which can be rewritten as

$$L_\beta\left(x^{k+1},y^{k+1},z^{k+1},u^{k+1}\right)+(\frac{\mu}{2}-\frac{5l_F^2+5\mu^2}{\beta})\|z^{k+1}-z^k\|^2$$
$$+(\tau-5\beta\|A\|^2-\frac{\beta\|A\|^2+\ell_{1,+}}{2}-\frac{\theta}{2})\|x^{k+1}-x^k\|^2+(\frac{\mu}{2}-\frac{\theta}{2}-\frac{\ell_{2,+}}{2})\|y^{k+1}-y^k\|^2$$
$$\le L_\beta(x^k,y^k,z^k,u^k)+\frac{5\mu^2}{\beta}\|z^k-z^{k-1}\|^2+(5\beta\|A\|^2+\frac{\theta}{2})\|x^k-x^{k-1}\|^2$$
$$+\frac{\theta}{2}\|y^k-y^{k-1}\|^2.$$

The proof is completed. □

Remark 3.3 Obviously, from Assumption A(iv) we have $C_1>0$, $C_2>0$, and $C_3>0$, since

$$\mu>\ell_{2,+}+2\theta,\beta>\frac{10l_F^2+20\mu^2}{\mu},\tau>10\beta\|A\|^2+\frac{\beta\|A\|^2+\ell_{1,+}}{2}+\theta.$$

Based on Lemma 3.1, we define the following regularized augmented Lagrangian function:

$$\hat{L}_\beta\left(x,y,z,u,x',y',z'\right)=L_\beta\left(x,y,z,u\right)+\frac{5\mu^2}{\beta}\|z-z'\|^2$$
$$+(5\beta\|A\|^2+\frac{\theta}{2})\|x-x'\|^2+\frac{\theta}{2}\|y-y'\|^2. \tag{3.8}$$

Let $\hat{\omega}=\left(x,y,z,u,x',y',z'\right)$, $\hat{\omega}^k=\left(x^k,y^k,z^k,u^k,x^{k-1},y^{k-1},z^{k-1}\right)$, and $\omega^k=(x^k,y^k,z^k,u^k)$. Then the following lemma implies that the sequence $\left\{\hat{L}_\beta(\hat{\omega}^k)\right\}_{k\ge1}$ is decreasing. It is important for our convergence analysis.

Lemma 3.2 (*Descent property*) *Suppose that Assumption* A *holds. Let* $\hat{L}_\beta(\hat{\omega}^k)$ *be defined as in* (3.8). *Then there exist* $C_1,C_2,C_3>0$ *such that*

$$\hat{L}_\beta\left(\hat{\omega}^{k+1}\right)+C_1\|x^{k+1}-x^k\|^2+C_2\|y^{k+1}-y^k\|^2+C_3\|z^{k+1}-z^k\|^2\le\hat{L}_\beta\left(\hat{\omega}^k\right). \tag{3.9}$$

Proof The result follows directly from Lemma 3.1 and Remark 3.1. □

Lemma 3.3 *Let*

$$\beta>3l_F.$$

Then there exists γ *such that*

$$\frac{1}{\gamma}-\frac{l_F}{2\gamma^2}=\frac{3}{2\beta}. \tag{3.10}$$

Proof We notice that the reduced discriminant of the quadratic equations in (3.10) (in γ) is

$$\Delta_\gamma := 4-\frac{12}{\beta}l_F.$$

Since

$$\beta > 3l_F,$$

it follows that $\Delta_\gamma > 0$, and hence the equation has a nonzero real solution. □

Theorem 3.1 (*Convergence*) *Suppose that Assumption A holds. If* $\{\hat{\omega}^k\}_{k\geq 0}$ *is a sequence generated by Algorithm 3.1, then the following statements are true*:

(*i*) *The sequence* $\left\{\hat{L}_\beta(\hat{\omega}^k)\right\}_{k\geq 1}$ *is bounded from below and convergent*;

(*ii*)

$$x^{k+1}-x^k\to 0, y^{k+1}-y^k\to 0, z^{k+1}-z^k\to 0 \quad \text{and} \quad u^{k+1}-u^k\to 0 \quad \text{as} \quad k\to +\infty;$$

(*iii*) *The sequence* $\{L_\beta(\omega^k)\}_{k\geq 1}$ *is convergent.*

Proof First, we show that $\underline{L}$ is a lower bound of $\left\{\hat{L}_\beta\left(\hat{\omega}^k\right)\right\}_{k\geq 2}$. Suppose on the contrary that there exists $k_0\geq 2$ such that $\hat{L}_\beta\left(\hat{\omega}^{k_0}\right)-\underline{L}\leq 0$. Since $\left\{\hat{L}_\beta\left(\hat{\omega}^k\right)\right\}_{k\geq 1}$ is a nonincreasing sequence, we have that for all $N\geq k_0$,

$$\sum_{k=1}^{N}\left(\hat{L}_\beta\left(\hat{\omega}^k\right)-\underline{L}\right)\leq \sum_{k=1}^{k_0-1}\left(\hat{L}_\beta\left(\hat{\omega}^k\right)-\underline{L}\right)+(N-k_0+1)\left(\hat{L}_\beta\left(\hat{\omega}^{k_0}\right)-\underline{L}\right),$$

which implies that

$$\lim_{N\to+\infty}\sum_{k=1}^{N}\left(\hat{L}_\beta\left(\hat{\omega}^k\right)-\underline{L}\right)=-\infty.$$

On the other hand, for $k\geq 1$,

$$\begin{aligned}\hat{L}_\beta\left(\hat{\omega}^k\right)-\underline{L}&\geq F(z^k)+G(y^k)+H(x^k,y^k)+\langle u^k,Ax^k-z^k\rangle-\underline{L}\\&\geq\langle u^k,Ax^k-z^k\rangle\\&=\frac{1}{\beta}\langle u^k,u^k-u^{k-1}\rangle\\&=\frac{1}{2\beta}\left\|u^k\right\|^2+\frac{1}{2\beta}\left\|u^k-u^{k-1}\right\|^2-\frac{1}{2\beta}\left\|u^{k-1}\right\|^2.\end{aligned} \tag{3.11}$$

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

$$\sum_{k=1}^{N} \left(\widehat{L}_{\beta}\left(\widehat{\omega}^{k}\right) - \underline{L}\right) \geq \frac{1}{2\beta} \sum_{k=1}^{N} \left\|u^{k} - u^{k-1}\right\|^{2} + \frac{1}{2\beta}\left\|u^{N}\right\|^{2} - \frac{1}{2\beta}\left\|u^{0}\right\|^{2} \geq -\frac{1}{2\beta}\left\|u^{0}\right\|^{2}$$

which leads to a contradiction. From Lemma 3.2 we have that

$$\hat{L}_{\beta}\left(\hat{\omega}^{k+1}\right) + C_1\left\|x^{k+1} - x^{k}\right\|^{2} + C_2\left\|y^{k+1} - y^{k}\right\|^{2} + C_3\left\|z^{k+1} - z^{k}\right\|^{2} \leq \hat{L}_{\beta}\left(\hat{\omega}^{k}\right).$$

As $\left\{\hat{L}_{\beta}(\hat{\omega}^{k})\right\}_{k\geq 1}$ is bounded from below, we obtain that $\left\{\hat{L}_{\beta}(\hat{\omega}^{k})\right\}_{k\geq 1}$ is convergent by Lemma 2.2 and also that

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

Then, according to (3.7), it follows that $u^{k+1} - u^{k} \to 0$ as $k \to \infty$. By the definition of $\left\{\hat{L}_{\beta}(\hat{\omega}^{k})\right\}_{k\geq 1}$ we obtain that $\left\{L_{\beta}\left(\omega^{k}\right)\right\}$ is convergent. □

Remark 3.4 (i) Thanks to (ii) of Theorem 3.1, it is easy to see that $\left\{x^{k+1} - x^{k}\right\}$, $\left\{y^{k+1} - y^{k}\right\}$, $\left\{z^{k+1} - z^{k}\right\}$, and $\left\{u^{k+1} - u^{k}\right\}$ are bounded. Define

$$S_{*} := \sup_{k\geq 0}\left\{\left\|x^{k+1} - x^{k}\right\|, \left\|y^{k+1} - y^{k}\right\|, \left\|z^{k+1} - z^{k}\right\|, \left\|u^{k+1} - u^{k}\right\|\right\} < +\infty.$$

Theorem 3.2 (*The boundedness of sequences*) *Suppose that Assumption A holds. Let $\left\{\left(x^{k}, y^{k}, z^{k}, u^{k}\right)\right\}_{k\geq 0}$ be a sequence generated by Algorithm 3.1, and suppose that there exists $\gamma \in R\backslash\{0\}$ such that (3.10) holds. Suppose that the function H is coercive. Then every sequence $\left\{\left(x^{k}, y^{k}, z^{k}, u^{k}\right)\right\}_{k\geq 0}$ generated by Algorithm 3.1 is bounded.*

Proof Let $k \geq 1$ be fixed. According to Lemma 3.2, we have that

$$\begin{aligned} \hat{L}_{\beta}\left(\hat{\omega}^{1}\right) \geq \cdots \geq \hat{L}_{\beta}\left(\hat{\omega}^{k}\right) \geq & \hat{L}_{\beta}\left(\hat{\omega}^{k+1}\right) \\ & \geq F\left(z^{k+1}\right) + G\left(y^{k+1}\right) + H\left(x^{k+1}, y^{k+1}\right) - \frac{1}{2\beta}\left\|u^{k+1}\right\|^{2} \\ & + \frac{\beta}{2}\left\|Ax^{k+1} - z^{k+1} + \frac{1}{\beta}u^{k+1}\right\|^{2}. \end{aligned} \tag{3.12}$$

From (3.3) we have

$$\begin{aligned} \left\|u^{k+1}\right\|^{2} &= \left\|\nabla F(z^{k+1}) + \beta(Ax^{k+1} - Ax^{k}) + \mu(z^{k+1} - z^{k})\right\|^{2} \\ &\leq 3\left\|\nabla F(z^{k+1})\right\|^{2} + 3\beta^{2}\|A\|^{2}\left\|x^{k+1} - x^{k}\right\|^{2} + 3\mu^{2}\left\|z^{k+1} - z^{k}\right\|^{2} \\ &\leq 3\left\|\nabla F(z^{k+1})\right\|^{2} + (3\beta^{2}\|A\|^{2} + 3\mu^{2})S_{*}^{2}. \end{aligned}$$

Multiplying this relation by $\frac{1}{2\beta}$ and combining it with (3.12), we get

$$\hat{L}_{\beta}\left(\hat{\omega}^{1}\right) \geq F\left(z^{k+1}\right) + G\left(y^{k+1}\right) + H\left(x^{k+1}, y^{k+1}\right)$$

$$-\frac{3\beta^2\|A\|^2+3\mu^2}{2\beta}S_*^2-\frac{3}{2\beta}\left\|\nabla F(z^{k+1})\right\|^2+\frac{\beta}{2}\left\|Ax^{k+1}-z^{k+1}+\frac{1}{\beta}u^{k+1}\right\|^2. \tag{3.13}$$

We will prove the boundedness of $\{(x^k,y^k,z^k,u^k)\}_{k\geq 0}$. According to (3.13) and Proposition 2.2, we have that for all $k \geq 1$,

$$\begin{aligned}
&H(x^{k+1},y^{k+1}) + \frac{\beta}{2}\left\|Ax^{k+1} - z^{k+1} + \frac{1}{\beta}u^{k+1}\right\|^2 \\
\leq&\widehat{L}_\beta(\widehat{\omega}^1) + \frac{3\beta^2\|A\|^2 + 3\mu^2}{2\beta}S_*^2 - \inf\{F(z) - \frac{3}{2\beta}\|\nabla F(z)\|^2\} - \inf\{G(y)\} \\
=&\widehat{L}_\beta(\widehat{\omega}^1)+\frac{3\beta^2\|A\|^2+3\mu^2}{2\beta}S_*^2 - \inf\{F(z)-(\frac{1}{\gamma}-\frac{l_F}{2\gamma^2})\|\nabla F(z)\|^2\}-\inf\{G(y)\} \\
\leq&\widehat{L}_\beta(\widehat{\omega}^1) + \frac{3\beta^2\|A\|^2 + 3\mu^2}{2\beta}S_*^2 - \inf\{F(z)\} - \inf\{G(y)\}.
\end{aligned} \tag{3.14}$$

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

$$\{(x^k,y^k)\}_{k\geq 0} \quad \text{and} \quad \left\{Ax^k - z^k + \frac{1}{\beta}u^k\right\}_{k\geq 0}$$

are bounded. As, according to (3.1d) and Remark 3.2, $\{Ax^k - z^k\}_{k\geq 0}$ is bounded, it follows that $\{u^k\}_{k\geq 0}$ and $\{z^k\}_{k\geq 0}$ are also bounded. □

The next lemma provides upper estimates for the square of the limiting subgradients of the regularization of the augmented Lagrangian function $\hat{L}_\beta(\hat{\omega}^k)$.

Lemma 3.4 *Suppose that Assumption A holds. Let $\{(x^k,y^k,z^k,u^k)\}_{k\geq 0}$ be the sequence generated by Algorithm 3.1, which is assumed to be bounded. We denote $v^k = (x^k,y^k,z^k)$. Then there exists $\zeta > 0$ such that*

$$dist\left(0,\partial\hat{L}_\beta\left(\hat{\omega}^{k+1}\right)\right) \leq \zeta\left(\left\|v^{k+1}-v^k\right\| + \left\|v^k - v^{k-1}\right\|\right).$$

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

$$\begin{aligned}
\partial_x\hat{L}_\beta\left(\hat{\omega}^{k+1}\right) =&\nabla_x H\left(x^{k+1},y^{k+1}\right) + A^T u^{k+1} + \beta A^T\left(Ax^{k+1} - z^{k+1}\right) \\
&+ (10\beta\|A\|^2 + \theta)\left(x^{k+1} - x^k\right),
\end{aligned} \tag{3.15a}$$

$$\partial_y\hat{L}_\beta\left(\hat{\omega}^{k+1}\right) =\partial G\left(y^{k+1}\right) + \nabla_y H\left(x^{k+1},y^{k+1}\right) + \theta\left(y^{k+1} - y^k\right), \tag{3.15b}$$

$$\begin{aligned}
\partial_z\hat{L}_\beta\left(\hat{\omega}^{k+1}\right) =&\nabla F\left(z^{k+1}\right) - u^{k+1} - \beta\left(Ax^{k+1} - z^{k+1}\right) \\
&+ \frac{10\mu^2}{\beta}\left(z^{k+1} - z^k\right),
\end{aligned} \tag{3.15c}$$

$$\partial_u\hat{L}_\beta\left(\hat{\omega}^{k+1}\right) =Ax^{k+1} - z^{k+1} = \frac{1}{\beta}\left(u^{k+1} - u^k\right), \tag{3.15d}$$

$$\partial_{x'}\hat{L}_\beta\left(\hat{\omega}^{k+1}\right) = -(10\beta\|A\|^2 + \theta)\left(x^{k+1} - x^k\right), \tag{3.15e}$$

$$\partial_{y'}\hat{L}_\beta\left(\hat{\omega}^{k+1}\right) = -\theta\left(y^{k+1} - y^k\right), \tag{3.15f}$$

$$\partial_{z'}\hat{L}_\beta\left(\hat{\omega}^{k+1}\right) = -\frac{10\mu^2}{\beta}\left(z^{k+1} - z^k\right). \tag{3.15g}$$

After combining (3.15a) with (3.1c), we have

$$\begin{aligned}\partial_x\hat{L}_\beta(\hat{\omega}^{k+1}) &= \nabla_x H(x^{k+1}, y^{k+1}) - \nabla_x H(x^k, y^{k+1}) + A^T u^{k+1} - A^T u^k \\ &+ \beta A^T A(x^{k+1} - x^k) + (10\beta\|A\|^2 + \theta - \tau)(x^{k+1} - x^k) - \theta(x^k - x^{k-1}).\end{aligned}$$

Substituting (3.2a) and (3.2b) into (3.15b) and (3.15c), respectively, leads to

$$\begin{aligned}&\nabla_y H(x^{k+1}, y^{k+1}) - \nabla_y H(x^k, y^k) + (\theta - \mu)(y^{k+1} - y^k) + \theta(y^{k+1} - y^k) \in \partial_y\hat{L}_\beta(\hat{\omega}^{k+1}) \\ &-(u^{k+1} - u^k) - \beta A(x^{k+1} - x^k) + (\frac{10\mu^2}{\beta} - \mu)(z^{k+1} - z^k) \in \partial_z\hat{L}_\beta(\hat{\omega}^{k+1})\end{aligned}$$

Let $D^k = \left(d_x^{k+1}, d_y^{k+1}, d_z^{k+1}, d_u^{k+1}, d_{x'}^{k+1}, d_{y'}^{k+1}, d_{z'}^{k+1}\right)$, where

$$\begin{aligned}d_x^{k+1} &= \nabla_x H(x^{k+1}, y^{k+1}) - \nabla_x H(x^k, y^{k+1}) + A^T u^{k+1} - A^T u^k \\ &\quad + \beta A^T A(x^{k+1} - x^k) + (10\beta\|A\|^2 + \theta - \tau)(x^{k+1} - x^k) - \theta(x^k - x^{k-1}), \\ d_y^{k+1} &= \nabla_y H(x^{k+1}, y^{k+1}) - \nabla_y H(x^k, y^k) + (\theta - \mu)(y^{k+1} - y^k) + \theta(y^k - y^{k-1}), \\ d_z^{k+1} &= -\beta A(x^{k+1} - x^k) + (\frac{10\mu^2}{\beta} - \mu)(z^{k+1} - z^k), \\ d_u^{k+1} &= \frac{1}{\beta}\left(u^{k+1} - u^k\right), \\ d_{x'}^{k+1} &= -(10\beta\|A\|^2 + \theta)\left(x^{k+1} - x^k\right), \\ d_{y'}^{k+1} &= -\theta\left(y^{k+1} - y^k\right), \\ d_{z'}^{k+1} &= -\frac{10\mu^2}{\beta}\left(z^{k+1} - z^k\right).\end{aligned}$$

Then it follows that $D^{k+1} \in \partial\hat{L}_\beta\left(\hat{\omega}^{k+1}\right)$ and $\left(d_x^{k+1}, d_y^{k+1}, d_z^{k+1}, d_u^{k+1}\right) \in \partial L_\beta\left(\omega^{k+1}\right)$.
Thus $dist^2\left(0, \partial\hat{L}_\beta\left(\omega^{k+1}\right)\right) \le \left\|D^{k+1}\right\|^2$. By Assumption A(iii) we have

$$\left\|\nabla_y H\left(x^{k+1}, y^{k+1}\right) - \nabla_y H\left(x^k, y^k\right)\right\|^2 \le \ell_h^2\left(\left\|x^{k+1} - x^k\right\|^2 + \left\|y^{k+1} - y^k\right\|^2\right).$$

Then there exists $\zeta_1 > 0$ such that

$$\begin{aligned}dist^2\left(0, \partial\hat{L}_\beta\left(\hat{\omega}^{k+1}\right)\right) &\le \left\|D^{k+1}\right\|^2 \\ &\le \zeta_1^2\left(\left\|x^{k+1} - x^k\right\|^2 + \left\|y^{k+1} - y^k\right\|^2 + \left\|z^{k+1} - z^k\right\|^2\right. \\ &\quad \left. + \left\|u^{k+1} - u^k\right\|^2 + \left\|y^k - y^{k-1}\right\|^2 + \left\|x^k - x^{k-1}\right\|^2\right).\end{aligned}$$

Thus by (3.7) there exists $\zeta_2 > 0$ such that

$$dist^2\left(0, \partial\hat{L}_\beta\left(\hat{\omega}^{k+1}\right)\right) \le \zeta^2\left(\left\|x^{k+1} - x^k\right\|^2 + \left\|y^{k+1} - y^k\right\|^2 + \left\|z^{k+1} - z^k\right\|^2\right.$$

$$+\left\|x^k - x^{k-1}\right\|^2 + \left\|y^k - y^{k-1}\right\|^2 + \left\|z^k - z^{k-1}\right\|^2\Big). \tag{3.16}$$

Then by $v^k = \left(x^k, y^k, z^k\right)$ it follows that

$$\left\|v^k - v^{k-1}\right\|^2 = \left\|x^k - x^{k-1}\right\|^2 + \left\|y^k - y^{k-1}\right\|^2 + \left\|z^k - z^{k-1}\right\|^2.$$

Combining with (3.16) gives

$$\begin{aligned} dist\left(0, \partial\hat{L}_\beta\left(\hat{\omega}^{k+1}\right)\right) &\le \sqrt{\zeta^2\left(\left\|v^{k+1} - v^k\right\|^2 + \left\|v^k - v^{k-1}\right\|^2\right)} \\ &\le \zeta\left(\left\|v^{k+1} - v^k\right\| + \left\|v^k - v^{k-1}\right\|\right). \end{aligned}$$

The proof is completed. □

Now we give the convergence analysis of the sequence in a general framework by proving that any cluster point of $\left\{\left(x^k, y^k, z^k, u^k\right)\right\}_{k\ge 0}$ is a KKT point of the optimization problem (1.1). Let Ω and $\hat{\Omega}$ denote the cluster point sets of the sequences $\left\{\omega^k\right\}$ and $\left\{\hat{\omega}^k\right\}$, respectively.

Theorem 3.3 (*Global convergence*) *Suppose that Assumption* A *holds. Suppose that the sequence generated by Algorithm* 3.1 *is bounded. Then we have that*

(*i*) $\hat{\Omega}$ *is nonempty, compact, and connected*;

(*ii*) $dist\left(\hat{L}_\beta\left(\hat{\omega}^k\right), \hat{\Omega}\right) \to 0$ *as* $k \to \infty$;

(*iii*) *If* $\left\{\left(x^{k_j}, y^{k_j}, z^{k_j}, u^{k_j}\right)\right\}_{j\ge 0}$ *is a subsequence of* $\left\{\left(x^k, y^k, z^k, u^k\right)\right\}_{k\ge 0}$ *that converges to* (x^*, y^*, z^*, u^*) *as* $k \to +\infty$, *then*

$$\lim_{k\to+\infty} \hat{L}_\beta\left(\hat{\omega}^{k_j}\right) = L_\beta\left(x^*, y^*, z^*, u^*\right); \tag{3.17}$$

(*iv*) $\hat{\Omega} \subset \mathrm{crit}\hat{L}_\beta(\hat{\omega})$;

(*v*) *The function* $\hat{L}_\beta$ *takes on* $\hat{\Omega}$ *the value*

$$\hat{L}_\beta^* = \lim_{k\to+\infty} \hat{L}_\beta\left(\hat{\omega}^k\right) = \lim_{k\to+\infty}\left\{F\left(z^k\right) + G\left(y^k\right) + H\left(x^k, y^k\right)\right\}.$$

Proof By the definition of Ω and $\hat{\Omega}$, (i) and (ii) are trivial.

(iii) Let $\left\{\omega^{k_j}\right\}$ be a subsequence of $\left\{\omega^k\right\}$ such that $\omega^{k_j} \to \omega^*$, $j \to \infty$. Since F and G are lower semicontinuous, so is L_β, which follows from

$$\liminf_{j\to\infty} L_\beta\left(\omega^{k_j}\right) \ge L_\beta\left(\omega^*\right). \tag{3.18}$$

On the other hand, the definition of z^{k+1} shows that

$$L_\beta\left(x^k, y^k, z^{k+1}, u^k\right) + \frac{\mu}{2}\left\|z^{k+1} - z^k\right\|^2 \le L_\beta\left(x^k, y^k, z^*, u^k\right) + \frac{\mu}{2}\left\|z^* - z^k\right\|^2,$$

from which we get

$$L_\beta\left(x^k, y^k, z^{k+1}, u^k\right) + \frac{\mu}{2}\left\|z^{k+1} - z^k\right\|^2 - \frac{\mu}{2}\left\|z^* - z^k\right\|^2 \le L_\beta\left(x^k, y^k, z^*, u^k\right).$$

Replacing $x^k, y^k, z^k, z^{k+1}, u^k$ by $x^{k_j}, y^{k_j}, z^{k_j}, z^{k_j+1}, u^{k_j}$, we get

$$L_\beta\left(x^{k_j},y^{k_j},z^{k_j+1},u^{k_j}\right)+\frac{\mu}{2}\left\|z^{k_j+1}-z^{k_j}\right\|^2-\frac{\mu}{2}\left\|z^*-z^{k_j}\right\|^2\leq L_\beta\left(x^{k_j},y^{k_j},z^*,u^{k_j}\right).$$

Combining this with Theorem 3.1(ii) gives

$$\left\|\omega^{k+1}-\omega^k\right\|\to 0 \text{ as } k\to\infty,$$

and then we have

$$\left\|\omega^{k_j+1}-\omega^{k_j}\right\|\to 0 \text{ and } \left\|\omega^{k_j}-\omega^*\right\|\to 0 \text{ as } j\to\infty,$$

which implies that

$$\limsup_{j\to\infty} L_\beta\left(x^{k_j},y^{k_j},z^{k_j+1},u^{k_j}\right)\leq L_\beta\left(\omega^*\right).$$

Since $z^{k+1}-z^k\to 0$ as $k\to\infty$, it is easy to get

$$\lim_{j\to\infty} L_\beta\left(x^{k_j},y^{k_j},z^{k_j+1},u^{k_j}\right)=\lim_{j\to\infty} L_\beta\left(\omega^{k_j}\right).$$

Then we have

$$\limsup_{j\to\infty} L_\beta\left(\omega^{k_j}\right)\leq L_\beta\left(\omega^*\right). \tag{3.19}$$

Therefore from (3.18) and (3.19) it follows that

$$\lim_{j\to+\infty} L_\beta\left(\omega^{k_j}\right)=L_\beta\left(\omega^*\right).$$

By the definition of $\hat{L}_\beta\left(\hat{\omega}^k\right)$, since $\left\|\omega^k-\omega^{k-1}\right\|\to 0$ as $k\to\infty$, the desired statement follows.

(iv) For the sequence D^k defined in Lemma 3.4, for $j\geq 1$, we have $D^{k_j}\in\partial\hat{L}_\beta\left(\hat{\omega}^{k_j}\right)$. Then

$$D^{k_j}\to 0 \text{ as } j\to\infty,$$

and thus

$$\hat{\omega}^{k_j}\to\hat{\omega}^* \text{ and } \hat{L}_\beta\left(\hat{\omega}^{k_j}\right)\to\hat{L}_\beta\left(\hat{\omega}^*\right) \text{ as } j\to\infty.$$

The closedness criterion of the limiting subdifferential guarantees that $0\in\partial\hat{L}_\beta\left(\hat{\omega}^{k_j}\right)$ or, in other words, that $\hat{\omega}^*\in\text{crit}(\hat{L}_\beta)$.

(v) Due to Theorem 3.1(ii) and the boundedness of $\{u_n\}_{n\geq 0}$, the sequences $\left\{\hat{L}_\beta(\hat{\omega}^k)\right\}_{k\geq 0}$ and $\left\{F\left(z^k\right)+G\left(y^k\right)+H\left(x^k,y^k\right)\right\}_{k\geq 0}$ have the same limit:

$$\hat{L}_\beta^*=\lim_{k\to+\infty}\hat{L}_\beta\left(\hat{\omega}^k\right)=\lim_{k\to+\infty}\left\{F\left(z^k\right)+G\left(y^k\right)+H\left(x^k,y^k\right)\right\}.$$

The conclusion now follows by statements (iii) and (iv). □

Next, we will prove global convergence for $\left\{\left(x^k, y^k, z^k, u^k\right)\right\}_{k\geq 0}$ generated by Algorithm 3.1 in the context of the Kurdyka–Łojasiewicz property. Suppose that $\hat{L}_\beta(\hat{\omega}^k)$ is a KŁ function with desingularization function

$$\varphi(s) := cs^{1-\theta}, \theta \in [0,1), c > 0.$$

Theorem 3.4 (*Strong convergence*) *Let* $v^k = \left(x^k, y^k, z^k\right)$. *Assume that* $\hat{L}_\beta(\hat{\omega}^k)$ *is a KŁ function and Assumption A is satisfied. Then we have*
(*i*) $\sum_{k=1}^{\infty} \left\|\omega^k - \omega^{k-1}\right\| < \infty$,
(*ii*) $\left\{\omega^k\right\}$ *converges to a critical point of* L_β.

Proof From the proof of Theorem 3.3 it follows that $\lim_{k\to+\infty} \hat{L}_\beta\left(\hat{\omega}^k\right) = \hat{L}_\beta\left(\hat{\omega}^*\right)$. We consider two cases.

Case 1. There exists an integer $k_0 > 0$ such that $\hat{L}_\beta\left(\hat{\omega}^{k_0}\right) = \hat{L}_\beta\left(\hat{\omega}^*\right)$. Then since $\left\{\hat{L}_\beta\left(\hat{\omega}^k\right)\right\}$ is decreasing, we know that for all $k > k_0$,

$$\begin{aligned} &C_1\left\|x^{k+1}-x^k\right\|^2+C_2\left\|y^{k+1}-y^k\right\|^2+C_3\left\|z^{k+1}-z^k\right\|^2 \\ &\quad\leq\hat{L}_\beta\left(\hat{\omega}^k\right)-\hat{L}_\beta\left(\hat{\omega}^{k+1}\right)\leq\hat{L}_\beta\left(\hat{\omega}^*\right)-\hat{L}_\beta\left(\hat{\omega}^*\right) = 0, \end{aligned}$$

which implies that $x^{k+1} = x^k$, $y^{k+1} = y^k$, $z^{k+1} = z^k$ for all $k > k_0$. Then from (3.7) and (3.8) we have $u^{k+1} = u^k$ for any $k > k_0$. Thus, for all $k > k_0 + 1$, we have $\omega^{k+1} = \omega^k$, and the desired results follow.

Case 2. $\hat{L}_\beta\left(\hat{\omega}^k\right) > \hat{L}_\beta\left(\hat{\omega}^*\right)$ for all k. Then since $dist\left(\hat{\omega}^k, \Omega\right) \to 0$, we have that for arbitrary $\varepsilon_1 > 0$, there exists $k_1 > 0$ such that $dist\left(\hat{\omega}^k, \Omega\right) < \varepsilon_1$ for all $k > k_1$. Since $\lim_{k_j\to+\infty} \hat{L}_\beta\left(\hat{\omega}^{k_j}\right) = \hat{L}_\beta\left(\hat{\omega}^*\right)$, we have that for arbitrary $\varepsilon_2 > 0$, there exists $k_2 > 0$ such that $\hat{L}_\beta\left(\hat{\omega}^k\right) < \hat{L}_\beta\left(\hat{\omega}^*\right) + \varepsilon_2$ for all $k > k_2$. Therefore, for any $\varepsilon_1, \varepsilon_2 > 0$, when $k > \widetilde{k} = \max\{k_1, k_2\}$, we have $dist\left(\hat{\omega}^k, \Omega\right) < \varepsilon_1$, $\hat{L}_\beta\left(\hat{\omega}^*\right) < \hat{L}_\beta\left(\hat{\omega}^k\right) < \hat{L}_\beta\left(\hat{\omega}^*\right) + \varepsilon_2$. Since $\left\{\omega^k\right\}$ is bounded, by Theorem 3.3 we know that $\hat{\Omega}$ is nonempty compact set and $\hat{L}_\beta$ is constant on $\hat{\Omega}$. Applying Lemma 2.4, we deduce that for all $k > \tilde{k}$,

$$\varphi'\left(\hat{L}_\beta\left(\hat{\omega}^k\right) - \hat{L}_\beta\left(\hat{\omega}^*\right)\right) dist\left(0, \partial\hat{L}_\beta\left(\hat{\omega}^k\right)\right) \geq 1.$$

Due to $\varphi'\left(\hat{L}_\beta\left(\hat{\omega}^k\right) - \hat{L}_\beta\left(\hat{\omega}^*\right)\right) > 0$, we obtain

$$\frac{1}{\varphi'\left(\hat{L}_\beta\left(\hat{\omega}^k\right) - \hat{L}_\beta\left(\hat{\omega}^*\right)\right)} \leq dist\left(0, \partial\hat{L}_\beta\left(\hat{\omega}^k\right)\right).$$

Using the concavity of φ, we get that

$$\begin{aligned} &\varphi\left(\hat{L}_\beta\left(\hat{\omega}^k\right) - \hat{L}_\beta\left(\hat{\omega}^*\right)\right) - \varphi\left(\hat{L}_\beta\left(\hat{\omega}^{k+1}\right) - \hat{L}_\beta\left(\hat{\omega}^*\right)\right) \\ &\quad\geq \varphi'\left(\hat{L}_\beta\left(\hat{\omega}^k\right) - \hat{L}_\beta\left(\hat{\omega}^*\right)\right)\left(\hat{L}_\beta\left(\hat{\omega}^k\right) - \hat{L}_\beta\left(\hat{\omega}^{k+1}\right)\right). \end{aligned}$$

Combining this with the KŁ property gives

$$\hat{L}_\beta\left(\hat{\omega}^k\right) - \hat{L}_\beta\left(\hat{\omega}^{k+1}\right) \leq \frac{\varphi\left(\hat{L}_\beta\left(\hat{\omega}^k\right) - \hat{L}_\beta\left(\hat{\omega}^*\right)\right) - \varphi\left(\hat{L}_\beta\left(\hat{\omega}^{k+1}\right) - \hat{L}_\beta\left(\hat{\omega}^*\right)\right)}{\varphi'\left(\hat{L}_\beta\left(\hat{\omega}^k\right) - \hat{L}_\beta\left(\hat{\omega}^*\right)\right)}$$

$$\leq dist\left(0, \partial \hat{L}_\beta\left(\hat{\omega}^k\right)\right)\left(\varphi\left(\hat{L}_\beta\left(\hat{\omega}^k\right)-\hat{L}_\beta\left(\hat{\omega}^*\right)\right)-\varphi\left(\hat{L}_\beta\left(\hat{\omega}^{k+1}\right)-\hat{L}_\beta\left(\hat{\omega}^*\right)\right)\right). \tag{3.20}$$

By Lemma 3.2 there exists $\eta = \min(C_1, C_2, C_3)$ such that

$$\begin{aligned}\hat{L}_\beta\left(\hat{\omega}^k\right)-\hat{L}_\beta\left(\hat{\omega}^{k+1}\right) &\geq C_1\left\|x^{k+1}-x^k\right\|^2+C_2\left\|y^{k+1}-y^k\right\|^2+C_2\left\|z^{k+1}-z^k\right\|^2\\ &\geq \eta\left\|v^{k+1}-v^k\right\|^2.\end{aligned} \tag{3.21}$$

By Lemma 3.4 we get

$$dist\left(0, \partial \hat{L}_\beta\left(\hat{\omega}^k\right)\right) \leq \zeta\left(\left\|v^k-v^{k-1}\right\|+\left\|v^{k-1}-v^{k-2}\right\|\right). \tag{3.22}$$

Putting (3.21) and (3.22) into (3.20), we obtain

$$\begin{aligned}\eta\left\|v^{k+1}-v^k\right\|^2 \leq{}& \zeta\left(\left\|v^k-v^{k-1}\right\|+\left\|v^{k-1}-v^{k-2}\right\|\right)\\ &\cdot\left(\varphi\left(\hat{L}_\beta\left(\hat{\omega}^k\right)-\hat{L}_\beta\left(\hat{\omega}^*\right)\right)-\varphi\left(\hat{L}_\beta\left(\hat{\omega}^{k+1}\right)-\hat{L}_\beta\left(\hat{\omega}^*\right)\right)\right).\end{aligned} \tag{3.23}$$

Set $b_k = \frac{\zeta}{\eta}\left(\varphi\left(\hat{L}_\beta\left(\hat{\omega}^k\right)-\hat{L}_\beta\left(\hat{\omega}^*\right)\right)-\varphi\left(\hat{L}_\beta\left(\hat{\omega}^{k+1}\right)-\hat{L}_\beta\left(\hat{\omega}^*\right)\right)\right) \geq 0$ and $a_k = \left\|v^k-v^{k-1}\right\| \geq 0$. Then (3.23) can be equivalently rewritten as

$$a_{k+1}^2 \leq b_k\left(a_k+a_{k-1}\right). \tag{3.24}$$

Since $\varphi \geq 0$, we know that

$$\sum_{k=1}^{\infty} b_k \leq \frac{\zeta}{\eta}\varphi\left(\hat{L}_\beta\left(\hat{\omega}^1\right)-\hat{L}_\beta\left(\hat{\omega}^*\right)\right),$$

and hence $\sum\limits_{k=1}^{\infty} b_k < \infty$. Note that from (3.24) we have

$$a_{k+1} \leq \sqrt{b_k\left(a_k+a_{k-1}\right)} \leq \frac{1}{4}\left(a_k+a_{k-1}\right)+b_k.$$

By Lemma 2.3 this gives that $\sum_{k=1}^{\infty} a_k < \infty$. Therefore

$$\sum_{k=1}^{\infty}\left\|x^k-x^{k-1}\right\| < \infty, \sum_{k=1}^{\infty}\left\|y^k-y^{k-1}\right\| < \infty, \sum_{k=1}^{\infty}\left\|z^k-z^{k-1}\right\| < \infty.$$

Combining this with (3.7), we have

$$\sum_{k=1}^{\infty}\left\|u^k-u^{k-1}\right\| < \infty.$$

This indicates that $\left\{\omega^k\right\}$ is a Cauchy sequence. Therefore $\left\{\omega^k\right\}$ is convergent. Let $\omega^k \to \omega^*$, $k \to \infty$. According to Theorem 3.3(iv), it is clear that ω^* is a critical point of $L_\beta(\omega^*)$. The proof is completed. □

4 Numerical experiments

In this section, we present a numerical example to compare the performance of our algorithm with PMA in [8] and BIPCA in [34]. We consider the following optimization problem:

$$\min_{x,y} \frac{1}{2}\|Ax-b\|^2 + c_1 \|y\|_{\frac{1}{2}}^{\frac{1}{2}} + \frac{c_2}{2}\|Bx-y\|^2,$$

which can be rewritten as

$$\min_{x,y,z} \frac{1}{2}\|z-b\|^2 + c_1 \|y\|_{\frac{1}{2}}^{\frac{1}{2}} + \frac{c_2}{2}\|Bx-y\|^2$$
$$\text{s.t.} \quad Ax = z$$

We select A and B as random matrices with $A = (a_{ij})_{p\times m}$ and $B = (b_{ij})_{q\times m}$, where $a_{ij}, b_{ij} \in (0,1)$. Let m, p, q be three positive integers with $m = q$. We take the initial points of Algorithm 3.1, $x^0 = zeros(m,1)$, $y^0 = zeros(q,1)$, $z^0 = zeros(p,1)$, $u^0 = zeros(p,1)$, $x^{-1} = rand(n,1)$, and $y^{-1} = rand(n,1)$. The parameters are set as $\mu = 4$, $\beta = 182.5$, $\tau = 1.92e + 7$, and $c_1 = c_2 = 1$. The initial points of PMA in [8] are also set as x^0, y^0, z^0, u^0, and the parameter $\sigma = 0.01$. The initial points of BIPCA in [34] are set as $x^0 = x^{-1}$ and $y^0 = y^{-1}$, and the parameter $\sigma = 0.01$, Define $\|Ax-z\|^2$ as the error, and select $\|Ax-z\|^2 < 10^{-4}$ as the stopping criterion. The numerical experiment is carried out in 64-bit MATLAB R2019b on 64-bit PC with Intel(R) core(TM)i7-6700HQ CPU@2.6 GHz anf 32 GB of RAM.

The numerical results are shown in Tables 1 and 2. To make it explicit, we also measure the performances of the algorithm by plotting the curve of error. The corresponding results are presented in Figs. 1 and 2. In the tables, k denotes the number of iterations, and s denotes the computing time.

In view of Table 1 and Fig. 1, we find that the inertial factor has a positive effect on the convergence of Algorithm 3.1. The larger the value of the inertial parameter, the faster the convergence speed. In addition, by observing Table 2 (where θ_{1k} and θ_{2k} are the two inertial

Table 1 Numerical results of two algorithms under various inertia values and dimensions

	$m = q = 100; p = 200$	$m = q = 100; p = 300$	$m = q = 100; p = 400$
Algorithm 3.1 $\theta_k = 0.1$	$k = 10; s = 0.3281$	$k = 12; s = 0.3750$	$k = 14; s = 0.3238$
Algorithm 3.1 $\theta_k = 0.5$	$k = 9; s = 0.3438$	$k = 11; s = 0.3594$	$k = 12; s = 0.3450$
Algorithm 3.1 $\theta_k = 0.8$	$k = 8; s = 0.3281$	$k = 10; s = 0.2969$	$k = 11; s = 0.3594$
PMA	$k = 305; s = 0.7813$	$k = 308; s = 1.0469$	$k = 310; s = 2.0313$

Table 2 Numerical results for Example 5.1 with various θ

θ	Algorithms 3.1		BIPCA		PMA	
	k	s	k	s	k	s
$\theta_k = 0.9$ for Algorithm 3.1 $\theta_{1k} = 0.05, \theta_{2k} = 0.03$ for BIPCA	85	3.9426	143	4.2761	162	4.3871
$\theta_k = 0.8$ for Algorithm 3.1 $\theta_{1k} = 0.04, \theta_{2k} = 0.02$ for BIPCA	106	4.5488	$k188$	4.7906	198	5.2917
$\theta_k = 0.7$ for Algorithm 3.1 $\theta_{1k} = 0.03, \theta_{2k} = 0.01$ for BIPCA	142	4.7215	234	5.8531	261	6.7946

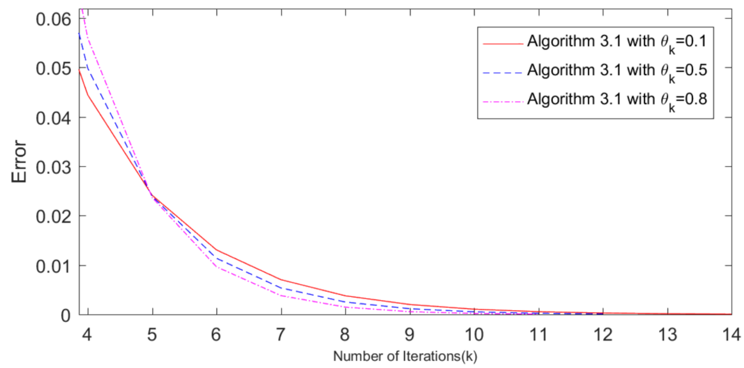

Figure 1 The performance of our algorithm with $m = q = 100$ and $p = 400$

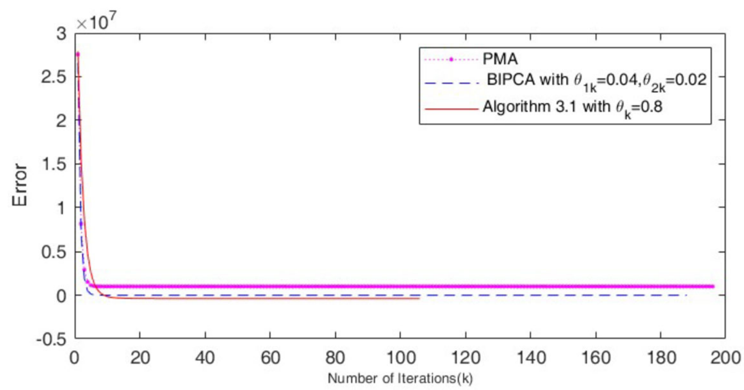

Figure 2 Errors for three algorithms

parameters in BIPCA) and Fig. 2, it appears that our algorithm needs fewer iterations and converges more quickly than PMA and BIPCA. In short, experimental results show that our algorithm is effective and performs better than PMA in [8] and BIPCA in [34] because of employing the inertial technique and ADMM.

5 Conclusions

An efficient modified inertial proximal minimization algorithm is presented for solving a nonconvex and nonsmooth problem that is the sum of a smooth function and a linear operator or of a nonsmooth function and a function that couples two variables. The proposed algorithm updates the x-subproblem and y-subproblem with inertial effect. The parameters are selected in a simple way. The numerical experiment reveals that the algorithm is feasibile and effective.

Acknowledgements
The authors are grateful to the editor and reviewers for their valuable comments and suggestions.

Author contributions
Zhonghui Xue wrote the main manuscript text and make the ideas, conception and design of study, analysis of data in numerical experiment, formal analysis, programming. Qianfeng Ma rewrote, review and editing, validation. All authors carefully reviewed the manuscript.

Funding
This work is supported by National Natural Science Foundation of China under grants 72071130 and 71901145.

Data Availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 9 June 2024 Accepted: 16 September 2024 Published online: 20 September 2024

References
1. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Lojasiewicz inequality. Math. Oper. Res. **35**(2), 438–457 (2010)
2. Attouch, H., Bolte, J., Svaiter, B.F.: Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods. Math. Program. **137**, 91–129 (2013)
3. Auslender, A.: Méthodes numériques pour la décomposition et la minimisation de fonctions nondifférentiables. Numer. Math. **18**(3), 213–223 (1971)
4. Bian, F.M., Liang, J.W., Zhang, X.Q.: A stochastic alternating direction method of multipliers for non-smooth and non-convex optimization. Inverse Probl. **37**(7), 075009 (2021)
5. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. **146**, 459–494 (2014)
6. Bot, R.I., Csetnek, E.R.: An inertial alternating direction method of multipliers (2014). Eprint arXiv:1404.4582
7. Bot, R.I., Csetnek, E.R.: An inertial Tseng's type proximal algorithm for nonsmooth and nonconvex optimization problems. J. Optim. Theory Appl. **171**(2), 600–616 (2016)
8. Bot, R.I., Csetnek, E.R., Nguyen, D.K.: A proximal minimization algorithm for structured nonconvex and nonsmooth problems. SIAM J. Optim. **29**(2), 1300–1328 (2019)
9. Bot, R.I., Kanzler, L.: A forward-backward dynamical approach for nonsmooth problems with block structure coupled by a smooth function. Appl. Math. Comput. **394**(10), 125822 (2021)
10. Bot, R.I., Nguyen, D.K.: The proximal alternating direction method of multipliers in the nonconvex setting: convergence analysis and rates. Math. Oper. Res. **45**(2), 682–712 (2020)
11. Chao, M.T., Zhang, Y., Jian, J.B.: An inertial proximal alternating direction method of multipliers for nonconvex optimization. Int. J. Comput. Math. **98**(6), 1199–1217 (2021)
12. Chen, C.H., Ma, S.Q., Yang, J.F.: A general inertial proximal point algorithm for mixed variational inequality problem. SIAM J. Optim. **25**(4), 2120–2142 (2015)
13. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward–backward splitting. Multiscale Model. Simul. **4**(4), 1168–1200 (2005)
14. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
15. Driggs, D., Tang, J., Liang, J.: SPRING: a fast stochastic proximal alternating method for non-smooth non-convex optimization (2020). Eprint arXiv:2002.12266
16. Gao, X., Cai, X.J., Han, D.R.: A Gauss–Seidel type inertial proximal alternating linearized minimization for a class of nonconvex optimization problems. J. Glob. Optim. **76**(4), 863–887 (2020)
17. Guo, C., Zhao, J., Dong, Q.L.: A stochastic two-step inertial Bregman proximal alternating linearized minimization algorithm for nonconvex and nonsmooth problems
18. Hong, M., Luo, Z.Q., Razaviyayn, M.: Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems. In: IEEE, pp. 3836–3840 (2015)
19. Li, G., Pong, T.K.: Global convergence of splitting methods for nonconvex composite optimization. SIAM J. Optim. **25**(4), 2434–2460 (2015)
20. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979)
21. Liu, Q., Shen, X., Gu, Y.: Linearized ADMM for non-convex non-smooth optimization with convergence analysis. IEEE Access **7**, 76131–76144 (2017)
22. Ochs, P., Brox, T., Pock, T.: iPiasco: inertial proximal algorithm for strongly convex optimization. J. Math. Imaging Vis. **53**(2), 171–181 (2015)
23. Ortega, J.M., Rheinboldt, W.C.: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, San Diego (1970)
24. Ouyang, Y., Chen, Y., Lan, G., Pasiliao, E.: An accelerated linearized alternating direction method of multipliers. SIAM J. Imaging Sci. **8**(1), 644–681 (2015)
25. Pock, T., Sabach, S.: Inertial proximal alternating linearized minimization (iPALM) for nonconvex and nonsmooth problems. SIAM J. Imaging Sci. **9**(4), 1756–1787 (2016)
26. Polyak, B.T.: Some methods of speeding up the convergence of iteration methods. USSR Comput. Math. Math. Phys. **4**(5), 1–17 (1964)
27. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D, Nonlinear Phenom. **60**(1–4), 259–268 (1992)

28. Tseng, P.: A modified forward–backward splitting method for maximal monotone mappings. SIAM J. Control Optim. **38**(2), 431–446 (2000)
29. Tseng, P.: Convergence of a block coordinate descent method for nondifferentiable minimization. J. Optim. Theory Appl. **109**(3), 475–494 (2001)
30. Wang, Y., Yin, W., Zeng, J.: Global convergence of ADMM in nonconvex nonsmooth optimization. J. Sci. Comput. **78**, 29–63 (2019)
31. Yang, J., Yuan, X.: Linearized augmented Lagrangian and alternating direction methods for nuclear norm minimization. Math. Comput. **82**(281), 301–329 (2013)
32. Zhang, X.Q., Burger, M., Osher, S.: A unified primal-dual algorithm framework based on Bregman iteration. J. Sci. Comput. **46**, 20–46 (2010)
33. Zhang, Y., He, S.: Inertial proximal alternating minimization for nonconvex and nonsmooth problems. J. Inequal. Appl. **2017**, 232 (2017)
34. Zhao, J., Dong, Q.L., Rassias, M.T., Wang, F.: Two-step inertial Bregman alternating minimization algorithm for nonconvex and nonsmooth problems. J. Glob. Optim. **84**, 941–966 (2022)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.