Hindawi
Journal of Mathematics
Volume 2023, Article ID 4316267, 12 pages
https://doi.org/10.1155/2023/4316267

Research Article

Convergence Analysis of Multiblock Inertial ADMM for Nonconvex Consensus Problem

Yang Liu [1,2] and Yazheng Dang [3]

[1] *Department of Information Science and Technology, East China University of Political Science and Law, Shanghai 200237, China*
[2] *China Institute for Smart Court, Shanghai Jiao Tong University, Shanghai 200030, China*
[3] *School of Management, University of Shanghai for Science and Technology, Shanghai 200093, China*

Correspondence should be addressed to Yazheng Dang; jgdyz@163.com

Received 29 July 2022; Revised 13 December 2022; Accepted 4 March 2023; Published 28 March 2023

Academic Editor: Qiang Wu

The alternating direction method of multipliers (ADMM) is one of the most powerful and successful methods for solving various nonconvex consensus problem. The convergence of the conventional ADMM (i.e., 2-block) for convex objective functions has been stated for a long time. As an accelerated technique, the inertial effect was used by many authors to solve 2-block convex optimization problem. This paper combines the ADMM and the inertial effect to construct an inertial alternating direction method of multipliers (IADMM) to solve the multiblock nonconvex consensus problem and shows the convergence under some suitable conditions. Simulation experiment verifies the effectiveness and feasibility of the proposed method.

1. Introduction

The nonconvex global consensus problem with regularization [1] has the following form:

$$\begin{aligned} &\min \sum_{i=1}^{N} f_i(x) + g(x) \\ &\text{s.t. } x \in X, \end{aligned} \tag{1}$$

where $f_i\colon R^n \longrightarrow R \cup \{+\infty\}, i = 1, 2, \cdots, N$ are smooth, possibly nonconvex functions, while $g\colon R^n \longrightarrow R$ is a convex nonsmooth regularization term and X is a closed convex set. This problem is related to the convex global consensus problem discussed heavily [2], but it is possible that f_i's are nonconvex.

In many practical applications, f_i's need to be handled by a single agent, such as a thread or a processor. Now, we transform problem (1) into the following equivalent linearly constrained problem under the help of new variables $\{x_i\}_{i=0}^{N}$:

$$\begin{aligned} &\min \sum_{i=1}^{N} f_i(x_i) + g(x_0) \\ &\text{s.t. } x_i = x_0 \forall i = 1, 2, \cdots, N, x_0 \in X. \end{aligned} \tag{2}$$

Note that the problem (2) owns N blocks with different variables $\{x_1, \ldots, x_N\}$ and one globe variable. Then, each distributed agent can handle a single local variable x_i and a local function f_i, respectively.

The augmented Lagrangian function with multipliers $y_i \in R^n, i = 1, 2, \cdots, N$ of problem (2) is defined as follows:

$$\begin{aligned} L_\rho(\{x_i\}, x_0, y) = &\sum_{i=1}^{N} f_i(x_i) + g(x_0) + \sum_{i=1}^{N} \langle y_i, x_i - x_0 \rangle \\ &+ \frac{\rho}{2} \sum_{i=1}^{N} \|x_i - x_0\|^2, \end{aligned} \tag{3}$$

where $\rho > 0$ is a penalty parameter and problem (2) can be solved distributively by the following classical ADMM procedure:

$$\begin{cases} x_0^{k+1} = \underset{x_0\in X}{\operatorname{argmin}}\left\{L_\rho\left(\left\{x_i^k\right\}, x_0, y^k\right)\right\}, \\ x_i^{k+1} = \underset{x_i}{\operatorname{argmin}}\left\{f_i(x_i) + \langle y_i^k, x_i - x_0^{k+1}\rangle + \frac{\rho}{2}\left\|x_i - x_0^{k+1}\right\|^2\right\}, \\ y_i^{k+1} = y_i^k + \rho\left(x_i^{k+1} - x_0^{k+1}\right), \\ i = 1, 2, \cdots, N. \end{cases} \tag{4}$$

ADMM was initially introduced in the 1970s [3, 4], and its convergence properties for convex case have been extensively studied. However, ADMM or its directly extended version may not converge when there is a nonconvex function in the objective. Yang et al. [5] studied the convergence of the ADMM for the nonconvex optimization model which come from the background/foreground extraction. Hong et al. [6] analyzed the convergence of alternating direction method of multipliers for a family of nonconvex problems. Guo et al. [7] studied the convergence of ADMM for multiblock nonconvex separable optimization models.

Recently, some scholars studied the inertial type of ADMM for convex optimization. For example, Chen et al. [8] analyzed a class of inertial ADMM for linearly constrained separable convex optimization, and Moudafi and Elissabeth [9] extended the inertial technique to solve the maximal monotone operator inclusion problem. The research interests for the nonconvex cases are increasing in recent years; e.g., Chao et al. [10] proposed and analyzed an inertial proximal ADMM for a class of nonconvex optimization problems while all the above inertial ADMM algorithms were presented for solving only two-block optimization problem (not for multiple-block case). Whether the convergence of the inertial ADMM is assured when the involved number of blocks is more than two? It is an important problem to research.

The purpose of the present study is to examine the convergence of inertial ADMM with multiblocks for nonconvex consensus problem under the assumption that the potential function satisfies the Kurdyka–Lojasiewicz property. The preliminary numerical results show the effectiveness of the proposed algorithm.

The rest of this paper is organized as follows. In Section 2, some necessary preliminaries for further analysis are summarized. Section 3 proposes a multiblock nonconvex inertial ADMM algorithm and analyzes its convergence under some conditions. In Section 4, we prove the validity of the algorithm by the numerical experiment. Finally, some conclusions are drawn in Section 5.

2. Preliminaries

Let R^n denote the n-dimensional Euclidean space, $R\cup\{+\infty\}$ denote the extended real number set, and N denote the natural number set. $\|.\|$ represents the Euclidean norm. Let $\operatorname{dom} f\colon = \{x \in R^n\colon f(x) < +\infty\}$ denote the domain of function $f\colon R^n \longrightarrow R\cup\infty$ and $\langle x, y\rangle = \sum_{i=1}^n x_i y_i$ denote the inner product. For function if f if $f(\overline{x}) \le \liminf_{x\longrightarrow\overline{x}} f(x)$, we say that f is lower semicontinuous at $\overline{x}$. If f is lower semicontinuous at every point $x \in R^n$, we say that f is lower semicontinuous function.

For a set $S \subset R^n$ and a point $x \in R^n$, let $d(x, S) = \inf_{y\in S} \|y - x\|^2$. If $S = \varnothing$, we set $d(x, S) = +\infty$ for all $x \in R^n$.

The Lagrangian function of (2), with multiplier $y = (y_1, y_2, \cdots, y_N)^T$, is defined as

$$L(\{x_i\}, x_0, y) = \sum_{i=1}^N f_i(x_i) + g(\mathrm{x}_0) + \sum_{i=1}^N \langle y_i, x_i - x_0\rangle. \tag{5}$$

Definition 1. If $w^* = (\{x_i^*\}, x_0^*, y^*)^{\mathrm{T}}$ such that

$$\begin{cases} \nabla_{x_i} f_i(x_i^*) = -y_i^*, \\ \sum_{i=1}^N y_i^* \in \partial g(x^*), \\ x_i^* - x_0^* = 0, \end{cases} \tag{6}$$

then w^* is called a critical point or stationary point of the Lagrange function $L(\{x_i\}, x_0, y)$.

A very important technique to prove the convergence of the ADMM for nonconvex optimization problems relies on the assumption that the potential function satisfying the following Kurdyka–Lojasiewicz property (KL property) [11–14].

For notational simplicity, we use $\Psi_{\varepsilon_2}(\varepsilon_2 > 0)$ to denote the set of concave functions $\phi\colon [0, \varepsilon_2) \longrightarrow (0, +\infty)$ such that

(i) $\phi(0) = 0$, ϕ is continuous differentiable on $(0, \varepsilon_2)$ and continuous at 0

(ii) $\phi'(s) > 0, \forall s \in (0, \varepsilon_2)$

Definition 2 (see [14]) (KL property). Let $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n \cup +\infty$ be a proper lower semicontinuous function. If there exists $\varepsilon_2 \in (0, +\infty)$, a neighborhood U of x^*, and a function $\phi \in \Psi_{\varepsilon_2}$, such that for all $x \in U \cap [f\ (x^*) < f < f(x^*) + \varepsilon_2]$, it holds that

$$\phi'(f(x) - f(x^*))d(0, \partial f(x)) \ge 1, \tag{7}$$

then f is said to have the KL property at x^*.

3. Algorithm and Convergence Analysis

For convenience, we fix the following notations: $w^k = (\{x_i^k\}, y^k, x_0^k)^{\mathrm{T}}$, $\widehat{w}^k = (\{x_i^k\}, y^k, x_0^k, \{x_i^{k-1}\}, y^{k-1}, x_0^{k-1})^{\mathrm{T}}$. Basis on (4), we propose the following algorithm for solving problem (2).

Algorithm 1. Inertial ADMM (IADMM). Choose $x_0^0 \in R^n$, $y_i^0 \in R^n, x_i^0 \in R^n, i = 1, 2, \cdots, N, \tau_k > 0, \rho > 0$ and $\theta_k \in [0, 1)$, $\forall k \ge 1$. For the given point $w^k = (\{x_i^k\}, y^k, x_0^k)^{\mathrm{T}}$, consider the iterative scheme:

$$\begin{cases} x_0^{k+1} = \underset{x_0 \in X}{\operatorname{argmin}} \left\{ g(x_0) + \sum_{i=1}^{N} \langle y_i^k, x_i^k - x_0 \rangle + \frac{\rho}{2} \sum_{i=1}^{N} \left\| x_i^k - x_0 \right\|^2 + \frac{\tau_k}{2} \left\| x_0 - z_0^k \right\|^2 \right\} (a), \\ x_i^{k+1} = \underset{x_i}{\operatorname{argmin}} \left\{ f_i(x_i) + \langle y_i^k, x_i - x_0^{k+1} \rangle + \frac{\rho}{2} \left\| x_i - x_0^{k+1} \right\|^2 + \frac{\tau_k}{2} \left\| x_i - z_i^k \right\|^2 \right\} (b), \\ y_i^{k+1} = y_i^k + \rho\left(x_i^{k+1} - x_0^{k+1}\right) + \tau_k\left(x_i^{k+1} - z_i^k\right) (c), \end{cases} \tag{8}$$

where

$$\begin{cases} z_0^k = x_0^k + \theta_k\left(x_0^k - x_0^{k-1}\right), \\ z_i^k = x_i^k + \theta_k\left(x_i^k - x_i^{k-1}\right), \end{cases} \tag{9}$$

associated with $i = 1, 2, \cdots, N$.

From the optimality conditions of (8) (a) and (8) (b), we have

$$0 \in \partial g\left(x_0^{k+1}\right) - \sum_{i=1}^{N} y_i^k - \rho \sum_{i=1}^{N} \left(x_i^k - x_0^{k+1}\right) + \tau_k\left(x_0^{k+1} - z_0^k\right), \tag{10}$$

$$0 = \nabla_{x_i} f_i\left(x_i^{k+1}\right) + y_i^k + \rho\left(x_i^{k+1} - x_0^{k+1}\right) + \tau_k\left(x_i^{k+1} - z_i^k\right), \quad i = 1, 2, \cdots, N. \tag{11}$$

Remark 1. Compared with the inertial ADMM in [10], each subproblem in our algorithm has the inertial term, and we handle multiblock case here.

Subsequently, we will discuss the convergence of Algorithm 1 under the following assumptions.

Assumption 1

(i) $g(x)$ is proper lower semicontinuous, and $\nabla_{x_i} f_i(x_i)$ is l_{f_i} Lipschitz continuous; i.e.,

$$\left\| \nabla_{x_i} f_i\left(x_i^{k+1}\right) - \nabla_{x_i} f_i\left(x_i^k\right) \right\| \le l_{f_i} \left\| x_i^{k+1} - x_i^k \right\|. \tag{12}$$

(ii) ρ is large enough, such that $0 \le \theta_k < \rho - l_f^2/2\rho + 2$, $\tau_k > 2l_f^2/(\rho - 2\rho\theta_k - l_f^2 - 2\theta_k)$.

Lemma 1. *For each $k \in \mathrm{N}$, define $l_f = \max\{l_{f_i}\}_{i=1,2,\cdots,N}$, we have*

$$\left\| y_i^{k+1} - y_i^k \right\|^2 \le l_{f_i}^2 \left\| x_i^{k+1} - x_i^k \right\|^2 \le l_f^2 \left\| x_i^{k+1} - x_i^k \right\|^2. \tag{13}$$

Proof. Since $y_i^{k+1} = y_i^k + \rho(x_i^{k+1} - x_0^{k+1}) + \tau_k(x_i^{k+1} - z_i^k)$, from (11), one has

$$y_i^{k+1} = -\nabla_{x_i} f_i\left(x_i^{k+1}\right). \tag{14}$$

Thus,

$$\begin{aligned} \left\| y_i^{k+1} - y_i^k \right\|^2 &= \left\| \nabla_{x_i} f_i\left(x_i^{k+1}\right) - \nabla_{x_i} f_i\left(x_i^k\right) \right\|^2 \\ &\le l_{f_i}^2 \left\| x_i^{k+1} - x_i^k \right\|^2 \\ &\le l_f^2 \left\| x_i^{k+1} - x_i^k \right\|^2. \end{aligned} \tag{15}$$

Hence, the result is obtained. □

Lemma 2. *Select ρ large enough, suppose that Assumption 1 holds. Then, for each $k \in \mathrm{N}$,*

$$\begin{aligned} &L_\rho\left(w^{k+1}\right) + \sum_{i=1}^{N} \gamma_1 \left\| x_i^{k+1} - x_i^k \right\|^2 + \gamma_1 \left\| x_0^{k+1} - x_0^k \right\|^2 \\ &\le L_\rho\left(w^k\right) + \sum_{i=1}^{N} \gamma_2 \left\| x_i^k - x_i^{k-1} \right\|^2 + \gamma_2 \left\| x_0^k - x_0^{k-1} \right\|^2, \end{aligned} \tag{16}$$

where $\gamma_1 = \tau_k/2(1 - \theta_k) - (1/\rho) + (\tau_k/2\rho)l_f^2 - (\tau_k/2\rho)$ and $\gamma_2 = (\tau_k\theta_k/2) + (\tau_k\theta_k/\rho)$.

Proof. By the definition of the augmented Lagrangian function, (8) (c) and (15), we have

$$\begin{aligned} &L_\rho\left(\left\{x_i^{k+1}\right\}, x_0^{k+1}, y^{k+1}\right) - L_\rho\left(\left\{x_i^{k+1}\right\}, x_0^{k+1}, y^k\right), \\ &= \sum_{i=1}^{N} \langle y_i^{k+1} - y_i^k, x_i^{k+1} - x_0^{k+1} \rangle, \\ &= \frac{1}{\rho} \sum_{i=1}^{N} \langle y_i^{k+1} - y_i^k, y_i^{k+1} - y_i^k - \tau_k\left(x_i^{k+1} - z_i^k\right) \rangle \\ &\le \left(\frac{1}{\rho} + \frac{\tau_k}{2\rho}\right) \sum_{i=1}^{N} \left\| y_i^{k+1} - y_i^k \right\|^2 + \frac{\tau_k}{2\rho} \sum_{i=1}^{N} \left\| x_i^{k+1} - z_i^k \right\|^2 \\ &\le \left(\frac{1}{\rho} + \frac{\tau_k}{2\rho}\right) l_f^2 \sum_{i=1}^{N} \left\| x_i^{k+1} - x_i^k \right\|^2 + \frac{\tau_k}{2\rho} \sum_{i=1}^{N} \left\| x_i^{k+1} - z_i^k \right\|^2. \end{aligned} \tag{17}$$

From (8) (a) and (8) (b), we obtain

$$\begin{aligned}
&g\left(x_0^{k+1}\right)+\sum_{i=1}^N\langle y_i^k, x_i^k-x_0^{k+1}\rangle+\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^k-x_0^{k+1}\right\|^2+\frac{\tau_k}{2}\left\|x_0^{k+1}-z_0^k\right\|^2\\
&\quad\le g\left(x_0^k\right)+\sum_{i=1}^N\langle y_i^k, x_i^k-x_0^k\rangle+\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^k-x_0^k\right\|^2+\frac{\tau_k}{2}\left\|x_0^k-z_0^k\right\|^2,\\
&\text{and}\\
&\sum_{i=1}^N f_i\left(x_i^{k+1}\right)+\sum_{i=1}^N\langle y_i^k, x_i^{k+1}-x_0^{k+1}\rangle+\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^{k+1}-x_0^{k+1}\right\|^2+\frac{\tau_k}{2}\sum_{i=1}^N\left\|x_i^{k+1}-z_i^k\right\|^2\\
&\quad\le\sum_{i=1}^N f_i\left(x_i^k\right)+\sum_{i=1}^N\langle y_i^k, x_i^k-x_0^{k+1}\rangle+\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^k-x_0^{k+1}\right\|^2+\frac{\tau_k}{2}\sum_{i=1}^N\left\|x_i^k-z_i^k\right\|^2,
\end{aligned}\tag{18}$$

respectively. Then, it is easy to get

$$\begin{aligned}
&g\left(x_0^{k+1}\right)-g\left(x_0^k\right)+\sum_{i=1}^N\langle y_i^k, x_0^k-x_0^{k+1}\rangle-\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^k-x_0^k\right\|^2\\
&\quad\le\frac{\tau_k}{2}\left\|x_0^k-z_0^k\right\|^2-\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^k-x_0^{k+1}\right\|^2-\frac{\tau_k}{2}\left\|x_0^{k+1}-z_0^k\right\|^2,\\
&\text{and}\\
&\sum_{i=1}^N f_i\left(x_i^{k+1}\right)-\sum_{i=1}^N f_i\left(x_i^k\right)+\sum_{i=1}^N\langle y_i^k, x_i^{k+1}-x_i^k\rangle+\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^{k+1}-x_0^{k+1}\right\|^2\\
&\quad\le\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^k-x_0^{k+1}\right\|^2+\frac{\tau_k}{2}\sum_{i=1}^N\left\|x_i^k-z_i^k\right\|^2-\frac{\tau_k}{2}\sum_{i=1}^N\left\|x_i^{k+1}-z_i^k\right\|^2.
\end{aligned}\tag{19}$$

Therefore, we have

$$\begin{aligned}
&L_\rho\left(\left\{x_i^{k+1}\right\}, x_0^{k+1}, y^k\right)-L_\rho\left(\left\{x_i^k\right\}, x_0^k, y^k\right)\\
&\quad=L_\rho\left(\left\{x_i^{k+1}\right\}, x_0^{k+1}, y^k\right)-L_\rho\left(\left\{x_i^k\right\}, x_0^{k+1}, y^k\right)+L_\rho\left(\left\{x_i^k\right\}, x_0^{k+1}, y^k\right)-L_\rho\left(\left\{x_i^k\right\}, x_0^k, y^k\right)\\
&\quad=\sum_{i=1}^N f_i\left(x_i^{k+1}\right)-\sum_{i=1}^N f_i\left(x_i^k\right)+\sum_{i=1}^N\langle y_i^k, x_i^{k+1}-x_i^k\rangle+\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^{k+1}-x_0^{k+1}\right\|^2\\
&\qquad+g\left(x_0^{k+1}\right)-g\left(x_0^k\right)-\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^k-x_0^k\right\|^2+\sum_{i=1}^N\langle y_i^k, x_0^k-x_0^{k+1}\rangle\\
&\quad\le\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^k-x_0^{k+1}\right\|^2+\frac{\tau_k}{2}\sum_{i=1}^N\left\|x_i^k-z_i^k\right\|^2-\frac{\tau_k}{2}\sum_{i=1}^N\left\|x_i^{k+1}-z_i^k\right\|^2\\
&\qquad+g\left(x_0^{k+1}\right)-g\left(x_0^k\right)-\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^k-x_0^k\right\|^2+\sum_{i=1}^N\langle y_i^k, x_0^k-x_0^{k+1}\rangle\\
&\quad\le\frac{\tau_k}{2}\sum_{i=1}^N\left\|x_i^k-z_i^k\right\|^2-\frac{\tau_k}{2}\sum_{i=1}^N\left\|x_i^{k+1}-z_i^k\right\|^2+\frac{\tau_k}{2}\left\|x_0^k-z_0^k\right\|^2-\frac{\tau_k}{2}\left\|x_0^{k+1}-z_0^k\right\|^2\\
&\quad\le-\frac{\tau_k}{2}\left(1-\theta_k\right)\sum_{i=1}^N\left\|x_i^{k+1}-x_i^k\right\|^2+\frac{\tau_k}{2}\theta_k\sum_{i=1}^N\left\|x_i^k-x_i^{k-1}\right\|^2-\frac{\tau_k}{2}\left(1-\theta_k\right)\left\|x_0^{k+1}-x_0^k\right\|^2\\
&\qquad+\frac{\tau_k}{2}\theta_k\left\|x_0^k-x_0^{k-1}\right\|^2.
\end{aligned}\tag{20}$$

Adding up (17) and (20), by the Assumption 1 (ii), we have

$$\begin{aligned} L_\rho\left(w^{k+1}\right) &\le L_\rho\left(w^k\right) - \frac{\tau_k}{2}\left(1-\theta_k\right)\sum_{i=1}^N \left\|x_i^{k+1}-x_i^k\right\|^2 + \frac{\tau_k}{2}\theta_k \sum_{i=1}^N \left\|x_i^k - x_i^{k-1}\right\|^2 - \frac{\tau_k}{2}\left(1-\theta_k\right)\left\|x_0^{k+1}-x_0^k\right\|^2 \\ &\quad + \frac{\tau_k}{2}\theta_k\left\|x_0^k - x_0^{k-1}\right\|^2 + \left(\frac{1}{\rho}+\frac{\tau_k}{2\rho}\right)\sum_{i=1}^N\left\|y_i^{k+1}-y_i^k\right\|^2 + \frac{\tau_k}{2\rho}\sum_{i=1}^N\left\|x_i^{k+1}-z_i^k\right\|^2 \\ &\le L_\rho\left(w^k\right) - \left(\frac{\tau_k}{2}\left(1-\theta_k\right) - \left(\frac{1}{\rho}+\frac{\tau_k}{2\rho}\right)l_f^2 - \frac{\tau_k}{\rho}\right)\sum_{i=1}^N\left\|x_i^{k+1}-x_i^k\right\|^2 + \left(\frac{\tau_k\theta_k}{2}+\frac{\tau_k\theta_k}{\rho}\right)\sum_{i=1}^N\left\|x_i^k-x_i^{k-1}\right\|^2 \\ &\quad - \frac{\tau_k}{2}\left(1-\theta_k\right)\left\|x_0^{k+1}-x_0^k\right\|^2 + \frac{\tau_k}{2}\theta_k\left\|x_0^k-x_0^{k-1}\right\|^2, \end{aligned} \tag{21}$$

which implies that

$$\begin{aligned} &L_\rho\left(w^{k+1}\right) + \sum_{i=1}^N \gamma_1\left\|x_i^{k+1}-x_i^k\right\|^2 + \gamma_1\left\|x_0^{k+1}-x_0^k\right\|^2 \\ &\le L_\rho\left(w^k\right) + \sum_{i=1}^N \gamma_2\left\|x_i^k - x_i^{k-1}\right\|^2 + \gamma_2\left\|x_0^k - x_0^{k-1}\right\|^2. \end{aligned} \tag{22}$$

Then, the results are obtained. □

Remark 2. From Assumption 1 (ii), we know that $\gamma_1 > \gamma_2$. Define the following potential regularized augmented Lagrangian function:

$$\widehat{L}_\rho\left(\{x_i\}, x_0, y, \{\widehat{x}_i\}, \widehat{x}_0\right) = L_\rho\left(\{x_i\}, x_0, y\right) + \sum_{i=1}^N \gamma_2\left\|x_i-\widehat{x}_i\right\|^2 + \gamma_2\left\|x_0-\widehat{x}_0\right\|^2, \tag{23}$$

where $\widehat{w} = (\{x_i\}, x_0, y, \{\widehat{x}_i\}, \widehat{x}_0)$.

If we take $\eta = \gamma_1 - \gamma_2 > 0, \widehat{w}^k = (\{x_i^k\}, x_0^k, y^k, \{x_i^{k-1}\}, x_0^{k-1})$, then

$$\widehat{L}_\rho\left(\widehat{w}^k\right) = L_\rho\left(w^k\right) + \sum_{i=1}^N \gamma_2\left\|x_i^k - x_i^{k-1}\right\|^2 + \gamma_2\left\|x_0^k - x_0^{k-1}\right\|^2. \tag{24}$$

From Lemma 2, we have

$$\widehat{L}_\rho\left(\widehat{w}^{k+1}\right) + \sum_{i=1}^N \eta\left\|x_i^{k+1}-x_i^k\right\|^2 + \eta\left\|x_0^{k+1}-x_0^k\right\|^2 \le \widehat{L}_\rho\left(\widehat{w}^k\right), \tag{25}$$

which implies that the whole sequence $\left\{\widehat{L}_\rho\left(\widehat{w}^k\right)\right\}_{k\ge1}$ is monotonically nonincreasing. It is importance for our convergence analysis.

Lemma 3. *If the sequence $\left\{w^k = (\{x_i^k\}, x_0^k, y^k)^T\right\}$ is bounded, then $\sum_{k=0}^{+\infty}\|w^{k+1}-w^k\|^2 < +\infty$.*

Proof. Since the sequence $\{w^k\}$ is bounded, there exists a subsequence $\left\{\widehat{w}^{k_j}\right\}$ such that $\lim_{j\longrightarrow+\infty}(\widehat{w}^{k_j})^{j\longrightarrow+\infty} = \widehat{w}^*$. Since $g(x)$ is lower semicontinuous, $f_i\colon R^n \longrightarrow R\cup\{+\infty\}$ is Lipschitz differentiable, and the function $\widehat{L}_\rho(\cdot)$ is lower semicontinuous, which leads to $\liminf_{j\longrightarrow+\infty}\widehat{L}_\rho(\widehat{w}^{k_j}) \ge \widehat{L}_\rho(\widehat{w}^*)$; thus, $\widehat{L}_\rho(\widehat{w}^{k_j})$ is bounded from below. From Lemma 2, we know that $\widehat{L}_\rho(\widehat{w}^k)$ is nonincreasing; thus, $\left\{\widehat{L}_\rho(\widehat{w}^{k_j})\right\}$ is convergent and $\widehat{L}_\rho(\widehat{w}^k) \ge \widehat{L}_\rho(\widehat{w}^*)$ for each k.

From Lemma 2, it yields

$$\eta\left(\sum_{i=1}^N\left\|x_i^{k+1}-x_i^k\right\|^2 + \left\|x_0^{k+1}-x_0^k\right\|^2\right) \le \widehat{L}_\rho\left(\widehat{w}^k\right) - \widehat{L}_\rho\left(\widehat{w}^{k+1}\right). \tag{26}$$

Hence,

$$\sum_{k=1}^t \eta\left(\sum_{i=1}^N\left\|x_i^{k+1}-x_i^k\right\|^2 + \left\|x_0^{k+1}-x_0^k\right\|^2\right) \le \widehat{L}_\rho\left(\widehat{w}^1\right) - \widehat{L}_\rho\left(\widehat{w}^{t+1}\right) \le \widehat{L}_\rho\left(\widehat{w}^1\right) - \widehat{L}_\rho\left(\widehat{w}^*\right). \tag{27}$$

Consequently, $\sum_{k=0}^{+\infty}\|w^{k+1}-w^k\|^2 < +\infty$. □

Lemma 4. *There exists* $\delta>0$ *such that* $d(0,\partial\widehat{L}_\rho(\widehat{w}^{k+1}))\le\delta\beta_k$ *for each* $k\in\mathrm{N}$*, where*

$$\beta_k=\sum_{i=1}^{N}\left\|x_i^{k+1}-x_i^k\right\|+\left\|x_0^{k+1}-x_0^k\right\|+\sum_{i=1}^{N}\left\|x_i^k-x_i^{k-1}\right\|+\left\|x_0^k-x_0^{k-1}\right\|. \tag{28}$$

Proof. From the definition of $\widehat{L}_\rho(\widehat{w})$, we have

$$\begin{cases}\partial_{x_i}\widehat{L}_\rho(\widehat{w}^{k+1})=\nabla_{x_i}f_i(x_i^{k+1})+y_i^{k+1}+\rho(x_i^{k+1}-x_0^{k+1})+2\gamma_2(x_i^{k+1}-x_i^k),\\ \partial_{x_0}\widehat{L}_\rho(\widehat{w}^{k+1})=\partial g(x_0^{k+1})-\sum_{i=1}^{N}(y_i^{k+1}+\rho(x_i^{k+1}-x_0^{k+1}))+2\gamma_2(x_0^{k+1}-x_0^k),\\ \partial_y\widehat{L}_\rho(\widehat{w}^{k+1})=\dfrac{1}{\rho}\sum_{i=1}^{N}(y_i^{k+1}-y_i^k)-\dfrac{\tau_k}{\rho}\sum_{i=1}^{N}(x_i^{k+1}-x_i^k)+\dfrac{\tau_k}{\rho}\theta_k\sum_{i=1}^{N}(x_i^k-x_i^{k-1}),\\ \partial_{\widehat{x}_i}\widehat{L}_\rho(\widehat{w}^{k+1})=-2\gamma_2\sum_{i=1}^{N}(x_i^{k+1}-x_i^k),\\ \partial_{\widehat{x}_0}\widehat{L}_\rho(\widehat{w}^{k+1})=-2\gamma_2(x_0^{k+1}-x_0^k).\end{cases} \tag{29}$$

From Lemma 1 and the optimality conditions, we get

$$\begin{cases}0=\nabla_{x_i}f_i(x_i^{k+1})+y_i^k+\rho(x_i^{k+1}-x_0^{k+1})+\tau_k(x_i^{k+1}-z_i^k),\\ 0\in\partial g(x_0^{k+1})-\sum_{i=1}^{N}y_i^k-\rho\sum_{i=1}^{N}(x_i^k-x_0^{k+1})+\tau_k(x_0^{k+1}-z_0^k),\\ y_i^{k+1}=y_i^k+\rho(x_i^{k+1}-x_0^{k+1})+\tau_k(x_i^{k+1}-z_i^k).\end{cases} \tag{30}$$

From (29) and (30), we obtain

$$(\alpha_1^k,\alpha_2^k,\alpha_3^k,\alpha_4^k,\alpha_5^k)^{\mathrm{T}}\in\partial\widehat{L}_\rho(\widehat{w}^{k+1}), \tag{31}$$

where

$$\begin{cases}\alpha_1^k=y_i^{k+1}-y_i^k+(2\gamma_2-\tau_k)(x_i^{k+1}-x_i^k)+\tau_k\theta_k(x_i^k-x_i^{k-1}),\\ \alpha_2^k=-\sum_{i=1}^{N}(y_i^{k+1}-y_i^k)-\rho\sum_{i=1}^{N}(x_i^{k+1}-x_i^k)+(2\gamma_2-\tau_k)(x_0^{k+1}-x_0^k)+\tau_k\theta_k(x_0^k-x_0^{k-1}),\\ \alpha_3^k=\dfrac{1}{\rho}\sum_{i=1}^{N}(y_i^{k+1}-y_i^k)-\dfrac{\tau_k}{\rho}\sum_{i=1}^{N}(x_i^{k+1}-x_i^k)+\dfrac{\tau_k}{\rho}\theta_k\sum_{i=1}^{N}(x_i^k-x_i^{k-1}),\\ \alpha_4^k=-2\gamma_2\sum_{i=1}^{N}(x_i^{k+1}-x_i^k),\\ \alpha_5^k=-2\gamma_2(x_0^{k+1}-x_0^k).\end{cases} \tag{32}$$

Thus,

$$d\left(0,\partial\widehat{L}_\rho\left(\widehat{w}^{k+1}\right)\right)\leq\left\|\left(\alpha_1^k,\alpha_2^k,\alpha_3^k,\alpha_4^k,\alpha_5^k\right)^{\mathrm{T}}\right\|. \quad (33)$$

It follows from Assumption 1 and Lemma 1 that there exists $\delta>0$ such that $d(0,\partial\widehat{L}_\rho(\widehat{w}^{k+1}))\leq\delta\beta_k$, for each $k\in\mathrm{N}$. □

Lemma 5. *Let $\Gamma(\{\widehat{w}^k\})$ denote the cluster point set of $\{\widehat{w}^k\}$. Then, $\Gamma(\{\widehat{w}^k\})$ is a nonempty compact set, and $\lim_{k\longrightarrow+\infty}d(\widehat{w}^k,\Gamma(\{\widehat{w}^k\}))=0$.*

And if $\widehat{w}^*=(\{x_i^*\},x_0^*,y^*,\{\widehat{x}_i^*\},\widehat{x}_0^*)^{\mathrm{T}}\in\Gamma(\{\widehat{w}^k\})$, then $w^*=(\{x_i^*\},x_0^*,y^*)^{\mathrm{T}}$ is a critical point of the Lagrangian function L of the problem (2). Moreover, $\widehat{L}_\rho(\cdot)$ is finite and constant on $\Gamma(\{\widehat{w}^k\})$ and $\inf_{k\in\mathrm{N}}\widehat{L}_\rho(\widehat{w}^k)=\lim_{k\longrightarrow+\infty}\widehat{L}_\rho(\widehat{w}^k)$.

Proof. In view of the definition of $\Gamma(\{\widehat{w}^k\})$, it is true that $\Gamma(\{\widehat{w}^k\})$ is nonempty and compact, and $\lim_{k\longrightarrow+\infty}d(\widehat{w}^k,\Gamma(\{\widehat{w}^k\}))=0$.

Let $\widehat{w}^*=(\{x_i^*\},x_0^*,y^*,\{\widehat{x}_i^*\},\widehat{x}_0^*)^{\mathrm{T}}\in\Gamma(\{\widehat{w}^k\})$. Then, there exists a subsequence $\{\widehat{w}^{k_j+1}\}$ of $\{\widehat{w}^k\}$ converging to $\{\widehat{w}^*\}$. Since $\|w^{k+1}-w^k\|\longrightarrow0$ $(k\longrightarrow+\infty)$, we have $\lim_{j\longrightarrow+\infty}\widehat{w}^{k_j+1}=\widehat{w}^*$.

Since $y_i^{k_j+1}=y_i^{k_j}+\rho(x_i^{k_j+1}-x_0^{k_j+1})+\tau_k(x_i^{k_j+1}-z_i^{k_j})$, we have $x_i^*-x_0^*=0$.

Let $m_k(x_0)=L_\rho(\{x_i^k\},x_0,y^k)+\tau_k/2\|x_0-z_0^k\|^2$. From Lemma 2, we have

$$\begin{aligned}&g\left(x_0^{k_j+1}\right)+\sum_{i=1}^N\langle y_i^{k_j},x_i^{k_j}-x_0^{k_j+1}\rangle+\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^{k_j}-x_0^{k_j+1}\right\|^2+\frac{\tau_k}{2}\left\|x_0^{k_j+1}-z_0^{k_j}\right\|^2\\&\leq g\left(x_0^*\right)+\sum_{i=1}^N\langle y_i^{k_j},x_i^{k_j}-x_0^*\rangle+\frac{\rho}{2}\sum_{i=1}^N\left\|x_i^{k_j}-x_0^*\right\|^2+\frac{\tau_k}{2}\left\|x_0^*-z_0^{k_j}\right\|^2.\end{aligned} \quad (34)$$

That is, $m_{k_j}(x_0^{k_j+1})\leq m_{k_j}(x_0^*)$.

Thus,

$$\begin{aligned}\limsup_{j\longrightarrow+\infty}m_{k_j}\left(x_0^{k_j+1}\right)&=\limsup_{j\longrightarrow+\infty}g\left(x_0^{k_j+1}\right)\leq\limsup_{j\longrightarrow+\infty}m_{k_j}\left(x_0^*\right)\\&=g\left(x_0^*\right).\end{aligned} \quad (35)$$

Since $g(x)$ is proper lower semicontinuous, we obtain

$$\liminf_{j\longrightarrow+\infty}g\left(x_0^{k_j+1}\right)\geq g\left(x_0^*\right). \quad (36)$$

From above, we get

$$\lim_{j\longrightarrow+\infty}g\left(x_0^{k_j+1}\right)=g\left(x_0^*\right). \quad (37)$$

Together with the continuity of $f_i\,(i=1,2,\ldots,N)$ and the closeness of ∂g, we obtain

$$\begin{cases}\nabla_{x_i}f_i\left(x_i^*\right)=-y_i^*,\\ \sum_{i=1}^N y_i^*\in\partial g\left(x^*\right),\\ x_i^*-x_0^*=0.\end{cases} \quad (38)$$

Thus, $\widehat{w}^*$ is a critical point of the Lagrange function L of the problem (2).

From (37) and Lemma 5, we have

$$\begin{aligned}\lim_{j\longrightarrow+\infty}\widehat{L}_\rho\left(\widehat{w}^{k_j+1}\right)&=\widehat{L}_\rho\left(\widehat{w}^*\right)\\&=L\left(w^*\right).\end{aligned} \quad (39)$$

Therefore, from (39) and the descent of $\{\widehat{L}_\rho(\widehat{w}^k)\}_{k\in N}$, we obtain

$$\lim_{k\longrightarrow+\infty}\widehat{L}_\rho\left(\widehat{w}^k\right)=\widehat{L}_\rho\left(\widehat{w}^*\right). \quad (40)$$

Thus, $\widehat{L}_\rho(\cdot)$ is constant on $\Gamma(\{\widehat{w}^k\})$. Moreover, $\inf_{k\in N}\widehat{L}_\rho(\widehat{w}^k)=\lim_{k\longrightarrow+\infty}\widehat{L}_\rho(\widehat{w}^k)$. □

Theorem 1. *Let $\widehat{L}_\rho(\widehat{w}^k)$ be the KL property at each point of $\Gamma(\{\widehat{w}^k\})$. Then, the bounded sequences $\{w^k\}$ converges to a critical point of $L(\cdot)$. Moreover,*

$$\sum_{k=0}^{+\infty}\left\|w^{k+1}-w^k\right\|<+\infty. \quad (41)$$

Proof. By Lemma 5, we have $\lim_{k\longrightarrow+\infty}\widehat{L}_\rho(\widehat{w}^k)=\widehat{L}_\rho(\widehat{w}^*)$, for all $w^*\in\Gamma(\{\widehat{w}^k\})$. We consider the following two cases:

(i) If there exist an integer k_0 such that $\widehat{L}_\rho(\widehat{w}^{k_0})=\widehat{L}_\rho(\widehat{w}^*)$. From Lemma 2, for all $k>k_0$, we have

$$\begin{aligned}&\eta\left(\sum_{i=1}^N\left\|x_i^{k+1}-x_i^k\right\|^2+\left\|x_0^{k+1}-x_0^k\right\|^2\right)\\&\leq\widehat{L}_\rho\left(\widehat{w}^k\right)-\widehat{L}_\rho\left(\widehat{w}^{k+1}\right)\\&\leq\widehat{L}_\rho\left(\widehat{w}^{k_0}\right)-\widehat{L}_\rho\left(\widehat{w}^*\right).\end{aligned} \quad (42)$$

Thus, for any $k > k_0$, we have $x_i^{k+1} = x_i^k, i = 1, 2, \cdots, N, x_0^{k+1} = x_0^k$; therefore, for any $k > k_0 + 1$, it follows that $\widehat{w}^{k+1} = \widehat{w}^k$ and the assertion holds.

(ii) Assume that $\widehat{L}_\rho(\widehat{w}^k) > \widehat{L}_\rho(\widehat{w}^*)$ for all $k \in N$. Since $\lim_{k\longrightarrow+\infty} d(\widehat{w}^k, \Gamma(\{\widehat{w}^k\})) = 0$, it follows that for any give $\varepsilon_1 > 0$, there exists $k_1 > 0$, such that $d(\widehat{w}^k, \Gamma(\{\widehat{w}^k\})) < \varepsilon_1$. Again since $\lim_{k\longrightarrow+\infty} \widehat{L}_\rho(\widehat{w}^k) = \widehat{L}_\rho(\widehat{w}^*)$, for give $\varepsilon_2 > 0$, there exists $k_2 > 0$, such that $\widehat{L}_\rho(\widehat{w}^k) < \widehat{L}_\rho(\widehat{w}^*) + \varepsilon_2$, for all $k > k_2$.

Thus, when $k > \widehat{k} = \max\{k_1, k_2\}$, we have

$$d\left(\widehat{w}^k, \Gamma\left(\left\{\widehat{w}^k\right\}\right)\right) < \varepsilon_1, \widehat{L}_\rho(\widehat{w}^*) < \widehat{L}_\rho\left(\widehat{w}^k\right) < \widehat{L}_\rho(\widehat{w}^*) + \varepsilon_2. \tag{43}$$

In view of $\Gamma(\{\widehat{w}^k\})$ is nonempty compact set, $\widehat{L}_\rho(\cdot)$ is constant on $\Gamma(\{\widehat{w}^k\})$. By Definition 2, we have $\varphi'(C) d(0, \partial\widehat{L}_\rho(\widehat{w}^k)) \geq 1$, for all $k > \widehat{k}$.

$$\frac{1}{\varphi'(C)} \leq d\left(0, \partial\widehat{L}_\rho\left(\widehat{w}^k\right)\right). \tag{44}$$

From the concavity of ϕ, we have

$$\begin{aligned} &\varphi\left(\widehat{L}_\rho\left(\widehat{w}^k\right) - \widehat{L}_\rho(\widehat{w}^*)\right) - \varphi\left(\widehat{L}_\rho\left(\widehat{w}^{k+1}\right) - \widehat{L}_\rho(\widehat{w}^*)\right) \\ &\quad \geq \varphi'\left(\widehat{L}_\rho\left(\widehat{w}^k\right) - \widehat{L}_\rho(\widehat{w}^*)\right)\left(\widehat{L}_\rho\left(\widehat{w}^k\right) - \widehat{L}_\rho\left(\widehat{w}^{k+1}\right)\right). \end{aligned} \tag{45}$$

Since $\varphi'(\widehat{L}_\rho(\widehat{w}^k) - \widehat{L}_\rho(\widehat{w}^*)) > 0$ and Lemma 2, we obtain

$$\begin{aligned} &(\gamma_1 - \gamma_2)\left(\sum_{i=1}^N \left\|x_i^{k+1} - x_i^k\right\|^2 + \left\|x_0^{k+1} - x_0^k\right\|^2\right) \\ &\quad \leq \widehat{L}_\rho\left(\widehat{w}^k\right) - \widehat{L}_\rho\left(\widehat{w}^{k+1}\right) \\ &\quad \leq \frac{\varphi\left(\widehat{L}_\rho\left(\widehat{w}^k\right) - \widehat{L}_\rho(\widehat{w}^*)\right) - \varphi\left(\widehat{L}_\rho\left(\widehat{w}^{k+1}\right) - \widehat{L}_\rho(\widehat{w}^*)\right)}{\varphi'\left(\widehat{L}_\rho\left(\widehat{w}^k\right) - \widehat{L}_\rho(\widehat{w}^*)\right)} \\ &\quad \leq \delta\beta_k\left(\varphi\left(\widehat{L}_\rho\left(\widehat{w}^k\right) - \widehat{L}_\rho(\widehat{w}^*)\right) - \varphi\left(\widehat{L}_\rho\left(\widehat{w}^{k+1}\right) - \widehat{L}_\rho(\widehat{w}^*)\right)\right). \end{aligned} \tag{46}$$

Let $\Phi_{a,b} = \varphi(\widehat{L}_\rho(\widehat{w}^a) - \widehat{L}_\rho(\widehat{w}^*)) - \varphi(\widehat{L}_\rho(\widehat{w}^b) - \widehat{L}_\rho(\widehat{w}^*))$. Thus,

$$\sum_{i=1}^N \left\|x_i^{k+1} - x_i^k\right\|^2 + \left\|x_0^{k+1} - x_0^k\right\|^2 \leq \frac{\delta}{\eta}\beta_k \Phi_{k,k+1}, \tag{47}$$

for all $k > \widehat{k}$. That is,

$$\begin{aligned} &(N+1)\left(\sum_{i=1}^N \left\|x_i^{k+1} - x_i^k\right\| + \left\|x_0^{k+1} - x_0^k\right\|\right) \\ &\quad \leq (N+1)\sqrt{N+1}\left(\sum_{i=1}^N \left\|x_i^{k+1} - x_i^k\right\|^2 + \left\|x_0^{k+1} - x_0^k\right\|^2\right)^{\frac{1}{2}} \\ &\quad \leq 2\sqrt{\beta_k}\sqrt{\frac{(N+1)^3\delta}{4(\gamma_1 - \gamma_2)}\Phi_{k,k+1}}, \end{aligned} \tag{48}$$

for all $k > \widehat{k}$.

Since $a + b \geq 2\sqrt{ab}\,(a, b > 0)$, we have

$$2\sqrt{\beta_k}\sqrt{\frac{(N+1)^3\delta}{4(\gamma_1 - \gamma_2)}\Phi_{k,k+1}} \leq \beta_k + \frac{(N+1)^3\delta}{4(\gamma_1 - \gamma_2)}\Phi_{k,k+1}. \tag{49}$$

From (48) and (49), we obtain

$$(N+1)\left(\sum_{i=1}^N \left\|x_i^{k+1} - x_i^k\right\| + \left\|x_0^{k+1} - x_0^k\right\|\right) \leq \beta_k + \frac{(N+1)^3\delta}{4(\gamma_1 - \gamma_2)}\Phi_{k,k+1}. \tag{50}$$

Summing up the above formula for $k = \widehat{k} + 1, \ldots, p$, we have

$$\sum_{k=\widehat{k}+1}^p (N+1)\left(\sum_{i=1}^N \left\|x_i^{k+1} - x_i^k\right\| + \left\|x_0^{k+1} - x_0^k\right\|\right) \leq \sum_{k=\widehat{k}+1}^p \left(\beta_k + \frac{(N+1)^3\delta}{4(\gamma_1 - \gamma_2)}\Phi_{k,k+1}\right). \tag{51}$$

Notice that $\phi(\widehat{L}(\widehat{w}^{p+1}) - \widehat{L}(\widehat{w}^*)) > 0$, it is easy to get

$$\begin{aligned}
&\sum_{k=\widehat{k}+1}^{p}\left(\sum_{i=1}^{N}\left\|x_i^{k+1} - x_i^k\right\| + \left\|x_0^{k+1} - x_0^k\right\|\right) \\
&\leq \gamma_k + \frac{(N+1)^3\delta}{4(\gamma_1 - \gamma_2)}\varphi\left(\widehat{L}_\rho\left(\widehat{w}^{\widehat{k}+1}\right) - \widehat{L}_\rho(\widehat{w}^*)\right) - \varphi\left(\widehat{L}_\rho\left(\widehat{w}^{p+1}\right) - \widehat{L}_\rho(\widehat{w}^*)\right) \\
&\leq \gamma_k + \frac{(N+1)^3\delta}{4(\gamma_1 - \gamma_2)}\left(\varphi\left(\widehat{L}_\rho\left(\widehat{w}^{\widehat{k}+1}\right) - \widehat{L}_\rho(\widehat{w}^*)\right)\right),
\end{aligned} \tag{52}$$

where

$$\gamma_k = \sum_{i=1}^{N}\left\|\widehat{x_i^{k+1}} - \widehat{x_i^k}\right\| + \left\|\widehat{x_0^{k+1}} - \widehat{x_0^k}\right\|. \tag{53}$$

Thus,

$$\sum_{k=0}^{+\infty}\left(\sum_{i=1}^{N}\left\|x_i^{k+1} - x_i^k\right\| + \left\|x_0^{k+1} - x_0^k\right\|\right) \leq +\infty. \tag{54}$$

From Lemma 1, we get

$$\sum_{k=0}^{+\infty}\left\|w^{k+1} - w^k\right\| < +\infty. \tag{55}$$

By Lemma 5, we conclude that the sequences $\{w^k\}$ converge to a critical point of $L(\cdot)$.

4. Numerical Experiment

In this section, we present the results of a simple numerical example to verify the effectiveness of Algorithm 1. We consider the following compressive sense problem, which takes the following form:

$$\begin{aligned}
&\min \lambda\left\|x_1\right\|_0 + \lambda\left\|x_2\right\|_1 + \frac{1}{2}\left\|Ax_0 - b\right\|^2, \\
&\text{s.t. } x_i - x_0 = 0, \forall i = 1, 2, x_0 \in X,
\end{aligned} \tag{56}$$

where A is a $m \times n$ feature matrix, $b \in R^m$ is a response vector, and λ is a regular parameter. In general, problem (56) is NP-hard. In order to overcome this difficulty, one may relax l_0 norm to the $l_{1/2}$ norm, considering the following nonconvex problem:

$$\begin{aligned}
&\min \lambda\left\|x_1\right\|_{1/2}^{1/2} + \lambda\left\|x_2\right\|_1 + \frac{1}{2}\left\|Ax_0 - b\right\|^2, \\
&\text{s.t. } x_i - x_0 = 0, \forall i = 1, 2, x_0 \in X.
\end{aligned} \tag{57}$$

Let $f_1(x_1) = \lambda\|x_1\|_{(1/2)}^{(1/2)}$, $f_2(x_2) = \lambda\|x_2\|_1$ and $g(x_0) = (1/2)\|Ax_0 - b\|^2$, $X = \mathrm{R}^{\mathrm{N}}$. We now focus on applying Algorithm 1 to solve problem (57) with the suitable parameters. The iterative processes are as follows:

$$\begin{cases}
x_0^{k+1} = \underset{x_0 \in X}{\operatorname{argmin}}\left\{g(x_0) + \sum_{i=1}^{N}\langle y_i^k, x_i^k - x_0\rangle + \frac{\rho}{2}\sum_{i=1}^{N}\left\|x_i^k - x_0\right\|^2 + \frac{\tau_k}{2}\left\|x_0 - z_0^k\right\|^2\right\}, \\
x_1^{k+1} = \underset{x_1}{\operatorname{argmin}}\left\{f_1(x_1) + \langle y_1^k, x_1 - x_0^{k+1}\rangle + \frac{\rho}{2}\left\|x_1 - x_0^{k+1}\right\|^2 + \frac{\tau_k}{2}\left\|x_1 - z_1^k\right\|^2\right\}, \\
x_2^{k+1} = \underset{x_2}{\operatorname{argmin}}\left\{f_2(x_2) + \langle y_2^k, x_2 - x_0^{k+1}\rangle + \frac{\rho}{2}\left\|x_2 - x_0^{k+1}\right\|^2 + \frac{\tau_k}{2}\left\|x_2 - z_2^k\right\|^2\right\}, \\
y_1^{k+1} = y_1^k + \rho\left(x_1^{k+1} - x_0^{k+1}\right) + \tau_k\left(x_1^{k+1} - z_1^k\right), \\
y_2^{k+1} = y_2^k + \rho\left(x_2^{k+1} - x_0^{k+1}\right) + \tau_k\left(x_2^{k+1} - z_2^k\right).
\end{cases} \tag{58}$$

Table 1: Comparison among two algorithms under different parameters.

θ	Iter.	Time (s)	ρ	τ_k
Case 1 (ADMM) 0	191	22.725124		
Case 2 (IADMM) $1/2 - l_f^2 + 1/\rho$	137	16.41164	600	$3l_f^2/\rho - 2\theta\rho - l_f^2 - 2\theta$
Case 3 (IADMM) $1/3 - l_f^2 + 1/\rho$	159	18.777039		
Case 4 (ADMM) 0	191	22.926591		
Case 5 (IADMM) $1/2 - l_f^2 + 1/\rho$	163	19.139843	500	$3l_f^2/\rho - 2\theta\rho - l_f^2 - 2\theta$
Case 6 (IADMM) $1/3 - l_f^2 + 1/\rho$	178	21.54476		

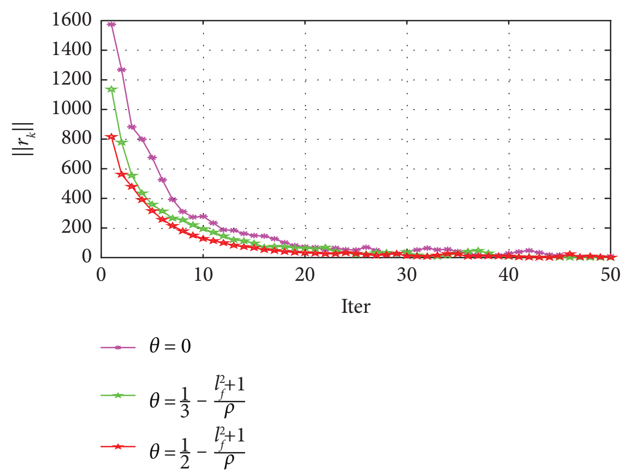

Figure 1: For $\rho = 600$ and $\tau_k = 3l_f^2/\rho - 2\theta\rho - l_f^2 - 2\theta$.

Simplifying the procedures (58), we obtain the closed-form iterative formulas:

$$\begin{cases} x_0^{k+1} = \left(A^T A + (\tau_k + \rho N) I\right)^{-1} \left(A^T b + \rho \sum_{i=1}^{N} x_i^k + \sum_{i=1}^{N} y_i^k + \tau_k z_0^k\right), \\ x_1^{k+1} = H\left(\dfrac{\tau_k z_1^k - y_1^k + \rho x_0^{k+1}}{\rho + \tau_k}, \dfrac{2\lambda}{\rho + \tau_k}\right), \\ x_2^{k+1} = S\left(\dfrac{\tau_k z_2^k - y_2^k + \rho x_0^{k+1}}{\rho + \tau_k}, \dfrac{\lambda}{\rho + \tau_k}\right), \\ y_1^{k+1} = y_1^k + \rho\left(x_1^{k+1} - x_0^{k+1}\right) + \tau_k\left(x_1^{k+1} - z_1^k\right), \\ y_2^{k+1} = y_2^k + \rho\left(x_2^{k+1} - x_0^{k+1}\right) + \tau_k\left(x_2^{k+1} - z_2^k\right), \end{cases} \quad (59)$$

where $z_0^k = x_0^k + \theta_k (x_0^k - x_0^{k-1}), z_i^k = x_i^k + \theta_k (x_i^k - x_i^{k-1})$, $i = 1, 2$, $H(\cdot,\cdot)$ is the half shrinkage operator[16], and $S(A,\cdot)$ indicates the soft shrinkage operator imposed on the entries of A.

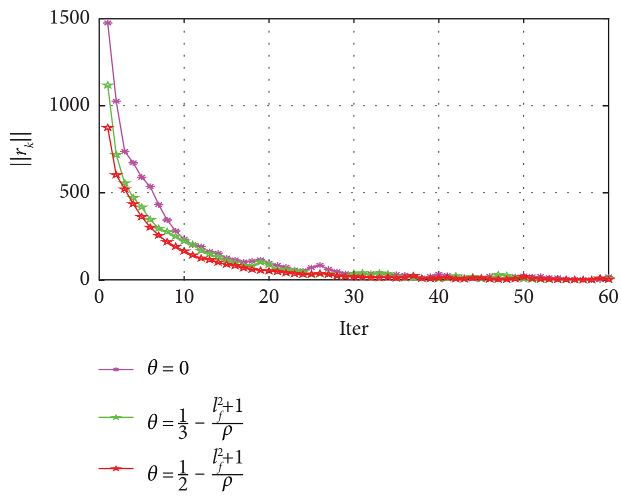

FIGURE 2: For $\rho = 500$ and $\tau_k = 3l_f^2/\rho - 2\theta\rho - l_f^2 - 2\theta$.

The experimental data are generated as follows. We use distributed computing toolbox in MATLAB, and the purpose is to achieve simple distributed computing. Suppose the feature matrix A is standard normal distribution $N(0, 1)\ m * n$. Select sparse vector $a \in R^n$ from the $N(0, 1)$ distribution. The parameters b and λ are set as $b_i = A_i a + \varepsilon$ and $\lambda = 0.01\|A_i^T * b_i\|$, where the noise vector $\varepsilon \in N(0, 0.01I)$. The variables x_0, x_i, y_i were initialized to be zero. The primal residual is defined as $r^k = \sum_{i=1}^N x_i^k$. We employ $r^k \leq \varepsilon$ as the stopping criteria, where $\varepsilon = 10^{-4}$. The numerical results are reported in Table 1. We report the number of iterations ("Iter.") and the computing time in seconds ("Time") for the algorithms with different parameters under the dimension $m = 2500$, $n = 1000$.

The values of $\|r^k\|$ with the iterations are plotted in Figures 1 and 2.

where $l_f = \max\{l_{f_i}\}_{i=1,2,\cdots,N}$.

From Table 1, and Figures 1 and 2, we can see that ADMM converges more slowly than IADMM since "Iter." of ADMM bigger than that of IADMM under the same conditions. Finally, numerical results show that the algorithm is feasible and effective.

5. Conclusion

In this paper, inspired by the application of nonconvex global consensus problem with regularization, we propose multiblock inertial ADMM algorithm for solving certain nonconvex global consensus problems. We have proven its convergence under some suitable conditions, and it turns out that any cluster point of the sequence generated by the proposed algorithm is a critical point. Numerical experiment is conducted to illustrate the effectiveness of the multiblock inertial ADMM (IADMM) algorithm. Its potential of the flexible multiblock inertial ADMM to analyze and design other types of nonconvex case, as well as a more thorough computational study, are topics of our further research.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (72071130; 71901145); National Social Science Fund Major Project of China (21&ZD200; 20&ZD199); Humanities and Social Sciences Research Project of the Ministry of Education (20YJC820030) and China Postdoctoral Science Foundation (2021M692047).

References

[1] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, pp. 1–122, 2010.

[2] G. Li and T. K. Pong, "Global convergence of splitting methods for nonconvex composite optimization," *SIAM Journal on Optimization*, vol. 25, no. 4, pp. 2434–2460, 2015.

[3] R. Glowinski and A. Marroco, "Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires," *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique*, vol. 9, no. 2, pp. 41–76, 1975.

[4] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40, 1976.

[5] L. Yang, T. K. Pong, and X. Chen, "Alternating direction method of multipliers for a class of nonconvex and nonsmooth problems with applications to background/foreground extraction," *SIAM Journal on Imaging Sciences*, vol. 10, no. 1, pp. 74–110, 2017.

[6] M. Hong, Z. Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 337–364, 2016.

[7] K. Guo, D. Han, D. Z. W. Wang, and T. Wu, "Convergence of ADMM for multi-block nonconvex separable optimization models," *Frontiers of Mathematics in China*, vol. 12, no. 5, pp. 1139–1162, 2017.

[8] C. H. Chen, R. H. Chan, S. Q. Ma, and J. F. Yang, "Inertial proximal ADMM for linearly constrained separable convex optimization," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2239–2267, 2015.

[9] A. Moudafi and E. Elissabeth, "Approximate inertial proximal methods using the enlargement of maximal monotone operators," *International Journal of Pure and Applied Mathematics*, vol. 5, no. 3, pp. 283–299, 2003.

[10] M. Chao, Y. Zhang, and J. Jian, "An inertial proximal alternating direction method of multipliers for nonconvex optimization," *International Journal of Computer Mathematics*, vol. 98, 2021.

[11] K. Kurdyka, "On gradients of functions definable in o-minimal structures," *Annales de l'Institut Fourier*, vol. 48, no. 3, pp. 769–783, 1998.

[12] S. Lojasiewicz, "Sur la geométrie semi-et sous-analytique," *Annales de l'Institut Fourier*, vol. 43, no. 5, pp. 1575–1595, 1993.

[13] S. Lojasiewicz, "Une propriété topologique des sous-ensembles analytiques réels," in *Les Équations Aux Dérivées Partielles*, Editions du centre National de la Recherche Scientifique, Paris, France, 1963.

[14] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet, "Characterizations of Łojasiewicz inequalities: subgradient flows, talweg, convexity," *Transactions of the American Mathematical Society*, vol. 362, no. 6, pp. 3319–3363, 2009.