Research Article

Convergence of Linear Bregman ADMM for Nonconvex and Nonsmooth Problems with Nonseparable Structure

Miantao Chao ¹, Zhao Deng,¹ and Jinbao Jian ²

¹Department of Mathematics, Guangxi University, Nanning 530004, China

²Guangxi University for Nationalities, Nanning, China

Correspondence should be addressed to Jinbao Jian; jianjb@gxu.edu.cn

Received 5 September 2019; Revised 6 January 2020; Accepted 27 January 2020; Published 26 February 2020

Academic Editor: Zhile Yang

Copyright © 2020 Miantao Chao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The alternating direction method of multipliers (ADMM) is an effective method for solving two-block separable convex problems and its convergence is well understood. When either the involved number of blocks is more than two, or there is a nonconvex function, or there is a nonseparable structure, ADMM or its directly extend version may not converge. In this paper, we proposed an ADMM-based algorithm for nonconvex multiblock optimization problems with a nonseparable structure. We show that any cluster point of the iterative sequence generated by the proposed algorithm is a critical point, under mild condition. Furthermore, we establish the strong convergence of the whole sequence, under the condition that the potential function satisfies the Kurdyka–Łojasiewicz property. This provides the theoretical basis for the application of the proposed ADMM in the practice. Finally, we give some preliminary numerical results to show the effectiveness of the proposed algorithm.

1. Introduction

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

$$\begin{aligned} \min \quad & \sum_{i=1}^{N-1} f_i(x_i) + g(x_1, \dots, x_{N-1}, y), \\ \text{s.t.} \quad & \sum_{i=1}^{N-1} A_i x_i + B y = b, \end{aligned} \quad (1)$$

where $x_i \in \mathbb{R}^{n_i}$ ($i = 1, 2, \dots, N-1$), $y \in \mathbb{R}^n$ are variables vectors, $g: \mathbb{R}^l \rightarrow \mathbb{R}$ ($l = n_1 + n_2 + \dots + n_{N-1} + n$) is differentiable, and each $f_i: \mathbb{R}^{n_i} \rightarrow \mathbb{R} \cup \{+\infty\}$ is proper and lower semicontinuous, $A_i \in \mathbb{R}^{m \times n_i}$ ($i = 1, 2, \dots, N-1$), $B \in \mathbb{R}^{m \times n}$ are given matrix, and $b \in \mathbb{R}^m$.

The alternating direction method of multipliers (ADMM) is a very effective method for solving the convex two-block optimization problem [1, 2]. A natural idea is to extend ADMM to solve problem (1). However, ADMM or its directly extend version may not converge, when either the involved number of blocks is more than two, or there is a nonconvex function, or there is a nonseparable structure.

Recently, there have been a few developments on it, e.g., [3–13].

Hong et al. [6] considered the sharing and consensus problem and showed that the classical ADMM converges to the set of stationary solutions, only provided that the penalty parameter in the augmented Lagrangian is chosen to be sufficiently large. Li and Pong [8] studied the convergence of ADMM for some special two-block nonconvex models, where one of the matrices A and B is an identity matrix. Wang et al. [9, 10] studied the convergence of the nonconvex Bregman ADMM algorithm, which includes ADMM as a special case. Wang et al. [11] studied the convergence of the ADMM for nonconvex nonsmooth optimization with a nonseparable structure. Guo et al. [4, 5] studied the convergence of classical ADMM for two-block and multiblock nonconvex models where one of the matrices is an identity matrix. Yang et al. [13] studied the convergence of the ADMM for a nonconvex optimization model which come from the background/foreground extraction.

The purpose and the main contribution of this paper is to propose and prove the convergence of a new variant ADMM

for nonconvex coupled problems (1). The novelty of this paper can be summarized as follows:

- (1) Compared to the existing literature, the model in this paper is more general. There is no nonseparable structure in the models considered by [4–10, 12, 13]. Wang et al. [11] considered two scenarios. If $g(x_1, \dots, x_{N-1}, y) = g_1(x_1, \dots, x_{N-1}) + h(y)$, then (1) is the scenario 1 in [11]. If $f_i(x) \equiv 0$, for $i = 1, 2, \dots, N-1$, then (1) becomes the scenario 2 in [11]. Furthermore, in this paper, the matrices A_i ($i = 1, 2, \dots, N-1$) and B are possibly not full column or row rank.
- (2) The proposed algorithm combines linearization technology with regularization technology. Linearization technology and regularization technology can effectively reduce the difficulty of the solving subproblems.

The rest of this paper is organized as follows. In Section 2, some basic concepts and necessary preliminaries for further analysis are summarized. In Section 3, we propose the algorithm and analyze the convergence of it for 3-block nonconvex and nonsmooth coupled problems. Finally, some conclusions are made in Section 4.

2. Preliminaries

$\mathbb{R}^n$ denotes the n -dimensional Euclidean space, $\mathbb{R} \cup \{+\infty\}$ denotes the extended real number set, and $\mathbb{N}$ denotes the natural number set. The image space of a matrix $Q \in \mathbb{R}^{m \times n}$ is defined as $\text{Im } Q := \{Qx : x \in \mathbb{R}^n\}$. $P_Q(\cdot)$ denotes the Euclidean projection onto $\text{Im } Q$. If matrix $Q \neq 0$, let μ_Q denote the smallest positive singular value of the matrix QQ^T . $\|\cdot\|$ represents the Euclidean norm. $\text{dom}(f) := \{x \in \mathbb{R}^n : f(x) < +\infty\}$ is the domain of a function $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$. $\langle x, y \rangle = x^T y = \sum_{i=1}^n x_i y_i$. $[\eta_1 < f < \eta_2] := \{x : \eta_1 < f(x) < \eta_2\}$. For a set $S \subset \mathbb{R}^n$ and a point $x \in \mathbb{R}^n$, let $d(x, S) = \inf_{y \in S} \|y - x\|^2$. If $S = \emptyset$, we set $d(x, S) = +\infty$ for all $x \in \mathbb{R}^n$. For a point-to-set mapping F , its graph is defined by $\text{Graph } F := \{(x, y) : y \in F(x)\}$.

Definition 1 (see [14]). Let $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper function. If there exists $\delta > 0$ such that

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - \left(\frac{\delta \lambda(1 - \lambda)}{2}\right) \|x - y\|^2, \quad (2)$$

for all $\lambda \in (0, 1)$ and $x, y \in \text{dom } f$, then f is called strongly convex with modulus δ .

Definition 2 (see [15]). For a convex differential function $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$, the associated Bregman distance is defined as

$$\Delta_\phi(x, y) = \phi(x) - \phi(y) - \langle \nabla \phi(y), x - y \rangle, \quad (3)$$

for all $x, y \in \mathbb{R}^n$.

The Bregman distance plays an important role in iterative algorithms. The Bregman distance share many similar nice properties of the Euclidean distance. However, the Bregman distance is not a metric, since it does not satisfy the triangle inequality nor symmetry. Some examples of Bregman distance include [16]

- (i) Classical Euclidean distance: if $\phi(x) = \|x\|^2$, then $\Delta_\phi(x, y) = \|x - y\|^2$
- (ii) Itakura–Saito distance: if $\phi(x) = \sum_{i=1}^m x_i (\log x_i)$, then $\Delta_\phi(x, y) = \sum_{i=1}^m x_i (\log(x_i/y_i)) - \sum_{i=1}^m (x_i - y_i)$
- (iii) Mahalanobis distance: if $\phi(x) = \|x\|_Q^2 = x^T Q x$ with Q a symmetric positive definite matrix, then $\Delta_\phi(x, y) = \|x - y\|_Q^2$

Let us now collect some useful properties about Bregman distance.

Proposition 1 (see [15]). Let ϕ be differentiable and strongly convex function with modulus δ , then

- (i) $\Delta_\phi(x, y) \geq 0$ and $\Delta_\phi(x, y) = 0$ if and only if $x = y$
- (ii) $\Delta_\phi(x, y) \geq (\delta/2)\|x - y\|^2$ for all x and y

The following notations and definitions are quite standard and can be founded in [14, 17].

Definition 3. Let $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function.

- (i) The Fréchet subdifferential, or regular subdifferential, of f at $x \in \text{dom } f$ is

$$\hat{\partial}f(x) = \left\{ x^* : \liminf_{y \neq x} \frac{f(y) - f(x) - \langle x^*, y - x \rangle}{\|y - x\|} \geq 0 \right\}. \quad (4)$$

When $x \notin \text{dom } f$, we set $\hat{\partial}f(x) = \emptyset$.

- (ii) The limiting subdifferential, or simply the subdifferential, of f at $x \in \text{dom } f$, written $\partial f(x)$, is defined as

$$\partial f(x) = \left\{ v \in \mathbb{R}^n : \exists x^k \rightarrow x, \text{ s.t. } f(x^k) \rightarrow f(x), v^k \in \hat{\partial}f(x^k), v^k \rightarrow v \right\}. \quad (5)$$

- (iii) A point that satisfies $0 \in \partial f(x)$ is called a critical point or a stationary point of the function f . The set of critical points of f is denoted by $\text{crit } f$.

The following proposition collects some properties of the subdifferential.

Proposition 2 (see [17]). Let $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper lower semicontinuous functions. Then, the following holds:

- (i) $\widehat{\partial}f(x) \subseteq \partial f(x)$ for each $x \in \text{dom } f$. Moreover, the first set is closed and convex, while the second is closed and not necessarily convex.
- (ii) Let $\{(x^k, u^k)\} \subset \text{Graph } \partial f$ be a sequence such that it converges to (x^*, u^*) . If $\lim_{k \rightarrow +\infty} f(x^k) = f(x^*)$, then $(x^*, u^*) \in \text{Graph } \partial f$.
- (iii) If $x \in \mathbb{R}^n$ is a local minimizer of f then $0 \in \widehat{\partial}f(x) \subset \partial f(x)$.
- (iv) If $g: \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous differentiable, then $\partial(f+g)(x) = \partial f(x) + \nabla g(x)$.

The Lagrangian function of (1), with multiplier $\lambda \in \mathbb{R}^m$, is defined as

$$\begin{aligned} L(x_1, \dots, x_{N-1}, y, \lambda) = & f_1(x_1) + \dots + f_{N-1}(x_{N-1}) \\ & + g(x_1, \dots, x_{N-1}, y) - \langle \lambda, A_1 x_1 \\ & + \dots + A_{N-1} x_{N-1} + B y - b \rangle. \end{aligned} \quad (6)$$

Definition 4. If $w^* = (x_1^*, \dots, x_{N-1}^*, y^*, \lambda^*)^T$ such that

$$\begin{cases} A_i^T \lambda^* \in \partial f_i(x_i^*) + \nabla_{x_i} g(x_1^*, \dots, x_{N-1}^*, y^*), & i = 1, 2, \dots, N-1, \\ B^T \lambda^* = \nabla_y g(x_1^*, \dots, x_{N-1}^*, y^*), \\ A_1 x_1^* + \dots + A_{N-1} x_{N-1}^* + B y^* - b = 0, \end{cases} \quad (7)$$

then w^* is called a critical point or stationary point of the Lagrange function $L(x_1, \dots, x_{N-1}, y, \lambda)$.

A very important technique to prove the strong convergence of the ADMM for nonconvex optimization problems relies on the assumption that the benefit function satisfying Kurdyka-Łojasiewicz property (KL property) [18–21]. There are many functions which satisfy this inequality. Especially, when the function belongs to some functional classes, e.g., semialgebraic, real subanalytic, and log-exp (see [22–24]). It is often elementary to check that such an inequality holds.

For notational simplicity, we use Φ_η ($\eta > 0$) to denote the set of concave functions $\varphi: [0, \eta] \rightarrow (0, +\infty)$ such that

- (i) $\varphi(0) = 0$
- (ii) φ is continuous differentiable on $(0, \eta)$ and continuous at 0
- (iii) $\varphi'(s) > 0, \forall s \in (0, \eta)$.

The KL property can be described as follows.

Definition 5 (see [18–21]) (KL property). Let $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. If there exists $\eta \in (0, +\infty]$, a neighborhood U of x^* , and a

function $\varphi \in \Phi_\eta$, such that for all $x \in U \cap [f(x^*) < f < f(x^*) + \eta]$, it holds that

$$\varphi'(f(x) - f(x^*))d(0, \partial f(x)) \geq 1, \quad (8)$$

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

Lemma 1 (see [22]) (uniformized KL property). Suppose that $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous function and Ω is a compact set. If $f(x) \equiv f^*$ for all $x \in \Omega$ and satisfies the KL property at each point of Ω . Then, there exist $\varepsilon > 0, \eta > 0$, and $\varphi \in \Phi_\eta$ such that

$$\varphi'(f(x) - f^*)d(0, \partial f(x)) \geq 1, \quad (9)$$

for all $x \in \{x \in \mathbb{R}^n: d(x, \Omega) < \varepsilon\} \cap [f^* < f < f^* + \eta]$.

Lemma 2 (see [25]) (Descent lemma). Let $h: \mathbb{R}^n \rightarrow \mathbb{R}$ be a continuous differentiable function where gradient ∇h is Lipschitz continuous with the modulus $l_h > 0$, then for any $x, y \in \mathbb{R}^n$, we have

$$|h(y) - h(x) - \langle \nabla h(x), y - x \rangle| \leq \frac{l_h}{2} \|y - x\|^2. \quad (10)$$

Lemma 3 (see [26]). Let $Q \in \mathbb{R}^{r \times p}$ be a nonzero matrix and let μ_Q denote the smallest positive eigenvalue of QQ^T . Then, for every $u \in \mathbb{R}^p$, there holds

$$\|P_{Q^T} u\| \leq \frac{1}{\sqrt{\mu_Q}} \|Qu\|. \quad (11)$$

3. Algorithm and Convergence

For the convenience of analysis, we only consider the case of $N = 2$. The obtained results could naturally be generalized to the case of $N > 2$. Thus, in the rest of this paper, we consider the following nonconvex and nonsmooth 3-block optimization problem:

$$\begin{aligned} \min \quad & f_1(x_1) + f_2(x_2) + g(x_1, x_2, y), \\ \text{s.t.} \quad & A_1 x_1 + A_2 x_2 + B y = b, \end{aligned} \quad (12)$$

where $f_i: \mathbb{R}^{n_i} \rightarrow \mathbb{R} \cup \{+\infty\}$ ($i = 1, 2$) is proper and lower semicontinuous but possibly nonconvex, $g: \mathbb{R}^{n_1+n_2+n} \rightarrow \mathbb{R}$ is differentiable, $A_i \in \mathbb{R}^{m \times n_i}$ ($i = 1, 2$), and $b \in \mathbb{R}^m$.

In this paper, we present the following algorithm for (12).

Algorithm 1. LBADMM: start with $(x_1^0, x_2^0, y^0)^T$ and λ^0 . With the given iteration point $w^k = (x_1^k, x_2^k, y^k, \lambda^k)^T$, the new iteration point $w^{k+1} = (x_1^{k+1}, x_2^{k+1}, y^{k+1}, \lambda^{k+1})^T$ is given as follows:

$$\left\{ \begin{array}{l} x_1^{k+1} \in \arg \min \left\{ f_1(x_1) + (x_1 - x_1^k)^T \nabla_{x_1} g(x_1^k, x_2^k, y^k) + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k + B y^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \Delta_{\varphi_1}(x_1, x_1^k) \right\}, \\ x_2^{k+1} \in \arg \min \left\{ f_2(x_2) + (x_2 - x_2^k)^T \nabla_{x_2} g(x_1^k, x_2^k, y^k) + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2 + B y^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \Delta_{\varphi_2}(x_2, x_2^k) \right\}, \\ y^{k+1} = \arg \min \left\{ g(x_1^{k+1}, x_2^{k+1}, y) + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y - b - \frac{\lambda^k}{\beta} \right\|^2 + \Delta_{\phi}(y, y^k) \right\}, \\ \lambda^{k+1} = \lambda^k - \beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y^{k+1} - b), \end{array} \right. \quad (13)$$

where $\Delta_{\varphi_1}(x_1, x_1^k)$, $\Delta_{\varphi_2}(x_2, x_2^k)$, and $\Delta_{\phi}(y, y^k)$ are the Bregman distances associated with φ_1 , φ_2 , and ϕ , respectively.

Remark 1. Due to the different structures of the problem, the algorithm in this paper is different from the existing algorithms. In order to make use of the properties of differentiable blocks and simplify the calculation of each iteration, we linearize the differentiable part in the x_1 and x_2 subproblems. If the function $g(x_1, x_2, y)$ is only related to the variable y , that is, $g(x_1, x_2, y) = h(y)$, then the algorithm LBADMM will become the Bregman ADMM in [9, 10]. Different from [9, 10], we do not assume B is full row rank.

In this section, we always assume that the sequence $\{w^k = (x_1^k, x_2^k, y^k, \lambda^k)^T\}$ is generated by algorithm LBADMM. Let $\mu = (1/\sqrt{\mu_{B^T}})$, where μ_{B^T} denotes the smallest positive eigenvalue of $B^T B$.

Assumption 1

- (i) ∇g is l_g -Lipschitz continuous, i.e., $\|\nabla g(u) - \nabla g(v)\| \leq l_g \|u - v\|$ for all $u, v \in R^n$
- (ii) $B \neq 0$ and $\text{Im } B^{\{b\} \cup \text{Im } A}$
- (iii) $\nabla \varphi_1, \nabla \varphi_2, \nabla \phi$ are Lipschitz continuous with the modulus $l_{\varphi_1}, l_{\varphi_2}, l_{\phi}$, respectively
- (iv) $\varphi_1, \varphi_2, \phi$ strongly convex with the modulus $\delta_{\varphi_1}, \delta_{\varphi_2}, \delta_{\phi}$, and $\delta_{\varphi_1}, \delta_{\varphi_2} > l_g$
- (v) $\beta > \max\{(6\mu^2(2l_{\phi}^2 + l_g^2)/\delta_{\phi}), (6\mu^2 l_g^2/\delta_{\varphi_1} - l_g), (6\mu^2 l_g^2/\delta_{\varphi_2} - l_g)\}$

The following lemma establishes the relationship between the dual variable and the original variables.

Lemma 4. For each $k \in \mathbb{N}$,

$$\begin{aligned} \|\lambda^{k+1} - \lambda^k\|^2 &\leq 3\mu^2 l_g^2 (\|x_1^{k+1} - x_1^k\|^2 + \|x_2^{k+1} - x_2^k\|^2) \\ &\quad + 3\mu^2 (l_g^2 + l_{\phi}^2) \|y^{k+1} - y^k\|^2 + 3\mu^2 l_{\phi}^2 \|y^k - y^{k-1}\|^2. \end{aligned} \quad (14)$$

Proof. By Assumption 1 (ii) and Lemma 3, we have

$$\|\lambda^{k+1} - \lambda^k\| \leq \frac{1}{\sqrt{\mu_{B^T}}} \|B^T(\lambda^{k+1} - \lambda^k)\|. \quad (15)$$

From the optimality condition of y -subproblem in (14) yields

$$\begin{aligned} 0 &= \nabla_y g(x_1^{k+1}, x_2^{k+1}, y^{k+1}) - B^T \lambda^k + \beta B^T (A_1 x_1^{k+1} \\ &\quad + A_2 x_2^{k+1} + B y^{k+1} - b) + \nabla \phi(y^{k+1}) - \nabla \phi(y^k). \end{aligned} \quad (16)$$

Taking into account $\lambda^{k+1} = \lambda^k - \beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y^{k+1} - b)$, one has

$$B^T \lambda^{k+1} = \nabla_y g(x_1^{k+1}, x_2^{k+1}, y^{k+1}) + \nabla \phi(y^{k+1}) - \nabla \phi(y^k). \quad (17)$$

Thus,

$$\begin{aligned} \|B^T \lambda^{k+1} - B^T \lambda^k\|^2 &= \|\nabla_y g(x_1^{k+1}, x_2^{k+1}, y^{k+1}) - \nabla_y g(x_1^k, x_2^k, y^k) + \nabla \phi(y^{k+1}) - \nabla \phi(y^k) + \nabla \phi(y^{k-1}) - \nabla \phi(y^k)\|^2 \\ &\leq (\|\nabla_y g(x_1^{k+1}, x_2^{k+1}, y^{k+1}) - \nabla_y g(x_1^k, x_2^k, y^k)\| + \|\nabla \phi(y^{k-1}) - \nabla \phi(y^k)\| + \|\nabla \phi(y^{k-1}) - \nabla \phi(y^k)\|)^2 \\ &\leq 3(\|\nabla_y g(x_1^{k+1}, x_2^{k+1}, y^{k+1}) - \nabla_y g(x_1^k, x_2^k, y^k)\|^2 + \|\nabla \phi(y^{k-1}) - \nabla \phi(y^k)\|^2 + \|\nabla \phi(y^{k-1}) - \nabla \phi(y^k)\|^2) \\ &\leq 3(l_g^2 \|(x_1^{k+1}, x_2^{k+1}, y^{k+1}) - (x_1^k, x_2^k, y^k)\|^2 + l_{\phi}^2 \|y^{k+1} - y^k\|^2 + l_{\phi}^2 \|y^{k-1} - y^k\|^2). \end{aligned} \quad (18)$$

It follows from the abovementioned formula and (17) that

$$\begin{aligned}
\|\lambda^{k+1} - \lambda^k\|^2 &\leq \mu^2 \|B^T(\lambda^{k+1} - \lambda^k)\|^2 \\
&\leq 3\mu^2 \left(l_g^2 \|(x_1^{k+1}, x_2^{k+1}, y^{k+1}) - (x_1^k, x_2^k, y^k)\|^2 \right. \\
&\quad \left. + l_\phi^2 \|y^{k+1} - y^k\|^2 + l_\phi^2 \|y^{k-1} - y^k\|^2 \right) \\
&\leq 3\mu^2 l_g^2 \left(\|x_1^{k+1} - x_1^k\|^2 + \|x_2^{k+1} - x_2^k\|^2 \right) \\
&\quad + 3\mu^2 (l_g^2 + l_\phi^2) \|y^{k+1} - y^k\|^2 + 3\mu^2 l_\phi^2 \|y^k - y^{k-1}\|^2.
\end{aligned} \tag{19}$$

The proof is completed. $\square$

The augmented Lagrangian function with multiplier λ of (12) is defined as

$$L_\beta(x_1, x_2, y, \lambda) = L(x_1, x_2, y, \lambda) + \frac{\beta}{2} \|A_1 x_1 + A_2 x_2 + B y - b\|^2, \tag{20}$$

where $L(x_1, x_2, y, \lambda)$ is the Lagrangian function of (12). Let

$$\widehat{L}(x_1, x_2, y, \lambda, \widehat{y}) = L_\beta(x_1, x_2, y, \lambda) + \frac{3\mu^2 l_\phi^2}{\beta} \|y - \widehat{y}\|^2. \tag{21}$$

Let

$$\begin{aligned}
\widehat{w}^k &= (x_1^k, x_2^k, y^k, \lambda^k, y^{k-1}), w^k = (x_1^k, x_2^k, y^k, \lambda^k), u^k \\
&= (x_1^k, x_2^k, y^k)^T,
\end{aligned} \tag{22}$$

$$r_k = A_1 x_1^k + A_2 x_2^k + B y^k - b.$$

The following lemma implies the monotonicity of the sequence $\{\widehat{L}(\widehat{w}^k)\}_{k \in \mathbb{N}}$. $\square$

Lemma 5. For each $k \in \mathbb{N}$,

$$\widehat{L}(\widehat{w}^{k+1}) \leq \widehat{L}(\widehat{w}^k) - \sigma \left(\|x_1^{k+1} - x_1^k\|^2 + \|x_2^{k+1} - x_2^k\|^2 + \|y^{k+1} - y^k\|^2 \right), \tag{23}$$

where

$$\begin{aligned}
\sigma &= \min \left\{ \left(\frac{\delta_\phi}{2} \right) - \left(\frac{3\mu^2 (2l_\phi^2 + l_g^2)}{\beta} \right), \left(\frac{\delta_{\varphi_1} - l_g}{2} \right) \right. \\
&\quad \left. - \left(\frac{3\mu^2 l_g^2}{\beta} \right), \left(\frac{\delta_{\varphi_2} - l_g}{2} \right) - \left(\frac{3\mu^2 l_g^2}{\beta} \right) \right\}.
\end{aligned} \tag{24}$$

Proof. From (17), we have

$$\begin{aligned}
&f_1(x_1^{k+1}) + \langle x_1^{k+1} - x_1^k, \nabla_{x_1} g(u^k) \rangle - \langle \lambda^k, A_1 x_1^{k+1} + A_2 x_2^k + B y^k - b \rangle + \frac{\beta}{2} \|A_1 x_1^{k+1} + A_2 x_2^k + B y^k - b\|^2 \\
&\leq f_1(x_1^k) - \langle \lambda^k, r_k \rangle + \frac{\beta}{2} \|r_k\|^2 - \Delta_{\varphi_1}(x_1^{k+1}, x_1^k), \\
&f_2(x_2^{k+1}) + \langle x_2^{k+1} - x_2^k, \nabla_{x_2} g(u^k) \rangle - \langle \lambda^k, A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y^k - b \rangle + \frac{\beta}{2} \|A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y^k - b\|^2 \\
&\leq f_2(x_2^k) - \langle \lambda^k, A_1 x_1^{k+1} + A_2 x_2^k + B y^k - b \rangle + \frac{\beta}{2} \|A_1 x_1^{k+1} + A_2 x_2^k + B y^k - b\|^2 - \Delta_{\varphi_2}(x_2^{k+1}, x_2^k), \\
&g(u^{k+1}) - \langle \lambda^k, r_{k+1} \rangle + \frac{\beta}{2} \|r_{k+1}\|^2 + \Delta_\phi(y^{k+1}, y^k) \\
&\leq g(x_1^{k+1}, x_2^{k+1}, y^k) - \langle \lambda^k, A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y^k - b \rangle + \frac{\beta}{2} \|A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y^k - b\|^2.
\end{aligned} \tag{25}$$

Adding up the abovementioned three formulas, we have

$$\begin{aligned}
& f_1(x_1^{k+1}) + f_2(x_2^{k+1}) + g(u^{k+1}) - \langle \lambda^k, r_{k+1} \rangle + \frac{\beta}{2} \|r_{k+1}\|^2 \\
& \leq f_1(x_1^k) + f_2(x_2^k) + (x_1^{k+1}, x_2^{k+1}, y^k)g - \langle \lambda^k, r_k \rangle + \frac{\beta}{2} \|r_k\|^2 \\
& \quad - [\langle x_1^{k+1} - x_1^k, \nabla_{x_1} g(u^k) \rangle + \langle x_2^{k+1} - x_2^k, \nabla_{x_2} g(u^k) \rangle] \\
& \quad - \Delta_{\varphi_1}(x_1^{k+1}, x_1^k) - \Delta_{\varphi_2}(x_2^{k+1}, x_2^k) - \Delta_{\phi}(y^{k+1}, y^k),
\end{aligned} \tag{26}$$

and hence

$$\begin{aligned}
& f_1(x_1^{k+1}) + f_2(x_2^{k+1}) + g(u^{k+1}) - \langle \lambda^k, r_{k+1} \rangle + \frac{\beta}{2} \|r_{k+1}\|^2 \\
& \leq f_1(x_1^k) + f_2(x_2^k) + g(u^k) - \langle \lambda^k, r_k \rangle + \frac{\beta}{2} \|r_k\|^2 \\
& \quad + g(x_1^{k+1}, x_2^{k+1}, y^k) - g(u^k) - \langle (x_1^{k+1} - x_1^k, x_2^{k+1} - x_2^k, 0), \nabla g(u^k) \rangle \\
& \quad - \Delta_{\varphi_1}(x_1^{k+1}, x_1^k) - \Delta_{\varphi_2}(x_2^{k+1}, x_2^k) - \Delta_{\phi}(y^{k+1}, y^k),
\end{aligned} \tag{27}$$

that is,

$$\begin{aligned}
& L_{\beta}(x_1^{k+1}, x_2^{k+1}, y^{k+1}, \lambda^k) \\
& \leq L_{\beta}(w^k) + g(x_1^{k+1}, x_2^{k+1}, y^k) - g(u^k) \\
& \quad - \langle (x_1^{k+1} - x_1^k, x_2^{k+1} - x_2^k, 0), \nabla g(u^k) \rangle \\
& \quad - \Delta_{\varphi_1}(x_1^{k+1}, x_1^k) - \Delta_{\varphi_2}(x_2^{k+1}, x_2^k) - \Delta_{\phi}(y^{k+1}, y^k).
\end{aligned} \tag{28}$$

From Lemma 2, Assumption 1 (iv), and Proposition 1, we obtain

$$\begin{aligned}
& L_{\beta}(x_1^{k+1}, x_2^{k+1}, y^{k+1}, \lambda^k) \\
& \leq L_{\beta}(w^k) - \frac{\delta_{\varphi_1} - l_g}{2} \|x_1^{k+1} - x_1^k\|^2 - \frac{\delta_{\varphi_2} - l_g}{2} \|x_2^{k+1} - x_2^k\|^2 \\
& \quad - \frac{\delta_{\phi}}{2} \|y^{k+1} - y^k\|^2.
\end{aligned} \tag{29}$$

Recall that

$$\begin{aligned}
L_{\beta}(w^{k+1}) &= L_{\beta}(x_1^{k+1}, x_2^{k+1}, y^{k+1}, \lambda^k) + \langle \lambda^k - \lambda^{k+1}, r_{k+1} \rangle \\
&= L_{\beta}(x_1^{k+1}, x_2^{k+1}, y^{k+1}, \lambda^k) + \frac{1}{\beta} \|\lambda^k - \lambda^{k+1}\|^2.
\end{aligned} \tag{30}$$

Adding up (71) and (72), we have

$$\begin{aligned}
L_{\beta}(w^{k+1}) &\leq L_{\beta}(w^k) - \frac{\delta_{\varphi_1} - l_g}{2} \|x_1^{k+1} - x_1^k\|^2 \\
&\quad - \frac{\delta_{\varphi_2} - l_g}{2} \|x_2^{k+1} - x_2^k\|^2 - \frac{\delta_{\phi}}{2} \|y^{k+1} - y^k\|^2 + \frac{1}{\beta} \|\lambda^k - \lambda^{k+1}\|^2.
\end{aligned} \tag{31}$$

Together with (14), we obtain

$$\begin{aligned}
L_{\beta}(w^{k+1}) &\leq L_{\beta}(w^k) - \frac{\delta_{\varphi_1} - l_g}{2} \|x_1^{k+1} - x_1^k\|^2 - \frac{\delta_{\varphi_2} - l_g}{2} \|x_2^{k+1} - x_2^k\|^2 \\
&\quad - \frac{\delta_{\phi}}{2} (y^{k+1} - y^k)^2 + \frac{3\mu^2(l_{\phi}^2 + l_g^2)}{\beta} \|y^{k+1} - y^k\|^2 + \frac{3\mu^2 l_{\phi}^2}{\beta} \|y^k - y^{k-1}\|^2 \\
&\quad + \frac{3\mu^2 l_g^2}{\beta} \|x_1^{k+1} - x_1^k\|^2 + \frac{3\mu^2 l_g^2}{\beta} \|x_2^{k+1} - x_2^k\|^2 \leq L_{\beta}(w^k) - \left[\frac{\delta_{\varphi_1} - l_g}{2} - \frac{3\mu^2 l_g^2}{\beta} \right] \|x_1^{k+1} - x_1^k\|^2 \\
&\quad - \left[\frac{\delta_{\varphi_2} - l_g}{2} - \frac{3\mu^2 l_g^2}{\beta} \right] \|x_2^{k+1} - x_2^k\|^2 \\
&\quad - \left[\frac{\delta_{\phi}}{2} - \frac{3\mu^2(l_{\phi}^2 + l_g^2)}{\beta} \right] \|y^{k+1} - y^k\|^2 + \frac{3\mu^2 l_{\phi}^2}{\beta} \|y^k - y^{k-1}\|^2,
\end{aligned} \tag{32}$$

which implies that

$$\begin{aligned}
& L_\beta(w^{k+1}) + \frac{3\mu^2 l_\phi^2}{\beta} \|y^{k+1} - y^k\|^2 \\
& \leq L_\beta(w^k) + \frac{3\mu^2 l_\phi^2}{\beta} \|y^k - y^{k-1}\|^2 - \left[\frac{\delta_{\phi_1} - l_g}{2} - \frac{3\mu^2 l_g^2}{\beta} \right] \|x_1^{k+1} - x_1^k\|^2 \\
& \quad - \left[\frac{\delta_{\phi_2} - l_g}{2} - \frac{3\mu^2 l_g^2}{\beta} \right] \|x_2^{k+1} - x_2^k\|^2 \\
& \quad - \left[\frac{\delta_\phi}{2} - \frac{3\mu^2 (2l_\phi^2 + l_g^2)}{\beta} \right] \|y^{k+1} - y^k\|^2.
\end{aligned} \tag{33}$$

That is, (23) holds. $\square$

Remark 2. From Assumption 1 (iv), we have $\delta_\phi - l_g > 0$ and $\delta_{\phi_1} - l_g > 0$. Furthermore, from Assumption 1 (v), we have $\sigma > 0$.

Lemma 6. If the sequence $\{w^k = (x_1^k, x_2^k, y^k, \lambda^k)^T\}$ is bounded, then

$$\sum_{k=0}^{+\infty} \|w^{k+1} - w^k\|^2 < +\infty. \tag{34}$$

Proof. Since $\{w^k\}$ is bounded, the sequence $\{\hat{w}^k\}$ is bounded and there exists a subsequence $\{\hat{w}^{k_j}\}$ such that $\lim_{j \rightarrow +\infty} (\hat{w}^{k_j}) \hat{w}^*$. Since f_1, f_2 are lower semicontinuous and g is Lipschitz differentiable, the function $\hat{L}(\cdot)$ is lower semicontinuous, which leads to

$$\liminf_{j \rightarrow +\infty} \hat{L}(\hat{w}^{k_j}) \geq \hat{L}(\hat{w}^*). \tag{35}$$

Thus, $\{\hat{L}(\hat{w}^{k_j})\}$ is bounded from below. From Lemma 5, $\{\hat{L}(\hat{w}^{k_j})\}$ is nonincreasing. Thus, $\{\hat{L}(\hat{w}^{k_j})\}$ is convergent. Furthermore, $\{\hat{L}(\hat{w}^k)\}$ is also convergent and $\hat{L}(\hat{w}^k) \geq \hat{L}(\hat{w}^*)$ for each k . By Lemma 5, we have

$$\sigma \left(\|x_1^{k+1} - x_1^k\|^2 + \|x_2^{k+1} - x_2^k\|^2 + \|y^{k+1} - y^k\|^2 \right) \leq \hat{L}(\hat{w}^k) - \hat{L}(\hat{w}^{k+1}). \tag{36}$$

From the abovementioned formula, we obtain

$$\begin{aligned}
& \sigma \sum_{k=1}^t \left(\|x_1^{k+1} - x_1^k\|^2 + \|x_2^{k+1} - x_2^k\|^2 + \|y^{k+1} - y^k\|^2 \right) \\
& \leq \hat{L}(\hat{w}^1) - \hat{L}(\hat{w}^{t+1}).
\end{aligned} \tag{37}$$

Note that $\sigma > 0$ and the arbitrariness of t , we obtain

$$\sum_{k=1}^{+\infty} \|u^{k+1} - u^k\|^2 < +\infty. \tag{38}$$

In view of (14), we have

$$\sum_{k=0}^{+\infty} \|\lambda^{k+1} - \lambda^k\|^2 < +\infty. \tag{39}$$

Thus,

$$\sum_{k=0}^{+\infty} w^{k+1} - w^{k2} < +\infty. \tag{40}$$

$\square$

Lemma 7. There exists $\delta > 0$ such that

$$d(0, \partial \hat{L}(\hat{w}^{k+1})) \leq \delta a_k, \quad \text{for all } k \in \mathbb{N}, \tag{41}$$

where

$$\begin{aligned}
a_k = & \|x_1^{k+1} - x_1^k\| + \|x_2^{k+1} - x_2^k\| + \|y^{k+1} - y^k\| + \|x_1^k - x_1^{k-1}\| \\
& + \|x_2^k - x_2^{k-1}\| + \|y^k - y^{k-1}\|.
\end{aligned} \tag{42}$$

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

$$\left\{ \begin{aligned} \partial_{x_1} \hat{L}(\hat{w}^{k+1}) &= \partial f_1(x_1^{k+1}) + \nabla_{x_1} g(u^{k+1}) - A_1^T \lambda^{k+1} + \beta A_1^T r_{k+1}, \\ \partial_{x_2} \hat{L}(\hat{w}^{k+1}) &= \partial f_2(x_2^{k+1}) + \nabla_{x_2} g(u^{k+1}) - A_2^T \lambda^{k+1} + \beta A_2^T r_{k+1}, \\ \partial_y \hat{L}(\hat{w}^{k+1}) &= \nabla_y g(u^{k+1}) - B^T \lambda^{k+1} + \beta B^T r_{k+1} + \frac{6\mu^2 l_\phi^2}{\beta} (y^{k+1} - y^k), \\ \partial_{\hat{y}} \hat{L}(\hat{w}^{k+1}) &= -\frac{6\mu^2 l_\phi^2}{\beta} (y^{k+1} - y^k), \\ \partial_\lambda \hat{L}(\hat{w}^{k+1}) &= \frac{1}{\beta} (\lambda^{k+1} - \lambda^k). \end{aligned} \right. \tag{43}$$

From (14) and the optimality conditions, one has

$$\begin{cases} 0 \in \partial f_1(x_1^{k+1}) + \nabla_{x_1} g(u^k) - A_1^T \lambda^k + \beta A_1^T (A_1 x_1^{k+1} + A_2 x_2^k + B y^k - b) + \nabla \varphi_1(x_1^{k+1}) - \nabla \varphi_1(x_1^k), \\ 0 \in \partial f_2(x_2^{k+1}) + \nabla_{x_2} g(u^k) - A_2^T \lambda^k + \beta A_2^T (A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y^k - b) + \nabla \varphi_2(x_2^{k+1}) - \nabla \varphi_2(x_2^k), \\ 0 = \nabla_y g(u^{k+1}) - B^T \lambda^k + \beta B^T (A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y^{k+1} - b) + \nabla \phi(y^{k+1}) - \nabla \phi(y^k), \\ \lambda^{k+1} = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y^{k+1} - b). \end{cases} \quad (44)$$

That is,

$$\begin{cases} A_1^T \lambda^{k+1} - \nabla_{x_1} g(u^k) + \beta A_1^T A_2 (x_2^k - x_2^{k+1}) + \beta A_1^T B (y^k - y^{k+1}) \\ \quad + \nabla \varphi_1(x_1^k) - \nabla \varphi_1(x_1^{k+1}) \in \partial f_1(x_1^{k+1}), \\ A_2^T \lambda^{k+1} - \nabla_{x_2} g(u^k) + \beta A_2^T B (y^k - y^{k+1}) + \nabla \varphi_2(x_2^k) - \nabla \varphi_2(x_2^{k+1}) \in \partial f_2(x_2^{k+1}), \\ \nabla_y g(u^{k+1}) = B^T \lambda^{k+1} + \nabla \phi(y^k) - \nabla \phi(y^{k+1}), \\ \lambda^{k+1} = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y^{k+1} - b). \end{cases} \quad (45)$$

From (43) and (45), we have

where

$$(\rho_1^k, \rho_2^k, \rho_3^k, \rho_4^k, \rho_5^k)^T \in \partial \widehat{L}(\widehat{w}^{k+1}), \quad (46)$$

$$\begin{cases} \rho_1 = A_1^T (\lambda^k - \lambda^{k+1}) + \nabla_{x_1} g(u^{k+1}) - \nabla_{x_1} g(u^k) + \beta A_1^T A_2 (x_2^k - x_2^{k+1}) + \beta A_1^T B (y^k - y^{k+1}) \\ \quad + \nabla \varphi_1(x_1^k) - \nabla \varphi_1(x_1^{k+1}), \\ \rho_2 = A_2^T (\lambda^k - \lambda^{k+1}) - \nabla_{x_2} g(u^k) + \beta A_2^T B (y^k - y^{k+1}) + \nabla \varphi_2(x_2^k) - \nabla \varphi_2(x_2^{k+1}) + \nabla_{x_2} g(u^{k+1}), \\ \rho_3^k = \beta B^T (\lambda^k - \lambda^{k+1}) + \frac{6\mu^2 l_\phi^2}{\beta} (y^{k+1} - y^k) + \nabla \phi(y^k) - \nabla \phi(y^{k+1}), \\ \rho_4^k = -\frac{6\mu^2 l_\phi^2}{\beta} (y^{k+1} - y^k), \\ \rho_5^k = \frac{1}{\beta} (\lambda^k - \lambda^{k+1}). \end{cases} \quad (47)$$

Thus,

$$d(0, \partial \widehat{L}(\widehat{w}^{k+1})) \leq \|(\rho_1^k, \rho_2^k, \rho_3^k, \rho_4^k, \rho_5^k)^T\|. \quad (48)$$

It follows from Assumption 1 and Lemma 4 that there exists a $\delta > 0$ such that

$$d(0, \partial \widehat{L}(\widehat{w}^{k+1})) \leq \delta a_k, \quad \text{for all } k \in \mathbb{N}. \quad (49)$$

The following theorem shows that the algorithm LBADMM has global convergence. $\square$

Theorem 1. Let $S(\{\widehat{w}^k\})$ denote the cluster point set of $\{\widehat{w}^k\}$, then

(i) $S(\{\widehat{w}^k\})$ is a nonempty compact set, and $\lim_{k \rightarrow +\infty} d(\widehat{w}^k, S(\widehat{w}^k)) = 0$

- (ii) If $(x_1^*, x_2^*, y^d, \lambda^*, \hat{y}^*)^T \in S(\{\hat{w}^k\})$, then $(x_1^*, x_2^*, y^*, \lambda^*)^T \in \text{crit } L(\cdot)$;
- (iii) $\hat{L}(\cdot)$ is finite and constant on $S(\{\hat{w}^k\})$ and equal to $\inf_{k \in \mathbb{N}} \hat{L}_\beta(\hat{w}^k) = \lim_{k \rightarrow +\infty} \hat{L}_\beta(\hat{w}^k)$.

Proof

- (i) By the definition of $S(\{\hat{w}^k\})$, it is trivial.
- (ii) Let $\hat{w}^* = (x_1^*, x_2^*, y^*, \lambda^*, \hat{y}^*)^T \in S(\{\hat{w}^k\})$, then there exists a subsequence $\{\hat{w}^{k_j}\}$ of $\{\hat{w}^k\}$ converging to $\{\hat{w}^*\}$. Since $\|w^{k+1} - w^k\| \rightarrow 0 (k \rightarrow +\infty)$, $\lim_{j \rightarrow +\infty} \hat{w}^{k_j+1} = \hat{w}^*$. Since $\lambda^{k_j+1} = \lambda^{k_j} - \beta(A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y^{k+1} - b)$ and $\beta > 0$, $A_1 x_1^* + A_2 x_2^* + B y^* - b = 0$

Let

$$\begin{aligned} h_k(x_1) &= f_1(x_1) + (x_1 - x_1^k)^T \nabla_{x_1} g(x_1^k, x_2^k, y^k) \\ &\quad + \frac{\beta}{2} \left\| A_1 x_1 + A_2 x_2^k + B y^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \Delta_{\varphi_1}(x_1, x_1^k), \\ \theta_k(x_2) &= f_2(x_2) + (x_2 - x_2^k)^T \nabla_{x_2} g(x_1^k, x_2^k, y^k) \\ &\quad + \frac{\beta}{2} \left\| A_1 x_1^{k+1} + A_2 x_2 + B y^k - b - \frac{\lambda^k}{\beta} \right\|^2 + \Delta_{\varphi_2}(x_2, x_2^k). \end{aligned} \quad (50)$$

It follows from (14) that $h_{k_j}(x_1^{k_j+1}) \leq h_{k_j}(x_1^*)$ and $\theta_{k_j}(x_2^{k_j+1}) \leq \theta_{k_j}(x_2^*)$. Thus,

$$\begin{aligned} \limsup_{j \rightarrow +\infty} h_{k_j}(x_1^{k_j+1}) &= \limsup_{j \rightarrow +\infty} f_1(x_1^{k_j+1}) + \frac{(\lambda^*)^2}{2\beta} \\ &\leq \limsup_{j \rightarrow +\infty} h_{k_j}(x_1^*) = f_1(x_1^*) + \frac{(\lambda^*)^2}{2\beta}, \\ \limsup_{j \rightarrow +\infty} \theta_{k_j}(x_2^{k_j+1}) &= \limsup_{j \rightarrow +\infty} f_2(x_2^{k_j+1}) + \frac{(\lambda^*)^2}{2\beta} \\ &\leq \limsup_{j \rightarrow +\infty} \theta_{k_j}(x_2^*) = f_2(x_2^*) + \frac{(\lambda^*)^2}{2\beta}. \end{aligned} \quad (51)$$

Noting that f_1 and f_2 are lower semicontinuous, we have

$$\begin{aligned} \liminf_{j \rightarrow +\infty} f_1(x_1^{k_j+1}) &\geq f_1(x_1^*), \\ \liminf_{j \rightarrow +\infty} f_2(x_2^{k_j+1}) &\geq f_2(x_2^*). \end{aligned} \quad (52)$$

It follows from the abovementioned four formulas that

$$\begin{aligned} \lim_{j \rightarrow +\infty} f_1(x_1^{k_j+1}) &= f_1(x_1^*), \\ \lim_{j \rightarrow +\infty} f_2(x_2^{k_j+1}) &= f_2(x_2^*). \end{aligned} \quad (53)$$

Together with the continuity of ∇g and the closeness of ∂f , by taking limit in (45) along the subsequence $\{\hat{w}^{k_j+1}\}$ and $\{\hat{w}^{k_j}\}$ yields

$$\begin{cases} -\nabla_{x_1} g(x_1^*, x_2^*, y^*) + A_1^T \lambda^* \in \partial f_1(x_1^*), \\ -\nabla_{x_2} g(x_1^*, x_2^*, y^*) + A_2^T \lambda^* \in \partial f_2(x_2^*), \\ \nabla_y g(x_1^*, x_2^*, y^*) = B^T \lambda^*, \\ A_1 x_1^* + A_2 x_2^* + B y^* - b = 0. \end{cases} \quad (54)$$

That is, w^* is a critical point of the Lagrange function L of (12).

(iii) From (53) and Lemma 5, we have

$$\lim_{j \rightarrow +\infty} \hat{L}(\hat{w}^{k_j+1}) = \hat{L}(\hat{w}^*) = L(w^*). \quad (55)$$

From (55) and the descent of $\{\hat{L}(\hat{w}^k)\}_{k \in \mathbb{N}}$, we obtain

$$\lim_{k \rightarrow +\infty} \hat{L}(\hat{w}^k) = L(\hat{w}^*). \quad (56)$$

Therefore, $\hat{L}(\cdot)$ is constant on $S(\{\hat{w}^k\})$. Moreover, $\inf_{k \in \mathbb{N}} \hat{L}(\hat{w}^k) = \lim_{k \rightarrow +\infty} \hat{L}(\hat{w}^k)$.

The following theorem is the main result of this paper. $\square$

Theorem 2 (strong convergence). *Suppose that Assumption 1 holds, $\hat{L}(\hat{w})$ satisfies the KL property at each point of $S(\{\hat{w}^k\})$, then*

- (i) $\sum_{k=0}^{+\infty} \|w^{k+1} - w^k\| < +\infty$,
- (ii) $\{w^k\}$ converges to a critical point of $L(\cdot)$

Proof. From Theorem 1, we have $\lim_{k \rightarrow +\infty} \hat{L}(\hat{w}^k) = \hat{L}(\hat{w}^*)$ for all $\hat{w}^* \in S(\{\hat{w}^k\})$. We consider two cases.

- (i) If there exists an integer k_0 such that $\hat{L}(\hat{w}^{k_0}) = \hat{L}(\hat{w}^*)$. From Lemma 5, we have

$$\begin{aligned} \sigma\left(\|x_1^{k+1} - x_1^k\|^2 + \|x_2^{k+1} - x_2^k\|^2 + \|y^{k+1} - y^k\|^2\right) \\ \leq \hat{L}(\hat{w}^k) - \hat{L}(\hat{w}^{k+1}) \leq L_\beta(\hat{w}^{k_0}) - L_\beta(\hat{w}^*) = 0, \end{aligned} \quad (57)$$

for all $k > k_0$.

Thus, for any $k > k_0$, we have $x_1^{k+1} = x_1^k, x_2^{k+1} = x_2^k, y^{k+1} = y^k$. Hence, for any $k > k_0 + 1$, it follows that $\hat{w}^{k+1} = \hat{w}^k$ and the assertion holds.

- (ii) Assume that $\hat{L}(\hat{w}^k) > \hat{L}(\hat{w}^*)$ for all $k \in \mathbb{N}$. Since $\lim_{k \rightarrow +\infty} d(\hat{w}^k, S(\{\hat{w}^k\})) = 0$, it follows that for any given $\varepsilon > 0$, there exists $k_1 > 0$, such that $d(\hat{w}^k, S(\{\hat{w}^k\})) < \varepsilon$, for all $k > k_1$. Since $\lim_{k \rightarrow +\infty} \hat{L}(\hat{w}^k) = \hat{L}(\hat{w}^*)$, for given $\eta > 0$, there exists

$k_2 > 0$, such that $\hat{L}(\hat{w}^k) < \hat{L}(\hat{w}^*) + \eta$, for all $k > k_2$.
Consequently, when $k > \tilde{k} := \max\{k_1, k_2\}$,

$$d(\hat{w}^k, S(\{\hat{w}^k\})) < \varepsilon, \hat{L}(\hat{w}^*) < \hat{L}(\hat{w}^k) < \hat{L}(\hat{w}^*) + \eta. \quad (58)$$

Since $S(\{\hat{w}^k\})$ is a nonempty compact set and $\hat{L}(\cdot)$ is constant on $S(\{\hat{w}^k\})$, applying Lemma 1, we have

$$\varphi'(\hat{L}(\hat{w}^k) - \hat{L}(\hat{w}^*))d(0, \partial\hat{L}(\hat{w}^k)) \geq 1, \quad \text{for all } k > \tilde{k}. \quad (59)$$

From Lemma 7, one has

$$\frac{1}{\varphi'(\hat{L}(\hat{w}^k) - \hat{L}(\hat{w}^*))} \leq d(0, \partial\hat{L}(\hat{w}^k)) \leq \delta a_k, \quad \text{for all } k > \tilde{k}. \quad (60)$$

From the concavity of φ , we have

$$\begin{aligned} & \varphi(\hat{L}(\hat{w}^k) - \hat{L}(\hat{w}^*)) - \varphi(\hat{L}(\hat{w}^{k+1}) - \hat{L}(\hat{w}^*)) \\ & \geq \varphi'(\hat{L}(\hat{w}^k) - \hat{L}(\hat{w}^*))(\hat{L}(\hat{w}^k) - \hat{L}(\hat{w}^{k+1})), \end{aligned} \quad (61)$$

Thus, associating with Lemma 5 and $\varphi'(\hat{L}(\hat{w}^k) - \hat{L}(\hat{w}^*)) > 0$, we have

$$\begin{aligned} & \sigma(\|x_1^{k+1} - x_1^k\|^2 + \|x_2^{k+1} - x_2^k\|^2 + \|y^{k+1} - y^k\|^2) \\ & \leq \hat{L}(\hat{w}^k) - \hat{L}(\hat{w}^{k+1}) \leq \frac{\varphi(\hat{L}(\hat{w}^k) - \hat{L}(\hat{w}^*)) - \varphi(\hat{L}(\hat{w}^{k+1}) - \hat{L}(\hat{w}^*))}{\varphi'(\hat{L}(\hat{w}^k) - \hat{L}(\hat{w}^*))} \\ & \leq \delta a_k(\varphi(\hat{L}(\hat{w}^k) - \hat{L}(\hat{w}^*)) - \varphi(\hat{L}(\hat{w}^{k+1}) - \hat{L}(\hat{w}^*))), \quad \forall k > \tilde{k}. \end{aligned} \quad (62)$$

For convenience, we set $\Delta_{p,q} = \varphi(\hat{L}(\hat{w}^p) - \hat{L}(\hat{w}^*)) - \varphi(\hat{L}(\hat{w}^q) - \hat{L}(\hat{w}^*))$. Thus,

$$\begin{aligned} & \|x_1^{k+1} - x_1^k\|^2 + \|x_2^{k+1} - x_2^k\|^2 + \|y^{k+1} - y^k\|^2 \leq \frac{\delta}{\sigma} a_k \Delta_{k,k+1}, \\ & \text{for all } k > \tilde{k}. \end{aligned} \quad (63)$$

That is,

$$\begin{aligned} & 3(\|x_1^{k+1} - x_1^k\| + \|x_2^{k+1} - x_2^k\| + \|y^{k+1} - y^k\|) \\ & \leq 3\sqrt{3}(\|x_1^{k+1} - x_1^k\|^2 + \|x_2^{k+1} - x_2^k\|^2 + \|y^{k+1} - y^k\|^2)^{1/2} \\ & \leq 2\sqrt{a_k} \sqrt{\frac{27\delta}{4\sigma} \Delta_{k,k+1}}, \quad \text{for all } k > \tilde{k}. \end{aligned} \quad (64)$$

By the fact $2\sqrt{\alpha\beta} \leq \alpha + \beta$ ($\alpha, \beta > 0$), we obtain

$$2\sqrt{a_k} \sqrt{\frac{27\delta}{4\sigma} \Delta_{k,k+1}} \leq a_k + \frac{27\delta}{4\sigma} \Delta_{k,k+1}, \quad (65)$$

which along with (64) yields

$$3(\|x_1^{k+1} - x_1^k\| + \|x_2^{k+1} - x_2^k\| + \|y^{k+1} - y^k\|) \leq a_k + \frac{27\delta}{4\sigma} \Delta_{k,k+1}. \quad (66)$$

Summing up the abovementioned formula for $k = \tilde{k} + 1, \dots, m$ yields

$$\begin{aligned} & \sum_{k=\tilde{k}+1}^m 3(\|x_1^{k+1} - x_1^k\| + \|x_2^{k+1} - x_2^k\| + \|y^{k+1} - y^k\|) \\ & \leq \sum_{k=\tilde{k}+1}^m \left[a_k + \frac{27\delta}{4\sigma} \Delta_{k,k+1} \right]. \end{aligned} \quad (67)$$

Notice that $\varphi(\hat{L}_\beta(\hat{w}^{m+1}) - \hat{L}_\beta(\hat{w}^*)) > 0$; thus,

$$\begin{aligned} & \sum_{k=\tilde{k}+1}^m (\|x_1^{k+1} - x_1^k\| + \|x_2^{k+1} - x_2^k\| + \|y^{k+1} - y^k\|) \\ & \leq b_k + \frac{27\delta}{4\sigma} (\varphi(\hat{L}(\hat{w}^{\tilde{k}+1}) - \hat{L}(\hat{w}^*)) - \varphi(\hat{L}(\hat{w}^{m+1}) - \hat{L}(\hat{w}^*))) \\ & \leq b_k + \frac{27\delta}{4\sigma} \varphi(\hat{L}(\hat{w}^{\tilde{k}+1}) - \hat{L}(\hat{w}^*)), \end{aligned} \quad (68)$$

where

$$b_k = \|x_1^{\tilde{k}+1} - x_1^{\tilde{k}}\| + \|x_2^{\tilde{k}+1} - x_2^{\tilde{k}}\| + \|y^{\tilde{k}+1} - y^{\tilde{k}}\|. \quad (69)$$

Thus,

$$\sum_{k=0}^{+\infty} (\|x_1^{k+1} - x_1^k\| + \|x_2^{k+1} - x_2^k\| + \|y^{k+1} - y^k\|) < +\infty. \quad (70)$$

By Lemma 4, one has $\sum_{k=0}^{+\infty} \|\lambda^{k+1} - \lambda^k\| < +\infty$. Furthermore, $\sum_{k=0}^{+\infty} \|w^{k+1} - w^k\| < +\infty$. Consequently $\{w^k\}$ is a Cauchy sequence. The assertion then follows immediately from Theorem 1. $\square$

Remark 3. In this section, the main conclusions are based on the boundedness assumption of the sequences $\{w^k = (x_1^k, x_2^k, y^k, \lambda^k)^T\}$. The following conclusion shows that we only need to assume that the sequence $\{u^k = (x_1^k, x_2^k, y^k)^T\}$ is bounded.

Proposition 3. If $\lambda^0 \in \text{Im}B$ and the sequence $\{u^k = (x_1^k, x_2^k, y^k)^T\}$ is bounded, then $\{\lambda^k\}$ is bounded.

Proof. From (17), one has

$$\begin{aligned}
\|B^T \lambda^{k+1}\|^2 &= \|\nabla_y g(x_1^k, x_2^k, y^k) + \nabla \phi(y^{k+1}) - \nabla \phi(y^k)\|^2 \\
&\leq \left(\|\nabla_y g(x_1^k, x_2^k, y^k)\| + l_\phi \|y^{k+1} - y^k\| \right)^2 \\
&\leq 2 \left(\|\nabla_y g(x_1^k, x_2^k, y^k)\|^2 + l_\phi^2 \|y^{k+1} - y^k\|^2 \right).
\end{aligned} \tag{71}$$

Since $\{u^k = (x_1^k, x_2^k, y^k)^T\}$ is bounded, $\{B^T \lambda^k\}$ is bounded. By Assumption 1 (ii) and Lemma 3, we have

$$\|\lambda^{k+1}\| \leq \frac{1}{\sqrt{\mu_{B^T}}} \|B^T \lambda^{k+1}\|. \tag{72}$$

Thus, $\{\lambda^k\}$ is bounded.

Next, we present a sufficient condition of boundedness of the sequence $\{u^k\}$, which is similar with Lemma 8 in [4]. $\square$

Lemma 8. Let $\{w^k := (x_1^k, x_2^k, y^k, \lambda^k)\}$ be the sequence generated by Algorithm 1. Suppose that $\lambda^0 \in \text{Im} B$ and there exists $\theta > 0$ such that

$$\inf_u \left\{ g(u) - \theta \|\nabla_y g(u)\|^2 \right\} := \bar{g} > -\infty. \tag{73}$$

If

$$\liminf_{\|u\| \rightarrow +\infty} [f_1(x_1) + f_2(x_2) + g(u)] = +\infty, l_g \langle l_\phi, \beta \rangle \frac{2\mu^2}{\theta}, \tag{74}$$

then $\{w^k\}$ is bounded.

Proof. From Lemma 5, we know that

$$\widehat{L}(x_1^k, x_2^k, y^k, \lambda^k, y^{k-1}) \leq \widehat{L}(x_1^2, x_2^2, y^2, \lambda^2, y^1). \tag{75}$$

Then, combining with $B^T \lambda^k = \nabla_y g(u^{k-1}) + \nabla \phi(y^k) - \nabla \phi(y^{k-1})$, we obtain

$$\begin{aligned}
&\widehat{L}(x_1^2, x_2^2, y^2, \lambda^2, y^1) \\
&\geq f_1(x_1^k) f_2(x_2^k) + g(u^k) - \langle \lambda^k, r_k \rangle + \frac{\beta}{2} \|r_k\|^2 \\
&\quad + \frac{3\mu^2 l_g^2}{\beta} \|y^k - y^{k-1}\|^2 \\
&= f_1(x_1^k) + f_2(x_2^k) + g(u^k) - \frac{1}{2\beta} \|\lambda^k\|^2 + \frac{\beta}{2} \left\| r_k - \frac{\lambda^k}{\beta} \right\|^2 \\
&\quad + \frac{3\mu^2 l_g^2}{\beta} \|y^k - y^{k-1}\|^2.
\end{aligned} \tag{76}$$

Note that $\lambda^{k2} \leq \mu^2 \|B^T \lambda^k\|^2 \leq 2\mu^2 (\|\nabla_y g(u^k)\|^2 + l_\phi^2 \|y^k - y^{k-1}\|^2)$, we have

$$\begin{aligned}
&\widehat{L}(x_1^2, x_2^2, y^2, \lambda^2, y^1) \\
&\geq \frac{1}{2} [f_1(x_1^k) + f_2(x_2^k) + g(u^k)] + \frac{1}{2} [f_1(x_1^k) + f_2(x_2^k) \\
&\quad + g(u^k)] - \frac{1}{2} \theta \|\nabla_y g(u^k)\|^2 \\
&\quad \cdot \left(\frac{\mu^2}{\beta} - \frac{1}{2} \theta \right) \|\nabla_y g(u^k)\|^2 + \frac{\beta}{2} \left\| r_k - \frac{\lambda^k}{\beta} \right\|^2 + \frac{3\mu^2 (l_g^2 - l_\phi^2)}{\beta} \|y^k - y^{k-1}\|^2.
\end{aligned} \tag{77}$$

Under the assumptions, one can easily observe that $\{x_1^k\}, \{x_2^k\}, \{y^k\}, \{\|\nabla_y g(u^k)\|^2\}$, and $\{\beta/2r_k - \lambda^k/\beta^2\}$ are all bounded. Boundedness of $\{\lambda^k\}$ follows from Proposition 3. Therefore, $\{w^k\}$ is bounded. $\square$

4. Numerical Results

In compressed sensing, a fundamental problem is recovering an n -dimensional sparse signal x from a set of m incomplete measurements. In such a case one needs to find the sparsest solution of a linear system, which can be modeled as

$$\begin{aligned}
&\min \quad \|x_0\|, \\
&\text{s.t.} \quad Dx = b,
\end{aligned} \tag{78}$$

where $D \in \mathcal{R}^{m \times n}$ is the measurement matrix, $b \in \mathcal{R}^m$ is the observed data, $c > 0$ is a regularization parameter, and $\|x\|_0$ denotes the number of nonzero elements of x . In general, the abovementioned models are NP-hard. In order to overcome such a difficulty, one can relax the l_0 regularization to $l_{(1/2)}$ regularization. And, some scholars generally solve the following problems instead of solving problem (78) [10, 27]:

$$\begin{aligned}
&\min \quad c \|x\|_{(1/2)}^{(1/2)} + \frac{1}{2} \|y\|^2, \\
&\text{s.t.} \quad Dx - y = b,
\end{aligned} \tag{79}$$

where

$$\|x\|_{(1/2)} = \left(\sum_{i=1}^n |x_i|^{(1/2)} \right)^2 \tag{80}$$

Based on (79), we construct the following problems:

$$\begin{aligned}
&\min \quad c \|x_1\|_{(1/2)}^{(1/2)} + \frac{1}{2} \|x_2\|^2 + \frac{1}{2} \|B_1 x_1 + B_2 x_2 + y\|^2 \\
&\text{s.t.} \quad A_1 x_1 + A_2 x_2 + y = b.
\end{aligned} \tag{81}$$

In order to verify the effectiveness of the algorithm LBADMM, we now focus on applying the algorithm LBADMM to solve the nonconvex optimization problem (81). Applying the algorithm LBADMM to problem (81) with $\varphi_i(x_i) = (1/2)x_i^2$ ($i = 1, 2$) and $\phi(y) = (\mu_3/2)\|y\|^2$, we have

TABLE 1: Numerical results.

Data			IPPRSM		
$n1$	$n2$	m	Iter	Time (s)	f -val
1500	1500	1500	94	3.942479	402.69
1500	1500	2000	76	4.031269	650.11
2000	1500	2000	84	5.359473	636.09
1500	2000	2000	78	5.213782	523.09
2000	2000	1500	95	5.904308	346.26
2000	2000	2000	95	7.222839	514.67
3000	3000	3000	92	17.289544	727.81
3000	3000	4000	99	21.769152	1242.77
4000	4000	4000	96	32.752034	924.24
5000	5000	5000	106	58.176851	1138.52
6000	6000	5000	114	84.718137	1035.05
6000	6000	6000	113	104.601905	1366.83

FIGURE 1: $n1 = 1500$, $n2 = 1500$, and $m = 2000$.

$$\begin{cases} x_1^{k+1} = H\left(\frac{1}{\mu_1} \left[G_1 x_1^k - B_1^T \left(B_1 x_1^k + B_2 x_2^k + y^k + \beta A_1^T \left(A_2 x_2^k + y^k - b - \frac{\lambda^k}{\beta} \right) \right) \right]; \frac{2c}{\mu_1} \right), \\ x_2^{k+1} = \frac{1}{1 + \mu_2} \left[G_2 x_2^k - B_2^T \left(B_1 x_1^k + B_2 x_2^k + y^k + \beta A_2^T \left(A_1 x_1^{k+1} + y^k - b - \frac{\lambda^k}{\beta} \right) \right) \right], \\ y^{k+1} = \frac{1}{1 + \beta + \mu_3} \left[\mu_3 y^k - \left(B_1 x_1^{k+1} + B_2 x_2^{k+1} + \beta \left(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b - \frac{\lambda^k}{\beta} \right) \right) \right], \\ \lambda^{k+1} = \lambda^k - \beta (A_1 x_1^{k+1} + A_2 x_2^{k+1} + B y^{k+1} - b), \end{cases} \quad (82)$$

where $H(\cdot; \cdot)$ is the half shrinkage operator [27] defined as $H(x; \tau) = \{h_\tau(x_1), h_\tau(x_2), \dots, h_\tau(x_n)\}^T$ with

$$h_\tau(x_i) = \begin{cases} \frac{2x_i}{3} \left(1 + \cos \frac{2}{3} (\pi - \varphi(x_i)) \right), & |x_i| > \frac{\sqrt[3]{54}}{4} \tau^{2/3}; \\ 0, & \text{otherwise,} \end{cases} \quad (83)$$

with

$$\varphi(x_i) = \arccos \left(\frac{\tau}{8} \left(\frac{|x_i|}{3} \right)^{-(3/2)} \right). \quad (84)$$

In the experiment, we choose $A_i \sim \mathcal{N}(0, 1)$, to normalize the columns to have unit ($L_{1/2}$) norm. The variable x_i^0 ($i = 1, 2$) are generated with 100 nonzero entries, each sample from an Gaussian distribution. The variables x_1^0 , x_2^0 , y^0 , and λ^0 were initialized to be zero. The vector $b = A_1 x_1^0 + A_2 x_2^0 + y^0 + v$,

where $v \sim \mathcal{N}(0, 10^{-3}I)$. We set $\mu_1 = 30$, $\mu_2 = 30$, $\mu_3 = 1$, and $\beta = 20$, and the regularization parameter $c = 1$. It is easy to verify that the parameters meet Assumption 1 (v). Defining the residual at iteration k as $r_k = A_1 x_1^k + A_2 x_2^k + y - b$, a reasonable termination criterion is that the residual must be small, so we choose the stop criterion as

$$\|r_k\|_2 \leq \sqrt{m} 10^{-4}. \quad (85)$$

The numerical results are reported in Table 1. The codes were written by matlab R2016a, the computer running the program is configured as Windows 10 system, Inter (R) Core (TM) i7-6500U 2.5 GHz CPU, 8 GB memory. We report the number of iterations ("Iter."), the computing time in seconds ("Time") and the objective function value ("f-val"). Numerical results show that the Algorithm LBADMM is stable and effective.

A part of computational results are presented in Figures 1–3. In each figure, we plot the trend of the objective

FIGURE 2: $n1 = 3000$, $n2 = 300$, and $m = 3000$. FIGURE 3: $n1 = 6000$, $n2 = 6000$, and $m = 6000$.

value (“objective-value”) and the trend of the residual defined by $\|r_k\| = \|A_1 x_1^k + A_2 x_2^k + y^k - b\|$ (“ $\|r\|_2$ ”).

5. Conclusions

We propose a new algorithm called linear Bregman ADMM for the three-blocks optimization problem with the nonseparable structure. The proposed algorithm integrates the basic ideas of the linearization technology and regularization technology. We show that any cluster point of the sequence generated by the proposed algorithm is a critical point. Under the condition that the potential function satisfies the Kurdyka-Łojasiewicz property and the penalty parameter is larger than a constant, the strong convergence of the algorithm is proved. Preliminary numerical results show that the algorithm LBADMM is stable and effective.

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 Natural Science Foundation of China (nos. 11601095 and 11771383) and Natural Science Foundation of Guangxi Province (nos. 2016GXNSFBA380185 and 2016GXNSFDA380019).

References

- [1] 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. R2, pp. 41–76, 1975.
- [2] 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.
- [3] M. T. Chao, C. Z. Cheng, and D. Y. Liang, “A proximal block minimization method of multipliers with a substitution procedure,” *Optimization Methods and Software*, vol. 30, pp. 825–842, 2014.
- [4] K. Guo, D. R. Han, and T. T. Wu, “Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints,” *International Journal of Computer Mathematics*, vol. 94, no. 8, pp. 1653–1669, 2017.
- [5] 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.
- [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] B. Jiang, T. Lin, S. Ma, and S. Zhang, “Structured nonconvex and nonsmooth optimization: algorithms and iteration complexity analysis,” *Computational Optimization and Applications*, vol. 72, no. 1, pp. 115–157, 2019.
- [8] 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.

- [9] F. H. Wang, W. F. Cao, and Z. B. Xu, "Convergence of multi-block Bregman ADMM for nonconvex composite problems," *Science China Information Sciences*, vol. 61, p. 122101, 2018.
- [10] F. H. Wang, Z. B. Xu, and H. K. Xu, "Convergence of alternating direction method with multipliers for non-convex composite problems," 2014, <http://arxiv.org/abs/1410.8625>.
- [11] Y. Wang, W. Yin, and J. Zeng, "Global convergence of ADMM in nonconvex nonsmooth optimization," *Journal of Scientific Computing*, vol. 78, no. 1, pp. 29–63, 2019.
- [12] Z. M. Wu, M. Li, D. Z. W. Wang, and D. R. Han, "A symmetric alternating direction method of multipliers for separable nonconvex minimization problems," *Asia-Pacific Journal of Operational Research*, vol. 34, no. 6, Article ID 1750030, 2017.
- [13] L. Yang, T. K. Pong, and X. Chen, "Alternating direction method of multipliers for a class of nonconvex and non-smooth problems with applications to background/foreground extraction," *SIAM Journal on Imaging Sciences*, vol. 10, no. 1, pp. 74–110, 2017.
- [14] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, Springer, Berlin, Germany, 1998.
- [15] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 3, pp. 200–217, 1967.
- [16] A. Banerjee, S. Merugu, I. Dhillon et al., "Clustering with bregman divergences," *Journal of Machine Learning Research*, vol. 6, pp. 1705–1749, 2005.
- [17] B. Mordukhovich, *Variational Analysis and Generalized Differentiation, I: Basic Theory*, Springer-Verlag, Berlin, Germany, 2006.
- [18] J. Bolte, A. Daniilidis, O. Ley, and L. Mazet, "Characterizations of Lojasiewicz inequalities and applications: subgradient flows, talweg, convexity," *Transactions of the American Mathematical Society*, vol. 362, pp. 3319–3363, 2010.
- [19] K. Kurdyka, "On gradients of functions definable in o-minimal structures," *Annales de l'institut Fourier*, vol. 48, no. 3, pp. 769–783, 1998.
- [20] S. Lojasiewicz, "Une propriété topologique des sous-ensembles analytiques réels," in *Les Équations aux Dérivées Partielles*, pp. 87–89, Editions du centre National de la Recherche Scientifique, Paris, France, 1963.
- [21] S. Lojasiewicz, "Sur la géométrie semi-et sous-analytique," *Annales de l'institut Fourier*, vol. 43, pp. 575–1595, 1993.
- [22] H. Attouch, J. Bolte, and B. F. Svaiter, "Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods," *Mathematical Programming*, vol. 137, no. 1–2, pp. 91–129, 2013.
- [23] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for non-convex problems: an approach based on the kurdyka-lojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010.
- [24] J. Bolte, A. Daniilidis, and A. Lewis, "The lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems," *SIAM Journal on Optimization*, vol. 17, no. 4, pp. 1205–1223, 2007.
- [25] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course, Volume 87 of Applied Optimization*, Springer, Berlin, Germany, 2004.
- [26] M. L. N. Goncalves, J. G. Melo, and R. D. C. Monteiro, "Convergence rate bounds for a proximal ADMM with over-relaxation stepsize parameter for solving nonconvex linearly constrained problems," February 2017, <http://arxiv.org/abs/1702.01850>.
- [27] Z. B. Xu, X. Y. Chang, F. M. Xu, and H. Zhang, " $l_{1/2}$ regularization: a thresholding representation theory and a fast solver," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, pp. 1013–1027, 2012.