

# Convergence of Bregman Alternating Direction Method of Multipliers for Nonseparable Nonconvex Objective with Linear Constraints

Xiaotong Zeng, Junping Yao, Haoming Xia\*

Key Laboratory of Optimization Theory and Applications, School of Mathematics and Information, China West Normal University, Nanchong, China

Email: 1137869503@qq.com, 13684215256@163.com, \*haomingxia1999@163.com

**How to cite this paper:** Zeng, X.T., Yao, J.P. and Xia, H.M. (2024) Convergence of Bregman Alternating Direction Method of Multipliers for Nonseparable Nonconvex Objective with Linear Constraints. *Journal of Applied Mathematics and Physics*, 12, 639-660.

<https://doi.org/10.4236/jamp.2024.122042>

**Received:** January 25, 2024

**Accepted:** February 26, 2024

**Published:** February 29, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

In this paper, our focus lies on addressing a two-block linearly constrained nonseparable nonconvex optimization problem with coupling terms. The most classical algorithm, the alternating direction method of multipliers (ADMM), is employed to solve such problems typically, which still requires the assumption of the gradient Lipschitz continuity condition on the objective function to ensure overall convergence from the current knowledge. However, many practical applications do not adhere to the conditions of smoothness. In this study, we justify the convergence of variant Bregman ADMM for the problem with coupling terms to circumvent the issue of the global Lipschitz continuity of the gradient. We demonstrate that the iterative sequence generated by our approach converges to a critical point of the issue when the corresponding function fulfills the Kurdyka-Lojasiewicz inequality and certain assumptions apply. In addition, we illustrate the convergence rate of the algorithm.

## Keywords

Nonseparable Nonconvex Optimization, Bregman ADMM, Kurdyka-Lojasiewicz Inequality

## 1. Introduction

We consider the following two-block linearly constrained nonseparable nonconvex optimization problem:

$$\begin{aligned} \min \quad & f(x) + g(y) + H(x, y) \\ \text{s.t.} \quad & Ax + y = b, \end{aligned} \quad (1)$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper lower semicontinuous function,  $g: \mathbb{R}^m \rightarrow \mathbb{R}$  is a continuous differentiable function,  $H: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  is a smooth function,  $A \in \mathbb{R}^{m \times n}$  is a matrix and  $b \in \mathbb{R}^m$  is a vector. Problem (1) finds numerous applications in scenarios such as the control of a smart grid system [1] [2] [3], the appliance load model [4], cognitive radio network and other related domains [5] [6].

To solve this problem, the commonly employed method is the Alternating Direction Method of Multipliers (ADMM) [7] which stands out as a popular and well-established approach.

In 2017, Guo *et al.* studied the convergence of the ADMM to solve the problem (1) [7] and its iteration is as follows:

$$\begin{cases} x_{k+1} = \arg \min_x \{ \mathcal{L}_\beta(x, y_k, \lambda_k) \}, \\ y_{k+1} = \arg \min_y \{ \mathcal{L}_\beta(x_{k+1}, y, \lambda_k) \}, \\ \lambda_{k+1} = \lambda_k + \beta(Ax_{k+1} + y_{k+1} - b). \end{cases} \quad (2)$$

Here,  $\mathcal{L}_\beta(\cdot)$  denotes the augmented Lagrangian function defined as

$$\begin{aligned} \mathcal{L}_\beta(x, y, \lambda) = & f(x) + g(y) + H(x, y) + \lambda^T(Ax + y - b) \\ & + \frac{\beta}{2} \|Ax + y - b\|_2^2, \end{aligned}$$

where  $\lambda$  is the Lagrangian multiplier associated with the linear constraints and  $\beta > 0$  is the penalty parameter.

Note that, there is a critical condition that the function  $g(\cdot)$  is  $L$ -smooth in their convergence analysis. Indeed, for most algorithms, the absence of convexity makes the smoothness condition a necessary requirement for convergence analysis. Unfortunately, the same holds true for ADMM [8] [9] [10] [11]. However, there exists a multitude of applications that are nonsmooth. The prevalence of non-smooth problems necessitates contemplating how to relax smoothness in nonconvex situations.

To alleviate the  $L$ -smoothness condition, Tan and Guo introduced in 2023 the convergence of the following Bregman ADMM for the special case of problem (1) [12] where the coupling term  $H(\cdot) = 0$ :

$$\begin{aligned} \min \quad & f(x) + g(y) \\ \text{s.t.} \quad & Ax + y = b, \end{aligned} \quad (3)$$

where  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper lower semicontinuous function,  $g: \mathbb{R}^m \rightarrow \mathbb{R}$  is a continuous differentiable function,  $A \in \mathbb{R}^{m \times n}$  is a matrix and  $b \in \mathbb{R}^m$  is a vector. It also encompasses a range of significant applications in diverse areas, such as imaging processing [13] [14] [15], statistical learning [16] [17] [18], engineering [19] [20] [21] and so on. Problem (1) reduces to (3) when the coupled function  $H$  is absent.

The iterative format of Bregman ADMM is as follows:

$$\begin{cases} x_{k+1} = \arg \min_x \{ \mathcal{L}_\beta^h(x, y_k, \lambda_k) \}, \\ y_{k+1} = \arg \min_y \{ \mathcal{L}_\beta^h(x_{k+1}, y, \lambda_k) \}, \\ \lambda_{k+1} = \lambda_k + \beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})). \end{cases} \quad (4)$$

Here, the augmented Lagrangian function  $\mathcal{L}_\beta^h(\cdot)$  which is defined as

$$\mathcal{L}_\beta^h(x, y, \lambda) = f(x) + g(y) + \lambda^T(Ax + y - b) + \beta D_h(-y, Ax - b).$$

where  $\lambda$  is the Lagrangian multiplier associated with the linear constraints and  $\beta > 0$  is the penalty parameter. The above Bregman ADMM reduces to the classic ADMM when  $h = \frac{1}{2}\|\cdot\|_2^2$ . However, their proof only established the convergence of the separable problem (3), where there is no coupling term. Therefore, the motivation of this paper is to extend the algorithm to the nonseparable problem (1) with a coupling term.

The iterative format of Bregman ADMM for the problem (1) is as follows:

$$\begin{cases} x_{k+1} = \arg \min_x \{ \mathcal{L}_\beta^h(x, y_k, \lambda_k) \}, \\ y_{k+1} = \arg \min_y \{ \mathcal{L}_\beta^h(x_{k+1}, y, \lambda_k) \}, \\ \lambda_{k+1} = \lambda_k + \beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})). \end{cases} \quad (5)$$

Here, the augmented Lagrangian function  $\mathcal{L}_\beta^h(\cdot)$  which is defined as

$$\mathcal{L}_\beta^h(x, y, \lambda) = f(x) + g(y) + H(x, y) + \lambda^T(Ax + y - b) + \beta D_h(-y, Ax - b).$$

where  $\lambda$  is the Lagrangian multiplier associated with the linear constraints and  $\beta > 0$  is the penalty parameter.

Given our aim to relax the strict smoothness condition  $g(\cdot)$  to be relatively smooth, our focus is solely on modifying the Euclidean distance in the iterative formulation of the  $y$  variable to the Bregman distance, and the iterative format of  $x$  is the same as the classical ADMM. The iterative format called new Bregman ADMM for the problem (1) is as follows:

$$\begin{cases} x_{k+1} = \arg \min_x \left\{ f(x) + g(y_k) + H(x, y_k) + \lambda_k^T(Ax + y_k - b) + \frac{\beta_1}{2}\|Ax + y_k - b\|^2 \right\}, \\ y_{k+1} = \arg \min_y \{ \mathcal{L}_{\beta_2}^h(x_{k+1}, y, \lambda_k) \}, \\ \lambda_{k+1} = \lambda_k + \beta_2(\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})). \end{cases} \quad (6)$$

Here denotes  $\mathcal{L}_{\beta_2}^h(\cdot)$  the augmented Lagrangian function which is defined as

$$\mathcal{L}_{\beta_2}^h(x, y, \lambda) = f(x) + g(y) + H(x, y) + \lambda^T(Ax + y - b) + \beta_2 D_h(-y, Ax - b).$$

where  $\lambda$  is the Lagrangian multiplier associated with the linear constraints and  $\beta_1, \beta_2 > 0$  is the penalty parameter.

Building upon the aforementioned motivation, our attention is directed towards the Bregman ADMM applied to the nonseparable problem (1). We delve into the following aspects: For two-block nonseparable nonconvex optimization

problems with linear constraints, we develop an appropriate Lyapunov function and leverage the KL inequality to examine the global convergence of the Bregman ADMM. The paper is organized as follows. We first summarize some necessary preliminaries for further analysis in Section 2. In Section 3, we analyze the convergence of variant Bregman ADMM and its convergence rate. Finally, some conclusions are made in Section 4.

## 2. Preliminaries

In this section, we summarized some notations and preliminaries to be used for further analysis.

**Definition 2.1.** [22] For an extended real-valued function  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ , the effective domain or just the domain is the set

$$\text{dom}(f) = \{x \in \mathbb{R}^n : f(x) < \infty\}.$$

**Definition 2.2.** [22] A function  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is called proper if there at least one  $x \in \mathbb{R}^n$  such that  $f(x) < \infty$ .

**Definition 2.3.** [22] A function  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is called lower semicontinuous at  $x \in \mathbb{R}^n$  if

$$f(x) \leq \liminf_{k \rightarrow \infty} f(x_k)$$

for any sequence  $\{x_k\} \subseteq \mathbb{R}^n$  for which  $x_k \rightarrow x$  as  $k \rightarrow \infty$ . Moreover,  $f$  is called lower semicontinuous if it is lower semicontinuous at each point in  $\mathbb{R}^n$ .

**Definition 2.4.** ([23] kernel generating distance) Let  $C$  be a nonempty, convex, and open subset of  $\mathbb{R}^m$ . Associated with  $C$ , a function  $h: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  is called a kernel generating distance if it satisfies the following:

- 1)  $h$  is proper, lower semicontinuous, and convex, with  $\text{dom}(h) \subset \bar{C}$  and  $\text{dom}(\partial h) = C$ .
- 2)  $h$  is  $C^1$  on  $\text{int}(\text{dom}(h)) \equiv C$ .

We denote the class of kernel generating distance by  $\mathcal{G}(C)$ .

Given  $h \in \mathcal{G}(C)$ , define Bregman distance  $D_h: \text{dom}(h) \times \text{int}(\text{dom}(h)) \rightarrow \mathbb{R}^+$  by

$$D_h(x, y) := h(x) - h(y) - \langle \nabla h(y), x - y \rangle. \quad (7)$$

Since  $h$  is convex,  $D_h(x, y) \geq 0$ , and  $D_h(x, y) = 0$  only when  $x = y$ .

**Lemma 2.1.** [24] Let  $h: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  be a kernel generating distance. The following properties follow directly from (7). Let  $x, y \in \text{int}(\text{dom}(h))$  and  $z \in \text{dom}(h)$ , then

- 1)  $D_h(x, y) + D_h(y, x) = \langle \nabla h(x) - \nabla h(y), x - y \rangle$ .
- 2) Three points identity holds:

$$D_h(z, x) - D_h(z, y) - D_h(y, x) = \langle \nabla h(x) - \nabla h(y), y - z \rangle. \quad (8)$$

In the subsequent analysis, we will use a pair of functions  $(g, h)$  satisfying

- 1)  $h \in \mathcal{G}(C)$ .

2)  $g: \mathbb{R}^m \cup \{+\infty\}$  is a proper and lower semicontinuous function  $\text{dom}(h) \subset \text{dom}(g)$ , which is continuously differentiable on  $C = \text{int}(\text{dom}(h))$ .

**Definition 2.5.** ([23] L-smooth adaptable) A pair  $(g, h)$  is called L-smooth adaptable on  $C$  if there exists  $L > 0$  such that  $Lh - g$  and  $Lh + g$  are convex on  $C$ .

**Remark 2.1.** Since  $Lh - g$  and  $Lh + g$  is convex, then its gradient are monotonous, we obtain

$$\langle \nabla g(x) - \nabla g(y), x - y \rangle \leq L \langle \nabla h(x) - \nabla h(y), x - y \rangle = L(D_h(x, y) + D_h(y, x)),$$

and

$$\langle \nabla g(y) - \nabla g(x), x - y \rangle \leq L \langle \nabla h(x) - \nabla h(y), x - y \rangle = L(D_h(x, y) + D_h(y, x)).$$

If we assume,

$$\|\nabla g(x) - \nabla g(y)\| \leq L \frac{D_h(x, y) + D_h(y, x)}{\|x - y\|}, \text{ for all } x \neq y \in \text{int}(\text{dom}(h)), \quad (9)$$

then by Cauchy-Schwarz inequality we immediately obtain the above two inequalities. Thus, (9) provides a Lipschitz-like gradient property of  $g$  with respect to  $D_h$ .

**Remark 2.2.** Definition 2.5 naturally complements and extends the definition of “A Lipschitz-like/Convexity Condition” in [22], which allows to obtain the following two-sided descent lemma.

**Lemma 2.2.** ([23] extended descent lemma) The pair of functions  $(g, h)$  is L-smooth adaptable on  $C$  if and only if

$$|g(x) - g(y) - \langle \nabla g(y), x - y \rangle| \leq LD_h(x, y) \quad \forall x, y \in \text{int}(\text{dom}(h)). \quad (10)$$

In particular, when the set  $C = \mathbb{R}^m$  and  $h(\cdot) = (1/2)\|\cdot\|^2$ , (10) reduces to the classical descent lemma for a function  $g$  with a L-smooth gradient on  $\mathbb{R}^m$ , i.e.,

$$|g(x) - g(y) - \langle \nabla g(y), x - y \rangle| \leq \frac{L}{2} \|x - y\|^2 \quad \forall x, y \in \mathbb{R}^m.$$

**Definition 2.6.** [23] Let  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower semicontinuous function.

1) The Fréchet subdifferential, or regular subdifferential of  $f$  at  $x \in \text{dom}(f)$ , written  $\hat{\partial}f(x)$ , is the set of vectors  $x^* \in \mathbb{R}^n$  that satisfy

$$\liminf_{y \neq x, y \rightarrow x} \frac{f(y) - f(x) - \langle x^*, y - x \rangle}{\|y - x\|} \geq 0.$$

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

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

$$\partial f(x) = \left\{ x^* \in \mathbb{R}^n, \exists x_n \rightarrow x, f(x_n) \rightarrow f(x), x_n^* \in \hat{\partial}f(x_n), \text{ with } x_n^* \rightarrow x^* \right\}.$$

**Remark 2.3.** From the above definition, we note that

1) The above definition implies  $\hat{\partial}f(x) \subseteq \partial f(x)$  for each  $x \in \mathbb{R}^n$ , where the first set is closed convex while the second one is only closed.

2) Let  $(x_k, x_k^*) \in \text{Graph } \partial f$  be a sequence that converges to  $(x, x^*)$ . By the definition of  $\partial f$ , if  $f(x_k)$  converges to  $f(x)$  as  $k \rightarrow +\infty$ , then  $(x, x^*) \in \text{Graph } \partial f$ .

3) A necessary condition for  $x \in \mathbb{R}^n$  to be a minimizer of  $f$  is

$$0 \in \partial f(x). \tag{11}$$

4) If  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is a proper lower semicontinuous and  $g: \mathbb{R}^n \rightarrow \mathbb{R}$  is continuous differentiable, then  $\partial(f + g)(x) = \partial f(x) + \nabla g(x)$  for any  $x \in \text{dom}(f)$ .

A point satisfying Equation (11) is called a critical point or a station point. The critical points set of  $f$  is denoted by  $\text{crit } f$ .

Now, we recall an important property of subdifferential calculus.

**Lemma 2.3.** ([25] Suppose that  $F(x, y) = f(x) + g(x)$ , where  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  and  $g: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  are proper lower  $r$  semicontinuous functions. Then for all  $(x, y) \in \text{dom}(F) = \text{dom}(f) \times \text{dom}(g)$ , we have

$$\partial F(x, y) = \partial_x F(x, y) \times \partial_y F(x, y).$$

**Definition 2.7.** ([25] Kurdyka-Lojasiewicz inequality) Let  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper lower semicontinuous function. For  $-\infty < \eta_1 < \eta_2 \leq +\infty$ , set

$$[\eta_1 < f < \eta_2] = \{x \in \mathbb{R}^n : \eta_1 < f(x) < \eta_2\}.$$

We say that function  $f$  has the KL property at  $x^* \in \text{dom}(\partial f)$  if there exist  $\eta \in (0, +\infty]$ , a neighbourhood  $U$  of  $x^*$ , and a continuous concave function  $\varphi: [0, \eta] \rightarrow \mathbb{R}_+$ , such that

- 1)  $\varphi(0) = 0$ ;
- 2)  $\varphi$  is  $C^1$  on  $(0, \eta)$  and continuous at 0;
- 3)  $\varphi'(x) > 0, \forall x \in (0, \eta)$ ;
- 4) for all  $x$  in  $U \cap [f(x^*) < f < f(x^*) + \eta]$ , the Kurdyka-Lojasiewicz inequality holds

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

**Definition 2.8.** ([26] Kurdyka-Lojasiewicz function) Denote  $\phi_\eta$  be the set of functions that satisfy Definition 2.7 1) - 3). If  $f$  satisfies the KL property at each point of  $\text{dom}(\partial f)$ , then  $f$  is called a KL function.

**Lemma 2.4** ([27] Uniformized KL property) Let  $\Omega$  be a compact set and let  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a proper and lower semicontinuous function. Assume that  $f$  is constant on  $\Omega$  and satisfies the KL property at each point of  $\Omega$ . Then, there exist  $\varepsilon > 0, \eta > 0$ , and  $\varphi \in \phi_\eta$  such that for all  $\bar{x} \in \Omega$  and for all  $x$  in the following intersection:

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

one has

$$\varphi'(f(x) - f(\bar{x}))d(0, \partial f(x)) \geq 1.$$

**Definition 2.9.** We say that  $(x^*, y^*, \lambda^*)$  is a critical point of the Augmented

Lagrangian Function with Bregman distance  $\mathcal{L}_\beta(\cdot)$  if it satisfies

$$\begin{cases} -A^T \lambda^* \in \partial f(x^*), \\ -\lambda^* = \nabla g(y^*), \\ Ax^* + y^* - b = 0. \end{cases}$$

**Lemma 2.5.** [28] Let  $h: \mathbb{R}^n \rightarrow \mathbb{R}$  be a continuously differentiable function whose gradient  $\nabla h$  is Lipschitz continuous with constant  $L > 0$ , then for any  $x, y \in \mathbb{R}^n$ , we have

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

**Definition 2.10.** A proper lower semicontinuous function  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is called semi-convex with constant  $\omega \geq 0$  if the function

$$x \mapsto f(x) + \frac{\omega}{2} \|x\|^2$$

is convex. Specially, if  $\omega = 0$ , then  $g$  is convex.

**Remark 2.4.** Definition (2.10) is equivalent to

$$f(y) \geq f(x) + \langle p, y - x \rangle - \frac{\omega}{2} \|y - x\|^2,$$

for all  $x, y \in \mathbb{R}^n$  and all  $p \in \partial f(x)$ .

### 3. Convergence Analysis

We commence our analysis by examining the optimality condition. By invoking the optimality condition for Equation (6), we obtain

$$\begin{cases} 0 \in \partial f(x_{k+1}) + \nabla_x H(x_{k+1}, y_k) + A^T \lambda_k + \beta_1 A^T (Ax_{k+1} + y_k - b) \\ 0 = \nabla g(y_{k+1}) + \nabla_y H(x_{k+1}, y_{k+1}) + \lambda_k + \beta_2 (\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})) \\ \lambda_{k+1} = \lambda_k + \beta_2 (\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})). \end{cases} \quad (12)$$

By utilizing the previous equation and reorganizing terms, we derive

$$\begin{cases} -\nabla_x H(x_{k+1}, y_k) - A^T \lambda_k - \beta_1 A^T (Ax_{k+1} + y_k - b) \in \partial f(x_{k+1}) \\ -\nabla_y H(x_{k+1}, y_{k+1}) - \lambda_k - \beta_2 (\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})) = \nabla g(y_{k+1}) \\ \lambda_{k+1} = \lambda_k + \beta_2 (\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})). \end{cases} \quad (13)$$

In what follows, we assume:

**Assumption 3.1.** Let  $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  be a semiconvex function with constant  $\omega > 0$ ,  $(g, h)$  be  $L$ -smooth adaptable on  $\text{dom}h$  where  $h$  be  $C^2$  on the interior of  $\text{dom}h$ ,  $\mu_h$  strong-convex and  $\nabla h$  is Lipschitz continuous with  $L_h$  on any bounded subset of  $\mathbb{R}^m$ , and let  $H: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$  be a smooth function. Assume the following holds:

- 1)  $\inf_{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m} H(x, y) > -\infty$ ,  $\inf_{x \in \mathbb{R}^n} f_1(x) > -\infty$ ,  $\inf_{y \in \mathbb{R}^m} f_2(y) > -\infty$ ;
- 2) For any fixed  $x$ , the partial gradient  $\nabla_y H(x, y)$  is globally Lipschitz with constant  $L_2(x)$ , that is

$$\|\nabla_y H(x, y_1) - \nabla_y H(x, y_2)\| \leq L_2(x) \|y_1 - y_2\|, \quad \forall y_1, y_2 \in \mathbb{R}^m.$$

For any fixed  $y$ , the partial gradient  $\nabla_x H(x, y)$  is globally Lipschitz with constant  $L_3(y)$ , that is

$$\|\nabla_x H(x_1, y) - \nabla_x H(x_2, y)\| \leq L_3(y) \|x_1 - x_2\|, \quad \forall x_1, x_2 \in \mathbb{R}^n.$$

3) There exists  $L_2, L_3 > 0$  such that

$$\sup\{L_2(x_k) : k \in N\} \leq L_2, \quad \sup\{L_3(y_k) : k \in N\} \leq L_3.$$

4)  $\nabla H$  is Lipschitz continuous on bounded subsets of  $\mathbb{R}^n \times \mathbb{R}^m$ . In other words, for each bounded subset  $B_1 \times B_2 \subseteq \mathbb{R}^n \times \mathbb{R}^m$ , there exists  $M > 0$  such that for all  $(x_i, y_i) \in B_1 \times B_2, i = 1, 2$ :

$$\begin{aligned} & \|(\nabla_x H(x_1, y_1) - \nabla_x H(x_2, y_2), \nabla_y H(x_1, y_1) - \nabla_y H(x_2, y_2))\| \\ & \leq M \|(x_1 - x_2, y_1 - y_2)\|. \end{aligned}$$

5)  $A^T A \succeq \mu I$  for some  $\mu > 0$ .

6)  $\beta_1 = L_h \beta_2$

For the sake of convenience, in the following analysis, we frequently employ the notation  $\{w_k := (x_k, y_k, \lambda_k)\}_{k \in N}$  and  $\{v_k := (x_k, y_k)\}_{k \in N}$ . We commence our analysis with the subsequent lemma.

**Lemma 3.1.** Let  $\{w_k\}_{k \in N}$  be the sequence generated by the ADMM (1.3) which is assumed to be bounded, then we have

$$\mathcal{L}(w_{k+1}) \leq \mathcal{L}(w_k) - \delta \|v_{k+1} - v_k\|^2 \tag{14}$$

where

$$\begin{aligned} \mathcal{L}(x, y, \lambda) &= f(x) + g(y) + H(x, y) + \lambda^T (Ax + y - b) \\ & \quad + \frac{\beta_1}{2} \|Ax + y - b\|^2 + \beta_2 D_h(-y, Ax - b) \end{aligned}$$

**Proof.**

By considering the definition  $\mathcal{L}(\cdot)$ , we obtain

$$\begin{aligned} \mathcal{L}(x_{k+1}, y_{k+1}, \lambda_{k+1}) &= \mathcal{L}(x_{k+1}, y_{k+1}, \lambda_k) + \langle \lambda_{k+1} - \lambda_k, Ax_{k+1} + y_{k+1} - b \rangle. \\ & \leq \mathcal{L}(x_{k+1}, y_{k+1}, \lambda_k) + \|\lambda_{k+1} - \lambda_k\| \|Ax_{k+1} + y_{k+1} - b\|. \end{aligned} \tag{15}$$

By applying the fact that  $h$  is  $\mu_h$ -strong-convex, we get

$$\langle \nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1}), Ax_{k+1} + y_{k+1} - b \rangle \geq \mu_h \|Ax_{k+1} + y_{k+1} - b\|^2. \tag{16}$$

Using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} & \|\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})\| \|Ax_{k+1} + y_{k+1} - b\| \\ & \geq \langle \nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1}), Ax_{k+1} + y_{k+1} - b \rangle. \end{aligned} \tag{17}$$

Combining (16) and (17), we can conclude that

$$\|\nabla h(-y_{k+1}) - \nabla h(Ax_{k+1} - b)\| \geq \mu_h \|Ax_{k+1} + y_{k+1} - b\|. \tag{18}$$

Based on the iterative scheme of the algorithm, we can infer the following

$$\lambda_{k+1} = \lambda_k + \beta_2 (\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})).$$



Therefore,

$$\|\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})\| = \frac{1}{\beta_2} \|\lambda_{k+1} - \lambda_k\|. \quad (19)$$

By combining Equations (18) and (19), we can infer that

$$\|Ax_{k+1} + y_{k+1} - b\| \leq \frac{1}{\beta_2 \mu_h} \|\lambda_{k+1} - \lambda_k\|. \quad (20)$$

Substituting the previous inequality into Equation (15), which yields

$$\mathcal{L}(x_{k+1}, y_{k+1}, \lambda_{k+1}) \leq \mathcal{L}(x_{k+1}, y_{k+1}, \lambda_k) + \frac{1}{\beta_2 \mu_h} \|\lambda_{k+1} - \lambda_k\|^2. \quad (21)$$

As defined by  $\mathcal{L}(\cdot)$ , we have

$$\begin{aligned} & \mathcal{L}(x_{k+1}, y_{k+1}, \lambda_k) \\ &= \mathcal{L}(x_{k+1}, y_k, \lambda_k) + g(y_{k+1}) - g(y_k) + H(x_{k+1}, y_{k+1}) \\ & \quad - H(x_{k+1}, y_k) + \langle \lambda_k, y_{k+1} - y_k \rangle + \beta_2 D_h(-y_{k+1}, Ax_{k+1} - b) \\ & \quad - \beta_2 D_h(-y_k, Ax_{k+1} - b) + \frac{\beta_1}{2} \|Ax_{k+1} + y_{k+1} - b\|^2 \\ & \quad - \frac{\beta_1}{2} \|Ax_{k+1} + y_k - b\|^2. \end{aligned} \quad (22)$$

Assumption A (3.1) demonstrates  $(g, h)$  is L-smooth adaptable, which implies

$$g(y_{k+1}) - g(y_k) \leq -\langle \nabla g(y_{k+1}), y_k - y_{k+1} \rangle + LD_h(y_k, y_{k+1}). \quad (23)$$

For any fixed  $x$ , the partial gradient  $\nabla_y H(x, y)$  is globally Lipschitz with constant  $L_2(x)$ , which can be inferred from assumptions 2). Further applying the Lemma (2.5) yields

$$\begin{aligned} & H(x_{k+1}, y_{k+1}) - H(x_{k+1}, y_k) \\ & \leq -\langle \nabla_y H(x_{k+1}, y_{k+1}), y_k - y_{k+1} \rangle + \frac{L_2(x_{k+1})}{2} \|y_{k+1} - y_k\|^2. \end{aligned} \quad (24)$$

By three points identity (8), we obtain,

$$\begin{aligned} & \beta_2 D_h(-y_{k+1}, Ax_{k+1} - b) - \beta_2 D_h(-y_k, Ax_{k+1} - b) \\ &= \beta_2 D_h(-y_{k+1}, Ax_{k+1} - b) - \beta_2 D_h(-y_k, Ax_{k+1} - b) \\ & \quad + \beta_2 D_h(-y_k, -y_{k+1}) - \beta_2 D_h(-y_k, -y_{k+1}) \\ &= -\beta_2 (D_h(-y_k, Ax_{k+1} - b) - D_h(-y_k, -y_{k+1})) \\ & \quad - D_h(-y_{k+1}, Ax_{k+1} - b) - \beta_2 D_h(-y_k, -y_{k+1}) \\ &= -\beta_2 \langle \nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1}), y_k - y_{k+1} \rangle \\ & \quad - \beta_2 D_h(-y_k, -y_{k+1}) \\ &= \langle \lambda_{k+1} - \lambda_k, y_{k+1} - y_k \rangle - \beta_2 D_h(-y_k, -y_{k+1}). \end{aligned} \quad (25)$$

The above two equations use that  $\lambda_{k+1} = \lambda_k + \beta_2 (\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1}))$ .

Inserting the above three Equations (24), (25), (23) into (22) yields

$$\begin{aligned}
 & \mathcal{L}(x_{k+1}, y_{k+1}, \lambda_k) - \mathcal{L}(x_{k+1}, y_k, \lambda_k) \\
 & \leq LD_h(y_k, y_{k+1}) - \beta_2 D_h(-y_k, -y_{k+1}) + \frac{L_2}{2} \|y_{k+1} - y_k\|^2 \\
 & \quad + \frac{\beta_1}{2} \|Ax_{k+1} + y_{k+1} - b\|^2 - \frac{\beta_1}{2} \|Ax_{k+1} + y_k - b\|^2 + \langle \lambda_{k+1} - \lambda_k, y_{k+1} - y_k \rangle \\
 & \quad + \langle -(\nabla g(y_{k+1}) + \nabla_y H(x_{k+1}, y_{k+1})), y_k - y_{k+1} \rangle + \langle \lambda_k, y_{k+1} - y_k \rangle
 \end{aligned} \tag{26}$$

From optimal condition,  $\lambda_{k+1} = -(\nabla g(y_{k+1}) + \nabla_y H(x_{k+1}, y_{k+1}))$  (13) we get

$$\begin{aligned}
 & \mathcal{L}(x_{k+1}, y_{k+1}, \lambda_k) - \mathcal{L}(x_{k+1}, y_k, \lambda_k) \\
 & \leq LD_h(y_k, y_{k+1}) - \beta_2 D_h(-y_k, -y_{k+1}) + \frac{L_2}{2} \|y_{k+1} - y_k\|^2 \\
 & \quad + \frac{\beta_1}{2} \|Ax_{k+1} + y_{k+1} - b\|^2 - \frac{\beta_1}{2} \|Ax_{k+1} + y_k - b\|^2.
 \end{aligned} \tag{27}$$

From the condition that  $h$  is  $\mu_h$ -strong-convex and  $L_h$ -smooth, we get

$$\begin{aligned}
 & \frac{\mu_h}{2} \|y_{k+1} - y_k\|^2 \leq D_h(-y_k, -y_{k+1}) \leq \frac{L_h}{2} \|y_{k+1} - y_k\|^2, \\
 & \frac{\mu_h}{2} \|y_{k+1} - y_k\|^2 \leq D_h(y_k, y_{k+1}) \leq \frac{L_h}{2} \|y_{k+1} - y_k\|^2.
 \end{aligned} \tag{28}$$

Substituting the aforementioned inequality into (27) reveals

$$\begin{aligned}
 & \mathcal{L}(x_{k+1}, y_{k+1}, \lambda_k) - \mathcal{L}(x_{k+1}, y_k, \lambda_k) \\
 & \leq \frac{LL_h + L_2 - \beta_2 \mu_h}{2} \|y_{k+1} - y_k\|^2 + \frac{\beta_1}{2} \|Ax_{k+1} + y_{k+1} - b\|^2 - \frac{\beta_1}{2} \|Ax_{k+1} + y_k - b\|^2.
 \end{aligned} \tag{29}$$

By definition of  $\mathcal{L}(\cdot)$ , we obtain

$$\begin{aligned}
 & \mathcal{L}(x_{k+1}, y_k, \lambda_k) - \mathcal{L}(x_k, y_k, \lambda_k) \\
 & = f(x_{k+1}) - f(x_k) + H(x_{k+1}, y_k) - H(x_k, y_k) + \langle \lambda_k, Ax_{k+1} - Ax_k \rangle \\
 & \quad + \beta_2 D_h(-y_k, Ax_{k+1} - b) - \beta_2 D_h(-y_k, Ax_k - b) \\
 & \quad + \frac{\beta_1}{2} \|Ax_{k+1} + y_k - b\|^2 - \frac{\beta_1}{2} \|Ax_k + y_k - b\|^2.
 \end{aligned} \tag{30}$$

From the Assumption 3.1 that the partial gradient  $\nabla_x H(x, y)$  is globally Lipschitz with constant  $L_3(y)$ , which implies

$$\|\nabla_x H(x_1, y) - \nabla_x H(x_2, y)\| \leq L_3(y) \|x_1 - x_2\|, \quad \forall x_1, x_2 \in \mathbb{R}^n.$$

By Lemma 2.5, we know

$$H(x_{k+1}, y_k) - H(x_k, y_k) \leq -\langle \nabla_x H(x_{k+1}, y_k), x_k - x_{k+1} \rangle + \frac{L_3(y_k)}{2} \|x_{k+1} - x_k\|^2. \tag{31}$$

Besides,  $f$  is semiconvex with parameter  $\omega$ , which implies (2.4) as follows:

$$f(x_{k+1}) - f(x_k) \leq -\langle p, x_k - x_{k+1} \rangle + \frac{\omega}{2} \|x_{k+1} - x_k\|^2,$$

for all  $x, y \in \mathcal{R}^n$  and all  $p \in \partial f(x_{k+1})$ .

From optimality condition (13), we could know

$$-\nabla_x H(x_{k+1}, y_k) - A^T \lambda_k + \beta_1 A^T (-Ax_{k+1} - y_k + b) \in \partial f(x_{k+1}).$$

Therefore,

$$\begin{aligned}
 & f(x_{k+1}) - f(x_k) \\
 & \leq \langle -\nabla_x H(x_{k+1}, y_k) - A^T \lambda_k + \beta_1 A^T (-Ax_{k+1} - y_k + b), x_{k+1} - x_k \rangle \\
 & \quad + \frac{\omega}{2} \|x_{k+1} - x_k\|^2 \\
 & = \langle -\nabla_x H(x_{k+1}, y_k), x_{k+1} - x_k \rangle - \langle A^T \lambda_k, x_{k+1} - x_k \rangle \\
 & \quad + \langle \beta_1 A^T (-Ax_{k+1} - y_k + b), x_{k+1} - x_k \rangle + \frac{\omega}{2} \|x_{k+1} - x_k\|^2.
 \end{aligned} \tag{32}$$

Inserting Equation (32), (31) into (30) yields

$$\begin{aligned}
 & \mathcal{L}(x_{k+1}, y_k, \lambda_k) - \mathcal{L}(x_k, y_k, \lambda_k) \\
 & \leq \langle -\nabla_x H(x_{k+1}, y_k), x_{k+1} - x_k \rangle - \langle A^T \lambda_k, x_{k+1} - x_k \rangle \\
 & \quad + \langle \beta_1 A^T (-Ax_{k+1} - y_k + b), x_{k+1} - x_k \rangle + \frac{\omega}{2} \|x_{k+1} - x_k\|^2 \\
 & \quad - \langle \nabla_x H(x_{k+1}, y_k), x_k - x_{k+1} \rangle + \frac{L_3(y_k)}{2} \|x_{k+1} - x_k\|^2 \\
 & \quad + \langle \lambda_k, Ax_{k+1} - Ax_k \rangle + \beta_2 D_h(-y_k, Ax_{k+1} - b) - \beta_2 D_h(-y_k, Ax_k - b) \\
 & \quad + \frac{\beta_1}{2} \|Ax_{k+1} + y_k - b\|^2 - \frac{\beta_1}{2} \|Ax_k + y_k - b\|^2 \\
 & = -\beta_1 \langle A^T (-Ax_{k+1} - y_k + b), x_k - x_{k+1} \rangle + \frac{\omega}{2} \|x_{k+1} - x_k\|^2 + \frac{L_3}{2} \|x_{k+1} - x_k\|^2 \\
 & \quad + \beta_2 D_h(-y_k, Ax_{k+1} - b) - \beta_2 D_h(-y_k, Ax_k - b) + \frac{\beta_1}{2} \|Ax_{k+1} + y_k - b\|^2 \\
 & \quad - \frac{\beta_1}{2} \|Ax_k + y_k - b\|^2.
 \end{aligned} \tag{33}$$

From the three-point equation, we get

$$\begin{aligned}
 & \beta_2 D_h(-y_k, Ax_{k+1} - b) - \beta_2 D_h(-y_k, Ax_k - b) \\
 & = -\beta_2 (D_h(-y_k, Ax_k - b) - D_h(-y_k, Ax_{k+1} - b)) \\
 & \quad - D_h(Ax_{k+1} - b, Ax_k - b) - \beta_2 D_h(Ax_{k+1} - b, Ax_k - b) \\
 & = -\beta_2 \langle \nabla h(Ax_k - b) - \nabla h(Ax_{k+1} - b), Ax_{k+1} + y_k - b \rangle \\
 & \quad - \beta_2 D_h(Ax_{k+1} - b, Ax_k - b).
 \end{aligned} \tag{34}$$

Using (28), we could know

$$-D_h(Ax_{k+1} - b, Ax_k - b) \leq -\frac{\mu_h}{2} \|Ax_{k+1} - Ax_k\|^2. \tag{35}$$

Substituting above two equations (34), (35) into (33), we obtain

$$\begin{aligned}
 & \mathcal{L}(x_{k+1}, y_k, \lambda_k) - \mathcal{L}(x_k, y_k, \lambda_k) \\
 & \leq \frac{\omega + L_3}{2} \|x_{k+1} - x_k\|^2 - \frac{\beta_2 \mu_h}{2} \|Ax_{k+1} - Ax_k\|^2 + \frac{\beta_1}{2} \|Ax_{k+1} + y_k - b\|^2 \\
 & \quad - \frac{\beta_1}{2} \|Ax_k + y_k - b\|^2 - \beta_1 \langle A^T (-Ax_{k+1} - y_k + b), x_k - x_{k+1} \rangle \\
 & \quad - \beta_2 \langle \nabla h(Ax_k - b) - \nabla h(Ax_{k+1} - b), Ax_{k+1} + y_k - b \rangle.
 \end{aligned} \tag{36}$$

By the fact that  $\nabla h$  is Lipschitz continuous with  $L_h$  on any bounded subset

of  $\mathbb{R}^m$ , we get

$$\begin{aligned} & -\beta_2 \langle \nabla h(Ax_k - b) - \nabla h(Ax_{k+1} - b), Ax_{k+1} + y_k - b \rangle \\ & \leq \beta_2 L_h \|Ax_{k+1} + y_k - b\| \|Ax_k - Ax_{k+1}\|. \end{aligned} \tag{37}$$

Using Equation (21), (29) and (36), which yields

$$\begin{aligned} & \mathcal{L}(x_{k+1}, y_{k+1}, \lambda_{k+1}) - \mathcal{L}(x_k, y_k, \lambda_k) \\ & \leq \frac{\omega + L_3}{2} \|x_{k+1} - x_k\|^2 - \frac{\beta_2 \mu_h}{2} \|Ax_{k+1} - Ax_k\|^2 - \frac{\beta_1}{2} \|Ax_k + y_k - b\|^2 \\ & \quad + \frac{LL_h + L_2 - \beta_2 \mu_h}{2} \|y_{k+1} - y_k\|^2 + \frac{\beta_1}{2} \|Ax_{k+1} + y_{k+1} - b\|^2 + \frac{1}{\beta_2 \mu_h} \|\lambda_{k+1} - \lambda_k\|^2 \\ & \quad - \beta_1 \langle A^T(-Ax_{k+1} - y_k + b), x_k - x_{k+1} \rangle + \beta_2 L_h \|Ax_{k+1} + y_k - b\| \|Ax_k - Ax_{k+1}\|. \end{aligned} \tag{38}$$

It is not difficult to see

$$\begin{aligned} & -\frac{\beta_1}{2} \|Ax_k + y_k - b\|^2 - \beta_1 \langle A^T(-Ax_{k+1} - y_k + b), x_k - x_{k+1} \rangle \\ & = -\frac{\beta_1}{2} \|Ax_{k+1} + y_k - b + Ax_k - Ax_{k+1}\|^2 - \beta_1 \langle A^T(-Ax_{k+1} - y_k + b), x_k - x_{k+1} \rangle \\ & = -\frac{\beta_1}{2} (\|Ax_{k+1} + y_k - b\|^2 + \|Ax_k - Ax_{k+1}\|^2 + 2 \langle Ax_{k+1} + y_k - b, Ax_k - Ax_{k+1} \rangle) \\ & \quad - \beta_1 \langle A^T(-Ax_{k+1} - y_k + b), x_k - x_{k+1} \rangle \\ & = -\frac{\beta_1}{2} (\|Ax_{k+1} + y_k - b\|^2 + \|Ax_k - Ax_{k+1}\|^2) \\ & \leq -\beta_1 \|Ax_{k+1} + y_k - b\| \|Ax_k - Ax_{k+1}\| \\ & = -\beta_2 L_h \|Ax_{k+1} + y_k - b\| \|Ax_k - Ax_{k+1}\| (\beta_1 = L_h \beta_2). \end{aligned} \tag{39}$$

Using the condition that  $\beta_1 = L_h \beta_2$ , which yields

$$\begin{aligned} & -\frac{\beta_1}{2} \|Ax_k + y_k - b\|^2 - \beta_1 \langle A^T(-Ax_{k+1} - y_k + b), x_k - x_{k+1} \rangle \\ & \quad - \beta_2 \langle \nabla h(Ax_k - b) - \nabla h(Ax_{k+1} - b), Ax_{k+1} + y_k - b \rangle \leq 0 \end{aligned} \tag{40}$$

From the assumption that  $A^T A \succeq \mu I$  for some  $\mu > 0$ , we obtain

$$\|Ax_{k+1} - Ax_k\|^2 \geq \mu \|x_{k+1} - x_k\|^2. \tag{41}$$

Substituting (40), (41) and (20) into (38), which yields

$$\begin{aligned} & \mathcal{L}(x_{k+1}, y_{k+1}, \lambda_{k+1}) - \mathcal{L}(x_k, y_k, \lambda_k) \\ & \leq \frac{LL_h + L_2 - \beta_2 \mu_h}{2} \|y_{k+1} - y_k\|^2 + \frac{\omega + L_3}{2} \|x_{k+1} - x_k\|^2 \\ & \quad - \frac{\beta_2 \mu_h}{2} \|Ax_{k+1} - Ax_k\|^2 + \frac{\beta_1 + 2\beta_2 \mu_h}{2\beta_2^2 \mu_h^2} \|\lambda_{k+1} - \lambda_k\|^2 \\ & \leq \frac{LL_h + L_2 - \beta_2 \mu_h}{2} \|y_{k+1} - y_k\|^2 + \frac{\omega + L_3 - \beta_2 \mu_h \mu}{2} \|x_{k+1} - x_k\|^2 \\ & \quad + \frac{L_h + 2\mu_h}{2\beta_2 \mu_h^2} \|\lambda_{k+1} - \lambda_k\|^2. \end{aligned} \tag{42}$$

From the optimality condition (13), we have

$$\lambda_{k+1} = -(\nabla g(y_{k+1}) + \nabla_y H(x_{k+1}, y_{k+1})). \tag{43}$$

Then,

$$\begin{aligned} \|\lambda_{k+1} - \lambda_k\|^2 &= \|\nabla g(y_k) - \nabla g(y_{k+1}) + \nabla_y H(x_k, y_k) - \nabla_y H(x_{k+1}, y_{k+1})\|^2 \\ &\leq 2\|\nabla g(y_k) - \nabla g(y_{k+1})\|^2 + 2\|\nabla_y H(x_k, y_k) - \nabla_y H(x_{k+1}, y_{k+1})\|^2. \end{aligned} \tag{44}$$

Using assumption 4),

$$\begin{aligned} &\|(\nabla_x H(x_1, y_1) - \nabla_x H(x_2, y_2), \nabla_y H(x_1, y_1) - \nabla_y H(x_2, y_2))\| \\ &\leq M \|(x_1 - x_2, y_1 - y_2)\| \end{aligned}$$

We can obtain

$$\begin{aligned} &\|\nabla_x H(x_k, y_k) - \nabla_x H(x_{k+1}, y_{k+1})\|^2 + \|\nabla_y H(x_k, y_k) - \nabla_y H(x_{k+1}, y_{k+1})\|^2 \\ &\leq M^2 \|x_{k+1} - x_k\|^2 + M^2 \|y_{k+1} - y_k\|^2, \end{aligned}$$

which implies

$$2\|\nabla_y H(x_k, y_k) - \nabla_y H(x_{k+1}, y_{k+1})\|^2 \leq 2M^2 \|x_{k+1} - x_k\|^2 + 2M^2 \|y_{k+1} - y_k\|^2. \tag{45}$$

Next, we consider two different scenarios:

1)  $y_{k+1} \neq y_k$ ,

From the given assumption that  $(g, h)$  is L-smooth adaptable and Equation (9), which yields

$$\|\nabla g(x) - \nabla g(y)\| \leq L \frac{D_h(x, y) + D_h(y, x)}{\|x - y\|}, \text{ for all } x \neq y \in \text{int}(\text{dom}(h)),$$

then,

$$\begin{aligned} &\|\nabla g(y_{k+1}) - \nabla g(y_k)\| \\ &\leq L \frac{D_h(y_{k+1}, y_k) + D_h(y_k, y_{k+1})}{\|y_{k+1} - y_k\|} \\ &= L \frac{h(y_{k+1}) - h(y_k) - \langle \nabla h(y_k), y_{k+1} - y_k \rangle + h(y_k) - h(y_{k+1}) - \langle \nabla h(y_{k+1}), y_k - y_{k+1} \rangle}{\|y_{k+1} - y_k\|} \\ &= L \frac{\langle \nabla h(y_k) - \nabla h(y_{k+1}), y_k - y_{k+1} \rangle}{\|y_{k+1} - y_k\|} \\ &\leq L \frac{L_h \|y_k - y_{k+1}\|^2}{\|y_{k+1} - y_k\|} = LL_h \|y_{k+1} - y_k\|. \end{aligned} \tag{46}$$

2)  $y_{k+1} = y_k$ , it is easy to verify that (46) holds.

By substituting Equation (45) and (46) into (44), we obtain

$$\begin{aligned} \|\lambda_{k+1} - \lambda_k\|^2 &\leq 2L^2 L_h^2 \|y_{k+1} - y_k\|^2 + 2M^2 \|x_{k+1} - x_k\|^2 + 2M^2 \|y_{k+1} - y_k\|^2 \\ &= (2L^2 L_h^2 + 2M^2) \|y_{k+1} - y_k\|^2 + 2M^2 \|x_{k+1} - x_k\|^2. \end{aligned} \tag{47}$$

Combining (47) and (42), we have

$$\begin{aligned} & \mathcal{L}(x_{k+1}, y_{k+1}, \lambda_{k+1}) - \mathcal{L}(x_k, y_k, \lambda_k) \\ & \leq \frac{\beta_2 \mu_h^2 (LL_h + L_2 - \beta_2 \mu_h) + (L_h + 2\mu_h)(2L^2 L_h^2 + 2M^2)}{2\beta_2 \mu_h^2} \|y_{k+1} - y_k\|^2 \\ & \quad + \frac{\beta_2 \mu_h^2 (\omega + L_3 - \beta_2 \mu_h \mu) + 2M^2 (L_h + 2\mu_h)}{2\beta_2 \mu_h^2} \|x_{k+1} - x_k\|^2. \end{aligned} \tag{48}$$

By simple calculation, when

$$\beta_2 > \max \{ \beta_x, \beta_y \}$$

where

$$\beta_y := \frac{LL_h + L_2 + \sqrt{(LL_h + L_2)^2 + 4\mu_h^3 (L_h + 2\mu_h)(2L^2 L_h^2 + 2M^2)}}{2\mu_h^3}, \tag{49}$$

$$\beta_x := \frac{\omega + L_3 + \sqrt{(\omega + L_3)^2 + 8\mu_h^3 \mu M^2 (L_h + 2\mu_h)}}{2\mu_h^3 \mu}, \tag{50}$$

then

$$\mathcal{L}(w_{k+1}) \leq \mathcal{L}(w_k) - \delta \|v_{k+1} - v_k\|^2,$$

where

$$\begin{aligned} \delta & := \min \{ \delta_1, \delta_2 \} \\ \delta_1 & := \frac{-LL_h - L_2 + \beta \mu_h}{2} - \frac{2L^2 L_h^2 + 2M^2}{\beta_2 \mu_h} - \frac{L^2 L_h^3 + M^2}{\beta_2 \mu_h^2} \\ \delta_2 & := \frac{-\omega - L_3 + \beta_2 \mu_h \mu}{2} - \frac{M^2 (L_h + 2\mu_h)}{\beta_2 \mu_h^2}. \end{aligned}$$

Thus, the theorem is established.

**Remark 3.1.** If  $H = 0$ , then we have  $L_2 = L_3 = M = 0$ . Moreover, by  $L_h = \mu_h = 1$ , we have  $\beta_2 > \max \left\{ 3L, \frac{\omega}{\mu} \right\}$ . This result is larger than  $2L$  which is proved by Guo et al.'s classical result [8]. However, if we directly solve the separable problem, it can cover their conclusion. The reason of our weaker conclusion could be that the inequality bounds are not tight enough in our proof.

**Lemma 3.2.** Let  $\{\omega_k\}_{k \in \mathbb{N}}$  be the sequence generated by the Bragman ADMM (6) which is assumed to be bounded. Then the following holds

$$\sum_{k=0}^{+\infty} \|\omega_{k+1} - \omega_k\|^2 < +\infty. \tag{51}$$

**Proof.**

Since  $\{\omega_k\}$  is bounded, then there exists a subsequence  $\{\omega_{k_j}\}$  such that  $\omega_{k_j} \rightarrow \omega^*$ . By the condition that  $f$  is lower semi-continuous and  $g, H$  are continuous, we obtain  $L(\cdot)$  is semicontinuous, which implies

$$\mathcal{L}(\omega^*) \leq \liminf_{j \rightarrow +\infty} \mathcal{L}(\omega_{k_j}).$$

Consequently,  $\left\{ \mathcal{L}(\omega_{k_j}) \right\}_{j \in N}$  is bounded from below.

Note that, Lemma 3.1 implies that  $\left\{ \mathcal{L}(\omega_{k_j}) \right\}_{j \in N}$  is nonincreasing and thus  $\left\{ \mathcal{L}(\omega_{k_j}) \right\}_{j \in N}$  is convergent. Moreover, we have  $\left\{ \mathcal{L}(\omega_{k_j}) \right\}_{j \in N}$  is convergent and  $\mathcal{L}(\omega_k) \geq \mathcal{L}(\omega^*)$ . Reassanging terms of (3.1) leads to

$$\delta \|v_{k+1} - v_k\|^2 \leq \mathcal{L}(\omega_k) - \mathcal{L}(\omega_{k+1}).$$

Now, summing the above inequality over  $k = 0, \dots, n$  yields

$$\sum_{k=0}^n \delta \|v_{k+1} - v_k\|^2 \leq \mathcal{L}(\omega_0) - \mathcal{L}(\omega_{n+1}) \leq \mathcal{L}(\omega_0) - \mathcal{L}(\omega^*) < +\infty.$$

Since  $\delta > 0$ , we have  $\sum_{k=0}^{+\infty} \|v_{k+1} - v_k\|^2 < +\infty$ , thus

$$\sum_{k=0}^{+\infty} \|x_{k+1} - x_k\|^2 < +\infty, \quad \sum_{k=0}^{+\infty} \|y_{k+1} - y_k\|^2 < +\infty. \tag{52}$$

Moreover, it follows from (52) and (47) that  $\sum_{k=0}^{+\infty} \|\lambda_{k+1} - \lambda_k\|^2 < +\infty$ , therefore, we obtain  $\sum_{k=0}^{+\infty} \|\omega_{k+1} - \omega_k\|^2 < +\infty$ .

**Lemma 3.3.** *Let  $\{\omega_k\}_{k \in N}$  be the sequence generated by the Bragman ADMM (6) which is assumed to be bounded. Then there exists  $\zeta > 0$  such that*

$$d(0, \partial \mathcal{L}(\omega_{k+1})) \leq \zeta \|v_{k+1} - v_k\|. \tag{53}$$

**Proof.**

By the definition of function  $\mathcal{L}(\cdot)$ , it follows

$$\begin{cases} \partial_x \mathcal{L}(\omega_{k+1}) = \partial f(x_{k+1}) + \nabla_x H(x_{k+1}, y_{k+1}) + A^T \lambda_{k+1} + L_h \beta_2 A^T \langle Ax_{k+1} + y_{k+1} - b \rangle \\ \quad + \beta_2 A^T \langle \nabla h^2(Ax_{k+1} - b), Ax_{k+1} + y_{k+1} - b \rangle \\ \partial_y \mathcal{L}(\omega_{k+1}) = \nabla g(y_{k+1}) + \nabla_y H(x_{k+1}, y_{k+1}) + \lambda_{k+1} + L_h \beta_2 (Ax_{k+1} + y_{k+1} - b) \\ \quad + \beta_2 (\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})) \\ \partial_\lambda \mathcal{L}(\omega_{k+1}) = Ax_{k+1} + y_{k+1} - b. \end{cases} \tag{54}$$

This, together with Equation (13), which yields

$$\begin{cases} \partial_x \mathcal{L}(\omega_{k+1}) = \nabla_x H(x_{k+1}, y_{k+1}) - \nabla_x H(x_{k+1}, y_k) + A^T (\lambda_{k+1} - \lambda_k) \\ \quad + L_h \beta_2 A^T \langle y_{k+1} - y_k \rangle + \beta_2 A^T \langle \nabla h^2(Ax_{k+1} - b), Ax_{k+1} + y_{k+1} - b \rangle \\ \partial_y \mathcal{L}(\omega_{k+1}) = \lambda_{k+1} - \lambda_k + L_h \beta_2 (Ax_{k+1} + y_{k+1} - b) \\ \partial_\lambda \mathcal{L}(\omega_{k+1}) = Ax_{k+1} + y_{k+1} - b. \end{cases} \tag{55}$$

Let

$$\begin{aligned} & \left( \xi_{k+1}^1, \xi_{k+1}^2, \xi_{k+1}^3 \right) \\ & := \left( \nabla_x H(x_{k+1}, y_{k+1}) - \nabla_x H(x_{k+1}, y_k) + A^T (\lambda_{k+1} - \lambda_k) + L_h \beta_2 A^T \langle y_{k+1} - y_k \rangle \right. \\ & \quad \left. + \beta_2 A^T \langle \nabla h^2(Ax_{k+1} - b), Ax_{k+1} + y_{k+1} - b \rangle, \lambda_{k+1} - \lambda_k \right. \\ & \quad \left. + L_h \beta_2 (Ax_{k+1} + y_{k+1} - b), Ax_{k+1} + y_{k+1} - b \right). \end{aligned} \tag{56}$$

Then it follows from Lemma 2.3 that  $(\xi_{k+1}^1, \xi_{k+1}^2, \xi_{k+1}^3) \in \partial \mathcal{L}(\omega_{k+1})$ . Furthermore,

there exist  $\zeta_1, \zeta_2, \zeta_3 > 0$  such that

$$\begin{aligned} \left\| \left( \xi_{k+1}^1, \xi_{k+1}^2, \xi_{k+1}^3 \right) \right\| &\leq \zeta_1 \|y_{k+1} - y_k\| + \zeta_2 \|\lambda_{k+1} - \lambda_k\| \\ &+ \zeta_3 \left\| \nabla_x H(x_{k+1}, y_{k+1}) - \nabla_x H(x_{k+1}, y_k) \right\|. \end{aligned} \tag{57}$$

Again, since  $\nabla H$  is Lipschitz continuous on bounded subsets and  $\{(x_k, y_k)\}$  is bounded, it follows

$$\left\| \nabla_x H(x_{k+1}, y_{k+1}) - \nabla_x H(x_{k+1}, y_k) \right\| \leq M \|y_{k+1} - y_k\|. \tag{58}$$

From (47), we know

$$\|\lambda_{k+1} - \lambda_k\| \leq \sqrt{2L^2L_h^2 + 2M^2} \|y_{k+1} - y_k\| + \sqrt{2}M \|x_{k+1} - x_k\|. \tag{59}$$

By setting  $\zeta = \sqrt{\left(\zeta_1 + \sqrt{2L^2L_h^2 + 2M^2}\zeta_2 + M\zeta_3\right)^2 + \left(\sqrt{2}M\zeta_3\right)^2}$ , it follows above three Equations (57), (58), (59), we can obtain

$$\begin{aligned} d(0, \partial\mathcal{L}(\omega_{k+1})) &\leq \left\| \left( \xi_{k+1}^1, \xi_{k+1}^2, \xi_{k+1}^3 \right) \right\| \\ &\leq \left( \zeta_1 + \sqrt{2L^2L_h^2 + 2M^2}\zeta_2 + M\zeta_3 \right) \|y_{k+1} - y_k\| \\ &+ \sqrt{2}M\zeta_3 \|x_{k+1} - x_k\| \\ &\leq \zeta \|v_{k+1} - v_k\|, \end{aligned} \tag{60}$$

where the third inequality follows from the Cauchy inequality. Thus, the Lemma 3.3 is proven.

**Lemma 3.4.** Let  $\{\omega_k\}_{k \in \mathbb{N}}$  be the sequence generated by the Bragman ADMM (6) which is assumed to be bounded. Let  $S(\omega_0)$  denote the set of its limit point. Then,

- 1)  $S(\omega_0)$  is a nonempty compact set, and  $d(\omega_k, S(\omega_0)) \rightarrow 0$ , as  $k \rightarrow +\infty$ ;
- 2)  $S(\omega_0) \subseteq \text{crit } \mathcal{L}_\beta$ ;
- 3)  $\mathcal{L}(\cdot)$  is finite and constant on  $S(\omega_0)$ , equal to

$$\inf_{k \in \mathbb{N}} \mathcal{L}(\omega_k) = \lim_{k \rightarrow +\infty} \mathcal{L}(\omega_k).$$

**Proof.**

1) This item follows as an elementary consequence of the definition of limit points.

2) For any fixed  $(x^*, y^*, \lambda^*) \in S(\omega_0)$ , there exists a subsequence  $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}_{j \in \mathbb{N}}$  of  $\{(x^k, y^k, \lambda^k)\}_{k \in \mathbb{N}}$  converges to  $(x^*, y^*, \lambda^*)$ .

Since  $x^{k+1}$  is the global minimizer of  $\mathcal{L}(x, y^k, \lambda^k)$  for the variable  $x$ , it holds

$$\mathcal{L}(x^{k+1}, y^k, \lambda^k) \leq \mathcal{L}(x^*, y^k, \lambda^k).$$

Using the above inequality and the continuity of  $\mathcal{L}(\cdot)$  with respect to  $y$  and  $\lambda$  ensure

$$\limsup_{j \rightarrow +\infty} \mathcal{L}_\beta(x^{k_j+1}, y^{k_j}, \lambda^{k_j}) = \limsup_{j \rightarrow +\infty} \mathcal{L}(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1}) \leq \mathcal{L}(x^*, y^*, \lambda^*).$$

On the other hand, Lemma 3.3 implies  $\|w^{k+1} - w^k\| \rightarrow 0$ , which means that



the subsequence  $\left\{ \left( x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1} \right) \right\}_{j \in N}$  also converges to  $(x^*, y^*, \lambda^*)$ .

From the lower semicontinuity of  $\mathcal{L}(\cdot)$ , we have

$$\liminf_{j \rightarrow +\infty} \mathcal{L}(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1}) \geq \mathcal{L}(x^*, y^*, \lambda^*). \tag{61}$$

Then by combining above two equations together we can get

$$\lim_{j \rightarrow +\infty} \mathcal{L}(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1}) = \mathcal{L}(x^*, y^*, \lambda^*), \tag{62}$$

which implies

$$\lim_{j \rightarrow +\infty} f(x^{k_j+1}) = f(x^*). \tag{63}$$

Passing to the limit in (13) and (54) along the subsequence

$\left\{ \left( x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1} \right) \right\}_{j \in N}$  and invoking (63), the continuity of  $\nabla g(\cdot)$ ,  $\nabla_x H(\cdot, \cdot)$ ,  $\nabla_y H(\cdot, \cdot)$ , it follows that

$$\begin{cases} -\nabla_x H(x^*, y^*) - A^T \lambda^* \in \partial f(x^*) \\ -\nabla_y H(x^*, y^*) - \lambda^* = \nabla g(y^*) \\ \nabla h(Ax^* - b) - \nabla h(-y^*) = 0. \end{cases}$$

The last equation implies that  $Ax^* + y^* - b = 0$  due to the strong convexity of  $h(\cdot)$ .

By the definition of  $\mathcal{L}(\cdot)$ , we have

$$\begin{cases} 0 \in \partial_x \mathcal{L}(w^*) = \partial f(x^*) + \nabla_x H(x^*, y^*) + A^T \lambda^* \\ 0 \in \partial_y \mathcal{L}(w^*) = \nabla g(y^*) + \nabla_y H(x^*, y^*) + \lambda^* \\ 0 \in \partial_\lambda \mathcal{L}(w^*) = Ax^* + y^* - b. \end{cases} \tag{64}$$

By the definition of critical point of the augmented Lagrangian function  $\mathcal{L}_\beta(\cdot)$  (1.2), if it satisfies:

$$\begin{cases} A^T \lambda^* - \nabla_x H(x^*, y^*) \in \partial f(x^*) \\ \lambda^* - \nabla_y H(x^*, y^*) = \nabla g(y^*) \\ Ax^* + y^* - b = 0 \end{cases}$$

Then,  $(x^*, y^*, \lambda^*)$  is a critical point of  $\mathcal{L}_\beta, \mathcal{L}$ , i.e.,  $(x^*, y^*, \lambda^*) \in \text{crit } \mathcal{L}_\beta$ , and  $(x^*, y^*, \lambda^*) \in \text{crit } \mathcal{L}$ .

3) For any point  $(x^*, y^*, \lambda^*) \in S(w^0)$ , there exists a subsequence

$\left\{ \left( x^{k_j}, y^{k_j}, \lambda^{k_j} \right) \right\}_{j \in N}$  converges to  $(x^*, y^*, \lambda^*)$ .

Since  $\left\{ \mathcal{L}(w^k) \right\}_{k \in N}$  is nonincreasing, combining (61) and (62) leads to

$$\lim_{k \rightarrow +\infty} \mathcal{L}(x^k, y^k, \lambda^k) = \mathcal{L}(x^*, y^*, \lambda^*). \tag{65}$$

Therefore,  $\mathcal{L}(\cdot)$  is constant on  $S(w^0)$ . Moreover,

$$\inf_{k \in N} \mathcal{L}(w^k) = \lim_{k \rightarrow +\infty} \mathcal{L}(w^k).$$

Next, we will prove the convergence of the iterative sequence.

**Theorem 3.1.** Let  $\{w^k\}_{k \in \mathbb{N}}$  be the sequence generated by the ADMM (6) which is assumed to be bounded. Suppose that  $\mathcal{L}(\cdot)$  is a KL function, then  $\{w^k\}_{k \in \mathbb{N}}$  has finite length, that is

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

and as a consequence, we have  $\{w^k\}_{k \in \mathbb{N}}$  converged to a critical point of  $\mathcal{L}_\beta(\cdot)$ .

**Proof.**

Since from the proof of Lemma (3.4), it follows that  $\mathcal{L}(w^k) \rightarrow \mathcal{L}(w^*)$  for all  $w^* \in S(w^0)$ . We consider two cases:

1) If there exists an integer  $k_0$  for which  $\mathcal{L}(w^{k_0}) = \mathcal{L}(w^*)$ . Rearranging terms of (14) we have that for any  $k > k_0$ ,

$$\delta \|v^{k+1} - v^k\|^2 \leq \mathcal{L}(w^k) - \mathcal{L}(w^{k+1}) \leq \mathcal{L}(w^{k_0}) - \mathcal{L}(w^*) = 0,$$

implying that  $v^{k+1} = v^k$  for any  $k > k_0$ .

Associated with (47), for any  $k > k_0 + 1$ , it follows that  $w^{k+1} = w^k$  and the assertion holds.

2) Now, assume  $\mathcal{L}(w^k) > \mathcal{L}(w^*)$  for all  $k$ . We claim there exists  $\tilde{k} > 0$  such that for all  $k > \tilde{k}$ ,

$$\delta \|v^{k+1} - v^k\|^2 \leq \zeta \|v^k - v^{k-1}\| \Delta_{k,k+1}, \tag{66}$$

where  $\Delta_{p,q} := \varphi(\mathcal{L}(w^p) - \mathcal{L}(w^*)) - \varphi(\mathcal{L}(w^q) - \mathcal{L}(w^*))$ . To see this, note that  $d(w^k, S(w^0)) \rightarrow 0$  and  $\mathcal{L}_\beta(w^k) \rightarrow \mathcal{L}(w^*)$ , then for all  $\varepsilon, \eta > 0$ , there exists  $\tilde{k} > 0$  such that when  $k > \tilde{k}$ , we have

$$d(w^k, S(w^0)) < \varepsilon, \quad \mathcal{L}(w^*) < \mathcal{L}(w^k) < \mathcal{L}(w^*) + \eta.$$

Since  $S(w^0)$  is a nonempty compact set and  $\mathcal{L}(\cdot)$  is constant on  $S(w^0)$ , applying Lemma (2.4) with  $\Omega := S(w^0)$ , we deduce that for any  $k > \tilde{k}$

$$\varphi'(\mathcal{L}(w^k) - \mathcal{L}(w^*)) d(0, \partial \mathcal{L}(w^k)) \geq 1.$$

Since  $\mathcal{L}(w^k) - \mathcal{L}(w^{k+1}) = (\mathcal{L}(w^k) - \mathcal{L}(w^*)) - (\mathcal{L}(w^{k+1}) - \mathcal{L}_\beta(w^*))$ , making use of the concavity of  $\varphi$  we get that

$$\begin{aligned} & \varphi(\mathcal{L}(w^k) - \mathcal{L}(w^*)) - \varphi(\mathcal{L}_\beta(w^{k+1}) - \mathcal{L}_\beta(w^*)) \\ & \geq \varphi'(\mathcal{L}(w^k) - \mathcal{L}(w^*)) (\mathcal{L}_\beta(w^k) - \mathcal{L}_\beta(w^{k+1})) \end{aligned}$$

Thus, using the inequalities  $d(0, \partial \mathcal{L}(w^k)) \leq \zeta \|v^k - v^{k-1}\|$  and  $\varphi'(\mathcal{L}(w^k) - \mathcal{L}(w^*)) > 0$ , we obtain

$$\begin{aligned} & \mathcal{L}(w^k) - \mathcal{L}(w^{k+1}) \\ & \leq \frac{\varphi(\mathcal{L}(w^k) - \mathcal{L}(w^*)) - \varphi(\mathcal{L}(w^{k+1}) - \mathcal{L}(w^*))}{\varphi'(\mathcal{L}(w^k) - \mathcal{L}(w^*))} \\ & \leq \zeta \|v^k - v^{k-1}\| \left[ \varphi(\mathcal{L}(w^k) - \mathcal{L}(w^*)) - \varphi(\mathcal{L}(w^{k+1}) - \mathcal{L}(w^*)) \right]. \end{aligned}$$

Combining Lemma 3.1 with the above relation gives (66), as desired. Moreover, (66) implies

$$\|v^{k+1} - v^k\| \leq \sqrt{\frac{\zeta}{\delta} \Delta_{k,k+1}} \|v^k - v^{k-1}\|^{\frac{1}{2}}.$$

Notice that  $2\sqrt{\alpha\beta} \leq \alpha + \beta$ , for all  $\alpha, \beta \geq 0$ . Then we obtain

$$2\|v^{k+1} - v^k\| \leq \|v^k - v^{k-1}\| + \frac{\zeta}{\delta} \Delta_{k,k+1}.$$

Summing the inequality above over  $k = \tilde{k} + 1, \dots, m$  yields

$$2 \sum_{k=\tilde{k}+1}^m \|v^{k+1} - v^k\| \leq \sum_{k=\tilde{k}+1}^m \|v^k - v^{k-1}\| + \frac{\zeta}{\delta} \Delta_{\tilde{k}+1, m+1}.$$

Since  $\varphi(\mathcal{L}(w^{m+1}) - \mathcal{L}(w^*)) > 0$ , rearranging terms and letting  $m \rightarrow +\infty$  lead to

$$\sum_{k=\tilde{k}+1}^{+\infty} \|v^{k+1} - v^k\| \leq \|v^{\tilde{k}+1} - v^{\tilde{k}}\| + \frac{\zeta}{\delta} \varphi(\mathcal{L}(w^{\tilde{k}+1}) - \mathcal{L}(w^*)),$$

which implies

$$\sum_{k=0}^{+\infty} \|v^{k+1} - v^k\| < +\infty.$$

Thus, it follows that

$$\sum_{k=0}^{+\infty} \|x^{k+1} - x^k\| < +\infty, \quad \sum_{k=0}^{+\infty} \|y^{k+1} - y^k\| < +\infty.$$

These, together with (47), we obtain

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

Moreover, note that

$$\begin{aligned} \|w^{k+1} - w^k\| &= \left( \|x^{k+1} - \lambda^k\|^2 + \|y^{k+1} - y^k\|^2 + \|\lambda^{k+1} - \lambda^k\|^2 \right)^{\frac{1}{2}} \\ &\leq \|x^{k+1} - \lambda^k\| + \|y^{k+1} - y^k\| + \|\lambda^{k+1} - \lambda^k\|. \end{aligned}$$

Therefore,

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

and  $\{w^k\}_{k \in N}$  is a Cauchy sequence (see P.482 for a simple proof), which shows that the sequence generated by our algorithm converges. The assertion then follows immediately from Lemma 3.4.

We now establish the convergence rate for the ADMM (6). The proof of the following result is similar to that in [8] and hence omitted here.

**Theorem 3.2** *Let  $\{w^k\}_{k \in N}$  be the sequence generated by the ADMM (6) and converges to  $w^* := (x^*, y^*, \lambda^*)$ . Assume that  $\mathcal{L}(\cdot)$  has the KL property at  $(x^*, y^*, \lambda^*)$  with  $\varphi(s) := cs^{1-\theta}$ ,  $\theta \in [0, 1)$ ,  $c > 0$ . Then the following estimations hold:*

- 1) If  $\theta = 0$ , then the sequence  $\{w^k\}_{k \in N}$  converges in a finite number of

steps.

2) If  $\theta \in \left(0, \frac{1}{2}\right]$ , then there exist  $c > 0$  and  $\tau \in [0, 1)$ , such that

$$\|(x^k, y^k, \lambda^k) - (x^*, y^*, \lambda^*)\| \leq c\tau^k.$$

3) If  $\theta \in \left(\frac{1}{2}, 1\right)$ , then there exists  $c > 0$ , such that

$$\|(x^k, y^k, \lambda^k) - (x^*, y^*, \lambda^*)\| \leq ck^{\frac{\theta-1}{2\theta-1}}.$$

## 4. Conclusions

In this paper, we conducted an analysis of the convergence properties of the Bregman Alternating Direction Method of Multipliers (ADMM) for addressing linearly constrained nonconvex minimization problems with coupled objective functions. Our study operated under the assumption that the associated functions satisfy the Kurdyka-Lojasiewicz inequality. We successfully demonstrated that the iterative sequence generated by the Bregman ADMM converges to a critical point of the augmented Lagrangian function, given that the penalty parameter in the augmented Lagrangian surpasses a certain threshold. Additionally, with further conditions on the problem's data, we established the convergence rate of the algorithm. Looking forward, extending Algorithm (6) to handle multi-block non-separable linearly constrained nonconvex optimization problems and incorporating generalized ADMM represent meaningful avenues for future exploration.

## Funding

These authors were supported by the Natural Science Foundation of China (Grant Nos. 11801455), the Sichuan Science and Technology Program (Grant No. 2023NSFSC1922), the Innovation Team Funds of China West Normal University (Grant No. KCXTD2023-3), the Fundamental Research Funds of China West Normal University (Grant No. 23kc010).

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Alizadeh, M., Li, X., Wang, Z., Scaglione, A. and Melton, R. (2012) Demand-Side Management in the Smart Grid: Information Processing for the Power Switch. *IEEE Signal Processing Magazine*, **29**, 55-67. <https://doi.org/10.1109/MSP.2012.2192951>
- [2] Chang, T.H., Alizadeh, M. and Scaglione, A. (2012) Coordinated Home Energy Management for Real-Time Power Balancing. 2012 *IEEE Power and Energy Society General Meeting*, San Diego, 22-26 July 2012, 22-26.
- [3] Li, N., Chen, L. and Low, S.H. (2011) Optimal Demand Response Based on Utility

- Maximization in Power Networks. 2011 *IEEE Power and Energy Society General Meeting*, Detroit, 24-28 July 2011, 1-8. <https://doi.org/10.1109/PES.2011.6039082>
- [4] Paatero, J.V. and Lund, P.D. (2006) A Model for Generating Household Electricity Load Profiles. *International Journal of Energy Research*, **30**, 273-290. <https://doi.org/10.1002/er.1136>
- [5] Scutari, G. and Palomar, D.P. (2010) MIMO Cognitive Radio: A Game Theoretical Approach. *IEEE Transactions on Signal Processing*, **58**, 761-780. <https://doi.org/10.1109/TSP.2009.2032039>
- [6] Zhao, Q. and Sadler, B.M. (2007) A Survey of Dynamic Spectrum Access. *IEEE Signal Processing Magazine*, **24**, 79-89. <https://doi.org/10.1109/MSP.2007.361604>
- [7] Guo, K., Han, D.R. and Wu, T.T. (2018) Convergence Analysis for Optimization Problems with Nonseparable Nonconvex Objective and Linear Constraints. *Pacific Journal of Optimization*, **14**, 489-506.
- [8] Guo, K., Han, D.R. and Wu, T.T. (2017) Convergence of Alternating Method for Minimizing Sum of Two Nonconvex Functions with Linear Constraints. *International Journal of Computer Mathematics*, **94**, 1653-1669. <https://doi.org/10.1080/00207160.2016.1227432>
- [9] Kong, W.W. and Monteiro, R.D.C. (2024) Global Complexity Bound of a Proximal ADMM for Linearly-Constrained Nonseparable Nonconvex Composite Programming. *SIAM Journal on Optimization*, **34**.
- [10] Guo, K., Han, D., Wang, D.Z.W. and Wu, T.T. (2017) Convergence of ADMM for Multi-Block Nonconvex Separable Optimization Models. *Frontiers of Mathematics in China*, **12**, 1139-1162.
- [11] Guo, K. and Wang, X. (2018) Convergence of Generalized Alternating Direction Method of Multipliers for Nonseparable Nonconvex Objective with Linear Constraints. <https://doi.org/10.3770/j.issn:2095-2651.2018.05.010>
- [12] Tan, L. and Guo, K. (2024) Bregman ADMM: A New Algorithm for Nonconvex Optimization with Linear Constraints, Accepted. *Journal of Nonlinear and Variational Analysis*, **8**.
- [13] Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2010) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*, **3**, 1-122. <https://doi.org/10.1561/22000000016>
- [14] Figueiredo, M.A.T. and Bioucas-Dias, J.M. (2010) Restoration of Poissonian Images Using Alternating Direction Optimization. *IEEE Transactions on Image Processing*, **19**, 3133-3145. <https://doi.org/10.1109/TIP.2010.2053941>
- [15] Goldstein, T., Bresson, X. and Osher, S. (2010) Geometric Applications of the Split Bregman Method: Segmentation and Surface Reconstruction. *Journal of Scientific Computing*, **45**, 272-293. <https://doi.org/10.1007/s10915-009-9331-z>
- [16] Fu, Q., Wang, H. and Banerjee, A. (2013) Bethe-ADMM for Tree Decomposition Based Parallel MAP Inference. arXiv: 1309.6829.
- [17] Wang, H. and Banerjee, A. (2013) Online alternating Direction Method. arXiv: 1206.6448.
- [18] Yang, J. and Zhang, Y. (2011) Alternating Direction Algorithms for  $\ell_1$ -Problems in Compressive Sensing. *SIAM Journal on Scientific Computing*, **33**, 250-278. <https://doi.org/10.1137/09077761>
- [19] Gabay, D. (1983) Applications of the Method of Multipliers to Variational Inequalities. *Studies in Mathematics and Its Applications*, **15**, 299-331.

- [https://doi.org/10.1016/S0168-2024\(08\)70034-1](https://doi.org/10.1016/S0168-2024(08)70034-1)
- [20] Kellogg, R.B. (1969) Nonlinear Alternating Direction Algorithm. *Mathematics of Computation*, **23**, 23-38.
- [21] Marchuk, G.I. (1990) Splitting and Alternating Direction Methods. *Handbook of Numerical Analysis*, **1**, 197-462. [https://doi.org/10.1016/S1570-8659\(05\)80035-3](https://doi.org/10.1016/S1570-8659(05)80035-3)
- [22] Beck, A. (2017) First Order Methods In Optimization. SIAM, Philadelphia. <https://doi.org/10.1137/1.9781611974997>
- [23] Bolte, J., Sabach, S., Teboulle, M. and Vaisbourd, T. (2018) First Order Methods beyond Convexity and Lipschitz Gradient Continuity with Applications to Quadratic Inverse Problems. *SIAM Journal on Optimization*, **28**, 2131-2151. <https://doi.org/10.1137/17M1138558>
- [24] Bauschke, H.H., Borwein, J. and Combettes, P.L. (2003) Bregman Monotone Optimization Algorithms. *SIAM Journal on Control and Optimization*, **42**, 596-636. <https://doi.org/10.1137/S0363012902407120>
- [25] Attouch, H., Bolte, J., Redont, P. and Soubeyran, A. (2010) Proximal Alternating Minimization and Projection Methods for Nonconvex Problems: An Approach Based on the Kurdyka Lojasiewicz Inequality. *Mathematics of Operations Research*, **35**, 438-457. <https://doi.org/10.1287/moor.1100.0449>
- [26] Attouch, H., Bolte, J. and Svaiter, B.F. (2013) Convergence of Descent Methods for Semi-Algebraic and Tame Problems: Proximal Algorithms, Forward-Backward Splitting, and Regularized Gauss-Seidel Methods. *Mathematical Programming*, **137**, 91-129. <https://doi.org/10.1007/s10107-011-0484-9>
- [27] Bolte, J., Sabach, S. and Teboulle, M. (2014) Proximal Alternating Linearized Minimization for Nonconvex and Nonsmooth Problem. *Mathematical Programming*, **146**, 459-494. <https://doi.org/10.1007/s10107-013-0701-9>
- [28] Nesterov, Y. (2004) Introduction Lectures on Convex Optimization: A Basic Course. Springer, New York. <https://doi.org/10.1007/978-1-4419-8853-9>