

A Symmetric Bregman Alternating Direction Method of Multipliers for Separable Nonconvex Minimization Problem

Zhangyuan Zeng, Shilian Zhao*

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

Email: 884102450@qq.com, *14622959@qq.com

How to cite this paper: Zeng, Z.Y. and Zhao, S.L. (2025) A Symmetric Bregman Alternating Direction Method of Multipliers for Separable Nonconvex Minimization Problem. *Journal of Applied Mathematics and Physics*, 13, 889-913.

<https://doi.org/10.4236/jamp.2025.133047>

Received: February 23, 2025

Accepted: March 22, 2025

Published: March 25, 2025

Copyright © 2025 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

The alternating direction method of multipliers (ADMM) and its symmetric version are efficient for minimizing two-block separable problems with linear constraints. However, both ADMM and symmetric ADMM have limited versatility across various fields due to the requirement that the gradients of differentiable functions exhibit global Lipschitz continuity, a condition that is typically challenging to satisfy in nonconvex optimization problems. Recently, a novel Bregman ADMM that not only eliminates the need for global Lipschitz continuity of the gradient, but also ensures that Bregman ADMM can be degenerated to the classical ADMM has been proposed for two-block nonconvex optimization problems with linear constraints. Building on this, we propose a symmetric Bregman alternating direction method of multipliers, which can be degenerated into the symmetric ADMM and the Bregman ADMM, and thus further degenerated into the classical ADMM. Moreover, when solving separable nonconvex optimization problems, it does not require the global Lipschitz continuity of the gradients of differentiable functions. Furthermore, we demonstrate that under the Kurdyka-Lojasiewicz inequality and certain conditions, the iterative sequence generated by our algorithm converges to a critical point of the problem. In addition, we examine the convergence rate of the algorithm.

Keywords

Symmetric ADMM, Bregman ADMM, Nonconvex Optimization, Separable, Kurdyka-Lojasiewicz Inequality

1. Introduction

In this paper, we consider the following two-block separable optimization prob-

lem with linear constraint

$$\begin{aligned} \min_{x,y} \quad & f(x) + g(y), \\ \text{s.t.} \quad & Ax + y = b, \end{aligned} \quad (1.1)$$

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, $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Numerous valuable optimization problems can be expressed in the form (1), rendering it widely applicable across diverse fields, such as machine learning [1]-[3], image processing [4]-[6], and signal processing [7]-[10].

When both f and g are convex functions, a prominent approach for addressing problem (1) is the alternating direction method of multipliers (ADMM), proposed by Gabay, Mercier, Glowinski and Marroco [11] [12] in 1970s. This method fully harnesses the separable properties to their utmost potential, thus attracting considerable attention across diverse domains in recent years. The iterative scheme of the ADMM is as follows

$$\begin{cases} x_{k+1} \in \arg \min_x \{ \mathcal{L}_\beta(x, y_k, \lambda_k) \}, \\ y_{k+1} \in \arg \min_y \{ \mathcal{L}_\beta(x_{k+1}, y, \lambda_k) \}, \\ \lambda_{k+1} := \lambda_k - \beta(Ax_{k+1} + y_{k+1} - b). \end{cases} \quad (1.2)$$

Here, $\mathcal{L}_\beta(\cdot)$ denotes the augmented Lagrangian function for (1.1), given by

$$\mathcal{L}_\beta(x, y, \lambda) = f(x) + g(y) - \langle \lambda, Ax + y - b \rangle + \frac{\beta}{2} \|Ax + y - b\|^2,$$

where λ is the Lagrangian multiplier associated with the linear constraint, and $\beta > 0$ is the penalty parameter. The study of ADMM has a lengthy academic lineage, its convergence and convergence rate are well-understood [13]-[16] for convex objectives.

However, for scenarios where there is at least one nonconvex component in the objective function, many studies primarily focus on proving the convergence of the ADMM or its variant and analyzing problem scenarios under additional conditions, such as Li and Pong [17] and Hong *et al.* [18]. Particularly, in 2017, Guo *et al.* [19] demonstrated that under conditions less stringent than those outlined in [17] [18], the convergence and convergence rate of ADMM for nonconvex problems can be established, contingent upon the augmented Lagrangian function satisfying the Kurdyka-Lojasiewicz inequality. Motivated by the insights provided in the aforementioned article, Wu *et al.* [20] considered utilizing symmetric ADMM to address a two-block linearly constrained separable nonconvex optimization, which can revert back to the classical ADMM, and conducted an analysis on its convergence and convergence rate for the case $B = I$ (I is the identity matrix with proper dimension). Note that it can numerically accelerate ADMM with some values of $\alpha > 0$. Its iterative format is as follows

$$\begin{cases} x_{k+1} \in \arg \min_x \{ \mathcal{L}_\beta(x, y_k, \lambda_k) \}, \\ \lambda_{k+\frac{1}{2}} := \lambda_k - \alpha\beta(Ax_{k+1} + By_k - b), \\ y_{k+1} \in \arg \min_y \left\{ \mathcal{L}_\beta \left(x_{k+1}, y, \lambda_{k+\frac{1}{2}} \right) \right\}, \\ \lambda_{k+1} := \lambda_{k+\frac{1}{2}} - \beta(Ax_{k+1} + By_{k+1} - b), \end{cases} \quad (1.3)$$

The only difference from the classical ADMM lies in the addition of a relaxation factor $\alpha \in (-1, 1)$ in the update of the multiplier λ^{k+1} between the iterative formulas of x and y . In particular, the algorithm returns to the classical ADMM when $\alpha = 0$. Therefore, it presents the same limitations in certain areas [21]-[23]. Whether using the ADMM or the symmetric ADMM to tackle two-block separable nonconvex problems with linear constraints, both methods are constrained by the assumption of Lipschitz continuity of differentiable functions.

To relax the Lipschitz continuity constraint on the gradient of the objective function, Tan and Guo [24] introduced a novel version of Bregman ADMM, which is distinguished from that proposed by Wang *et al.* [25] by its ability to revert to the classical ADMM. Its iteration is as follows

$$\begin{cases} x_{k+1} \in \arg \min_x \{ \mathcal{L}_\beta^h(x, y_k, \lambda_k) \}, \\ y_{k+1} \in \arg \min_y \{ \mathcal{L}_\beta^h(x_{k+1}, y, \lambda_k) \}, \\ \lambda_{k+1} := \lambda_k - \beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})), \end{cases} \quad (1.4)$$

where $\mathcal{L}_\beta^h(\cdot)$ denotes the Bregman augmented Lagrangian function for (1.1)

$$\mathcal{L}_\beta^h(x, y, \lambda) = f(x) + g(y) - \langle \lambda, Ax + y - b \rangle + \beta D_h(-y, Ax - b), \quad (1.5)$$

where λ is the Lagrangian multiplier associated with the linear constraint, and $\beta > 0$ is the penalty parameter. And the Bregman distance $D_h(\cdot, \cdot)$ is defined as

$$D_h(x, y) = h(x) - h(y) - \nabla h(y)^T(x - y).$$

When $h(\cdot) = \frac{1}{2}\|\cdot\|^2$, the Bregman ADMM (1.4) reduces to the classical ADMM (1.2).

Drawing on the aforementioned concept, our aim is to alleviate the necessity for Lipschitz continuity of the gradient of differentiable functions in the symmetric ADMM while addressing two-block separable nonconvex problems with linear constraints. In this paper, we propose an iteration for symmetric version of the Bregman ADMM, whose iterative scheme is

$$x_{k+1} \in \arg \min_x \{ \mathcal{L}_\beta^h(x, y_k, \lambda_k) \}, \quad (1.6a)$$

$$\lambda_{k+\frac{1}{2}} := \lambda_k - \alpha\beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_k)), \quad (1.6b)$$

$$y_{k+1} \in \arg \min_y \left\{ \mathcal{L}_\beta^h \left(x_{k+1}, y, \lambda_{k+\frac{1}{2}} \right) \right\}, \quad (1.6c)$$

$$\lambda_{k+1} := \lambda_{k+\frac{1}{2}} - \beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})), \quad (1.6d)$$

The symmetric Bregman ADMM (1.6) can transition to the symmetric ADMM (1.3) by setting $h(\cdot) = \frac{1}{2}\|\cdot\|^2$, to the Bregman ADMM (1.4) by setting $\alpha = 0$, and to the classic ADMM (1.2) by setting $h(\cdot) = \frac{1}{2}\|\cdot\|^2$ and $\alpha = 0$. In essence, the ADMM framework investigated by Wu *et al.* [20] and Tan and Guo [24] is a particular instance of the approach we introduce. Under the same assumption as those in Tan and Guo [24], we can prove the convergence of the symmetric Bregman ADMM (1.6), providing that the associated function satisfies the Kurdyka-Lojasiewicz inequality. Moreover, we demonstrate that the iterative sequence produced by the symmetric Bregman ADMM (1.6) converges to a critical point of the problem (1.1), and we also analyze the convergence rate of the algorithm.

The remainder of this paper is organized as follows. In Section 2, we provide some necessary preliminaries for our subsequent analysis. In Section 3, we present the convergence analysis of the symmetric Bregman ADMM (1.6) and analyze its convergence rate. Finally, in Section 4, we summarize our findings and draw conclusions.

2. Preliminaries

In this section, we recall some definitions and basic results that will be used for further analysis.

Definition 2.1. [26] 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. [26] A function $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is called proper if there exists at least one $x \in \mathbb{R}^n$ such that $f(x) < \infty$.

Definition 2.3. [26] 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, $f(\cdot)$ is called lower semicontinuous if it is lower semicontinuous at each point in \mathbb{R}^n .

Definition 2.4. ([27], 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

(i) $h(\cdot)$ is proper, lower semicontinuous, and convex, with $\text{dom}(h) \subset \bar{C}$ and $\text{dom}(\partial h) = C$.

(ii) $h(\cdot)$ is C^1 on $\text{int}(\text{dom}(h)) \equiv C$.

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

Definition 2.5 ([27], L-smooth adaptable) Let $h \in \mathcal{G}(C)$, $g: \mathbb{R}^m \rightarrow \mathbb{R}$ continuously differentiable on $C = \text{int}(\text{dom}(h))$. 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. Definition 2.5 serves as a natural extension and complement to the definition of “A Lipschitz-like/Convexity Condition” as presented in reference [21]. This extension enables the derivation of the following two-sided descent lemma.

Lemma 2.1. ([27], extended descent lemma) The pair of functions (g, h) is L -smooth adaptable on C if and only if

$$|g(x) - g(y) - \langle \nabla g(y), x - y \rangle| \leq LD_h(x, y) \quad \forall x, y \in \text{int}(\text{dom}(h)). \quad (2.1)$$

Remark 2.2. In particular, when the set $C = \mathbb{R}^m$ and $h(\cdot) = \frac{1}{2} \|\cdot\|^2$, (2.1) reduces to the classical descent lemma for function g , 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. [27] Let $g: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. The gradient of $g(\cdot)$ is D -Lipschitz if there exists $L > 0$ satisfying

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

Remark 2.3. According to Cauchy-Schwarz inequality, we have

$$|\langle \nabla g(x) - \nabla g(y), x - y \rangle| \leq \|\nabla g(x) - \nabla g(y)\| \|x - y\|,$$

which combines with definition ??, we obtain that

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

Using the conclusion in Lemma 2.4, the above inequality is equivalent to

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

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

Then,

$$\begin{aligned} \langle (L\nabla h(x) - \nabla g(x)) - (L\nabla h(y) - \nabla g(y)), x - y \rangle &\geq 0, \\ \langle (L\nabla h(x) + \nabla g(x)) - (L\nabla h(y) + \nabla g(y)), x - y \rangle &\geq 0. \end{aligned} \quad (2.2)$$

Based on inequality (2.2), it can be concluded that the functions $Lh + g$ and $Lh - g$ exhibit convexity due to the monotonicity of their gradients on the set C . Hence, ensuring the gradient of function $g(\cdot)$ satisfies the D -Lipschitz continuity condition is adequate for establishing the (g, h) function pair as L -smooth adaptable. Considering the intricacy of the ADMM iterative procedure in this study, we find it necessary to presuppose the D -Lipschitz continuity of function $g(\cdot)$.

Remark 2.4. Certainly, the D -Lipschitz continuity characteristic is essentially a Lipschitz-like gradient property in the context of the Bregman distance for the function $g(\cdot)$, and it becomes equivalent to the gradient Lipschitz continuity of

$$g(\cdot) \text{ when } h(\cdot) = \frac{1}{2} \|\cdot\|^2.$$

Definition 2.7. [27] Let $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function.

(i) The Fréchet subdifferential, or regular subdifferential, of $f(\cdot)$ 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 $\hat{\partial}f(x) = \emptyset$.

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

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

Remark 2.5. From above definition, we note that

(i) It 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.

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

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

$$0 \in \partial f(x). \quad (2.3)$$

(iv) If $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous function and $g: \mathbb{R}^n \rightarrow \mathbb{R}$ is continuous differentiable function, then

$$\partial(f+g)(x) = \partial f(x) + \nabla g(x) \text{ for any } x \in \text{dom}(f).$$

A point that meets the condition of Equation (2.3) is referred to as a critical or a stationary point. The critical points set of f is denoted by $\text{crit } f$.

Next, we recall an important property of subdifferential calculus.

Lemma 2.2. [28] Suppose that $F(x, y) = f(x) + g(y)$, where $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ and $g: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$ are proper lower 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 2.8. ([28], 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 $f(\cdot)$ 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

- (i) $\varphi(0) = 0$;
- (ii) φ is C^1 on $(0, \eta)$ and continuous at 0;
- (iii) $\varphi'(x) > 0$, $\forall x \in (0, \eta)$;

(iv) for all x in $U \cap [f(x^*) < f < f(x^*) + \eta]$, the Kurdyka-Lojasiewicz inequality holds

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

where $d(x, \partial f(x)) = \inf_{y \in \partial f(x)} \|y - x\|$, is the distance from x to $\partial f(x)$.

Remark 2.6. Denote Φ_η be the set of all continuous functions $\varphi(\cdot)$ which satisfy (i)-(iii).

Definition 2.9. ([29], Kurdyka-Lojasiewicz function) If function $f(\cdot)$ satisfies the KL property at each point of $\text{dom}(\partial f)$, then $f(\cdot)$ is called a KL function.

Lemma 2.3. ([30], Uniformized KL property) Let Ω be a compact set and $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. Assume that $f(\cdot)$ is constant on Ω and satisfies the KL property at each point of Ω . 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

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

Lemma 2.4. [31] Let $h: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$. For any $x, y \in \text{int}(\text{dom}(h))$ and $z \in \text{dom}(h)$, then

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

(ii) Three points identity holds:

$$D_h(z, x) - D_h(z, y) - D_h(y, x) = \langle \nabla h(x) - \nabla h(y), y - z \rangle. \quad (2.4)$$

Definition 2.10. [32] Let $h \in \mathcal{G}(C)$. The Bregman distance $D_h: \text{dom}(h) \times \text{int}(\text{dom}(h)) \rightarrow \mathbb{R}^+$ is defined by

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

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

Definition 2.11. We say that (x^*, y^*, λ^*) is a critical point of the Augmented Lagrangian Function $\mathcal{L}_\beta^h(\cdot)$ (1.5) with Bregman distance if it satisfies

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

Lemma 2.5. Let $\{w_k := (x_k, y_k, \lambda_k)\}$ be the sequence generated by the symmetric Bregman ADMM (1.6). Then, we have

$$\begin{cases} \nabla h(Ax_{k+1} - b) - \nabla h(-y_k) = \frac{1}{(\alpha+1)\beta}(\lambda_k - \lambda_{k+1}) + \frac{1}{\alpha+1}(\nabla h(-y_{k+1}) - \nabla h(-y_k)), \\ \nabla g(y_{k+1}) = \lambda_{k+1}, \\ \nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1}) = \frac{1}{(\alpha+1)\beta}(\lambda_k - \lambda_{k+1}) - \frac{\alpha}{\alpha+1}(\nabla h(-y_{k+1}) - \nabla h(-y_k)). \end{cases}$$

Proof. Combining (1.6b) and (1.6d), we get

$$\begin{aligned}\lambda_{k+1} - \lambda_k &= -\alpha\beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_k)) - \beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})) \\ &= -\alpha\beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_k)) - \beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_k)) \\ &\quad + \beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_k)) - \beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})) \\ &= -(\alpha + 1)\beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_k)) + \beta(\nabla h(-y_{k+1}) - \nabla h(-y_k))\end{aligned}$$

and thus

$$\nabla h(Ax_{k+1} - b) - \nabla h(-y_k) = \frac{1}{(\alpha + 1)\beta}(\lambda_k - \lambda_{k+1}) + \frac{1}{\alpha + 1}(\nabla h(-y_{k+1}) - \nabla h(-y_k)).$$

Similarly, we combine (1.6b) and (1.6d) again

$$\begin{aligned}\lambda_{k+1} - \lambda_k &= -\alpha\beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_k)) - \beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})) \\ &= -\alpha\beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})) - \beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})) \\ &\quad - \alpha\beta(\nabla h(-y_{k+1}) - \nabla h(-y_k)) \\ &= -(\alpha + 1)\beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})) - \alpha\beta(\nabla h(-y_{k+1}) - \nabla h(-y_k)),\end{aligned}$$

thus, we obtain

$$\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1}) = \frac{1}{(\alpha + 1)\beta}(\lambda_k - \lambda_{k+1}) - \frac{\alpha}{\alpha + 1}(\nabla h(-y_{k+1}) - \nabla h(-y_k)).$$

From the optimality condition of (1.6c), we have

$$0 = \nabla g(y_{k+1}) - \lambda_{k+\frac{1}{2}} + \beta(\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})).$$

Substituting (1.6d) into the above equation yields

$$\nabla g(y_{k+1}) = \lambda_{k+1}.$$

This completes the proof.

3. Convergence Analysis

In this section, we analyze the convergence of the symmetric Bregman ADMM (1.6) and show that the sequence $\{w_k := (x_k, y_k, \lambda_k)\}$ generated by the symmetric Bregman ADMM (1.6) converges to a critical point $\{w^* := (x^*, y^*, \lambda^*)\}$ of $\mathcal{L}_h(\cdot)$ under the following assumptions.

Assumption A. Let $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function, $g: \mathbb{R}^m \rightarrow \mathbb{R}$ be a continuously differentiable function with ∇g being D -Lipschitz continuous. Additionally, let $h \in \mathcal{G}(C)$ be a twice differentiable function on $C = \text{int}(\text{dom}(h))$, which is 1-strong-convex, and whose ∇h is Lipschitz continuous with L_h on any bounded subset of \mathbb{R}^m . We assume that

$$(i) \quad \alpha \in \left[-\frac{1}{2L_h^2 + 1}, 0\right], \quad \beta > \frac{2LL_h^2 - 2\alpha LL_h^3}{1 + \alpha + 2\alpha L_h^2}, \text{ which implies}$$

$$\delta = \left(\frac{\beta}{L_h} - L + \frac{2\alpha LL_h + 2\alpha\beta L_h + 2\alpha LL_h^2}{\alpha + 1} - \frac{2L^2 L_h^2}{(\alpha + 1)\beta}\right) > 0.$$

$$(ii) \quad \alpha \in \left[0, \frac{1}{2L_h^2 - 1}\right), \quad \beta > \frac{2(1+\alpha)LL_h^2 + 2\alpha LL_h^3}{1 + \alpha - 2\alpha L_h^2}, \text{ which implies}$$

$$\delta = \left(\frac{\beta}{L_h} - L - \frac{2\alpha LL_h + 2\alpha\beta L_h + 2\alpha LL_h^2}{\alpha + 1} - \frac{2L^2 L_h^2}{(\alpha + 1)\beta}\right) > 0.$$

$$(iii) \quad A^T A \succeq MI \quad \text{for some } M > 0.$$

Next, we examine the optimality conditions of (1.6). Invoking the optimality conditions, we have that

$$\begin{cases} 0 \in \partial f(x_{k+1}) - A^T \lambda_k - \beta A^T (\nabla^2 h(Ax_{k+1} - b))^T (-Ax_{k+1} - y_k + b), \\ \lambda_{k+\frac{1}{2}} = \lambda_k - \alpha\beta (\nabla h(Ax_{k+1} - b) - \nabla h(-y_k)), \\ 0 = \nabla g(y_{k+1}) - \lambda_{k+\frac{1}{2}} + \beta (\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})), \\ \lambda_{k+1} = \lambda_{k+\frac{1}{2}} - \beta (\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})). \end{cases} \quad (3.1)$$

Lemma 3.1. Let $\{w_k = (x_k, y_k, \lambda_k)\}$ be the sequence generated by the symmetric Bregman ADMM (1.6), which is assumed to be bounded. Then we have

$$\mathcal{L}_\beta^h(w_{k+1}) \leq \mathcal{L}_\beta^h(w_k) - \delta \cdot D_h(y_k, y_{k+1}). \quad (3.2)$$

Proof. From the definition of \mathcal{L}_β^h in (1.5), it follows that

$$\begin{aligned} & \mathcal{L}_\beta^h\left(x_{k+1}, y_k, \lambda_{k+\frac{1}{2}}\right) - \mathcal{L}_\beta^h\left(x_{k+1}, y_{k+1}, \lambda_{k+\frac{1}{2}}\right) \\ &= g(y_k) - \left\langle \lambda_{k+\frac{1}{2}}, Ax_{k+1} + y_k - b \right\rangle + \beta D_h(-y_k, Ax_{k+1} - b) \\ & \quad - g(y_{k+1}) + \left\langle \lambda_{k+\frac{1}{2}}, Ax_{k+1} + y_{k+1} - b \right\rangle - \beta D_h(-y_{k+1}, Ax_{k+1} - b) \\ &= g(y_k) - g(y_{k+1}) + \left\langle \lambda_{k+\frac{1}{2}}, y_{k+1} - y_k \right\rangle + \beta D_h(-y_k, Ax_{k+1} - b) \\ & \quad - \beta D_h(-y_{k+1}, Ax_{k+1} - b). \end{aligned} \quad (3.3)$$

Since the gradient of the function g is D -Lipschitz continuous on $\text{int}(\text{dom}(h))$, it follows that the function pair (g, h) is L -smooth adaptable. Then, according to Lemma 2.1, we can obtain

$$g(y_k) - g(y_{k+1}) \geq \langle \nabla g(y_{k+1}), y_k - y_{k+1} \rangle - LD_h(y_k, y_{k+1}). \quad (3.4)$$

By substituting inequality (3.4) into relation (3.3), we obtain

$$\begin{aligned} & \mathcal{L}_\beta^h\left(x_{k+1}, y_k, \lambda_{k+\frac{1}{2}}\right) - \mathcal{L}_\beta^h\left(x_{k+1}, y_{k+1}, \lambda_{k+\frac{1}{2}}\right) \\ & \geq \langle \nabla g(y_{k+1}), y_k - y_{k+1} \rangle - LD_h(y_k, y_{k+1}) + \left\langle \lambda_{k+\frac{1}{2}}, y_{k+1} - y_k \right\rangle \\ & \quad + \beta D_h(-y_k, Ax_{k+1} - b) - \beta D_h(-y_{k+1}, Ax_{k+1} - b) \\ &= \left\langle \lambda_{k+1} - \lambda_{k+\frac{1}{2}}, y_k - y_{k+1} \right\rangle - LD_h(y_k, y_{k+1}) \end{aligned}$$

$$\begin{aligned}
& +\beta D_h(-y_k, Ax_{k+1}-b)-\beta D_h(-y_{k+1}, Ax_{k+1}-b) \\
& =\left\langle \lambda_{k+1}-\lambda_{k+\frac{1}{2}}, y_k-y_{k+1}\right\rangle-L D_h\left(y_k, y_{k+1}\right)+\beta D_h\left(-y_k,-y_{k+1}\right) \\
& \quad +\beta\left(D_h\left(-y_k, A x_{k+1}-b\right)-D_h\left(-y_k,-y_{k+1}\right)-D_h\left(-y_{k+1}, A x_{k+1}-b\right)\right) \\
& =\left\langle \lambda_{k+1}-\lambda_{k+\frac{1}{2}}, y_k-y_{k+1}\right\rangle-L D_h\left(y_k, y_{k+1}\right)+\beta D_h\left(-y_k,-y_{k+1}\right) \\
& \quad +\beta\left\langle \nabla h\left(A x_{k+1}-b\right)-\nabla h\left(-y_{k+1}\right), y_k-y_{k+1}\right\rangle \\
& =\left\langle \lambda_{k+1}-\lambda_{k+\frac{1}{2}}, y_k-y_{k+1}\right\rangle-L D_h\left(y_k, y_{k+1}\right)+\beta D_h\left(-y_k,-y_{k+1}\right) \\
& \quad +\left\langle \lambda_{k+\frac{1}{2}}-\lambda_{k+1}, y_k-y_{k+1}\right\rangle \\
& =\beta D_h\left(-y_k,-y_{k+1}\right)-L D_h\left(y_k, y_{k+1}\right),
\end{aligned}$$

where the first and the third equalities follow from Lemma (2.5) and the three points identity (2.4) respectively.

Based on the 1-strong convexity and L_h -smoothness of the function $h(\cdot)$, we can obtain the following two inequalities respectively

$$\begin{aligned}
D_h\left(y_k, y_{k+1}\right) & =h\left(y_k\right)-h\left(y_{k+1}\right)-\left\langle \nabla h\left(y_{k+1}\right), y_k-y_{k+1}\right\rangle \leq \frac{L_h}{2}\left\| y_k-y_{k+1}\right\|^2, \\
D_h\left(-y_k,-y_{k+1}\right) & =h\left(-y_k\right)-h\left(-y_{k+1}\right)-\left\langle \nabla h\left(-y_{k+1}\right), y_{k+1}-y_k\right\rangle \geq \frac{1}{2}\left\| y_k-y_{k+1}\right\|^2.
\end{aligned}$$

Combining the above two inequalities, we get

$$D_h\left(-y_k,-y_{k+1}\right) \geq \frac{1}{2}\left\| y_{k+1}-y_k\right\|^2 \geq \frac{1}{L_h} \cdot D_h\left(y_k, y_{k+1}\right),$$

thus, we can obtain

$$\mathcal{L}_{\beta}^h\left(x_{k+1}, y_k, \lambda_{k+\frac{1}{2}}\right)-\mathcal{L}_{\beta}^h\left(x_{k+1}, y_{k+1}, \lambda_{k+\frac{1}{2}}\right) \geq\left(\frac{\beta}{L_h}-L\right) D_h\left(y_k, y_{k+1}\right). \quad (3.5)$$

Next, by using (1.6b) and (1.5), we have

$$\begin{aligned}
& \mathcal{L}_{\beta}^h\left(x_{k+1}, y_k, \lambda_k\right)-\mathcal{L}_{\beta}^h\left(x_{k+1}, y_k, \lambda_{k+\frac{1}{2}}\right)+\mathcal{L}_{\beta}^h\left(x_{k+1}, y_{k+1}, \lambda_{k+\frac{1}{2}}\right)-\mathcal{L}_{\beta}^h\left(x_{k+1}, y_{k+1}, \lambda_{k+1}\right) \\
& =\left\langle \lambda_{k+\frac{1}{2}}-\lambda_k, A x_{k+1}+y_k-b\right\rangle+\left\langle \lambda_{k+1}-\lambda_{k+\frac{1}{2}}, A x_{k+1}+y_{k+1}-b\right\rangle \\
& =\left\langle -\alpha \beta\left(\nabla h\left(A x_{k+1}-b\right)-\nabla h\left(-y_k\right)\right), A x_{k+1}+y_k-b\right\rangle \\
& \quad +\left\langle \lambda_{k+1}-\lambda_k+\alpha \beta\left(\nabla h\left(A x_{k+1}-b\right)-\nabla h\left(-y_k\right)\right), A x_{k+1}+y_{k+1}-b\right\rangle \\
& =\left\langle \alpha \beta\left(\nabla h\left(A x_{k+1}-b\right)-\nabla h\left(-y_k\right)\right), y_{k+1}-y_k\right\rangle+\left\langle \lambda_{k+1}-\lambda_k, A x_{k+1}+y_{k+1}-b\right\rangle,
\end{aligned}$$

where the second equality follows from (1.6b). Then, according to Lemma 2.5

$$\nabla h\left(A x_{k+1}-b\right)-\nabla h\left(-y_{k+1}\right)=\frac{1}{(\alpha+1) \beta}\left(\lambda_k-\lambda_{k+1}\right)-\frac{\alpha}{\alpha+1}\left(\nabla h\left(-y_{k+1}\right)-\nabla h\left(-y_k\right)\right),$$

$$\nabla h(Ax_{k+1} - b) - \nabla h(-y_k) = \frac{1}{(\alpha+1)\beta}(\lambda_k - \lambda_{k+1}) + \frac{1}{\alpha+1}(\nabla h(-y_{k+1}) - \nabla h(-y_k)),$$

and, since $h(\cdot)$ is 1-strong-convex, we have

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

Now, we discuss the two cases based on the range of α .

(i) $\alpha \in \left[-\frac{1}{2L_h^2+1}, 0\right]$, combining the above formulas

$$\begin{aligned} & \mathcal{L}_\beta^h(x_{k+1}, y_k, \lambda_k) - \mathcal{L}_\beta^h\left(x_{k+1}, y_k, \lambda_{k+\frac{1}{2}}\right) + \mathcal{L}_\beta^h\left(x_{k+1}, y_{k+1}, \lambda_{k+\frac{1}{2}}\right) - \mathcal{L}_\beta^h(x_{k+1}, y_{k+1}, \lambda_{k+1}) \\ & \geq \left\langle \frac{\alpha\beta}{(\alpha+1)\beta}(\lambda_k - \lambda_{k+1}) + \frac{\alpha\beta}{\alpha+1}(\nabla h(-y_{k+1}) - \nabla h(-y_k)), y_{k+1} - y_k \right\rangle \\ & \quad - \|\lambda_{k+1} - \lambda_k\| \|Ax_{k+1} + y_{k+1} - b\| \\ & \geq \left\langle \frac{\alpha}{\alpha+1}(\lambda_k - \lambda_{k+1}) + \frac{\alpha\beta}{\alpha+1}(\nabla h(-y_{k+1}) - \nabla h(-y_k)), y_{k+1} - y_k \right\rangle \\ & \quad - \|\lambda_{k+1} - \lambda_k\| \|\nabla h(-y_{k+1}) - \nabla h(Ax_{k+1} - b)\| \\ & \geq \left\langle \frac{\alpha}{\alpha+1}(\lambda_k - \lambda_{k+1}) + \frac{\alpha\beta}{\alpha+1}(\nabla h(-y_{k+1}) - \nabla h(-y_k)), y_{k+1} - y_k \right\rangle \\ & \quad - \|\lambda_{k+1} - \lambda_k\| \left[\frac{1}{(\alpha+1)\beta} \|\lambda_{k+1} - \lambda_k\| - \frac{\alpha}{\alpha+1} \|\nabla h(-y_{k+1}) - \nabla h(-y_k)\| \right] \\ & \geq \frac{\alpha}{\alpha+1} \|\lambda_k - \lambda_{k+1}\| \|y_{k+1} - y_k\| + \frac{\alpha\beta}{\alpha+1} \|\nabla h(-y_{k+1}) - \nabla h(-y_k)\| \|y_{k+1} - y_k\| \\ & \quad - \frac{1}{(\alpha+1)\beta} \|\lambda_{k+1} - \lambda_k\|^2 + \frac{\alpha}{\alpha+1} \|\lambda_{k+1} - \lambda_k\| \|\nabla h(-y_{k+1}) - \nabla h(-y_k)\|. \end{aligned} \quad (3.6)$$

Subsequently, we claim that

$$\|\nabla g(y_{k+1}) - \nabla g(y_k)\| \leq LL_h \cdot \|y_{k+1} - y_k\|. \quad (3.7)$$

To prove (3.7), we consider two cases. When $y_k = y_{k+1}$, (3.7) holds obviously. Next, we assume $y_{k+1} \neq y_k$. Since $\nabla g(\cdot)$ is D -Lipschitz, we obtain

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

where the second inequality is a consequence of ∇h is Lipschitz continuous with L_h ($L_h \geq 1$) on any bounded subset of \mathbb{R}^m , that is

$$\|\nabla h(y_{k+1}) - \nabla h(y_k)\| \leq L_h \|y_{k+1} - y_k\|. \quad (3.8)$$

Since $\lambda_{k+1} = \nabla g(y_{k+1})$, (3.7) becomes

$$\|\lambda_{k+1} - \lambda_k\| \leq LL_h \|y_{k+1} - y_k\|. \quad (3.9)$$

Moreover, according to $h(\cdot)$ is 1-strong-convex, we get $D_h(y_k, y_{k+1}) \geq \frac{1}{2} \|y_k - y_{k+1}\|^2$, i.e.,

$$\|y_k - y_{k+1}\|^2 \leq 2D_h(y_k, y_{k+1}). \quad (3.10)$$

Substituting (3.8), (3.9) and (3.10) into (3.6), we conclude that

$$\begin{aligned} & \mathcal{L}_\beta^h(x_{k+1}, y_k, \lambda_k) - \mathcal{L}_\beta^h\left(x_{k+1}, y_k, \lambda_{k+\frac{1}{2}}\right) + \mathcal{L}_\beta^h\left(x_{k+1}, y_{k+1}, \lambda_{k+\frac{1}{2}}\right) - \mathcal{L}_\beta^h(x_{k+1}, y_{k+1}, \lambda_{k+1}) \\ & \geq \frac{2\alpha LL_h}{\alpha+1} D_h(y_k, y_{k+1}) + \frac{2\alpha\beta L_h}{\alpha+1} D_h(y_k, y_{k+1}) \\ & \quad - \frac{2L^2 L_h^2}{(\alpha+1)\beta} D_h(y_k, y_{k+1}) + \frac{2\alpha LL_h^2}{\alpha+1} D_h(y_k, y_{k+1}) \\ & = \left(\frac{2\alpha LL_h + 2\alpha\beta L_h + 2\alpha LL_h^2}{\alpha+1} - \frac{2L^2 L_h^2}{(\alpha+1)\beta} \right) D_h(y_k, y_{k+1}). \end{aligned} \quad (3.11)$$

Then, combining (3.5) and (3.11), we obtain that

$$\begin{aligned} & \mathcal{L}_\beta^h(x_{k+1}, y_k, \lambda_k) - \mathcal{L}_\beta^h(x_{k+1}, y_{k+1}, \lambda_{k+1}) \\ & \geq \left(\frac{\beta}{L_h} - L + \frac{2\alpha LL_h + 2\alpha\beta L_h + 2\alpha LL_h^2}{\alpha+1} - \frac{2L^2 L_h^2}{(\alpha+1)\beta} \right) D_h(y_k, y_{k+1}). \end{aligned} \quad (3.12)$$

Consequently, according to (1.6a), we have

$$\mathcal{L}_\beta^h(x_k, y_k, \lambda_k) - \mathcal{L}_\beta^h(x_{k+1}, y_k, \lambda_k) \geq 0. \quad (3.13)$$

Finally, summing the inequality (3.12) and (3.13), we conclude that

$$\begin{aligned} & \mathcal{L}_\beta^h(w^k) - \mathcal{L}_\beta^h(w^{k+1}) \\ & \geq \left(\frac{\beta}{L_h} - L + \frac{2\alpha LL_h + 2\alpha\beta L_h + 2\alpha LL_h^2}{\alpha+1} - \frac{2L^2 L_h^2}{(\alpha+1)\beta} \right) D_h(y_k, y_{k+1}). \end{aligned}$$

This completes the discussion for case (i), we now proceed to discuss case (ii).

(ii) $\alpha \in \left[0, \frac{1}{2L_h^2 - 1}\right]$, combining the same formulas

$$\begin{aligned} & \mathcal{L}_\beta^h(x_{k+1}, y_k, \lambda_k) - \mathcal{L}_\beta^h\left(x_{k+1}, y_k, \lambda_{k+\frac{1}{2}}\right) + \mathcal{L}_\beta^h\left(x_{k+1}, y_{k+1}, \lambda_{k+\frac{1}{2}}\right) - \mathcal{L}_\beta^h(x_{k+1}, y_{k+1}, \lambda_{k+1}) \\ & \geq \left\langle \frac{\alpha\beta}{(\alpha+1)\beta} (\lambda_k - \lambda_{k+1}) + \frac{\alpha\beta}{\alpha+1} (\nabla h(-y_{k+1}) - \nabla h(-y_k)), y_{k+1} - y_k \right\rangle \\ & \quad - \|\lambda_{k+1} - \lambda_k\| \|Ax_{k+1} + y_{k+1} - b\| \\ & \geq \left\langle \frac{\alpha}{\alpha+1} (\lambda_k - \lambda_{k+1}) + \frac{\alpha\beta}{\alpha+1} (\nabla h(-y_{k+1}) - \nabla h(-y_k)), y_{k+1} - y_k \right\rangle \end{aligned}$$

$$\begin{aligned}
& -\|\lambda_{k+1} - \lambda_k\| \|\nabla h(-y_{k+1}) - \nabla h(Ax_{k+1} - b)\| \\
& \geq \left\langle \frac{\alpha}{\alpha+1}(\lambda_k - \lambda_{k+1}) + \frac{\alpha\beta}{\alpha+1}(\nabla h(-y_{k+1}) - \nabla h(-y_k)), y_{k+1} - y_k \right\rangle \\
& \quad - \|\lambda_{k+1} - \lambda_k\| \left[\frac{1}{(\alpha+1)\beta} \|\lambda_{k+1} - \lambda_k\| + \frac{\alpha}{\alpha+1} \|\nabla h(-y_{k+1}) - \nabla h(-y_k)\| \right] \quad (3.14) \\
& \geq -\frac{\alpha}{\alpha+1} \|\lambda_k - \lambda_{k+1}\| \|y_{k+1} - y_k\| - \frac{\alpha\beta}{\alpha+1} \|\nabla h(-y_{k+1}) - \nabla h(-y_k)\| \|y_{k+1} - y_k\| \\
& \quad - \frac{1}{(\alpha+1)\beta} \|\lambda_{k+1} - \lambda_k\|^2 - \frac{\alpha}{\alpha+1} \|\lambda_{k+1} - \lambda_k\| \|\nabla h(-y_{k+1}) - \nabla h(-y_k)\|.
\end{aligned}$$

Similarly, substituting (3.8), (3.9) and (3.10) into (3.14), we conclude that

$$\begin{aligned}
& \mathcal{L}_\beta^h(x_{k+1}, y_k, \lambda_k) - \mathcal{L}_\beta^h\left(x_{k+1}, y_k, \lambda_{k+\frac{1}{2}}\right) + \mathcal{L}_\beta^h\left(x_{k+1}, y_{k+1}, \lambda_{k+\frac{1}{2}}\right) - \mathcal{L}_\beta^h(x_{k+1}, y_{k+1}, \lambda_{k+1}) \\
& \geq -\frac{2\alpha LL_h}{\alpha+1} D_h(y_k, y_{k+1}) - \frac{2\alpha\beta L_h}{\alpha+1} D_h(y_k, y_{k+1}) \\
& \quad - \frac{2L^2 L_h^2}{(\alpha+1)\beta} D_h(y_k, y_{k+1}) - \frac{2\alpha LL_h^2}{\alpha+1} D_h(y_k, y_{k+1}) \\
& = \left(-\frac{2\alpha LL_h + 2\alpha\beta L_h + 2\alpha LL_h^2}{\alpha+1} - \frac{2L^2 L_h^2}{(\alpha+1)\beta} \right) D_h(y_k, y_{k+1}). \quad (3.15)
\end{aligned}$$

Then, combining (3.5) and (3.15), we obtain that

$$\begin{aligned}
& \mathcal{L}_\beta^h(x_{k+1}, y_k, \lambda_k) - \mathcal{L}_\beta^h(x_{k+1}, y_{k+1}, \lambda_{k+1}) \\
& \geq \left(\frac{\beta}{L_h} - L - \frac{2\alpha LL_h + 2\alpha\beta L_h + 2\alpha LL_h^2}{\alpha+1} - \frac{2L^2 L_h^2}{(\alpha+1)\beta} \right) D_h(y_k, y_{k+1}). \quad (3.16)
\end{aligned}$$

Consequently, according to (1.6a), we have

$$\mathcal{L}_\beta^h(x_k, y_k, \lambda_k) - \mathcal{L}_\beta^h(x_{k+1}, y_{k+1}, \lambda_{k+1}) \geq 0. \quad (3.17)$$

Finally, summing the inequality (3.16) and (3.17), we conclude that

$$\begin{aligned}
& \mathcal{L}_\beta^h(w^k) - \mathcal{L}_\beta^h(w^{k+1}) \\
& \geq \left(\frac{\beta}{L_h} - L - \frac{2\alpha LL_h + 2\alpha\beta L_h + 2\alpha LL_h^2}{\alpha+1} - \frac{2L^2 L_h^2}{(\alpha+1)\beta} \right) D_h(y_k, y_{k+1}).
\end{aligned}$$

This completes the discussion for case (ii), and with this, the proof is complete. \square

Remark 3.1. Since $\delta > 0$, Lemma 3.1 implies that $\{\mathcal{L}_\beta^h(w_k)\}$ is monotonically nonincreasing. Note that when $\alpha = 0$, we have $\beta > 2LL_h^2$. This corresponds to the requirement in Tan and Guo [24]. Furthermore, when

$h(\cdot) = \frac{1}{2} \|\cdot\|^2$, we have $\alpha = 0$ and $L_h = 1$, thus $\beta > 2L$. This corresponds to the requirement in Guo et al. [19].

Lemma 3.2. Let $\{w_k = (x_k, y_k, \lambda_k)\}$ be the sequence generated by the symmetric Bregman ADMM (1.6), which is assumed to be bounded. Then we have

$$\sum_{k=0}^{+\infty} \|w_{k+1} - w_k\|^2 < +\infty.$$

Proof: Considering that $\{w_k\}$ is bounded, it has at least one limit point. Let w^* be a limit point of $\{w_k\}$ and let $\{w_{k_j}\}$ be the subsequence converging to it, i.e. $w_{k_j} \rightarrow w^*$. Given that $\mathcal{L}_\beta^h(\cdot)$ is a lower semicontinuous function, we can deduce that

$$\mathcal{L}_\beta^h(w^*) \leq \liminf_{j \rightarrow +\infty} \mathcal{L}_\beta^h(w_{k_j}).$$

Consequently, $\{\mathcal{L}_\beta^h(w_{k_j})\}$ is bounded from below. Besides, the fact that $\{\mathcal{L}_\beta^h(w_k)\}$ is nonincreasing implies that $\{\mathcal{L}_\beta^h(w_{k_j})\}$ is convergent. Moreover, $\{\mathcal{L}_\beta^h(w_k)\}$ is convergent, and $\mathcal{L}_\beta^h(w_k) \geq \mathcal{L}_\beta^h(w^*)$. Based on the Equation (3.2), we can obtain

$$\delta \cdot D_h(y_k, y_{k+1}) \leq \mathcal{L}_\beta^h(w_k) - \mathcal{L}_\beta^h(w_{k+1}).$$

Summing up the above inequality from $k=0$ to n , we conclude that

$$\sum_{k=0}^n \delta D_h(y_k, y_{k+1}) \leq \mathcal{L}_\beta^h(w_0) - \mathcal{L}_\beta^h(w_{n+1}) \leq \mathcal{L}_\beta^h(w_0) - \mathcal{L}_\beta^h(w^*) < +\infty.$$

Owing to $\delta > 0$, we have $\sum_{k=0}^{\infty} D_h(y_k, y_{k+1}) < +\infty$, which implies $\sum_{k=0}^{\infty} \|y_{k+1} - y_k\|^2 < +\infty$. Consequently, according to (3.9), we get that

$$\sum_{k=0}^{\infty} \|\lambda_{k+1} - \lambda_k\|^2 < +\infty.$$

Recall Lemma 2.5, we have

$$\begin{aligned} \lambda_{k+1} &= \lambda_k - \beta \left[(\alpha+1) (\nabla h(Ax_{k+1} - b) - \nabla h(-y_k)) - (\nabla h(-y_{k+1}) - \nabla h(-y_k)) \right], \\ \lambda_k &= \lambda_{k-1} - \beta \left[(\alpha+1) (\nabla h(Ax_k - b) - \nabla h(-y_{k-1})) - (\nabla h(-y_k) - \nabla h(-y_{k-1})) \right]. \end{aligned}$$

Combining the two equalities, we obtain

$$\begin{aligned} & (\alpha+1)\beta (\nabla h(Ax_{k+1} - b) - \nabla h(Ax_k - b)) \\ &= (\lambda_k - \lambda_{k-1}) - (\lambda_{k+1} - \lambda_k) - \alpha\beta (\nabla h(-y_{k-1}) - \nabla h(-y_k)) \\ & \quad - \beta (\nabla h(-y_k) - \nabla h(-y_{k+1})). \end{aligned}$$

Then, we use the 1-strong-convex of $h(\cdot)$ and Assumption A(iii), we can follow

$$M \|x_{k+1} - x_k\|^2 \leq \|Ax_{k+1} - Ax_k\|^2 \leq \|\nabla h(Ax_{k+1} - b) - \nabla h(Ax_k - b)\|^2.$$

Thus, combining the above two formulas, we obtain that

$$\begin{aligned} & (\alpha+1)^2 \beta^2 M \|x_{k+1} - x_k\|^2 \\ & \leq (\alpha+1)^2 \beta^2 \|Ax_{k+1} - Ax_k\|^2 \leq (\alpha+1)^2 \beta^2 \|\nabla h(Ax_{k+1} - b) - \nabla h(Ax_k - b)\|^2 \\ & \leq 4 \left(\|\lambda_k - \lambda_{k-1}\|^2 + \|\lambda_{k+1} - \lambda_k\|^2 + \alpha^2 \beta^2 \|\nabla h(-y_{k-1}) - \nabla h(-y_k)\|^2 \right. \\ & \quad \left. + \beta^2 \|\nabla h(-y_k) - \nabla h(-y_{k+1})\|^2 \right) \\ & \leq 4 \left(\|\lambda_k - \lambda_{k-1}\|^2 + \|\lambda_{k+1} - \lambda_k\|^2 + \alpha^2 \beta^2 L_h^2 \|y_k - y_{k-1}\|^2 + \beta^2 L_h^2 \|y_{k+1} - y_k\|^2 \right), \end{aligned} \quad (3.18)$$

where $M > 0$. Then, the last inequality implies $\sum_{k=0}^{\infty} \|x_{k+1} - x_k\|^2 < +\infty$.

Thus, $\sum_{k=0}^{\infty} \|w_{k+1} - w_k\|^2 < +\infty$. This completes the proof. \square

Lemma 3.3. Let $\{w_k = (x_k, y_k, \lambda_k)\}$ be the sequence generated by the symmetric Bregman ADMM (1.6), which is supposed to be bounded, and given that Assumption A holds. For any positive integer k , we define

$$\begin{cases} x_{k+1}^* := A^T \lambda_k - A^T \lambda_{k+1} + \beta A^T \langle \nabla h^2(Ax_{k+1} - b), y_{k+1} - y_k \rangle, \\ y_{k+1}^* := \lambda_{k+\frac{1}{2}} - \lambda_{k+1}, \\ \lambda_{k+1}^* := -Ax_{k+1} - y_{k+1} + b. \end{cases}$$

Then, $(x_{k+1}^*, y_{k+1}^*, \lambda_{k+1}^*) \in \partial \mathcal{L}_\beta^h(w_{k+1})$ and there exists $\xi > 0$ such that

$$d(0, \partial \mathcal{L}_\beta^h(w_{k+1})) \leq \xi \|y_{k+1} - y_k\|.$$

Proof: By definition of function $\mathcal{L}_\beta^h(\cdot)$, we can obtain that

$$\begin{cases} \partial_x \mathcal{L}_\beta^h(w_{k+1}) = \partial f(x_{k+1}) - A^T \lambda_{k+1} - \beta A^T \langle \nabla^2 h(Ax_{k+1} - b), -Ax_{k+1} - y_{k+1} + b \rangle, \\ \partial_y \mathcal{L}_\beta^h(w_{k+1}) = \nabla g(y_{k+1}) - \lambda_{k+1} + \beta (\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})), \\ \partial_\lambda \mathcal{L}_\beta^h(w_{k+1}) = -(Ax_{k+1} + y_{k+1} - b). \end{cases} \quad (3.19)$$

Combining the optimality condition (3.1) with equation (3.19), we get

$$\begin{cases} A^T \lambda_k - A^T \lambda_{k+1} + \beta A^T \langle \nabla h^2(Ax_{k+1} - b), y_{k+1} - y_k \rangle \in \partial_x \mathcal{L}_\beta^h(w_{k+1}), \\ \lambda_{k+\frac{1}{2}} - \lambda_{k+1} \in \partial_y \mathcal{L}_\beta^h(w_{k+1}), \\ -Ax_{k+1} - y_{k+1} + b \in \partial_\lambda \mathcal{L}_\beta^h(w_{k+1}). \end{cases}$$

Then, according to Lemma 2.2, we obtain that $(x_{k+1}^*, y_{k+1}^*, \lambda_{k+1}^*) \in \partial \mathcal{L}_\beta^h(w_{k+1})$.

Furthermore,

$$\langle \nabla h^2(Ax_{k+1} - b), y_{k+1} - y_k \rangle \leq \|\nabla h^2(Ax_{k+1} - b)\| \|y_{k+1} - y_k\| \leq L_h \|y_{k+1} - y_k\|.$$

In addition,

$$\begin{aligned} \|Ax_{k+1} + y_{k+1} - b\| &\leq \|\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})\| \\ &\leq \frac{1}{(\alpha+1)\beta} \|\lambda_k - \lambda_{k+1}\| + \frac{\alpha}{\alpha+1} \|\nabla h(-y_{k+1}) - \nabla h(-y_k)\| \\ &\leq \frac{1}{(\alpha+1)\beta} \|\lambda_k - \lambda_{k+1}\| + \frac{\alpha L_h}{\alpha+1} \|y_k - y_{k+1}\|. \end{aligned}$$

On the other hand,

$$\begin{aligned} \left\| \lambda_{k+\frac{1}{2}} - \lambda_{k+1} \right\| &= \left\| \beta (\nabla h(Ax_{k+1} - b) - \nabla h(-y_{k+1})) \right\| \\ &\leq \frac{1}{\alpha+1} \|\lambda_k - \lambda_{k+1}\| + \frac{\alpha\beta}{\alpha+1} \|\nabla h(-y_{k+1}) - \nabla h(-y_k)\| \\ &\leq \frac{1}{\alpha+1} \|\lambda_k - \lambda_{k+1}\| + \frac{\alpha\beta L_h}{\alpha+1} \|y_{k+1} - y_k\|. \end{aligned}$$

Based on the above relationship, we know that there exist $\xi_1, \xi_2 > 0$ such that

$$\left\| \begin{pmatrix} x_{k+1}^*, y_{k+1}^*, \lambda_{k+1}^* \end{pmatrix} \right\| \leq \xi_1 \|y_{k+1} - y_k\| + \xi_2 \|\lambda_{k+1} - \lambda_k\|.$$

Defining $\xi := \xi_1 + \xi_2 LL_h$ and using (3.9), we obtain

$$d(0, \partial \mathcal{L}_\beta^h(w_{k+1})) \leq \left\| \begin{pmatrix} x_{k+1}^*, y_{k+1}^*, \lambda_{k+1}^* \end{pmatrix} \right\| \leq \xi \|y_{k+1} - y_k\|.$$

This completes the proof.

Lemma 3.4. Let $\{w_k = (x_k, y_k, \lambda_k)\}$ be the sequence generated by the symmetric Bregman ADMM (1.6), which is supposed to be bounded. Let $S(w_0)$ denote the set of its limit points. Then

(i) $S(w_0)$ is a nonempty compact set, and

$$d(w_k, S(w_0)) \rightarrow 0, \text{ as } k \rightarrow +\infty;$$

(ii) $S(w_0) \subset \text{crit } \mathcal{L}_\beta^h$, where $\text{crit } \mathcal{L}_\beta^h$ denotes the set of all stationary points of \mathcal{L}_β^h ;

(iii) $\mathcal{L}_\beta^h(\cdot)$ is finite and constant on $S(w_0)$, which equals to

$$\inf_{k \in \mathbb{N}} \mathcal{L}_\beta^h(w_k) = \lim_{k \rightarrow +\infty} \mathcal{L}_\beta^h(w_k).$$

Proof:

(i) The proposition is immediately derived from the definition of limit points.

(ii) We assume that $(x^*, y^*, \lambda^*) \in S(w_0)$, then there exists a subsequence $\{(x_{k_j}, y_{k_j}, \lambda_{k_j})\}$ that converges to (x^*, y^*, λ^*) . Note that Lemma 3.2, we have

$$\|w_{k+1} - w_k\| \rightarrow 0. \quad (3.20)$$

Consequently, we deduce that $\{(x_{k_j+1}, y_{k_j+1}, \lambda_{k_j+1})\}$ also converges to (x^*, y^*, λ^*) . Given that x_{k+1} is the minimizer of $\mathcal{L}_\beta^h(x, y_k, \lambda_k)$ concerning the variable x , we have

$$\mathcal{L}_\beta^h(x_{k+1}, y_k, \lambda_k) \leq \mathcal{L}_\beta^h(x^*, y_k, \lambda_k). \quad (3.21)$$

On one hand, according to (3.20), (3.21) and the continuity of $\mathcal{L}_\beta^h(\cdot)$ with respect to y and λ , we get

$$\limsup_{j \rightarrow +\infty} \mathcal{L}_\beta^h(x_{k_j+1}, y_{k_j}, \lambda_{k_j}) = \limsup_{j \rightarrow +\infty} \mathcal{L}_\beta^h(x_{k_j+1}, y_{k_j+1}, \lambda_{k_j+1}) \leq \mathcal{L}_\beta^h(x^*, y^*, \lambda^*). \quad (3.22)$$

On the other hand, using the lower semicontinuity of $\mathcal{L}_\beta^h(\cdot)$, we have

$$\limsup_{j \rightarrow +\infty} \mathcal{L}_\beta^h(x_{k_j+1}, y_{k_j+1}, \lambda_{k_j+1}) \geq \mathcal{L}_\beta^h(x^*, y^*, \lambda^*). \quad (3.23)$$

The above two relations (3.22) and (3.23) imply that

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

Taking the limit in the optimality conditions (3.1) along the subsequence $\{(x_{k_j+1}, y_{k_j+1}, \lambda_{k_j+1})\}$, and utilizing (3.20) yields

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

Thus, (x^*, y^*, λ^*) is a critical point of (5), which implies that $w^* \in \text{crit } \mathcal{L}_\beta^h$.

(iii) For any point $(x^*, y^*, \lambda^*) \in S(w_0)$, there exists a subsequence $\{(x_{k_j}, y_{k_j}, \lambda_{k_j})\}$ that converges to (x^*, y^*, λ^*) as $j \rightarrow +\infty$. By merging equations (22) and (23) with the observation that the sequence $\{\mathcal{L}_\beta^h(w_k)\}$ is nonincreasing, we have

$$\lim_{k \rightarrow +\infty} \mathcal{L}_\beta^h(x_k, y_k, \lambda_k) = \mathcal{L}_\beta^h(x^*, y^*, \lambda^*).$$

Therefore, $\mathcal{L}_\beta^h(\cdot)$ is constant on $S(w_0)$. Furthermore, we also have $\inf_{k \in \mathbb{N}} \mathcal{L}_\beta^h(w_k) = \lim_{k \rightarrow +\infty} \mathcal{L}_\beta^h(w_k)$.

Hence, we have complete the proof.

Now we present the main convergence result for the symmetric Bregman ADMM (1).

Theorem 3.1. Let $\{w_k = (x_k, y_k, \lambda_k)\}$ be the sequence generated by the symmetric Bregman ADMM (1.6), which is assumed to be bounded. Suppose that $\mathcal{L}_\beta^h(\cdot)$ is a KL function, then $\{w_k\}$ has finite length, that is

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

and as a consequence, $\{w_k\}$ converges to a critical point of $\mathcal{L}_\beta^h(\cdot)$.

Proof: Based on the proof of Lemma 3.4, it implies that $\mathcal{L}_\beta^h(w_k) \rightarrow \mathcal{L}_\beta^h(w^*)$ for all $w^* \in S(w_0)$. Next, we will consider two cases.

(i) If there exists an integer k_0 for which $\mathcal{L}_\beta^h(w_{k_0}) = \mathcal{L}_\beta^h(w^*)$. Then according to Remark 3.1 and (2), we have

$$\delta D_h(y_k, y_{k+1}) \leq \mathcal{L}_\beta^h(w_k) - \mathcal{L}_\beta^h(w_{k+1}) \leq \mathcal{L}_\beta^h(w_{k_0}) - \mathcal{L}_\beta^h(w^*) = 0,$$

for any $k > k_0$. Consequently, we conclude that $y_{k+1} = y_k$ for any $k > k_0$. Combining (9) and (18), we further deduce that $\lambda_{k+1} = \lambda_k$ and $x_{k+1} = x_k$ for any $k > k_0 + 1$, which implies that $w_{k+1} = w_k$. As a result, the assertion is substantiated.

(ii) If $\mathcal{L}_\beta^h(w_k) > \mathcal{L}_\beta^h(w^*)$ for all k . Considering that $d(w_k, S(w_0)) \rightarrow 0$, there exists $k_1 > 0$, such that $d(w_k, S(w_0)) < \varepsilon$ ($\forall \varepsilon > 0$) for any $k > k_1$. Furthermore, with $\mathcal{L}_\beta^h(w_k) \rightarrow \mathcal{L}_\beta^h(w^*)$, it follows that there exists $k_2 > 0$, such that $\mathcal{L}_\beta^h(w_k) < \mathcal{L}_\beta^h(w^*) + \eta$ ($\forall \eta > 0$) for any $k > k_2$.

Thus, when $k > \tilde{k} = \max\{k_1, k_2\}$ for all $\varepsilon, \eta > 0$, we derive the following conclusions

$$d(w_k, S(w_0)) < \varepsilon, \quad \mathcal{L}_\beta^h(w^*) < \mathcal{L}_\beta^h(w_k) < \mathcal{L}_\beta^h(w^*) + \eta.$$

Since $\mathcal{L}_\beta^h(\cdot)$ is constant on $S(w_0)$ and $S(w_0)$ is a nonempty compact set,

we use Lemma 2.3 with $\Omega = S(w_0)$ to derive that for any $k > \tilde{k}$,

$$\varphi'(\mathcal{L}_\beta^h(w_k) - \mathcal{L}_\beta^h(w^*))d(0, \partial\mathcal{L}_\beta^h(w_k)) \geq 1. \quad (3.24)$$

Relying on the fact that $\mathcal{L}_\beta^h(w_k) - \mathcal{L}_\beta^h(w_{k+1}) = \mathcal{L}_\beta^h(w_k) - \mathcal{L}_\beta^h(w^*) - (\mathcal{L}_\beta^h(w_{k+1}) - \mathcal{L}_\beta^h(w^*))$, and the concavity of $\varphi(\cdot)$, it follows that

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

Combining $d(0, \partial\mathcal{L}_\beta^h(w_k)) \leq \xi \|y_k - y_{k-1}\|$ with the above inequality, $\varphi'(\mathcal{L}_\beta^h(w_k) - \mathcal{L}_\beta^h(w^*)) > 0$ and inequality (3.24), we can obtain

$$\begin{aligned} \mathcal{L}_\beta^h(w_k) - \mathcal{L}_\beta^h(w_{k+1}) & \leq \frac{\varphi(\mathcal{L}_\beta^h(w_k) - \mathcal{L}_\beta^h(w^*)) - \varphi(\mathcal{L}_\beta^h(w_{k+1}) - \mathcal{L}_\beta^h(w^*))}{\varphi'(\mathcal{L}_\beta^h(w_k) - \mathcal{L}_\beta^h(w^*))} \\ & \leq d(0, \partial\mathcal{L}_\beta^h(w_k)) \left[\varphi(\mathcal{L}_\beta^h(w_k) - \mathcal{L}_\beta^h(w^*)) - \varphi(\mathcal{L}_\beta^h(w_{k+1}) - \mathcal{L}_\beta^h(w^*)) \right] \\ & \leq \xi \|y_k - y_{k-1}\| \left[\varphi(\mathcal{L}_\beta^h(w_k) - \mathcal{L}_\beta^h(w^*)) - \varphi(\mathcal{L}_\beta^h(w_{k+1}) - \mathcal{L}_\beta^h(w^*)) \right]. \end{aligned} \quad (3.25)$$

For convenience, for all $p, q \in \mathbb{N}$, we define $\Delta_{p,q} := \varphi(\mathcal{L}_\beta^h(w_p) - \mathcal{L}_\beta^h(w^*)) - \varphi(\mathcal{L}_\beta^h(w_q) - \mathcal{L}_\beta^h(w^*))$.

Hence, (3.25) can be simplified as

$$\mathcal{L}_\beta^h(w_k) - \mathcal{L}_\beta^h(w_{k+1}) \leq \xi \|y_k - y_{k-1}\| \Delta_{k,k+1}. \quad (3.26)$$

Combining inequality (3.26) with Lemma 2, we get that for all $k > \tilde{k}$,

$$\frac{\delta}{2} \|y_k - y_{k+1}\|^2 \leq \delta D_h(y_k, y_{k+1}) \leq \xi \|y_k - y_{k-1}\| \Delta_{k,k+1}.$$

Then,

$$\|y_k - y_{k+1}\| \leq \sqrt{\frac{2\xi}{\delta} \Delta_{k,k+1}} \|y_k - y_{k-1}\|^{1/2}.$$

Applying the fact that $2\sqrt{\alpha\beta} \leq \alpha + \beta$, we obtain

$$2\|y_k - y_{k+1}\| \leq \|y_k - y_{k-1}\| + \frac{2\xi}{\delta} \Delta_{k,k+1}. \quad (3.27)$$

Summing up (3.27) over for $k = \tilde{k} + 1, \dots, m$, we have

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

Note that $\varphi(\mathcal{L}_\beta^h(w_{m+1}) - \mathcal{L}_\beta^h(w^*)) > 0$ from Definition 2.8. Taking $m \rightarrow +\infty$, we have

$$\sum_{k=\tilde{k}+1}^{+\infty} \|y_{k+1} - y_k\| \leq \|y_{\tilde{k}+1} - y_{\tilde{k}}\| + \frac{2\xi}{\delta} \varphi(\mathcal{L}_\beta^h(w_{\tilde{k}+1}) - \mathcal{L}_\beta^h(w^*)). \quad (3.28)$$

Therefore,

$$\sum_{k=0}^{+\infty} \|y_{k+1} - y_k\| < +\infty. \quad (3.29)$$

Combining (3.9) and (3.29), we get

$$\sum_{k=0}^{+\infty} \|\lambda_{k+1} - \lambda_k\| < +\infty. \quad (3.30)$$

Using (3.18), we obtain

$$\begin{aligned} \|x_{k+1} - x_k\| &\leq \frac{2}{(\alpha+1)\beta\sqrt{M}} \left(\|\lambda_k - \lambda_{k-1}\|^2 + \|\lambda_{k+1} - \lambda_k\|^2 \right. \\ &\quad \left. + \alpha^2 \beta^2 L_h^2 \|y_k - y_{k-1}\|^2 + \beta^2 L_h^2 \|y_{k+1} - y_k\|^2 \right)^{\frac{1}{2}} \\ &\leq \frac{2}{(\alpha+1)\beta\sqrt{M}} \left(\|\lambda_k - \lambda_{k-1}\| + \|\lambda_{k+1} - \lambda_k\| \right. \\ &\quad \left. + \alpha\beta L_h \|y_k - y_{k-1}\| + \beta L_h \|y_{k+1} - y_k\| \right). \end{aligned}$$

Combining above inequality with (3.29) and (3.30), we have

$$\sum_{k=0}^{+\infty} \|x_{k+1} - x_k\| < +\infty. \quad (3.31)$$

Furthermore, we note that

$$\begin{aligned} \|w_{k+1} - w_k\| &= \left(\|x_{k+1} - x_k\|^2 + \|y_{k+1} - y_k\|^2 + \|\lambda_{k+1} - \lambda_k\|^2 \right)^{1/2} \\ &\leq \|x_{k+1} - x_k\| + \|y_{k+1} - y_k\| + \|\lambda_{k+1} - \lambda_k\|. \end{aligned}$$

Using (3.29), (3.30) and (3.31), it follows that

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

which implies that $\{w_k\}$ is a Cauchy sequence and thus convergent. The assertion follows from Lemma 3.4 immediately. \square

Next, we provide the essential sufficient conditions to establish that the sequence $\{w_k = (x_k, y_k, \lambda_k)\}$ generated by the symmetric Bregman ADMM (1.6) is bounded.

Lemma 3.5. *Let $\{w_k = (x_k, y_k, \lambda_k)\}$ be the sequence generated by the symmetric Bregman ADMM (1.6). Suppose that*

$$\inf_y \left\{ g(y) - \frac{1}{2L} \|\nabla g(y)\|^2 \right\} =: \bar{g} > -\infty.$$

If at least one of the following statements is true:

- (i) $\liminf_{\|x\| \rightarrow +\infty} f(x) = +\infty$.
- (ii) $\inf_x f(x) > -\infty$ and $\liminf_{\|y\| \rightarrow +\infty} g(y) = +\infty$.

Then, the sequence $\{w_k = (x_k, y_k, \lambda_k)\}$ is bounded.

Proof. Suppose that condition (i) holds. From Lemma 3.1, we have

$$\mathcal{L}_\beta^h(x_k, y_k, \lambda_k) \leq \mathcal{L}_\beta^h(x_1, y_1, \lambda_1).$$

which implies

$$\mathcal{L}_\beta^h(x_1, y_1, \lambda_1) \geq f(x_k) + g(y_k) - \langle \lambda_k, Ax_k + y_k - b \rangle + \beta D_h(-y_k, Ax_k - b). \quad (3.32)$$

Using the 1-strong convexity of the function $h(\cdot)$ and the definition of $D_h(\cdot, \cdot)$, we have

$$\begin{aligned} D_h(-y_k, Ax_k - b) &= h(-y_k) - h(Ax_k - b) - \langle \nabla h(Ax_k - b), -y_k - Ax_k - b \rangle \\ &\geq \frac{1}{2} \|Ax_k + y_k - b\|^2. \end{aligned} \quad (3.33)$$

Combining (3.32) with (3.33) and using the fact $\lambda_k = \nabla g(y_k)$, we get

$$\begin{aligned} \mathcal{L}_\beta^h(x_1, y_1, \lambda_1) &\geq f(x_k) + g(y_k) - \langle \lambda_k, Ax_k + y_k - b \rangle + \frac{\beta}{2} \|Ax_k + y_k - b\|^2 \\ &= f(x_k) + g(y_k) - \frac{1}{2\beta} \|\lambda_k\|^2 + \frac{\beta}{2} \left\| Ax_k + y_k - b - \frac{1}{\beta} \lambda_k \right\|^2 \\ &= f(x_k) + \left(g(y_k) - \frac{1}{2L} \|\nabla g(y_k)\|^2 \right) + \left(\frac{1}{2L} - \frac{1}{2\beta} \right) \|\lambda_k\|^2 \\ &\quad + \frac{\beta}{2} \left\| Ax_k + y_k - b - \frac{1}{\beta} \lambda_k \right\|^2 \\ &\geq f(x_k) + \bar{g} + \left(\frac{1}{2L} - \frac{1}{2\beta} \right) \|\lambda_k\|^2 + \frac{\beta}{2} \left\| Ax_k + y_k - b - \frac{1}{\beta} \lambda_k \right\|^2. \end{aligned} \quad (3.4)$$

Note that (i) implies that $\inf_x f(x) > -\infty$. When $\alpha \in \left(-\frac{1}{2L_h^2 + 1}, 0 \right]$, we have

$$\beta > \frac{2LL_h^2 - 2\alpha LL_h^3}{1 + \alpha + 2\alpha L_h^2} > 2L, \text{ besides, when } \alpha \in \left[0, \frac{1}{2L_h^2 - 1} \right), \text{ we have}$$

$$\beta > \frac{2(1 + \alpha)LL_h^2 + 2\alpha LL_h^3}{1 + \alpha - 2\alpha L_h^2} > 2L. \text{ Therefore, we can derive that } \{x_k\}, \{\lambda_k\} \text{ and}$$

$$\left\{ \frac{\beta}{2} \left\| Ax_k + y_k - b - \frac{1}{\beta} \lambda_k \right\|^2 \right\} \text{ are bounded. Consequently, } \{y_k\} \text{ is also bounded,}$$

and hence $\{w_k\}$ is bounded.

Next, suppose condition (ii) holds. Using $\lambda_k \in \partial g(y_k)$ and (3.34), we get that

$$\begin{aligned} \mathcal{L}_\beta^h(x_1, y_1, \lambda_1) &\geq f(x_k) + g(y_k) - \frac{1}{2\beta} \|\lambda_k\|^2 + \frac{\beta}{2} \left\| Ax_k + y_k - b - \frac{1}{\beta} \lambda_k \right\|^2 \\ &= f(x_k) + \frac{1}{2} g(y_k) + \left(\frac{1}{2} g(y_k) - \frac{1}{2L} \|\nabla g(y_k)\|^2 \right) \\ &\quad + \left(\frac{1}{4L} - \frac{1}{2\beta} \right) \|\lambda_k\|^2 + \frac{\beta}{2} \left\| Ax_k + y_k - b - \frac{1}{\beta} \lambda_k \right\|^2 \\ &\geq f(x_k) + \frac{1}{2} g(y_k) + \frac{1}{2} \bar{g} + \left(\frac{1}{4L} - \frac{1}{2\beta} \right) \|\lambda_k\|^2 \\ &\quad + \frac{\beta}{2} \left\| Ax_k + y_k - b - \frac{1}{\beta} \lambda_k \right\|^2. \end{aligned}$$

Note that $\liminf_{\|y\| \rightarrow \infty} g(y) = +\infty$ implies that $\inf_y g(y) > -\infty$. When

$\alpha \in \left(-\frac{1}{2L_h^2+1}, 0\right]$, we have $\beta > \frac{2LL_h^2-2\alpha LL_h^3}{1+\alpha+2\alpha L_h^2} > 2L$, besides, when $\alpha \in \left[0, \frac{1}{2L_h^2-1}\right)$, we have $\beta > \frac{2(1+\alpha)LL_h^2+2\alpha LL_h^3}{1+\alpha-2\alpha L_h^2} > 2L$. Therefore, we con-

clude that the sequences $\{y_k\}$, $\{\lambda_k\}$ and $\left\{\frac{\beta}{2}\left\|Ax_k + y_k - b - \frac{1}{\beta}\lambda_k\right\|^2\right\}$ are

bounded. Then, from (3.18), it follows that $\{x_k\}$ is also bounded, and as a consequence, $\{w_k\}$ is bounded. This completes the proof.

Theorem 3.2. (Convergence rate) Let $\{w_k = (x_k, y_k, \lambda_k)\}$ be the sequence generated by the symmetric Bregman ADMM (1.6) and converge to $\{w^* = (x^*, y^*, \lambda^*)\}$. Assuming that $\mathcal{L}_\beta^h(\cdot)$ has the KL property at (x^*, y^*, λ^*) with $\varphi(s) = cs^{1-\theta}$, $\theta \in [0, 1)$, $c > 0$. Then, the following results hold

(i) If $\theta = 0$, then the sequence $\{w_k = (x_k, y_k, \lambda_k)\}$ converges in a finite number of steps.

(ii) If $\theta \in \left(0, \frac{1}{2}\right]$, then there exists $c_1 > 0$ and $\tau \in [0, 1)$ such that

$$\|(x_k, y_k, \lambda_k) - (x^*, y^*, \lambda^*)\| \leq c_1 \tau^k.$$

(iii) If $\theta \in \left(\frac{1}{2}, 1\right)$, then there exists $c_2 > 0$ such that

$$\|(x_k, y_k, \lambda_k) - (x^*, y^*, \lambda^*)\| \leq c_2 k^{(\theta-1)/(2\theta-1)}.$$

Proof: Firstly, consider the case that $\theta = 0$, we have $\varphi(s) = cs$ and $\varphi'(s) = c$. Proof by contradiction, suppose that $\{w_k = (x_k, y_k, \lambda_k)\}$ does not converge in a finite number of steps, and then, the KL property at (x^*, y^*, λ^*) yields $c \cdot d(0, \partial \mathcal{L}_\beta^h(w_k)) \geq 1$ for any sufficiently large k , which is contrary to Lemma 3.1.

Secondly, consider that $\theta > 0$ and set $\Delta_k = \sum_{i=k}^{+\infty} \|y_{i+1} - y_i\|$ for $k \geq 0$. By the triangle inequality, we derive that $\Delta_k \geq \|y_k - y^*\|$, and hence it is able to estimate Δ_k . With these notations, it follows from (3.28) that

$$\Delta_{\tilde{k}+1} \leq \Delta_{\tilde{k}} - \Delta_{\tilde{k}+1} + \frac{2\xi}{\delta} \varphi(\mathcal{L}_\beta^h(w_{\tilde{k}+1}) - \mathcal{L}_\beta^h(w^*)).$$

Invoking the KL property of $\mathcal{L}_\beta^h(\cdot)$ at (x^*, y^*, λ^*) , we conclude that

$$\varphi'(\mathcal{L}_\beta^h(w_{\tilde{k}+1}) - \mathcal{L}_\beta^h(w^*)) d(0, \partial \mathcal{L}_\beta^h(w_{\tilde{k}+1})) \geq 1.$$

This can be taken to imply

$$(\mathcal{L}_\beta^h(w_{\tilde{k}+1}) - \mathcal{L}_\beta^h(w^*))^\theta \leq c \cdot (1-\theta) d(0, \partial \mathcal{L}_\beta^h(w_{\tilde{k}+1})). \quad (3.35)$$

According to Lemma 3.1, we get

$$d(0, \partial \mathcal{L}_\beta^h(w_{\tilde{k}+1})) \leq \xi \|y_{\tilde{k}+1} - y_{\tilde{k}}\| = \xi (\Delta_{\tilde{k}} - \Delta_{\tilde{k}+1}). \quad (3.36)$$

Combining (3.35) and (3.36), it follows that there exists $\gamma > 0$ such that

$$\varphi\left(\mathcal{L}_{\beta}^h(w_{\tilde{k}+1}) - \mathcal{L}_{\beta}^h(w^*)\right) = c \cdot \left(\mathcal{L}_{\beta}^h(w_{\tilde{k}+1}) - \mathcal{L}_{\beta}^h(w^*)\right)^{1-\theta} \leq \gamma(\Delta_{\tilde{k}} - \Delta_{\tilde{k}+1})^{(1-\theta)/\theta}.$$

Therefore,

$$\Delta_{\tilde{k}+1} \leq \Delta_{\tilde{k}} - \Delta_{\tilde{k}+1} + \frac{2\xi}{\delta} \gamma (\Delta_{\tilde{k}} - \Delta_{\tilde{k}+1})^{(1-\theta)/\theta}.$$

Sequences satisfying such inequalities have been studied in [33]. It is shown that

- If $\theta \in \left(0, \frac{1}{2}\right]$, then there exists $c_1 > 0$ and $\tau \in [0, 1)$, such that

$$\|y_k - y^*\| \leq c_1 \tau^k. \quad (3.37)$$

- If $\theta \in \left(\frac{1}{2}, 1\right)$, then there exists $c_2 > 0$, such that

$$\|y_k - y^*\| \leq c_2 k^{(\theta-1)/(2\theta-1)}. \quad (3.38)$$

Recalling that

$$\|\lambda_{k+1} - \lambda_k\| \leq LL_h \|y_{k+1} - y_k\|,$$

consequently,

$$\|\lambda_k - \lambda^*\| \leq LL_h \|y_k - y^*\|. \quad (3.39)$$

Furthermore, from the relations

$$\lambda_k = \lambda_{k-1} - \alpha\beta(\nabla h(Ax_k - b) - \nabla h(-y_{k-1})) - \beta(\nabla h(Ax_k - b) - \nabla h(-y_k)),$$

and

$$\nabla h(Ax^* - b) - \nabla h(-y^*) = 0,$$

it follows that

$$\begin{aligned} & (\alpha+1)\beta(\nabla h(Ax_k - b) - \nabla h(Ax^* - b)) \\ &= \beta(\nabla h(-y_k) - \nabla h(-y^*)) + (\lambda_{k-1} - \lambda^*) + (\lambda^* - \lambda_k) \\ & \quad + \alpha\beta(\nabla h(-y_{k-1}) - \nabla h(-y^*)). \end{aligned}$$

i.e.,

$$\begin{aligned} & (\nabla h(Ax_k - b) - \nabla h(Ax^* - b)) \\ &= \frac{1}{\alpha+1}(\nabla h(-y_k) - \nabla h(-y^*)) + \frac{1}{(\alpha+1)\beta}(\lambda_{k-1} - \lambda^*) \\ & \quad + \frac{1}{(\alpha+1)\beta}(\lambda^* - \lambda_k) + \frac{\alpha}{\alpha+1}(\nabla h(-y_{k-1}) - \nabla h(-y^*)). \end{aligned}$$

Subsequently, combining the 1-strong-convexity of $h(\cdot)$ and above equality, it follows that

$$\begin{aligned}
\|x_k - x^*\| &\leq \frac{1}{M} \|\nabla h(Ax_k - b) - \nabla h(Ax^* - b)\| \\
&\leq \frac{1}{M} \left(\frac{1}{(\alpha+1)\beta} \|\lambda_{k-1} - \lambda^*\| + \frac{1}{(\alpha+1)\beta} \|\lambda^* - \lambda_k\| \right. \\
&\quad \left. + \frac{1}{\alpha+1} \|\nabla h(-y_k) - \nabla h(-y^*)\| + \frac{\alpha}{\alpha+1} \|\nabla h(-y_{k-1}) - \nabla h(-y^*)\| \right) \\
&\leq \frac{1}{M} \left(\frac{1}{(\alpha+1)\beta} \|\lambda_{k-1} - \lambda^*\| + \frac{1}{(\alpha+1)\beta} \|\lambda^* - \lambda_k\| \right. \\
&\quad \left. + \frac{L_h}{(\alpha+1)\beta} \|y_k - y^*\| + \frac{\alpha L_h}{(\alpha+1)\beta} \|y_{k-1} - y^*\| \right) \\
&\leq \frac{1}{M} \left(\frac{LL_h}{(\alpha+1)\beta} \|y_{k-1} - y^*\| + \frac{LL_h}{(\alpha+1)\beta} \|y_k - y^*\| \right. \\
&\quad \left. + \frac{L_h}{(\alpha+1)\beta} \|y_k - y^*\| + \frac{\alpha L_h}{(\alpha+1)\beta} \|y_{k-1} - y^*\| \right) \\
&\leq \frac{1}{M} \left(\frac{(1+L)L_h}{(\alpha+1)\beta} \|y_k - y^*\| + \frac{(\alpha+L)L_h}{(\alpha+1)\beta} \|y_{k-1} - y^*\| \right). \tag{3.40}
\end{aligned}$$

The desired inequalities follow from (3.37)-(3.40) immediately. \square

4. Conclusions

In this paper, we proposed a symmetric Bregman ADMM, which can return to the symmetric ADMM, the Bregman ADMM and classical ADMM while circumventing the requirement for global Lipschitz continuity of the gradient when minimizing a linearly constrained nonconvex minimization problem whose objective function is the sum of two separable nonconvex functions. Moreover, we analyze its convergence, under certain assumptions and when the associated function satisfies the Kurdyka-Lojasiewicz inequality, we prove that the iterative sequence generated by the symmetric Bregman ADMM converges to a critical point of the problem. Finally, we establish the convergence rate of the algorithm.

Conflicts of Interest

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

References

- [1] Goldstein, T., O'Donoghue, B., Setzer, S. and Baraniuk, R. (2014) Fast Alternating Direction Optimization Methods. *SIAM Journal on Imaging Sciences*, **7**, 1588-1623. <https://doi.org/10.1137/120896219>
- [2] Boyd, S. (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>
- [3] Yin, W., Osher, S., Goldfarb, D. and Darbon, J. (2008) Bregman Iterative Algorithms for ℓ_1 -Minimization with Applications to Compressed Sensing. *SIAM Journal on Imaging Sciences*, **1**, 143-168. <https://doi.org/10.1137/070703983>

- [4] 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>
- [5] Goldstein, T., Bresson, X. and Osher, S. (2009) 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>
- [6] Ochs, P., Chen, Y., Brox, T. and Pock, T. (2014) iPiano: Inertial Proximal Algorithm for Nonconvex Optimization. *SIAM Journal on Imaging Sciences*, **7**, 1388-1419. <https://doi.org/10.1137/130942954>
- [7] Candes, E.J. and Tao, T. (2006) Near-Optimal Signal Recovery from Random Projections: Universal Encoding Strategies? *IEEE Transactions on Information Theory*, **52**, 5406-5425. <https://doi.org/10.1109/tit.2006.885507>
- [8] Beck, A. and Teboulle, M. (2009) Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems. *IEEE Transactions on Image Processing*, **18**, 2419-2434. <https://doi.org/10.1109/tip.2009.2028250>
- [9] Dai, W. and Milenkovic, O. (2009) Subspace Pursuit for Compressive Sensing Signal Reconstruction. *IEEE Transactions on Information Theory*, **55**, 2230-2249. <https://doi.org/10.1109/tit.2009.2016006>
- [10] Aharon, M., Elad, M. and Bruckstein, A. (2006) K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Transactions on Signal Processing*, **54**, 4311-4322. <https://doi.org/10.1109/tsp.2006.881199>
- [11] Gabay, D. and Mercier, B. (1976) A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation. *Computers & Mathematics with Applications*, **2**, 17-40. [https://doi.org/10.1016/0898-1221\(76\)90003-1](https://doi.org/10.1016/0898-1221(76)90003-1)
- [12] Glowinski, R. and Marroco, A. (1975) 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*, **9**, 41-76. <https://doi.org/10.1051/m2an/197509r200411>
- [13] Boley, D. (2013) Local Linear Convergence of the Alternating Direction Method of Multipliers on Quadratic or Linear Programs. *SIAM Journal on Optimization*, **23**, 2183-2207. <https://doi.org/10.1137/120878951>
- [14] Han, D. and Yuan, X. (2013) Local Linear Convergence of the Alternating Direction Method of Multipliers for Quadratic Programs. *SIAM Journal on Numerical Analysis*, **51**, 3446-3457. <https://doi.org/10.1137/120886753>
- [15] He, B. and Yuan, X. (2012) On the $O(1/n)$ Convergence Rate of the Douglas-Rachford Alternating Direction Method. *SIAM Journal on Numerical Analysis*, **50**, 700-709. <https://doi.org/10.1137/110836936>
- [16] Yang, W.H. and Han, D. (2016) Linear Convergence of the Alternating Direction Method of Multipliers for a Class of Convex Optimization Problems. *SIAM Journal on Numerical Analysis*, **54**, 625-640. <https://doi.org/10.1137/140974237>
- [17] Li, G. and Pong, T.K. (2015) Global Convergence of Splitting Methods for Nonconvex Composite Optimization. *SIAM Journal on Optimization*, **25**, 2434-2460. <https://doi.org/10.1137/140998135>
- [18] Hong, M., Luo, Z. and Razaviyayn, M. (2016) Convergence Analysis of Alternating Direction Method of Multipliers for a Family of Nonconvex Problems. *SIAM Journal on Optimization*, **26**, 337-364. <https://doi.org/10.1137/140990309>
- [19] Guo, K., Han, D.R. and Wu, T.T. (2016) Convergence of Alternating Direction 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>
- [20] Wu, Z., Li, M., Wang, D.Z.W. and Han, D. (2017) A Symmetric Alternating Direction Method of Multipliers for Separable Nonconvex Minimization Problems. *Asia-Pacific Journal of Operational Research*, **34**, Article ID: 1750030.
<https://doi.org/10.1142/s0217595917500300>
 - [21] Bauschke, H.H., Bolte, J. and Teboulle, M. (2017) A Descent Lemma Beyond Lipschitz Gradient Continuity: First-Order Methods Revisited and Applications. *Mathematics of Operations Research*, **42**, 330-348.
<https://doi.org/10.1287/moor.2016.0817>
 - [22] Dragomir, R., d'Aspremont, A. and Bolte, J. (2021) Quartic First-Order Methods for Low-Rank Minimization. *Journal of Optimization Theory and Applications*, **189**, 341-363. <https://doi.org/10.1007/s10957-021-01820-3>
 - [23] Nesterov, Y. (2019) Implementable Tensor Methods in Unconstrained Convex Optimization. *Mathematical Programming*, **186**, 157-183.
<https://doi.org/10.1007/s10107-019-01449-1>
 - [24] Tan, L. and Guo, K. (2025) Bregman ADMM: A New Algorithm for Nonconvex Optimization with Linear Constraints. *Journal of Nonlinear and Variational Analysis*, **9**, 176-196. <https://doi.org/10.23952/jnva.9.2025.2.02>
 - [25] Wang, H. and Banerjee, A. (2014) Bregman Alternating Direction Method of Multipliers. arXiv: 1306.3203.
 - [26] Beck, A. (2017) First-Order Methods in Optimization. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611974997>
 - [27] Bolte, J., Sabach, S., Teboulle, M. and Vaisbourd, Y. (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>
 - [28] 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-Łojasiewicz Inequality. *Mathematics of Operations Research*, **35**, 438-457. <https://doi.org/10.1287/moor.1100.0449>
 - [29] Attouch, H., Bolte, J. and Svaiter, B.F. (2011) 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>
 - [30] Bolte, J., Sabach, S. and Teboulle, M. (2013) Proximal Alternating Linearized Minimization for Nonconvex and Nonsmooth Problems. *Mathematical Programming*, **146**, 459-494. <https://doi.org/10.1007/s10107-013-0701-9>
 - [31] Bauschke, H.H., Borwein, J.M. and Combettes, P.L. (2003) Bregman Monotone Optimization Algorithms. *SIAM Journal on Control and Optimization*, **42**, 596-636.
<https://doi.org/10.1137/s0363012902407120>
 - [32] Bregman, L.M. (1967) 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*, **7**, 200-217.
[https://doi.org/10.1016/0041-5553\(67\)90040-7](https://doi.org/10.1016/0041-5553(67)90040-7)
 - [33] Attouch, H. and Bolte, J. (2007) On the Convergence of the Proximal Algorithm for Nonsmooth Functions Involving Analytic Features. *Mathematical Programming*, **116**, 5-16. <https://doi.org/10.1007/s10107-007-0133-5>