Journal of Applied Mathematics and Physics, 2025, 13(4), 1138-1162
https://www.scirp.org/journal/jamp
ISSN Online: 2327-4379
ISSN Print: 2327-4352

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

Junping Yao, Mei Lu*

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: 2986962794@qq.com, *1339975549@qq.com

How to cite this paper: Yao, J.P. and Lu, M. (2025) Convergence of Generalized Bregman Alternating Direction Method of Multipliers for Nonconvex Objective with Linear Constraints. *Journal of Applied Mathematics and Physics*, **13**, 1138-1162.
https://doi.org/10.4236/jamp.2025.134060

Received: March 3, 2025
Accepted: April 7, 2025
Published: April 10, 2025

Abstract

In this paper, we investigate the convergence of the generalized Bregman alternating direction method of multipliers (ADMM) for solving nonconvex separable problems with linear constraints. This algorithm relaxes the requirement of global Lipschitz continuity of differentiable functions that is often seen in many researches, and it incorporates the acceleration technique of the proximal point algorithm (PPA). As a result, the scope of application of the algorithm is broadened and its performance is enhanced. Under the assumption that the augmented Lagrangian function satisfies the Kurdyka-Lojasiewicz inequality, we demonstrate that the iterative sequence generated by the algorithm converges to a critical point of its augmented Lagrangian function when the penalty parameter in the augmented Lagrangian function is sufficiently large. Finally, we analyze the convergence rate of the algorithm.

Keywords

Generalized Bregman Alternating Direction Method of Multipliers, Nonconvex Optimization, Lipschitz-Like Convexity Condition, Kurdyka-Lojasiewicz Inequality

1. Introduction

In this paper, what we consider is the two-block separable optimization problem model with linear constraints:

*Corresponding author.

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

where $f:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ is a proper lower semicontinuous function, $g:\mathbb{R}^m \to \mathbb{R}$ is a continuously differentiable function, $A\in\mathbb{R}^{m\times n}$ is a matrix, and $b\in\mathbb{R}^m$ is a vector. Many valuable optimization problems can be formulated in the form of (1.2), making it applicable across a wide range of fields, such as image and signal processing [1]-[4], statistical learning [5], and compressed sensing [6] [7].

Among the many methods to solve this kind of problem (1.1), the Alternating Direction Method of Multipliers (ADMM) is one of the most classic methods. The iterative scheme of the ADMM is as follows:

$$\begin{cases} x^{k+1}=\arg\min\limits_{x}\left\{\mathcal{L}_\beta\left(x,y^k,\lambda^k\right)\right\},\\ y^{k+1}=\arg\min\limits_{y}\left\{\mathcal{L}_\beta\left(x^{k+1},y,\lambda^k\right)\right\},\\ \lambda^{k+1}=\lambda^k+\beta\left(Ax^{k+1}+y^{k+1}-b\right).\end{cases} \tag{1.2}$$

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

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

where λ is the Lagrangian multiplier associated with the linear constraint, and $\beta>0$ is the penalty parameter. ADMM has been known since the mid-1970s when it was introduced by Gabay, Mercier, Glowinski, and Marrocco [8] [9]. When both f and g are convex functions, ADMM has produced a number of well-understood results for both convergence and rate of convergence in problem (1.1) [10]-[15]. However, when the objective function contains a nonconvex part, many subsequent studies have focused on variants of ADMM, often adding conditions to prove the corresponding convergence. For instance, Li and Pong [16] proposed a proximal ADMM, and in their convergence analysis, they required the constraint to be $Ax=y$. Subsequently, Hong, Luo, and Razaviyayn [17] proved the convergence of ADMM for solving consensus and sharing problems by assuming that the penalty parameter in the augmented Lagrangian is chosen to be sufficiently large (in 2016). In 2017, Guo *et al.* [18] further improved on these results, demonstrating that under more concise conditions than those in [16] [17], provided that the augmented Lagrangian function satisfies the Kurdyka-Lojasiewicz inequality.

Reference [19] indicates that ADMM (1.2) is actually the dual of the well-known Douglas-Rachford splitting method (DRSM) [20] applied to problem (1.1). In the literature [21], DRSM is further interpreted as a special case of the proximity point algorithm (PPA). Additionally, literature [21] uses the acceleration form of PPA to accelerate the original ADMM (1.2). With this acceleration technique, Guo *et al.* proposed a Generalized Alternating Direction Method of Multipliers (GADMM) in 2018 to accelerate the original ADMM for both separable and inseparable problems [22]. The iteration format for one of them is as follows:

$$\begin{cases} x^{k+1} \in \arg\min_x \left\{ f(x) + H(x, y^k) - \langle \lambda^k, Ax \rangle + \frac{\beta}{2} \left\| Ax + y^k - b \right\|^2 \right\} \\ y^{k+1} \in \arg\min_y \left\{ g(y) + H(x^{k+1}, y) - \langle \lambda^k, y \rangle + \frac{\beta}{2} \left\| \alpha A x^{k+1} + (1-\alpha)(b - y^k) + y - b \right\|^2 \right\} \\ \lambda^{k+1} = \lambda^k - \beta \left(\alpha A x^{k+1} + (1-\alpha)(b - y^k) + y^{k+1} - b \right) \end{cases} \tag{1.3}$$

Obviously, the GADMM (1.3) reduces to the classic ADMM (1.2) when $\alpha = 1$, and it reduces to the classic GADMM [21] when $H \equiv 0$. We can prove the acceleration performance of the GADMM [21] through references [23]-[25]. Nevertheless, investigations into the application of ADMM within the realm of nonconvex optimization frequently hinge on the premise that the gradient of the differentiable function adheres to global Lipschitz continuity. However, this condition is not universally satisfied in pivotal problem formulations across domains such as Poisson inverse problems [26], quadratic inverse problems [27], and rank minimization [28] [29]. Consequently, this stringent assumption unduly constrains the versatility of ADMM in these critical areas.

In 2016, Bauschke, Bolte, and Teboulle [26] introduced the Lipschitz-like convexity condition as a relaxation of the global Lipschitz continuous gradient assumption in optimization problems. In 2018, Bolte, Sabach, and Teboulle [27] further introduced the L-smooth adaptivity condition as a supplement to the Lipschitz-like convexity condition. These conditions allowed ADMM to be extended to problems where the gradient of the differentiable function does not satisfy global Lipschitz continuity. Recently, Guo and Tan [30] proposed a "real" Bregman ADMM based on these weakened conditions, which can reduce to the classical ADMM and override its results [18]. The iterative format for Guo and Tan's Bregman ADMM is as follows:

$$\begin{cases} x^{k+1} = \arg\min_x \left\{ \mathcal{L}_\beta^h (x, y^k, \lambda^k) \right\} \\ y^{k+1} = \arg\min_y \left\{ \mathcal{L}_\beta^h (x^{k+1}, y, \lambda^k) \right\} \\ \lambda^{k+1} = \lambda^k + \beta \left(\nabla h (A x^{k+1} - b) - \nabla h (-y^{k+1}) \right) \end{cases} \tag{1.4}$$

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

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

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

Combining the aforementioned content, we aim to integrate the acceleration technique from the Proximal Point Algorithm (PPA) while relaxing the requirement of gradient Lipschitz continuity for differentiable functions in ADMM. To this end, we propose a generalized Bregman ADMM, whose iterative format is as follows:

$$\begin{cases} x^{k+1} = \arg\min_x \left\{ f(x) + g(y^k) - \langle \lambda^k, Ax + y^k - b \rangle + \frac{\beta}{2} \|Ax + y^k - b\|^2 \right\}, \\ y^{k+1} = \arg\min_y \left\{ f(x^{k+1}) + g(y) - \langle \lambda^k, Ax^{k+1} + y - b \rangle + \beta D_h \left(-y - (\alpha - 1) y^k, \alpha \left(Ax^{k+1} - b \right) \right) \right\}, \\ \lambda^{k+1} = \lambda^k - \beta \left(\nabla h \left(\alpha \left(Ax^{k+1} - b \right) \right) - \nabla h \left(-y^{k+1} - (\alpha - 1) y^k \right) \right). \end{cases} \tag{1.6}$$

Here, the parameter α is the relaxation factor. The generalized Bregman ADMM (1.6) elegantly transitions to the generalized ADMM (1.3) by setting $h(\cdot) = \frac{1}{2}\|\cdot\|^2$, and to the Bregman ADMM (1.4) by setting $\alpha = 1$.

Remark 1.1 It is worth noting that, since we have only weakened the conditions on the function $g(y)$, the iterative format for the variable x remains unchanged (consistent with the classical method). Only the Bregman distance D_h has been introduced in the iterative format for the variable y.

We know that a very important technique for proving the convergence of nonconvex optimization problems depends on assuming that the objective function satisfies the Kurdyka-Lojasiewicz (KL) inequality. This assumption is also used in many previous articles [16] [18] [30]. Therefore, we also assume that the function satisfies the KL inequality. It is further proved that when the augmented Lagrangian function is a KL function, the sequence generated by the generalized Bregman ADMM converges to a KKT point of the problem (1.1). Ultimately, we analyze the convergence rate of the proposed algorithm under the specified parameter configurations. The structure of the remainder of this paper is outlined as follows. In Section 2, we establish the necessary theoretical foundations for our subsequent analysis. In Section 3, we conduct a detailed convergence analysis of the generalized Bregman Alternating Direction Method of Multipliers (ADMM) and determine its convergence rate. Lastly, in Section 4, we encapsulate our key findings and present our conclusions.

2. Preliminaries

In this section, we review some definitions and fundamental results that will be utilized in our subsequent analysis.

Definition 2.1 [31] For an extended real-valued function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$, the effective domain, or simply the domain, is the set

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

Definition 2.2 [31] A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called proper if there exists at least one $x \in \mathbb{R}^n$ such that $f(x) < \infty$.

Definition 2.3 [31] A function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is called lower semicontinuous at $x \in \mathbb{R}^n$ if

$$f(x) \le \liminf_{k \to \infty} f(x^k)$$

for any sequence $\{x^k\} \subseteq \mathbb{R}^n$ such that $x_k \to x$ as $k \to \infty$. Moreover, f 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 $\mathbb{R}^m$. A function $h:\mathbb{R}^m \to \mathbb{R}\cup\{+\infty\}$ associated with C is called a kernel generating distance if it satisfies the following conditions:

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

ii) h is C^1 on $int(dom(h))\equiv C$.

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

Definition 2.5 ([32]) Let $h\in\mathcal{G}(C)$. The Bregman distance $D_h: dom(h)\times int(dom(h))\to\mathbb{R}^+$ is defined by

$$D_h(x,y):=h(x)-h(y)-\langle \nabla h(y), x-y\rangle. \tag{2.1}$$

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

Lemma 2.1 ([33]) Let $h\in\mathcal{G}(C)$. For any $x,y\in int(dom(h))$ and $z\in dom(h)$, the following properties hold:

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

ii) The three-point identity holds:

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

Definition 2.6 ([27], L-smooth adaptable) Let $h\in\mathcal{G}(C)$ and $g:\mathbb{R}^m\to\mathbb{R}$ be continuously differentiable on $C=int(dom(h))$. 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 2.6 naturally complements and extends the definition of "A Lipschitz-like/Convexity Condition" in [26], which allows us to obtain the following two-sided descent lemma.

Lemma 2.2 ([27], extended descent lemma) The pair of functions (g,h) is L-smooth adaptable on C if and only if

$$\left|g(x)-g(y)-\langle \nabla g(y), x-y\rangle\right|\le LD_h(x,y)\quad \forall x,y\in int(dom(h)). \tag{2.2}$$

Remark 2.2 In particular, when the set $C=\mathbb{R}^m$ and $h(\cdot)=\frac{1}{2}\|\cdot\|^2$, (2.2) reduces to the classical descent lemma for the function g, i.e.,

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

Definition 2.7 ([26]) Let $g:\mathbb{R}^m\to\mathbb{R}\cup\{+\infty\}$ be a proper and lower semicontinuous function. The gradient of g is D-Lipschitz if there exists $L>0$ satisfying

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

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

$$\left|\langle \nabla g(x)-\nabla g(y), x-y\rangle\right|\le \|\nabla g(x)-\nabla g(y)\|\|x-y\|,$$

which, combined with Definition 2.7, yields

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

Using the conclusion in Lemma 2.1, the above inequality is equivalent to

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

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

Thus,

$$\langle (L\nabla h(x)-\nabla g(x))-(L\nabla h(y)-\nabla g(y)), x-y\rangle \ge 0,$$

$$\langle (L\nabla h(x)+\nabla g(x))-(L\nabla h(y)+\nabla g(y)), x-y\rangle \ge 0. \quad (2.3)$$

According to inequality (2.3), the functions $Lh+g$ and $Lh-g$ are convex functions since their gradients are monotone on the set C. Therefore, the D-Lipschitz continuity property of the gradient of the function g is a sufficient condition for the function pair (g,h) to be L-smooth adaptable. In this paper, given the complexity of the iterative scheme of ADMM, we need to assume that the gradient of g is D-Lipschitz continuous.

Remark 2.4 Indeed, the D-Lipschitz continuity property is a sort of Lipschitz-like gradient property of the function $g(\cdot)$ with respect to the Bregman distance, which reduces to gradient Lipschitz continuity of function $g(\cdot)$ when $h(\cdot)=\frac{1}{2}\|\cdot\|^2$.

Definition 2.8 [18] Let $f:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be a proper lower semicontinuous function.

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

$$\lim_{y\neq x}\inf_{y\to x}\frac{f(y)-f(x)-\langle x^*, y-x\rangle}{\|y-x\|}\ge 0.$$

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

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

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

Remark 2.5 From the above definition, we note that

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

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

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

$$0\in\partial f(x). \quad (2.4)$$

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

$x \in dom(f)$.

A point satisfying (2.4) is called a critical point or a stationary point. The critical points set of f is denoted by $\operatorname{crit} f$.

Now, we recall an important property of subdifferential calculus.

Lemma 2.3 [34] Suppose that $F(x,y) = f(x) + g(x)$, where $f:\mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $g:\mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ are proper lower semicontinuous functions. Then for all $(x,y) \in dom(F) = dom(f) \times dom(g)$, we have

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

Definition 2.9 ([34], Kurdyka-Lojasiewicz inequality) Let $f:\mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. For $-\infty < \eta_1 < \eta_2 \le +\infty$, 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 $x^* \in dom(\partial f)$ if there exist $\eta \in (0,+\infty]$, a neighbourhood U of x^*, and a continuous concave function $\varphi:[0,\eta) \to \mathbb{R}_+$, 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

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

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 2.10 ([35], Kurdyka-Lojasiewicz function) If $f(\cdot)$ satisfies the KL property at each point of $dom(\partial f)$, then $f(\cdot)$ is called a KL function.

Lemma 2.4 ([36], Uniformized KL property) Let Ω be a compact set and $f:\mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ 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 $\epsilon > 0, \eta > 0$, and $\varphi \in \Phi_\eta$ such that for all $\bar{x} \in \Omega$ and for all x in the following intersection:

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

one has

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

Definition 2.11 We say that (x^*, y^*, λ^*) is a critical point of the Augmented Lagrangian Function with Bregman distance $\mathcal{L}_\beta^h(\cdot)$ (3.3) if it satisfies

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

3. Convergence Analysis

To ensure that the Generalized Bregman ADMM (1.6) is well-defined and generates an infinite iterative sequence $\left\{\left(x^k, y^k, \lambda^k\right)\right\}$, we assume that the two minimization subproblems in (1.6) have solutions throughout the analysis. The optimality conditions for (1.6) are:

$$\begin{cases} 0 \in \partial f\left(x^{k+1}\right) - A^{\mathrm{T}}\lambda^k + \beta A^{\top}\left(Ax^{k+1} + y^k - b\right) \\ 0 = \nabla g\left(y^{k+1}\right) - \lambda^k + \beta\left(\nabla h\left(\alpha\left(Ax^{k+1} - b\right)\right) - \nabla h\left(-y^{k+1} - (\alpha - 1)y^k\right)\right) \\ \lambda^{k+1} = \lambda^k - \beta\left(\nabla h\left(\alpha\left(Ax^{k+1} - b\right)\right) - \nabla h\left(-y^{k+1} - (\alpha - 1)y^k\right)\right) \end{cases} \tag{3.1}$$

In terms of rearrangement of (3.1), it is equivalent to the following relation:

$$\begin{cases} A^{\mathrm{T}}\lambda^k - \beta A^{\top}\left(Ax^{k+1} + y^k - b\right) \in \partial f\left(x^{k+1}\right) & \text{(3.2a)} \\ \lambda^k - \beta\left(\nabla h\left(\alpha\left(Ax^{k+1} - b\right)\right) - \nabla h\left(-y^{k+1} - (\alpha - 1)y^k\right)\right) = \nabla g\left(y^{k+1}\right) & \text{(3.2b)} \\ \lambda^{k+1} = \lambda^k - \beta\left(\nabla h\left(\alpha\left(Ax^{k+1} - b\right)\right) - \nabla h\left(-y^{k+1} - (\alpha - 1)y^k\right)\right). & \text{(3.2c)} \end{cases}$$

To analyze the Generalized Bregman ADMM (1.6), we make the following basic assumptions. **Assumption A.** Assuming that $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is a proper lower semicontinuous function, $g : \mathbb{R}^m \to \mathbb{R}$ is a continuously differentiable function with ∇g being D-Lipschitz continuous and $h \in \mathcal{G}(C)$ is a twice differentiable function function on $C = int(dom(h))$, 1-strong-convex, and ∇h is Lipschitz continuous with L_h on any bounded subset of $\mathbb{R}^m$. Assume the following conditions hold:

i) if $\alpha \in \left(\frac{L_h + L_h^2}{1 + L_h + L_h^2}, 1\right]$, then $\beta \geq \frac{7LL_h^3 + \sqrt{49L^2L_h^6 + 16L^2L_h^3}}{\alpha\left(4 + 4L_h + 4L_h^2\right) - \left(4L_h + 4L_h^2\right)}$ which implies

$$\delta = \left[\frac{2\beta}{L_h} - \frac{2\beta(1-\alpha)(1+L_h)}{\alpha} - \frac{7LL_h^2}{\alpha} - \frac{2L^2L_h^2}{\alpha\beta}\right] \geq 0$$

ii) if $\alpha \in \left[1, \frac{-L_h - L_h^2}{1 - L_h - L_h^2}\right)$, then $\beta \geq \frac{(\alpha+6)LL_h^3 + \sqrt{\left((\alpha+6)LL_h^3\right)^2 + 16\alpha L^2L_h^3}}{\alpha\left(4 - 4L_h - 4L_h^2\right) + \left(4L_h + 4L_h^2\right)}$

which implies

$$\delta = \left[\frac{2\beta}{L_h} - \frac{2\beta(\alpha-1)(1+L_h)}{\alpha} - \frac{(\alpha+6)LL_h^2}{\alpha} - \frac{2L^2L_h^2}{\alpha\beta}\right] \geq 0$$

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

The Bregman augmented Lagrangian function of problem (1.1) is defined by

$$\mathcal{L}_\beta^h(x, y, \lambda) = f(x) + g(y) - \langle\lambda, Ax + y - b\rangle + \frac{\beta}{2}\|Ax + y - b\|^2 + \beta D_h(-y, Ax - b). \tag{3.3}$$

Here, λ is the Lagrangian multiplier associated with the linear constraints, and $\beta > 0$ is the penalty parameter. Moreover, we set

$$\hat{\mathcal{L}}_\beta^h(x,y,\lambda;\alpha,v) := f(x)+g(y)-\langle\lambda, Ax+y-b\rangle+\frac{\beta}{2}\|Ax+y-b\|^2 \\ +\beta D_h(-y-(\alpha-1)v,\alpha(Ax-b)). \tag{3.4}$$

Now, we begin our analysis with the following technical lemma.

Lemma 3.1 Let $\{w^k=(x^k,y^k,\lambda^k)\}$ be the sequence generated by the Generalized Bregman ADMM (1.6), which is assumed to be bounded. Then we have

$$\mathcal{L}_\beta^h(w^k)-\mathcal{L}_\beta^h(w^{k+1}) \ge \delta D_h(y^k,y^{k+1}). \tag{3.5}$$

Proof. From the definition of $\hat{\mathcal{L}}_\beta^h$ in (3.4), it follows that

$$\begin{aligned}
&\hat{\mathcal{L}}_\beta^h(x^{k+1},y^k,\lambda^k;\alpha,y^k)-\hat{\mathcal{L}}_\beta^h(x^{k+1},y^{k+1},\lambda^k;\alpha,y^k)\\
&= f(x^{k+1})+g(y^k)-\langle\lambda^k, Ax^{k+1}+y^k-b\rangle+\beta D_h(-\alpha y^k,\alpha(Ax^{k+1}-b))\\
&\quad+\frac{\beta}{2}\|Ax^{k+1}+y^k-b\|^2-\Big[f(x^{k+1})+g(y^{k+1})-\langle\lambda^k, Ax^{k+1}+y^{k+1}-b\rangle\\
&\quad+\beta D_h(-y^{k+1}-(\alpha-1)y^k,\alpha(Ax^{k+1}-b))+\frac{\beta}{2}\|Ax^{k+1}+y^{k+1}-b\|^2\Big]\\
&= g(y^k)-g(y^{k+1})-\langle\lambda^k, y^k-y^{k+1}\rangle+\beta D_h(-\alpha y^k,\alpha(Ax^{k+1}-b))\\
&\quad+\frac{\beta}{2}\|Ax^{k+1}+y^k-b\|^2-\beta D_h(-y^{k+1}-(\alpha-1)y^k,\alpha(Ax^{k+1}-b))\\
&\quad-\frac{\beta}{2}\|Ax^{k+1}+y^{k+1}-b\|^2
\end{aligned} \tag{3.6}$$

Given that the gradient of the function g is D-Lipschitz continuous within $int(dom(h))$, it can be inferred that the function pair (g,h) possesses L-smooth adaptability. Consequently, by referring to Lemma2.2, we are able to deduce a certain result

$$g(y^k)-g(y^{k+1}) \ge \langle\nabla g(y^{k+1}), y^k-y^{k+1}\rangle - LD_h(y^k,y^{k+1}). \tag{3.7}$$

Utilizing the optimal condition (3.2) $\lambda^{k+1}=\nabla g(y^{k+1})$, and by inequality (3.7) into identity (3.6), we achieve a certain outcome

$$\begin{aligned}
&\hat{\mathcal{L}}_\beta^h(x^{k+1},y^k,\lambda^k;\alpha,y^k)-\hat{\mathcal{L}}_\beta^h(x^{k+1},y^{k+1},\lambda^k;\alpha,y^k)\\
&= \langle\lambda^{k+1}-\lambda^k, y^k-y^{k+1}\rangle - LD_h(y^k,y^{k+1})+\frac{\beta}{2}\|Ax^{k+1}+y^k-b\|^2\\
&\quad-\frac{\beta}{2}\|Ax^{k+1}+y^{k+1}-b\|^2+\beta D_h(-\alpha y^k,\alpha(Ax^{k+1}-b))\\
&\quad-\beta D_h(-y^{k+1}-(\alpha-1)y^k,\alpha(Ax^{k+1}-b))\\
&\quad-\beta D_h(-\alpha y^k,-y^{k+1}-(\alpha-1)y^k)+\beta D_h(-\alpha y^k,-y^{k+1}-(\alpha-1)y^k)\\
&\ge -LD_h(y^k,y^{k+1})+\beta D_h(-\alpha y^k,-y^{k+1}-(\alpha-1)y^k)+\frac{\beta}{2}\|y^k-y^{k+1}\|^2\\
&\quad+\beta\langle Ax^{k+1}+y^{k+1}-b, y^k-y^{k+1}\rangle
\end{aligned} \tag{3.8}$$

In which the final inequality is established by applying the three-point identity and (3.9).

$$\begin{aligned}&\frac{\beta}{2}\left\|Ax^{k+1}+y^{k}-b\right\|^{2}-\frac{\beta}{2}\left\|Ax^{k+1}+y^{k+1}-b\right\|^{2}\\&=\frac{\beta}{2}\left\|y^{k}-y^{k+1}\right\|^{2}+\beta\left\langle Ax^{k+1}+y^{k+1}-b,y^{k}-y^{k+1}\right\rangle\end{aligned}\tag{3.9}$$

Subsequently, we proceed to estimate the remaining terms.

$$\begin{aligned}&\left(\hat{\mathcal{L}}_{\beta}^{h}\left(x^{k+1},y^{k},\lambda^{k};1,\cdot\right)-\hat{\mathcal{L}}_{\beta}^{h}\left(x^{k+1},y^{k},\lambda^{k};\alpha,y^{k}\right)\right)\\&+\left(\hat{\mathcal{L}}_{\beta}^{h}\left(x^{k+1},y^{k+1},\lambda^{k};\alpha,y^{k}\right)-\hat{\mathcal{L}}_{\beta}^{h}\left(x^{k+1},y^{k+1},\lambda^{k+1};1,\cdot\right)\right)\end{aligned}$$

Indeed,

$$\begin{aligned}&\hat{\mathcal{L}}_{\beta}^{h}\left(x^{k+1},y^{k},\lambda^{k};1,\cdot\right)-\hat{\mathcal{L}}_{\beta}^{h}\left(x^{k+1},y^{k},\lambda^{k};\alpha,y^{k}\right)\\&=f\left(x^{k+1}\right)+g\left(y^{k}\right)-\left\langle\lambda^{k},Ax^{k+1}+y^{k}-b\right\rangle+\beta D_{h}\left(-y^{k},Ax^{k+1}-b\right)\\&\quad-\left[f\left(x^{k+1}\right)+g\left(y^{k}\right)-\left\langle\lambda^{k},Ax^{k+1}+y^{k}-b\right\rangle+\beta D_{h}\left(-\alpha y^{k},\alpha\left(Ax^{k+1}-b\right)\right)\right]\\&=\beta D_{h}\left(-y^{k},Ax^{k+1}-b\right)-\beta D_{h}\left(-\alpha y^{k},\alpha\left(Ax^{k+1}-b\right)\right)\end{aligned}$$

and

$$\begin{aligned}&\hat{\mathcal{L}}_{\beta}^{h}\left(x^{k+1},y^{k+1},\lambda^{k};\alpha,y^{k}\right)-\hat{\mathcal{L}}_{\beta}^{h}\left(x^{k+1},y^{k+1},\lambda^{k+1};1,\cdot\right)\\&=f\left(x^{k+1}\right)+g\left(y^{k+1}\right)-\left\langle\lambda^{k},Ax^{k+1}+y^{k+1}-b\right\rangle+\beta D_{h}\left(-y^{k+1}-(\alpha-1)y^{k},\alpha\left(Ax^{k+1}-b\right)\right)\\&\quad-\left[f\left(x^{k+1}\right)+g\left(y^{k+1}\right)-\left\langle\lambda^{k+1},Ax^{k+1}+y^{k+1}-b\right\rangle+\beta D_{h}\left(-y^{k+1},Ax^{k+1}-b\right)\right]\\&=\left\langle\lambda^{k+1}-\lambda^{k},Ax^{k+1}+y^{k+1}-b\right\rangle+\beta D_{h}\left(-y^{k+1}-(\alpha-1)y^{k},\alpha\left(Ax^{k+1}-b\right)\right)\\&\quad-\beta D_{h}\left(-y^{k+1},Ax^{k+1}-b\right)\end{aligned}$$

By merging the two equalities, we obtain:

$$\begin{aligned}&\left(\hat{\mathcal{L}}_{\beta}^{h}\left(x^{k+1},y^{k},\lambda^{k};1,\cdot\right)-\hat{\mathcal{L}}_{\beta}^{h}\left(x^{k+1},y^{k},\lambda^{k};\alpha,y^{k}\right)\right)\\&+\left(\hat{\mathcal{L}}_{\beta}^{h}\left(x^{k+1},y^{k+1},\lambda^{k};\alpha,y^{k}\right)-\hat{\mathcal{L}}_{\beta}^{h}\left(x^{k+1},y^{k+1},\lambda^{k+1};1,\cdot\right)\right)\\&=\left\langle\lambda^{k+1}-\lambda^{k},Ax^{k+1}+y^{k+1}-b\right\rangle+\beta D_{h}\left(-y^{k},Ax^{k+1}-b\right)-\beta D_{h}\left(-y^{k+1},Ax^{k+1}-b\right)\\&\quad+\beta D_{h}\left(-y^{k+1}-(\alpha-1)y^{k},\alpha\left(Ax^{k+1}-b\right)\right)-\beta D_{h}\left(-\alpha y^{k},\alpha\left(Ax^{k+1}-b\right)\right)\\&=\beta D_{h}\left(-y^{k},-y^{k+1}\right)-\beta D_{h}\left(-\alpha y^{k},-y^{k+1}-(\alpha-1)y^{k}\right)\\&\quad+\left\langle\lambda^{k+1}-\lambda^{k},Ax^{k+1}+y^{k}-b\right\rangle+\beta\left\langle\nabla h\left(Ax^{k+1}-b\right)-\nabla h\left(-y^{k+1}\right),y^{k}-y^{k+1}\right\rangle\end{aligned}\tag{3.10}$$

The final equation is derived by employing the three-point equation (3.11) (3.121) and incorporating the optimality conditions (3.2c).

$$\begin{aligned}&\beta D_{h}\left(-y^{k}-(\alpha-1)y^{k},\alpha\left(Ax^{k+1}-b\right)\right)-\beta D_{h}\left(-y^{k+1}-(\alpha-1)y^{k},\alpha\left(Ax^{k+1}-b\right)\right)\\&=\beta D_{h}\left(-\alpha y^{k},-y^{k+1}-(\alpha-1)y^{k}\right)\\&\quad+\beta\left\langle\nabla h\left(\alpha\left(Ax^{k+1}-b\right)\right)-\nabla h\left(-y^{k+1}-(\alpha-1)y^{k}\right),y^{k}-y^{k+1}\right\rangle\\&=\beta D_{h}\left(-\alpha y^{k},-y^{k+1}-(\alpha-1)y^{k}\right)+\left\langle\lambda^{k}-\lambda^{k+1},y^{k}-y^{k+1}\right\rangle\end{aligned}\tag{3.11}$$

$$\begin{aligned}&\beta D_h\left(-y^{k+1},Ax^{k+1}-b\right)-\beta D_h\left(-y^k,Ax^{k+1}-b\right)\\&=\beta D_h\left(-y^{k+1},-y^k\right)+\beta\left\langle\nabla h\left(Ax^{k+1}-b\right)-\nabla h\left(-y^k\right),y^{k+1}-y^k\right\rangle\\&=-\beta D_h\left(-y^k,-y^{k+1}\right)-\beta\left\langle\nabla h\left(Ax^{k+1}-b\right)-\nabla h\left(-y^{k+1}\right),y^k-y^{k+1}\right\rangle\end{aligned}\tag{3.12}$$

Given that x^{k+1} is the minimizer of $\mathcal{L}_\beta^h\left(x,y^k,\lambda^k\right)$ with respect to the variable x, we have:

$$\hat{\mathcal{L}}_\beta^h\left(x^k,y^k,\lambda^k;1,\cdot\right)-\hat{\mathcal{L}}_\beta^h\left(x^{k+1},y^k,\lambda^k;1,\cdot\right)=\mathcal{L}_\beta^h\left(x^k,y^k,\lambda^k\right)-\mathcal{L}_\beta^h\left(x^{k+1},y^k,\lambda^k\right)\geq 0\tag{3.13}$$

Obsevre that

$$\begin{aligned}\mathcal{L}_\beta^h\left(\omega^k\right)-\mathcal{L}_\beta^h\left(\omega^{k+1}\right)&=\mathcal{L}_\beta^h\left(x^k,y^k,\lambda^k\right)-\mathcal{L}_\beta^h\left(x^{k+1},y^{k+1},\lambda^{k+1}\right)\\&=\hat{\mathcal{L}}_\beta^h\left(x^k,y^k,\lambda^k;1,\cdot\right)-\hat{\mathcal{L}}_\beta^h\left(x^{k+1},y^{k+1},\lambda^{k+1};1,\cdot\right)\\&=\left(\hat{\mathcal{L}}_\beta^h\left(x^k,y^k,\lambda^k;1,\cdot\right)-\hat{\mathcal{L}}_\beta^h\left(x^{k+1},y^k,\lambda^k;1,\cdot\right)\right)\\&\quad+\left(\hat{\mathcal{L}}_\beta^h\left(x^{k+1},y^k,\lambda^k;1,\cdot\right)-\hat{\mathcal{L}}_\beta^h\left(x^{k+1},y^k,\lambda^k;\alpha,y^k\right)\right)\\&\quad+\left(\hat{\mathcal{L}}_\beta^h\left(x^{k+1},y^k,\lambda^k;\alpha,y^k\right)-\hat{\mathcal{L}}_\beta^h\left(x^{k+1},y^{k+1},\lambda^k;\alpha,y^k\right)\right)\\&\quad+\left(\hat{\mathcal{L}}_\beta^h\left(x^{k+1},y^{k+1},\lambda^k;\alpha,y^k\right)-\hat{\mathcal{L}}_\beta^h\left(x^{k+1},y^{k+1},\lambda^{k+1};1,\cdot\right)\right)\end{aligned}$$

Consequently, by adding up the inequalities (3.8), (3.10) and (3.13), we arrive at the conclusion that:

$$\begin{aligned}&\mathcal{L}_\beta^h\left(\omega^k\right)-\mathcal{L}_\beta^h\left(\omega^{k+1}\right)\\&\geq -LD_h\left(y^k,y^{k+1}\right)+\left\langle\lambda^{k+1}-\lambda^k,Ax^{k+1}+y^k-b\right\rangle\\&\quad+\frac{\beta}{2}\left\|y^k-y^{k+1}\right\|^2+\beta\left\langle Ax^{k+1}+y^{k+1}-b,y^k-y^{k+1}\right\rangle\\&\quad+\beta D_h\left(-y^k,-y^{k+1}\right)+\beta\left\langle\nabla h\left(Ax^{k+1}-b\right)-\nabla h\left(-y^{k+1}\right),y^k-y^{k+1}\right\rangle\\&\geq -LD_h\left(y^k,y^{k+1}\right)-\left\|\lambda^{k+1}-\lambda^k\right\|\cdot\left\|Ax^{k+1}+y^k-b\right\|+\frac{\beta}{2}\left\|y^k-y^{k+1}\right\|^2\\&\quad+\beta D_h\left(-y^k,-y^{k+1}\right)-\beta\cdot\left(L_h+1\right)\left\|Ax^{k+1}+y^{k+1}-b\right\|\cdot\left\|y^{k+1}-y^{k+1}\right\|\end{aligned}\tag{3.14}$$

Considering the 1-strong convexity and L_h -smoothness of the function $h(\cdot)$, we can respectively derive the following inequalities:

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

$$\frac{1}{2}\left\|y^k-y^{k+1}\right\|^2\leq h\left(-y^k\right)-h\left(-y^{k+1}\right)-\left\langle\nabla h\left(-y^{k+1}\right),-y^k+y^{k+1}\right\rangle=D_h\left(-y^k,-y^{k+1}\right)$$

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

Given that h is 1-strong convex, we can deduce that:

$$\begin{aligned}&\left\|\nabla h\left(\alpha\left(Ax^{k+1}-b\right)\right)-\nabla h\left(-y^{k+1}-(\alpha-1)y^k\right)\right\|\cdot\left\|\alpha\left(Ax^{k+1}-b\right)-\left(-y^{k+1}-(\alpha-1)y^k\right)\right\|\\&\geq\left\langle\nabla h\left(\alpha\left(Ax^{k+1}-b\right)\right)-\nabla h\left(-y^{k+1}-(\alpha-1)y^k\right),\alpha\left(Ax^{k+1}-b\right)-\left(-y^{k+1}-(\alpha-1)y^k\right)\right\rangle\\&\geq\left\|\alpha\left(Ax^{k+1}-b\right)-\left(-y^{k+1}-(\alpha-1)y^k\right)\right\|^2\end{aligned}\tag{3.16}$$

Hence, from the aforementioned equation (3.15), we are able to derive the following:

$$\begin{aligned}&\left\|\nabla h\left(\alpha\left(Ax^{k+1}-b\right)\right)-\nabla h\left(-y^{k+1}-(\alpha-1)y^k\right)\right\|\\&\geq\left\|\alpha\left(Ax^{k+1}-b\right)-\left(-y^{k+1}-(\alpha-1)y^k\right)\right\|\\&=\left\|\alpha\left(Ax^{k+1}+y^k-b\right)+\left(y^{k+1}-y^k\right)\right\|\end{aligned}\tag{3.16}$$

Combining optimimal condition (3.2c)

$$\nabla h\left(\alpha\left(Ax^{k+1}-b\right)\right)-\nabla h\left(-y^{k+1}-(\alpha-1)y^k\right)=\frac{\lambda^k-\lambda^{k+1}}{\beta}\tag{3.17}$$

Additionally, by applying the triangle inequality, we can know that:

$$\left\|\alpha\left(Ax^{k+1}+y^k-b\right)\right\|\leq\left\|y^k-y^{k+1}\right\|+\left\|\alpha\left(Ax^{k+1}+y^k-b\right)+\left(y^{k+1}-y^k\right)\right\|\tag{3.18}$$

So combining (3.6), (3.17) and (3.18) we can derive the following result

$$\begin{aligned}\left\|Ax^{k+1}+y^k-b\right\|&\leq\frac{\left\|y^k-y^{k+1}\right\|+\left\|\alpha\left(Ax^{k+1}+y^k-b\right)+\left(y^{k+1}-y^k\right)\right\|}{\alpha}\\&\leq\frac{\left\|y^k-y^{k+1}\right\|}{\alpha}+\frac{\left\|\lambda^k-\lambda^{k+1}\right\|}{\alpha\cdot\beta}\end{aligned}\tag{3.19}$$

And by the same token, we can get an inequality for $Ax^{k+1}+y^{k+1}-b$

$$\begin{aligned}&\left\|Ax^{k+1}+y^{k+1}-b\right\|\\&\leq\frac{\left\|(1-\alpha)\left(y^k-y^{k+1}\right)\right\|+\left\|\alpha\left(Ax^{k+1}+y^k-b\right)+\left(y^{k+1}-y^k\right)\right\|}{\alpha}\\&\leq\frac{\left\|(1-\alpha)\left(y^k-y^{k+1}\right)\right\|}{\alpha}+\frac{\left\|\lambda^k-\lambda^{k+1}\right\|}{\alpha\cdot\beta}\end{aligned}\tag{3.20}$$

Next, we declare that

$$\left\|\nabla g\left(y^{k+1}\right)-\nabla g\left(y^k\right)\right\|\leq LL_h\cdot\left\|y^{k+1}-y^k\right\|.\tag{3.21}$$

To prove (3.21), we consider two cases. When $y^k=y^{k+1}$, (3.21) holds trivially. Now, we assume $y^{k+1}\neq y^k$. Since $\nabla g(\cdot)$ is D-Lipschitz, we have

$$\begin{aligned}\left\|\nabla g\left(y^{k+1}\right)-\nabla g\left(y^k\right)\right\|&\leq L\cdot\frac{D_h\left(y^{k+1},y^k\right)+D_h\left(y^k,y^{k+1}\right)}{\left\|y^{k+1}-y^k\right\|}\\&=L\cdot\frac{\left\langle\nabla h\left(y^k\right)-\nabla h\left(y^{k+1}\right),y^k-y^{k+1}\right\rangle}{\left\|y^{k+1}-y^k\right\|}\end{aligned}$$

$$\leq L \cdot \frac{\left\|\nabla h\left(y^k\right)-\nabla h\left(y^{k+1}\right)\right\| \cdot \left\|y^k - y^{k+1}\right\|}{\left\|y^{k+1}-y^k\right\|}$$

$$\leq L \cdot \frac{L_h \left\|y^k - y^{k+1}\right\|^2}{\left\|y^{k+1}-y^k\right\|} = LL_h \cdot \left\|y^{k+1}-y^k\right\|,$$

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

$$\left\|\nabla h\left(y^{k+1}\right)-\nabla h\left(y^k\right)\right\| \leq L_h \left\|y^{k+1}-y^k\right\|.$$

Since $\lambda^{k+1}=\nabla g\left(y^{k+1}\right)$, (3.21) becomes

$$\left\|\lambda^{k+1}-\lambda^k\right\| \leq LL_h \left\|y^{k+1}-y^k\right\|. \tag{3.22}$$

And we also know

$$\left\|\lambda^k-\lambda^{k+1}\right\|^2 = \left\|\nabla g\left(y^k\right)-\nabla g\left(y^{k+1}\right)\right\|^2 \leq L^2 L_h^2 \left\|y^{k+1}-y^k\right\|^2 \tag{3.23}$$

And

$$\left\|\lambda^k-\lambda^{k+1}\right\|\left\|y^{k+1}-y^k\right\| = \left\|\nabla g\left(y^k\right)-\nabla g\left(y^{k+1}\right)\right\|\left\|y^{k+1}-y^k\right\| \leq LL_h \left\|y^{k+1}-y^k\right\|^2$$

Since $\alpha > 1$ and $\alpha < 1$ will affect the result of our calculation, we will classify and discuss them based on the different values of α.

i) if $\alpha \in \left(\frac{L_h+L_h^2}{1+L_h+L_h^2}, 1\right]$, then we put (3.19) (3.20) and (3.22) into (3.14), we can get

$$\begin{aligned}
&\mathcal{L}_\beta^h\left(\omega^k\right)-\mathcal{L}_\beta^h\left(\omega^{k+1}\right) \\
&\geq -LD_h\left(y^k, y^{k+1}\right)-\frac{\left\|y^k-y^{k+1}\right\|\left\|\lambda^k-\lambda^{k+1}\right\|}{\alpha}-\frac{\left\|\lambda^k-\lambda^{k+1}\right\|^2}{\alpha\beta}+\frac{\beta}{2}\left\|y^k-y^{k+1}\right\|^2 \\
&\quad +\frac{\beta}{L_h} D_h\left(y^k, y^{k+1}\right)-\beta\left(L_h+1\right)\left[\frac{\left\|(1-\alpha)\left(y^k-y^{k+1}\right)\right\|\left\|y^k-y^{k+1}\right\|}{\alpha}+\frac{\left\|\lambda^k-\lambda^{k+1}\right\|\left\|y^k-y^{k+1}\right\|}{\alpha\beta}\right] \\
&\geq -LD_h\left(y^k, y^{k+1}\right)+\frac{\beta}{L_h} D_h\left(y^k, y^{k+1}\right)+\frac{\beta}{2}\left\|y^k-y^{k+1}\right\|^2-\frac{2LL_h}{\alpha}\left\|y^k-y^{k+1}\right\|^2 \\
&\quad -\frac{L^2L_h^2\left\|y^k-y^{k+1}\right\|^2}{\alpha\beta}-\beta\frac{\left(1+L_h\right)(1-\alpha)\left\|y^k-y^{k+1}\right\|^2}{\alpha}-\frac{LL_h^2\left\|y^k-y^{k+1}\right\|^2}{\alpha} \\
&\geq \left(\frac{2\beta}{L_h}-L\right)D_h\left(y^k, y^{k+1}\right)+\left[-\frac{\beta\cdot(1-\alpha)\left(1+L_h\right)}{\alpha}-\frac{2LL_h}{\alpha}-\frac{LL_h^2}{\alpha}-\frac{L^2\cdot L_h^2}{\alpha\beta}\right]\left\|y^k-y^{k+1}\right\|^2 \\
&\geq \left[\frac{2\beta}{L_h}-\frac{2\beta(1-\alpha)\left(1+L_h\right)}{\alpha}-\frac{\alpha L+4LL_h+2LL_h^2}{\alpha}-\frac{2L^2L_h^2}{\alpha\beta}\right]D_h\left(y^k, y^{k+1}\right) \\
&\geq \left[\frac{2\beta}{L_h}-\frac{2\beta(1-\alpha)\left(1+L_h\right)}{\alpha}-\frac{7LL_h^2}{\alpha}-\frac{2L^2L_h^2}{\alpha\beta}\right]D_h\left(y^k, y^{k+1}\right)
\end{aligned} \tag{3.24}$$

ii) if $\alpha \in \left[1, \frac{-L_h-L_h^2}{1-L_h-L_h^2}\right)$, and from the formula above we can get

$$
\begin{aligned}
&\mathcal{L}_\beta^h\left(\omega^k\right)-\mathcal{L}_\beta^h\left(\omega^{k+1}\right)\\
&\geq -LD_h\left(y^k,y^{k+1}\right)-\frac{\left\|y^k-y^{k+1}\right\|\left\|\lambda^k-\lambda^{k+1}\right\|}{\alpha}-\frac{\left\|\lambda^k-\lambda^{k+1}\right\|^2}{\alpha\beta}+\frac{\beta}{2}\left\|y^k-y^{k+1}\right\|^2\\
&\quad+\frac{\beta}{L_h}D_h\left(y^k,y^{k+1}\right)-\beta\left(L_h+1\right)\left[\frac{\left\|(1-\alpha)\left(y^k-y^{k+1}\right)\right\|\left\|y^k-y^{k+1}\right\|}{\alpha}+\frac{\left\|\lambda^k-\lambda^{k+1}\right\|\left\|y^k-y^{k+1}\right\|}{\alpha\beta}\right]\\
&\geq -LD_h\left(y^k,y^{k+1}\right)+\frac{\beta}{L_h}D_h\left(y^k,y^{k+1}\right)+\frac{\beta}{2}\left\|y^k-y^{k+1}\right\|^2-\frac{2LL_h}{\alpha}\left\|y^k-y^{k+1}\right\|^2\\
&\quad-\frac{L^2L_h^2\left\|y^k-y^{k+1}\right\|^2}{\alpha\beta}-\beta\frac{(1+L_h)(\alpha-1)\left\|y^k-y^{k+1}\right\|^2}{\alpha}-\frac{LL_h^2\left\|y^k-y^{k+1}\right\|^2}{\alpha} \qquad (3.25)\\
&\geq\left(\frac{2\beta}{L_h}-L\right)D_h\left(y^k,y^{k+1}\right)+\left[-\frac{\beta\cdot(\alpha-1)(1+L_h)}{\alpha}-\frac{2LL_h}{\alpha}-\frac{LL_h^2}{\alpha}-\frac{L^2\cdot L_h^2}{\alpha\beta}\right]\left\|y^k-y^{k+1}\right\|^2\\
&\geq\left[\frac{2\beta}{L_h}-\frac{2\beta(\alpha-1)(1+L_h)}{\alpha}-\frac{\alpha L+4LL_h+2LL_h^2}{\alpha}-\frac{2L^2L_h^2}{\alpha\beta}\right]D_h\left(y^k,y^{k+1}\right)\\
&\geq\left[\frac{2\beta}{L_h}-\frac{2\beta(\alpha-1)(1+L_h)}{\alpha}-\frac{(\alpha+6)LL_h^2}{\alpha}-\frac{2L^2L_h^2}{\alpha\beta}\right]D_h\left(y^k,y^{k+1}\right)
\end{aligned}
$$

The proof is complete.

Lemma 3.2 Let $\left\{w^k=\left(x^k,y^k,\lambda^k\right)\right\}$ be the sequence generated by the Generalized Bregman ADMM (1.6), which is assumed to be bounded. Then we have

$$\sum_{k=0}^{+\infty}\left\|w^{k+1}-w^k\right\|^2<+\infty. \qquad (3.26)$$

Proof: Given that the sequence $\left\{w^k\right\}$ is bounded, it follows that there exists a subsequence $\left\{w^{k_j}\right\}$ such that $w^{k_j}\to w^*$. As $f(\cdot)$ is lower semicontinuous and $g(\cdot)$ is continuous, it can be deduced that the function $\mathcal{L}_\beta^h(\cdot)$ is also lower semicontinuous. Therefore,

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

As a result, $\left\{\mathcal{L}_\beta^h\left(w^{k_j}\right)\right\}$ is bounded from below. Additionally, since $\left\{\mathcal{L}_\beta^h\left(w^k\right)\right\}$ is nonincreasing, it follows that $\left\{\mathcal{L}_\beta^h\left(w^{k_j}\right)\right\}$ is convergent. Moreover, $\left\{\mathcal{L}_\beta^h\left(w^k\right)\right\}$ is convergent, and $\mathcal{L}_\beta^h\left(w^k\right)\geq\mathcal{L}_\beta^h\left(w^*\right)$. According to equation (3.5), we have

$$\delta\cdot D_h\left(y^k,y^{k+1}\right)\leq\mathcal{L}_\beta^h\left(w^k\right)-\mathcal{L}_\beta^h\left(w^{k+1}\right).$$

Summing over $k=0,\cdots,n$, it follows

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

Since $\delta>0$, we have $\sum_{k=0}^{\infty}D_h\left(y^k,y^{k+1}\right)<+\infty$, which implies $\sum_{k=0}^{\infty}\left\|y^{k+1}-y^k\right\|^2<+\infty$. Hence, it follows from (3.22) that $\sum_{k=0}^{\infty}\left\|\lambda^{k+1}-\lambda^k\right\|^2<+\infty$. Recall that

$$\lambda^k = \lambda^{k-1} - \beta\left(\nabla h\left(\alpha\left(Ax^k - b\right)\right) - \nabla h\left(-y^k - (\alpha-1)y^{k-1}\right)\right),$$

$$\lambda^{k+1} = \lambda^k - \beta\left(\nabla h\left(\alpha\left(Ax^{k+1} - b\right)\right) - \nabla h\left(-y^{k+1} - (\alpha-1)y^k\right)\right)$$

Subtracting the first equality from the second equality, we obtain

$$\begin{aligned}\lambda^{k+1} - \lambda^k = \lambda^k - \lambda^{k-1} &+ \beta\left(\nabla h\left(\alpha\left(Ax^k - b\right)\right) - \nabla h\left(\alpha\left(Ax^{k+1} - b\right)\right)\right) \\ &+ \beta\left(\nabla h\left(-y^{k+1} - (\alpha-1)y^k\right) - \nabla h\left(-y^k - (\alpha-1)y^{k-1}\right)\right).\end{aligned}$$

Rearranging the above equation and taking the square of the ℓ_2-norm, it follows

$$\begin{aligned}&\left\|\beta\left(\nabla h\left(\alpha\left(Ax^{k+1} - b\right)\right) - \nabla h\left(\alpha\left(Ax^k - b\right)\right)\right)\right\|^2 \\ &= \left\|\lambda^{k+1} - \lambda^k - \left(\lambda^k - \lambda^{k-1}\right) - \beta\left(\nabla h\left(-y^{k+1} - (\alpha-1)y^k\right) - \nabla h\left(-y^k - (\alpha-1)y^{k-1}\right)\right)\right\|^2 \\ &\le 3\left(\left\|\lambda^{k+1} - \lambda^k\right\|^2 + \left\|\lambda^k - \lambda^{k-1}\right\|^2 + \beta^2\left\|\nabla h\left(-y^{k+1} - (\alpha-1)y^k\right) - \nabla h\left(-y^k - (\alpha-1)y^{k-1}\right)\right\|^2\right) \\ &\le 3\left(\left\|\lambda^{k+1} - \lambda^k\right\|^2 + \left\|\lambda^k - \lambda^{k-1}\right\|^2 + \beta^2 L_h^2\left\|y^{k+1} - y^k + (\alpha-1)\left(y^k - y^{k+1}\right)\right\|^2\right) \\ &\le 3(\left\|\lambda^{k+1} - \lambda^k\right\|^2 + \left\|\lambda^k - \lambda^{k-1}\right\|^2 + 2\beta^2 L_h^2\left\|y^{k+1} - y^k\right\|^2 + 2\left(\alpha-1\right)^2\beta^2 L_h^2\left\|y^k - y^{k+1}\right\|^2\right).\end{aligned} \tag{3.27}$$

On the other hand,

$$\beta^2\alpha^2\left\|Ax^{k+1} - Ax^k\right\|^2 \le \left\|\beta\left(\nabla h\left(\alpha\left(Ax^{k+1} - b\right)\right) - \nabla h\left(\alpha\left(Ax^k - b\right)\right)\right)\right\|^2. \tag{3.28}$$

From (ii) of Assumption 3.1, we note that

$$\left\|Ax^{k+1} - Ax^k\right\|^2 \ge M\left\|x^{k+1} - x^k\right\|^2. \tag{3.29}$$

Combining (3.27) - (3.29) together, we get

$$\begin{aligned}\alpha^2\beta^2 M\left\|x^{k+1} - x^k\right\|^2 &\le \alpha^2\beta^2\left\|Ax^{k+1} - Ax^k\right\|^2 \\ &\le \left\|\beta\left(\nabla h\left(\alpha\left(Ax^{k+1} - b\right)\right) - \nabla h\left(\alpha\left(Ax^k - b\right)\right)\right)\right\|^2 \\ &\le 3\left(\left\|\lambda^{k+1} - \lambda^k\right\|^2 + \left\|\lambda^k - \lambda^{k-1}\right\|^2 + 2\beta^2 L_h^2\left\|y^{k+1} - y^k\right\|^2\right. \\ &\quad \left. + 2\left(\alpha-1\right)^2\beta^2 L_h^2\left\|y^k - y^{k+1}\right\|^2\right)\end{aligned} \tag{3.30}$$

where $M > 0$. Then, (3.30) implies $\sum_{k=0}^{+\infty}\left\|x^{k+1} - x^k\right\|^2 < +\infty$. Thus, $\sum_{k=0}^{+\infty}\left\|w^{k+1} - w^k\right\|^2 < +\infty$. This completes the proof. □

Lemma 3.3 t $\left\{w^k = \left(x^k, y^k, \lambda^k\right)\right\}_k \in N$ be the sequence generated by the Generalized Bregman ADMM (1.3), which is assumed to be bounded. Furthermore, there exists $\eta > 0$ such that

$$d\left(0, \partial\mathcal{L}_\beta^h\left(w^{k+1}\right)\right) \le \eta\left\|y^{k+1} - y^k\right\|.$$

Proof: By definition of function $\mathcal{L}_\beta^h(\cdot)$, we have the following system of equations:

$$\begin{cases} \partial_x \mathcal{L}_\beta^h\left(\omega^{k+1}\right) = \partial f\left(x^{k+1}\right) - A^{\mathrm{T}}\lambda^{k+1} + \beta A^{\mathrm{T}}\left(Ax^{k+1} + y^{k+1} - b\right) + \beta A^{\mathrm{T}}\left\langle \nabla^2 h\left(Ax^{k+1} - b\right), Ax^{k+1} + y^{k+1} - b\right\rangle, \\ \partial_y \mathcal{L}_\beta^h\left(\omega^{k+1}\right) = \nabla g\left(y^{k+1}\right) - \lambda^{k+1} + \beta\left(Ax^{k+1} + y^{k+1} - b\right) + \beta\left(\nabla h\left(Ax^{k+1} - b\right) - \nabla h\left(-y^{k+1}\right)\right), \\ \partial_\lambda \mathcal{L}_\beta^h\left(\omega^{k+1}\right) = -\left(Ax^{k+1} + y^{k+1} - b\right). \end{cases} \tag{3.31}$$

Combining equation (3.31) with optimality condition (3.1), we obtain:

$$\begin{cases} A^{\mathrm{T}}\left(\lambda^k - \lambda^{k+1}\right) + \beta A^{\top}\left(y^{k+1} - y^k\right) + \beta A^{\mathrm{T}}\left\langle \nabla^2 h\left(Ax^{k+1} - b\right), Ax^{k+1} + y^{k+1} - b\right\rangle \in \partial_x \mathcal{L}_\beta^h\left(w^{k+1}\right), \\ \beta\left(Ax^{k+1} + y^{k+1} - b\right) + \beta\left(\nabla h\left(Ax^{k+1} - b\right) - \nabla h\left(-y^{k+1}\right)\right) \in \partial_y \mathcal{L}_\beta^h\left(w^{k+1}\right), \\ -\left(Ax^{k+1} + y^{k+1} - b\right) \in \partial_\lambda \mathcal{L}_\beta^h\left(w^{k+1}\right). \end{cases}$$

In addition, By the L_h -smoothness of the function $h(\cdot)$, we can deduce that

$$\left\|\nabla h\left(Ax^{k+1} - b\right) - \nabla h\left(-y^{k+1}\right)\right\| \le L_h \cdot \left\|Ax^{k+1} + y^{k+1} - b\right\|$$

From formula (3.20) we know the following result

$$\left\|Ax^{k+1} + y^{k+1} - b\right\| \le \frac{\left\|(1-\alpha)\left(y^k - y^{k+1}\right)\right\|}{\alpha} + \frac{\left\|\lambda^k - \lambda^{k+1}\right\|}{\alpha \cdot \beta}$$

In addition,

$$\begin{aligned} &\left\langle \nabla h^2\left(Ax^{k+1} - b\right), Ax^{k+1} + y^{k+1} - b\right\rangle \\ &\le \left\|\nabla h^2\left(Ax^{k+1} - b\right)\right\|\left\|Ax^{k+1} + y^{k+1} - b\right\| \le L_h\left\|Ax^{k+1} + y^{k+1} - b\right\|. \end{aligned}$$

Thus, if we set

$$\begin{aligned} &\left(\xi_1^{k+1}, \xi_2^{k+1}, \xi_3^{k+1}\right) \\ &:= \left(A^{\mathrm{T}}\left(\lambda^k - \lambda^{k+1}\right) + \beta A^{\top}\left(y^{k+1} - y^k\right) + \beta A^{\mathrm{T}}\left\langle \nabla^2 h\left(Ax^{k+1} - b\right), Ax^{k+1} + y^{k+1} - b\right\rangle,\right. \\ &\qquad \left.\beta\left(Ax^{k+1} + y^{k+1} - b\right) + \beta\left(\nabla h\left(Ax^{k+1} - b\right) - \nabla h\left(-y^{k+1}\right)\right), -\left(Ax^{k+1} + y^{k+1} - b\right)\right) \end{aligned}$$

Then it follows from lemma 2.3 that $\left(\xi_1^{k+1}, \xi_2^{k+1}, \xi_3^{k+1}\right) \in \partial \mathcal{L}\beta\left(\omega^{k+1}\right)$. Moreover, there exist $\eta_1, \eta_2 > 0$ such that

$$\left\|\left(\xi_1^{k+1}, \xi_2^{k+1}, \xi_3^{k+1}\right)\right\| \le \eta_1\left\|y^{k+1} - y^k\right\| + \eta_2\left\|\lambda^{k+1} - \lambda^k\right\|$$

Notice that, we can deduce from (3.22) that

$$\left\|\lambda^{k+1} - \lambda^k\right\| \le L \cdot L_n\left\|y^{k+1} - y^k\right\|$$

We define $\eta := \eta_1 + \eta_2 L L_n$, it follows from above... that

$$\begin{aligned} d\left(0, \partial \mathcal{L}_\beta^h\left(w^{k+1}\right)\right) &\le \left\|\left(\xi_1^{k+1}, \xi_2^{k+1}, \xi_3^{k+1}\right)\right\| \le \eta_1\left\|y^{k+1} - y^k\right\| + \eta_2\left\|\lambda^{k+1} - \lambda^k\right\| \\ &\le \left(\eta_1 + \eta_2 L L_n\right) \cdot \left\|y^{k+1} - y^k\right\| \\ &\le \eta\left\|y^{k+1} - y^k\right\| \end{aligned}$$

This completes the proof. □

Lemma 3.4 Let $\left\{w^k = \left(x^k, y^k, \lambda^k\right)\right\}$ be the sequence generated by the Generalized

Bregman ADMM (1.3), which is assumed 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

$$d\left(w^k, S\left(w^0\right)\right) \to 0, \text{ as } k \to +\infty;$$

ii) $S(w^0) \subset crit\,\mathcal{L}_\beta^h$, where $crit\,\mathcal{L}_\beta^h$ denotes the set of all stationary points of $\mathcal{L}_\beta^h$;

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

$$\inf_{k\in\mathbb{N}} \mathcal{L}_\beta^h\left(w^k\right) = \lim_{k\to+\infty} \mathcal{L}_\beta^h\left(w^k\right).$$

Proof: We proof the results item by item.

i) The item follows as an elementary consequence of the definition of limit points.

ii) For any fixed $(x^*, y^*, \lambda^*) \in S(w_0)$, then there exists a subsequence $\left\{\left(x^{k_j}, y^{k_j}, \lambda^{k_j}\right)\right\}_j \in N$ that converges to (x^*, y^*, λ^*). By the definition of the augmented Lagrangian function (3.3), the x-subproblem of (1.6) is equivalent to

$$x^{k+1} \in \arg\min_x \left\{\mathcal{L}_\beta^h\left(x, y^k, \lambda^k\right)\right\}$$

that means x^{k+1} is the global minimizer of $\mathcal{L}_\beta^h\left(x, y^k, \lambda^k\right)$ for the variable x, then it holds that

$$\mathcal{L}_\beta^h\left(x^{k+1}, y^k, \lambda^k\right) \le \mathcal{L}_\beta^h\left(x^*, y^k, \lambda^k\right). \tag{3.32}$$

On one hand, using (3.32) and the continuity of $\mathcal{L}_\beta^h(\cdot)$ with respect to y and λ, we have

$$\limsup_{j\to+\infty} \mathcal{L}_\beta^h\left(x^{k_j+1}, y^{k_j}, \lambda^{k_j}\right) = \limsup_{j\to+\infty} \mathcal{L}_\beta^h\left(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1}\right) \le \mathcal{L}_\beta^h\left(x^*, y^*, \lambda^*\right). \tag{3.33}$$

On the other hand, (3.26) implies $\left\|w^{k+1} - w^k\right\| \to 0$, 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^*, λ^*). From the lower semicontinuity of $\mathcal{L}_\beta^h(\cdot)$, we have

$$\limsup_{j\to+\infty} \mathcal{L}_\beta^h\left(x^{k_j+1}, y^{k_j}, \lambda^{k_j}\right) \ge \mathcal{L}_\beta^h\left(x^*, y^*, \lambda^*\right). \tag{3.34}$$

Then by combining (3.33) and (3.34) together we can get

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

which implies

$$\lim_{j\to+\infty} f\left(x^{k_j+1}\right) = f\left(x^*\right) \tag{3.35}$$

Passing to the limit in (3.2) along the subsequence $\left\{\left(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1}\right)\right\}_{j\in N}$ and invoking (3.35) and the continuity of ∇g, it follows that

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

The last equation implies that $Ax^{*}+y^{*}=b$ due to the strong convexity of $h(\cdot)$. Thus, $\left(x^{*},y^{*},\lambda^{*}\right)$ is a critical point of (3.3), which implies that $w^{*} \in \operatorname{crit} \mathcal{L}_{\beta}^{h}$.

iii) For any point $\left(x^{*},y^{*},\lambda^{*}\right) \in S\left(w_0\right)$, there exists a subsequence $\left\{\left(x^{k_j},y^{k_j},\lambda^{k_j}\right)\right\}$ that converges to $\left(x^{*},y^{*},\lambda^{*}\right)$. Combining equations (3.33), (3.34) and the fact that $\left\{\mathcal{L}_{\beta}^{h}\left(w^{k}\right)\right\}$ is nonincreasing, we can get

$$\lim_{k\to+\infty} \mathcal{L}_{\beta}^{h}\left(x^{k},y^{k},\lambda^{k}\right) = \mathcal{L}_{\beta}^{h}\left(x^{*},y^{*},\lambda^{*}\right).$$

Therefore, $\mathcal{L}_{\beta}^{h}(\cdot)$ is finite and constant on $S\left(w_0\right)$, Moreover,

$$\inf_{k\in\mathbb{N}} \mathcal{L}_{\beta}^{h}\left(w^{k}\right) = \lim_{k\to+\infty} \mathcal{L}_{\beta}^{h}\left(w^{k}\right)$$

The proof is completed. □

In the following, we will present an important result of this paper, which provides a detailed analysis of the convergence of Generalized Bregman ADMM (1.4).

Theorem 3.1 Let $\left\{w^{k}=\left(x^{k},y^{k},\lambda^{k}\right)\right\}$ be the sequence generated by the Generalized Bregman ADMM (1.3), which is assumed to be bounded. Suppose that $\mathcal{L}_{\beta}^{h}(\cdot)$ is a KL function, then $\left\{w^{k}\right\}$ has finite length, that is

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

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

Proof: From the proof of Lemma 3.4, we know that $\mathcal{L}_{\beta}^{h}\left(w^{k}\right) \to \mathcal{L}_{\beta}^{h}\left(w^{*}\right)$ for all $w^{*} \in S\left(w_0\right)$. Let us now consider two cases.

i) If there exists an integer k_0 such that $\mathcal{L}_{\beta}^{h}\left(w^{k_0}\right) = \mathcal{L}_{\beta}^{h}\left(w^{*}\right)$, then using (3.5), we have

$$\delta D_h\left(y^{k},y^{k+1}\right) \le \mathcal{L}_{\beta}^{h}\left(w^{k}\right) - \mathcal{L}_{\beta}^{h}\left(w^{k+1}\right) \le \mathcal{L}_{\beta}^{h}\left(w^{k_0}\right) - \mathcal{L}_{\beta}^{h}\left(w^{*}\right) = 0,$$

for any $k > k_0$. Thus, we obtain $y^{k+1} = y^{k}$ for any $k > k_0$. Combining (3.22) and (3.30), we further derive that $\lambda^{k+1} = \lambda^{k}$ and $x^{k+1} = x^{k}$ for any $k > k_0 + 1$, which implies that $w^{k+1} = w^{k}$. Hence, the assertion holds.

ii) If $\mathcal{L}_{\beta}^{h}\left(w^{k}\right) > \mathcal{L}_{\beta}^{h}\left(w^{*}\right)$ for all k, then since $d\left(w^{k},S\left(w_0\right)\right) \to 0$, there exists $k_1 > 0$, such that for any $k > k_1$, we have $d\left(w^{k},S\left(w_0\right)\right) < \varepsilon$ for all $\varepsilon > 0$. Moreover, with $\mathcal{L}_{\beta}^{h}\left(w^{k}\right) \to \mathcal{L}_{\beta}^{h}\left(w^{*}\right)$, it follows that there exists $k_2 > 0$ such that for any $k > k_2$, $\mathcal{L}_{\beta}^{h}\left(w^{k}\right) < \mathcal{L}_{\beta}^{h}\left(w^{*}\right) + \eta$ for all $\eta > 0$. Therefore, when $k > \tilde{k} = \max\left\{k_1,k_2\right\}$ for all $\varepsilon,\eta > 0$, we can obtain the following:

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

Since $S\left(w_0\right)$ is a nonempty compact set and $\mathcal{L}_{\beta}^{h}(\cdot)$ is constant on $S\left(w_0\right)$, we can apply Lemma 2.4 with $\Omega = S\left(w_0\right)$ to deduce that for any $k > \tilde{k}$,

$$\varphi'\left(\mathcal{L}_\beta^h\left(w^k\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)d\left(0,\partial\mathcal{L}_\beta^h\left(w^k\right)\right)\geq 1. \tag{3.36}$$

Using the fact that $\mathcal{L}_\beta^h\left(w^k\right)-\mathcal{L}_\beta^h\left(w^{k+1}\right)=\mathcal{L}_\beta^h\left(w^k\right)-\mathcal{L}_\beta^h\left(w^*\right)-\left(\mathcal{L}_\beta^h\left(w^{k+1}\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)$, and the concavity of $\varphi(\cdot)$, we can show that

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

Combining the above inequality with $d\left(0,\partial\mathcal{L}_\beta^h\left(w^k\right)\right)\leq\xi\left\|y^k-y^{k-1}\right\|$, $\varphi'\left(\mathcal{L}_\beta^h\left(w^k\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)>0$ and relation (3.36), we obtain

$$\begin{aligned}&\mathcal{L}_\beta^h\left(w^k\right)-\mathcal{L}_\beta^h\left(w^{k+1}\right)\leq\frac{\varphi\left(\mathcal{L}_\beta^h\left(w^k\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)-\varphi\left(\mathcal{L}_\beta^h\left(w^{k+1}\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)}{\varphi'\left(\mathcal{L}_\beta^h\left(w^k\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)}\\&\leq d\left(0,\partial\mathcal{L}_\beta^h\left(w^k\right)\right)\left[\varphi\left(\mathcal{L}_\beta^h\left(w^k\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)-\varphi\left(\mathcal{L}_\beta^h\left(w^{k+1}\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)\right]\\&\leq\xi\left\|y^k-y^{k-1}\right\|\left[\varphi\left(\mathcal{L}_\beta^h\left(w^k\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)-\varphi\left(\mathcal{L}_\beta^h\left(w^{k+1}\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)\right].\end{aligned} \tag{3.37}$$

For convenience, we define $\Delta_{p,q}:=\varphi\left(\mathcal{L}_\beta^h\left(w^p\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)-\varphi\left(\mathcal{L}_\beta^h\left(w^q\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)$. Then, (3.37) can be simplified as

$$\mathcal{L}_\beta^h\left(w^k\right)-\mathcal{L}_\beta^h\left(w^{k+1}\right)\leq\xi\left\|y^k-y^{k-1}\right\|\Delta_{k,k+1}. \tag{3.38}$$

According to the 1-strong convexity of the function h, and combining Lemma 3.1 with inequality (3.38), we get that for all $k>\tilde{k}$,

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

Then

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

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

$$2\left\|y^k-y^{k+1}\right\|\leq\left\|y^k-y^{k-1}\right\|+\frac{2\xi}{\delta}\Delta_{k,k+1}. \tag{3.39}$$

Summing (3.39) over for $k=\tilde{k}+1,\cdots,m$ yields

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

Notice that $\varphi\left(\mathcal{L}_\beta^h\left(w^{m+1}\right)-\mathcal{L}_\beta^h\left(w^*\right)\right)>0$ from Definition 2.9. Rearranging terms and taking $m\to+\infty$ yield

$$\sum_{k=\tilde{k}+1}^{+\infty}\left\|y^{k+1}-y^k\right\|\leq\left\|y^{\tilde{k}+1}-y^{\tilde{k}}\right\|+\frac{2\xi}{\delta}\varphi\left(\mathcal{L}_\beta^h\left(w^{\tilde{k}+1}\right)-\mathcal{L}_\beta^h\left(w^*\right)\right), \tag{3.40}$$

Therefore,

$$\sum_{k=0}^{+\infty}\left\|y^{k+1}-y^k\right\|<+\infty. \tag{3.41}$$

Combining (3.22) and (3.41), we obtain

$$\sum_{k=0}^{+\infty}\left\|\lambda^{k+1}-\lambda^{k}\right\|<+\infty. \tag{3.42}$$

Using (3.30), we obtain

$$\begin{aligned}\left\|x^{k+1}-x^{k}\right\| &\le \sqrt{\frac{3}{\alpha\beta^2 M}}\Big(\left\|\lambda^{k+1}-\lambda^{k}\right\|^2+\left\|\lambda^{k}-\lambda^{k-1}\right\|^2+2\beta^2L_h^2\left\|y^{k+1}-y^{k}\right\|^2\\ &\quad +2(\alpha-1)^2\beta^2L_h^2\left\|y^{k}-y^{k+1}\right\|^2\Big)^{1/2}\\ &\le \sqrt{\frac{3}{\alpha\beta^2 M}}\Big(\left\|\lambda^{k+1}-\lambda^{k}\right\|+\left\|\lambda^{k}-\lambda^{k-1}\right\|+2\beta^2L_h^2\left\|y^{k+1}-y^{k}\right\|\\ &\quad +2(\alpha-1)^2\beta^2L_h^2\left\|y^{k}-y^{k+1}\right\|\Big)\end{aligned}$$

Combining this inequality with (3.41) and (3.42), we have

$$\sum_{k=0}^{+\infty}\left\|x^{k+1}-x^{k}\right\|<+\infty. \tag{3.43}$$

Additionally, we note that

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

Using (3.41) - (3.43), we can conclude that

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

implying that $\{w^k\}$ is a Cauchy sequence and thus convergent. By Lemma 3.4, we complete the proof. □

Theorem 3.2 (Convergence rate) Let $\{w^k=(x^k,y^k,\lambda^k)\}$ be the sequence generated by the Generalized 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

$$\left\|(x^k,y^k,\lambda^k)-(x^*,y^*,\lambda^*)\right\|\le c_1\tau^k.$$

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

$$\|(x^k,y^k,\lambda^k)-(x^*,y^*,\lambda^*)\|\le c_2k^{(\theta-1)/(2\theta-1)}.$$

Proof: When $\theta=0$, we have $\varphi(s)=cs$ and $\varphi'(s)=c$. Suppose, by contradiction, that $\{w^k=(x^k,y^k,\lambda^k)\}$ does not converge in a finite number of steps. Then, the KL property at (x^*,y^*,λ^*) yields, for any sufficiently large k,

$c \cdot d\left(0, \partial \mathcal{L}_\beta^h\left(w^k\right)\right) \geq 1$, which contradicts Lemma 3.3.

Next, let $\theta > 0$ and set $\Delta_k = \sum_{i=k}^{+\infty} \left\| y^{i+1} - y^i \right\|$ for $k \geq 0$. By the triangle inequality, we have $\Delta_k \geq \left\| y^k - y^* \right\|$, which allows us to estimate Δ_k. With these notations, it follows from (3.40) that

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

By invoking the KL property of $\mathcal{L}_\beta^h(\cdot)$ at $\left(x^*, y^*, \lambda^*\right)$, we obtain

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

which is equivalent to

$$\left(\mathcal{L}_\beta^h\left(w^{\tilde{k}+1}\right) - \mathcal{L}_\beta^h\left(w^*\right)\right)^\theta \leq c \cdot (1-\theta) d\left(0, \partial \mathcal{L}_\beta^h\left(w^{\tilde{k}+1}\right)\right). \tag{3.44}$$

Using Lemma 3.3, we get

$$d\left(0, \partial \mathcal{L}_\beta^h\left(w^{\tilde{k}+1}\right)\right) \leq \xi \left\| y^{\tilde{k}+1} - y^{\tilde{k}} \right\| = \xi\left(\Delta_{\tilde{k}} - \Delta_{\tilde{k}+1}\right). \tag{3.45}$$

Combining (3.44) and (3.45), we obtain that there exists $\gamma > 0$ such that

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

and then

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

Sequences satisfying such inequalities have been studied in Attouch and Bolte [37]. It follows that

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

$$\left\| y^k - y^* \right\| \leq c_1 \tau^k. \tag{3.46}$$

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

$$\left\| y^k - y^* \right\| \leq c_2 k^{\frac{\theta-1}{2\theta-1}}. \tag{3.47}$$

Recalling that

$$\left\| \lambda^{k+1} - \lambda^k \right\| \leq L L_h \left\| y^{k+1} - y^k \right\|,$$

we obtain

$$\left\| \lambda^k - \lambda^* \right\| \leq L L_h \left\| y^k - y^* \right\|. \tag{3.48}$$

Furthermore, from the relations

$$\lambda^k = \lambda^{k-1} - \beta\left(\nabla h\left(\alpha\left(Ax^k - b\right)\right) - \nabla h\left(-y^k - (\alpha-1) y^{k-1}\right)\right),$$

and

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

it follows that

$$\beta\left(\nabla h\left(\alpha\left(Ax^k-b\right)\right)-\nabla h\left(\alpha\left(Ax^*-b\right)\right)\right)$$
$$=\left(\lambda^{k-1}-\lambda^*\right)+\left(\lambda^*-\lambda^k\right)+\beta\left(\nabla h\left(-y^k-(\alpha-1)y^{k-1}\right)-\nabla h\left(-\alpha y^*\right)\right)$$

We multiply both sides of the above equation by $\frac{1}{\beta}$ at the same time

$$\nabla h\left(\alpha\left(Ax^k-b\right)\right)-\nabla h\left(\alpha\left(Ax^*-b\right)\right)$$
$$=\frac{1}{\beta}\left(\lambda^{k-1}-\lambda^*\right)+\frac{1}{\beta}\left(\lambda^*-\lambda^k\right)+\left(\nabla h\left(-y^k-(\alpha-1)y^{k-1}\right)-\nabla h\left(-\alpha y^*\right)\right)$$

Now combine the above equation with the 1-strong convexity of $h(\cdot)$, and then we can get the following

$$\begin{aligned}
\alpha\left\|x_k-x^*\right\| &\le \left\|\nabla h\left(\alpha\left(Ax_k-b\right)\right)-\nabla h\left(\alpha\left(Ax^*-b\right)\right)\right\| \\
&\le \frac{1}{\beta}\left\|\left(\lambda^{k-1}-\lambda^*\right)+\frac{1}{\beta}\left(\lambda^*-\lambda^k\right)+\left(\nabla h\left(-y^k-(\alpha-1)y^{k-1}\right)-\nabla h\left(-\alpha y^*\right)\right)\right\| \\
&\le \frac{1}{\beta}\left\|\lambda^{k-1}-\lambda^*\right\|+\frac{1}{\beta}\left\|\lambda^*-\lambda^k\right\|+L_h\left\|y^{k-1}-y^k\right\|+\alpha L_h\left\|y^{k-1}-y^*\right\| \\
&\le \frac{LL_h}{\beta}\left\|y^{k-1}-y^*\right\|+\frac{LL_h}{\beta}\left\|y^*-y^k\right\|+L_h\left\|y^{k-1}-y^k\right\|+\alpha L_h\left\|y^{k-1}-y^*\right\| \\
&\le \frac{LL_h}{\beta}\left\|y^{k-1}-y^*\right\|+\frac{LL_h}{\beta}\left\|y^*-y^k\right\|+L_h\left\|y^{k-1}-y^*\right\|+L_h\left\|y^*-y^k\right\|+\alpha L_h\left\|y^{k-1}-y^*\right\| \\
&= \left(\frac{LL_h}{\beta}+L_h+\alpha L_h\right)\left\|y_{k-1}-y^*\right\|+\left(\frac{LL_h}{\beta}+L_h\right)\left\|y^k-y^*\right\|.
\end{aligned} \tag{3.49}$$

Combining (3.48) and (3.49), we immediately obtain the desired inequalities from (3.46) and (3.47).

4. Conclusion

In this paper, we primarily analyze the generalized Bregman alternating direction method of multipliers (ADMM) for solving nonconvex separable problems subject to linear constraints. In contrast to the classical alternating direction method of multipliers, we modify the iterative format of the second subproblem. This modification relaxes the condition of global Lipschitz continuity for the gradient of differentiable functions. Additionally, we introduce a relaxation parameter α, inspired by the acceleration technique of the proximal point algorithm (PPA), to enhance the algorithm's performance. Under the assumption that the augmented Lagrangian function satisfies the Kurdyka-Lojasiewicz inequality, we prove that when the penalty parameters in the augmented Lagrangian function are sufficiently large, the iterative sequence generated by the algorithm converges to a critical point of the augmented Lagrangian function. Lastly, we set the corresponding parameters to further analyze the convergence rate of the algorithm.

Conflicts of Interest

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

References

[1] 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

[2] 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

[3] 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

[4] 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

[5] 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/10.1561/2200000016

[6] 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

[7] 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

[8] 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

[9] 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

[10] He, B. and Yuan, X. (2014) On Non-Ergodic Convergence Rate of Douglas–Rachford Alternating Direction Method of Multipliers. *Numerische Mathematik*, **130**, 567-577. https://doi.org/10.1007/s00211-014-0673-6

[11] 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

[12] Hong, M. and Luo, Z. (2016) On the Linear Convergence of the Alternating Direction Method of Multipliers. *Mathematical Programming*, **162**, 165-199. https://doi.org/10.1007/s10107-016-1034-2

[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] 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

[16] 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

[17] 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

[18] 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

[19] Gabay, D. (1983) Applications of the Method of Multipliers to Variational Inequalities. In: *Studies in Mathematics and Its Applications*, Elsevier, 299-331. https://doi.org/10.1016/s0168-2024(08)70034-1

[20] Lions, P.L. and Mercier, B. (1979) Splitting Algorithms for the Sum of Two Nonlinear Operators. *SIAM Journal on Numerical Analysis*, **16**, 964-979. https://doi.org/10.1137/0716071

[21] Eckstein, J. and Bertsekas, D.P. (1992) On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators. *Mathematical Programming*, **55**, 293-318. https://doi.org/10.1007/bf01581204

[22] Wang, X. and Guo, K. (2018) Convergence of the Generalized Alternating Direction Method of Multipliers for a Class of Nonconvex Optimization Problems. *Applied Mathematics and Mechanics*, **39**, 1410-1425.

[23] Bertsekas, D.P. (2014) Constrained Optimization and Lagrange Multiplier Methods. Academic Press.

[24] Cai, X., Chen, Y. and Han, D. (2013) Nonnegative Tensor Factorizations Using an Alternating Direction Method. *Frontiers of Mathematics in China*, **8**, 3-18. https://doi.org/10.1007/s11464-012-0264-8

[25] Eckstein, J. (1994) Parallel Alternating Direction Multiplier Decomposition of Convex Programs. *Journal of Optimization Theory and Applications*, **80**, 39-62. https://doi.org/10.1007/bf02196592

[26] 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

[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] 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

[29] Nesterov, Y. (2019) Implementable Tensor Methods in Unconstrained Convex Optimization. *Mathematical Programming*, **186**, 157-183. https://doi.org/10.1007/s10107-019-01449-1

[30] Tan, L. and Guo, K. (2025) Bregman ADMM: A New Algorithm for Nonconvex Optimization with Linear Constraints. *Journal of Nonlinear and Variational Analysis*, **9**,

179-196.

[31] Beck, A. (2017). First-Order Methods in Optimization. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611974997

[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

[33] 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

[34] 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

[35] 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

[36] Bolte, J., Sabach, S. and Teboulle, M. (2013) Proximal Alternating Linearized Minimization for Nonconvex and Non-Smooth Problems. *Mathematical Programming*, **146**, 459-494. https://doi.org/10.1007/s10107-013-0701-9

[37] Attouch, H. and Bolte, J. (2007) On the Convergence of the Proximal Algorithm for Non-Smooth Functions Involving Analytic Features. *Mathematical Programming*, **116**, 5-16. https://doi.org/10.1007/s10107-007-0133-5