Journal of Mathematical Research with Applications
Sept., 2018, Vol. 38, No. 5, pp. 523–540
DOI:10.3770/j.issn:2095-2651.2018.05.010
Http://jmre.dlut.edu.cn

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

Ke GUO*, **Xin WANG**
School of Mathematics and Information, China West Normal University, Sichuan 637002*, P. R. China*

Abstract In this paper, we consider the convergence of the generalized alternating direction method of multipliers (GADMM) for solving linearly constrained nonconvex minimization model whose objective contains coupled functions. Under the assumption that the augmented Lagrangian function satisfies the Kurdyka-Łojasiewicz inequality, we prove that the sequence generated by the GADMM converges to a critical point of the augmented Lagrangian function when the penalty parameter in the augmented Lagrangian function is sufficiently large. Moreover, we also present some sufficient conditions guaranteeing the sublinear and linear rate of convergence of the algorithm.

Keywords generalized alternating direction method of multipliers; Kurdyka-Łojasiewicz inequality; nonconvex optimization

MR(2010) Subject Classification 90C26; 65K10

1. Introduction

In this paper, we consider the nonconvex optimization problem with the following form

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

where $f:\mathcal{R}^n\to\mathcal{R}\cup\{+\infty\}$ is a proper lower semicontinuous function and $g:\mathcal{R}^m\to\mathcal{R}$ is continuously differentiable function whose gradient ∇g is Lipschitz continuous with constant $L_1>0, H:\mathcal{R}^n\times\mathcal{R}^m\to\mathcal{R}$ is a smooth function, $A\in\mathcal{R}^{m\times n}$ is a given matrix, and $b\in\mathcal{R}^m$ is a vector. A special case of problem (1.1) is when the coupled function H is absent, that is,

$$\begin{aligned} \min \quad & f(x)+g(y) \\ \text{s.t.} \quad & Ax+y=b. \end{aligned} \tag{1.2}$$

As we know, the alternating direction method of multipliers (ADMM) in [1] plays a fundamental theoretical and algorithmic role in solving problem (1.2). The iterative scheme of ADMM for

Received September 3, 2017; Accepted June 5, 2018
Supported by the National Natural Science Foundation of China (Grant Nos. 11571178; 11801455) and the Fundamental Research Funds of China West Normal University (Grant No. 17E084).
* Corresponding author
E-mail address: keguo2014@126.com (Ke GUO); wangxinwatermelon@163.com (Xin WANG)

(1.2) reads as:

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

where λ is the Lagrangian multiplier associated with the linear constraints and $\beta > 0$ is the penalty parameter. For the case both f and g are proper lower semicontinuous convex functions, the convergence of ADMM (1.3) is well-understood and there are recently some convergence rate analysis [2–5]. Without convexity assumption, the convergence analysis for ADMM (1.3) is much more challenging. Recently, there have been a few developments on it, e.g., [6–10].

Applying the classic ADMM to problem (1.1), we can get the following iterative process:

$$\begin{cases} x^{k+1} \in \arg\min_x\{f(x) + H(x, y^k) - \langle \lambda^k, Ax\rangle + \frac{\beta}{2}\|Ax + y^k - b\|^2\}, \\ y^{k+1} \in \arg\min_y\{g(y) + H(x^{k+1}, y) - \langle \lambda^k, y\rangle + \frac{\beta}{2}\|Ax^{k+1} + y - b\|^2\}, \\ \lambda^{k+1} = \lambda^k - \beta(Ax^{k+1} + y^{k+1} - b). \end{cases} \tag{1.4}$$

However, due to the existence of coupled function H, the convergence of ADMM (1.4) is still in its infancy, even the objective functions are assumed to be convex. Recently, Gao and Zhang [11] considered the case where H is a smooth convex function and f, g are convex functions. Under the assumptions that ∇H is Lipschitz continuous and g is strongly convex, they proved the sequence generated by the proximal version of ADMM (1.4) converges to an optimal solution of the problem (1.1). Chen et al. [12] analyzed the convergence of the ADMM (1.4) for the problem (1.1) when the coupled function H is a quadratic function. Guo et al. [13] studied the convergence of the classic ADMM (1.4) for the nonconvex problem (1.1), i.e., without assuming the convexity of f, g and the coupled term H. By using the important Kurdyka-Lojasiewicz (KL) inequality (see Definition 2.6), they proved that the sequence generated by the ADMM (1.4) converges to a critical point of the augmented Lagrangian function if the augmented Lagrangian function for problem (1.1) is a KL function. The importance of the KL inequality is due to the fact that many functions satisfy this inequality, especially when the functions belong to some functional classes, e.g., semi-algebraic (such as $\|\cdot\|_p^p$, $p \in [0, 1]$ is a rational number), real sub-analytic and so on (see also [14–17] and references therein).

As pointed in [18], the ADMM (1.3) is actually an application of the well-known Douglas-Rachford splitting method (DRSM) in [19] to the dual of (1.2); and in [20], the DRSM was further explained as an application of the proximal point algorithm (PPA) in [21]. Therefore, it was suggested in [20] to apply the acceleration scheme in [22] for the PPA to accelerate the original ADMM (1.3). Back to our problem (1.1), a generalized ADMM (GADMM) is thus proposed:

$$\begin{cases} x^{k+1} \in \arg\min_x\{f(x) + H(x, y^k) - \langle \lambda^k, Ax\rangle + \frac{\beta}{2}\|Ax + y^k - b\|^2\}, \\ y^{k+1} \in \arg\min_y\{g(y) + H(x^{k+1}, y) - \langle \lambda^k, y\rangle + \frac{\beta}{2}\|\alpha Ax^{k+1} + (1-\alpha)(b - y^k) + y - b\|^2\}, \\ \lambda^{k+1} = \lambda^k - \beta(\alpha Ax^{k+1} + (1-\alpha)(b - y^k) + y^{k+1} - b), \end{cases} \tag{1.5}$$

where the parameter $\alpha \in (0, 2)$ is a relaxation factor. Obviously, the GADMM (1.5) reduces to the classic ADMM (1.4) when $\alpha = 1$ and reduces to the classic GADMM [20] when $H \equiv 0$. We refer to [23–25] for empirical studies of the acceleration performance of the GADMM.

The purpose of this paper is to prove the convergence of the GADMM (1.5) for nonconvex optimization problem (1.1). By means of the Kurdyka-Łojasiewicz inequality, we prove that if the augmented Lagrangian function for problem (1.1) is a KL function, then the sequence generated by the GADMM (1.5) converges to a critical point of the augmented Lagrangian function (see Section 3). Under some further conditions on the problem's data, we prove the convergence rate of the GADMM (1.5). The paper is organized as follows. We first summarize some necessary preliminaries for further analysis in Section 2. In Section 3, we analyze the convergence of GADMM. Finally, some conclusions are made in Section 4.

2. Preliminaries

In this section, we give some preliminaries that will be frequently used in this paper. Let $F : \mathcal{R}^n \rightrightarrows \mathcal{R}^m$ be a point-to-set mapping. Then its graph is defined by

$$\text{Graph } F := \{(x, y) \in \mathcal{R}^n \times \mathcal{R}^m : y \in F(x)\}.$$

We define the distance of a point $x \in \mathcal{R}^n$ to a subset S of $\mathcal{R}^n$ by

$$d(x, S) := \inf_{y \in S} \|y - x\|.$$

When $S = \emptyset$, we set $d(x, S) := +\infty$, for all x.

Given a function $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$, the effective domain and the epigraph of f are defined by

$$\text{dom } f := \{x \in \mathcal{R}^n : f(x) < +\infty\} \quad \text{and} \quad \text{epi } f := \{(x, \alpha) \in \mathcal{R}^n \times \mathcal{R} : f(x) \leq \alpha\},$$

respectively. We say that the function f is proper (respectively, lower semicontinuous) if the dom f (respectively, epi f) set is nonempty (respectively, closed).

Definition 2.1 *Let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a proper lower semicontinuous function.*

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

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

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

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

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

Remark 2.2 From Definition 2.1 we can find that

(i) The above definition implies $\hat{\partial} f(x) \subseteq \partial f(x)$ for each $x \in \mathcal{R}^n$, 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 $\partial f(x)$, if $f(x_k)$ converges to $f(x)$ as $k \to +\infty$, then $(x, x^*) \in \text{Graph } \partial f$;

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

$$0 \in \partial f(x); \tag{2.1}$$

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

A point that satisfies (2.1) is called a critical point or a stationary point. The set of critical points of f is denoted by crit f.

Let us recall the important properties of subdifferential calculus.

Lemma 2.3 ([14]) *Suppose that $S(x, y) := s_1(x) + s_2(y)$, where $s_1 : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ and $s_2 : \mathcal{R}^m \to \mathcal{R} \cup \{+\infty\}$ are proper lower semicontinuous functions. Then for all $(x, y) \in$ dom $S =$ dom $s_1 \times$ dom s_2, we have $\partial S(x, y) = \partial_x S(x, y) \times \partial_y S(x, y)$.*

Definition 2.4 ([26]) *A proper lower semicontinuous function $g : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ is called weakly convex (or semiconvex) if for some $\omega > 0$, the function $x \mapsto g(x) + \frac{\omega}{2}\|x\|^2$ is convex.*

Remark 2.5 It is well known that the set of semiconvex functions contains several important classes of (nonsmooth) functions as special cases, for example, φ-convex functions [27] and primal-lower-nice functions [28]. Moreover, any twice continuously differentiable function with a bounded second-order derivative is semiconvex; see, e.g., [22]. In [29], the semiconvexity is also called hypoconvexity [29, Definition 3.10], and the proximal operator of a hypoconvex function is well studied therein.

Let $\eta \in (0, +\infty]$. We denote by Φ_η the class of all concave and continuous functions $\varphi : [0, \eta) \to \mathcal{R}_+$ which satisfies the following assumptions:

(i) $\varphi(0) = 0$;

(ii) φ is continuously differentiable on $(0, \eta)$ and continuous at 0;

(iii) $\varphi'(s) > 0, \forall s \in (0, \eta)$.

Definition 2.6 ([14] Kurdyka-Łojasiewicz inequality) *Let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a proper lower semicontinuous function. For $-\infty < \eta_1 < \eta_2 \le +\infty$, set*

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

We say that function f has the KL property at $x^* \in$ dom ∂f if there exists $\eta \in (0, +\infty]$, a neighborhood U of x^* and a function $\varphi \in \Phi_\eta$, such that for all $x \in U \cap [f(x^*) < f < f(x^*) + \eta]$, the Kurdyka-Łojasiewicz inequality holds:

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

Definition 2.7 ([14] Kurdyka-Łojasiewicz function) *If f satisfies the KL property at each point of dom ∂f, then f is called a KL function.*

Remark 2.8 One can easily check that the Kurdyka-Łojasiewicz property is automatically satisfied at any non-critical point $x^* \in \text{dom } f$ (see [14, Lemma 2.1, Remark 3.2 (b)]).

Lemma 2.9 ([30] Uniformized KL property) *Let Ω be a compact set and let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a proper and lower semicontinuous function. Assume that f 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*

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

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

The following lemma for smooth functions is very useful for the convergence analysis.

Lemma 2.10 ([31]) *Let $h : \mathcal{R}^n \to \mathcal{R}$ be a continuously differentiable function with gradient ∇h being Lipschitz continuous with constant $L > 0$. Then for any $x, y \in \mathcal{R}^n$, we have*

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

Definition 2.11 *We say that (x^*, y^*, λ^*) is a critical point of the augmented Lagrangian function $\mathcal{L}_\beta(\cdot)$ (3.4) of problem (1.1) if it satisfies*

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

The set of critical points of $\mathcal{L}_\beta(\cdot)$ is denoted by crit $\mathcal{L}_\beta$. It is easy to see that a critical point of the augmented Lagrangian function of problem (1.1) is exactly a KKT point associated with it.

3. Convergence analysis

Before the proof, let us present the variational characterization of scheme (1.5). By the optimality condition for (1.5), we have

$$\begin{cases} 0 \in \partial f(x^{k+1}) + \nabla_x H(x^{k+1}, y^k) - A^T\lambda^k + \beta A^T(Ax^{k+1} + y^k - b), \\ 0 = \nabla g(y^{k+1}) + \nabla_y H(x^{k+1}, y^{k+1}) - \lambda^k + \beta(\alpha Ax^{k+1} + (1-\alpha)(b - y^k) + y^{k+1} - b), \\ \lambda^{k+1} = \lambda^k - \beta(\alpha Ax^{k+1} + (1-\alpha)(b - y^k) + y^{k+1} - b). \end{cases} \tag{3.1}$$

Using the last equality of (3.1) and rearranging terms, we obtain

$$\begin{cases} A^T\lambda^k - \beta A^T(Ax^{k+1} + y^k - b) - \nabla_x H(x^{k+1}, y^k) \in \partial f(x^{k+1}), \\ \nabla g(y^{k+1}) = \lambda^{k+1} - \nabla_y H(x^{k+1}, y^{k+1}), \\ \lambda^{k+1} = \lambda^k - \beta(\alpha Ax^{k+1} + (1-\alpha)(b - y^k) + y^{k+1} - b). \end{cases} \tag{3.2}$$

Throughout this paper, we make the following assumptions.

Assumptions 3.1 Let $f : \mathcal{R}^n \to \mathcal{R} \cup \{+\infty\}$ be a weakly convex function with constant $\omega > 0, g : \mathcal{R}^m \to \mathcal{R}$ be a continuously differentiable function whose gradient ∇g is Lipschitz continuous with constant $L_1 > 0$, and let $H : \mathcal{R}^n \times \mathcal{R}^m \to \mathcal{R}$ be a smooth function. Assume the following conditions are satisfied:

(i) $\inf_{(x,y)\in\mathcal{R}^n\times\mathcal{R}^m} H(x,y) > -\infty, \inf_{x\in\mathcal{R}^n} f(x) > -\infty, \inf_{y\in\mathcal{R}^m} g(y) > -\infty$;

(ii) For any fixed x, the partial gradient $\nabla_y H(x,y)$ is global Lipschitz continuous with constant $L_2(x) > 0$, that is

$$\|\nabla_y H(x,y) - \nabla_y H(x,\tilde{y})\| \le L_2(x)\|y-\tilde{y}\|, \quad \forall y,\tilde{y}\in\mathcal{R}^m;$$

For any fixed y, the partial gradient $\nabla_x H(x,y)$ is global Lipschitz continuous with constant $L_3(y) > 0$, that is

$$\|\nabla_x H(x,y) - \nabla_x H(\tilde{x},y)\| \le L_3(y)\|x-\tilde{x}\|, \quad \forall x,\tilde{x}\in\mathcal{R}^n;$$

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

$$\|\nabla_x H(x_1,y_1) - \nabla_x H(x_2,y_2), \nabla_y H(x_1,y_1) - \nabla_y H(x_2,y_2)\| \le M\|(x_1-y_1,x_2-y_2)\|;$$

(iv) $A^TA \succeq \mu I$ for some $\mu>0$;

(v) There exist $L_2, L_3 > 0$ such that

$$\sup\{L_2(x^k) : k\in N\} \le L_2, \quad \sup\{L_3(y^k) : k\in N\} \le L_3;$$

(vi) $\beta > \widetilde{\beta}$, where

$$\widetilde{\beta} := \max\Big\{\frac{\alpha(L_3+\omega)+\sqrt{\alpha^2(L_3+\omega)^2+16\alpha\mu M^2}}{2\alpha\mu}, \frac{\alpha(L_1+L_2)+\sqrt{\alpha^2(L_1+L_2)^2+16(2-\alpha)(L_1^2+M^2)}}{2(2-\alpha)}\Big\} \tag{3.3}$$

and the parameter $\alpha\in(0,2)$ is a relaxation factor.

Note that, if we set

$$\delta := \min\Big\{\frac{\beta\mu-L_3-\omega}{2} - \frac{2M^2}{\alpha\beta}, \frac{\beta-L_1-L_2}{2} - \frac{2L_1^2+2M^2}{\alpha\beta} + \frac{(1-\alpha)\beta}{\alpha}\Big\},$$

we know $\delta>0$ in view of (vi) of Assumption 3.1. In the sequel for convergence, we often use the notations $\omega^k := (x^k,y^k,\lambda^k)$ and $v^k := (x^k,y^k)$. The augmented Lagrangian function of problem (1.1) is defined by

$$\mathcal{L}_\beta(x,y,\lambda) := f(x)+g(y)+H(x,y)-\langle\lambda, Ax+y-b\rangle + \frac{\beta}{2}\|Ax+y-b\|^2, \tag{3.4}$$

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

$$\widehat{\mathcal{L}}_\beta(x,y,\lambda;\alpha,\vartheta) := f(x)+g(y)+H(x,y)-\langle\lambda, Ax+y-b\rangle + \frac{\beta}{2}\|\alpha Ax-(1-\alpha)(\vartheta-b)+y-b\|^2.$$

We begin our analysis with the following lemma.

Lemma 3.2 *Let $\{\omega^k\}_{k\in N}$ be the sequence generated by the GADMM (1.5) which is assumed to be bounded. Then we have*

$$\mathcal{L}_\beta(\omega^{k+1}) \le \mathcal{L}_\beta(\omega^k) - \delta\|v^{k+1}-v^k\|^2. \tag{3.5}$$

Proof First, by definition we have

$$\begin{aligned}
&\widehat{\mathcal{L}}_\beta(x^k, y^k, \lambda^k; 1, \cdot) - \widehat{\mathcal{L}}_\beta(x^{k+1}, y^k, \lambda^k; 1, \cdot) \\
&= f(x^k) + g(y^k) + H(x^k, y^k) - \langle \lambda^k, Ax^k + y^k - b\rangle + \frac{\beta}{2}\|Ax^k + y^k - b\|^2 - \\
&\quad \big\{f(x^{k+1}) + g(y^k) + H(x^{k+1}, y^k) - \langle \lambda^k, Ax^{k+1} + y^k - b\rangle + \frac{\beta}{2}\|Ax^{k+1} + y^k - b\|^2\big\} \\
&= f(x^k) - f(x^{k+1}) + \langle \lambda^k, Ax^{k+1} - Ax^k\rangle + \beta\langle Ax^{k+1} + y^k - b, Ax^k - Ax^{k+1}\rangle + \\
&\quad \frac{\beta}{2}\|Ax^{k+1} - Ax^k\|^2 + H(x^k, y^k) - H(x^{k+1}, y^k). \qquad (3.6)
\end{aligned}$$

Since f is weakly convex with constant $\omega > 0$, it follows from the first relation of (3.2) that

$$f(x^k) \ge f(x^{k+1}) + \langle A^T\lambda^k - \beta A^T(Ax^{k+1} + y^k - b) - \nabla_x H(x^{k+1}, y^k), x^k - x^{k+1}\rangle - \frac{\omega}{2}\|x^{k+1} - x^k\|^2.$$

Again since $\nabla_x H(\cdot, y^k)$ is Lipschitz with constant $L_3(y^k)$, we know from Lemma 2.10 that

$$H(x^k, y^k) - H(x^{k+1}, y^k) \ge \langle \nabla_x H(x^{k+1}, y^k), x^k - x^{k+1}\rangle - \frac{L_3(y^k)}{2}\|x^{k+1} - x^k\|^2.$$

By (iv) of Assumption 3.1, we have

$$\|Ax^{k+1} - Ax^k\|^2 \ge \mu\|x^{k+1} - x^k\|^2.$$

Thus, substituting the above three inequalities into (3.6), we obtain

$$\widehat{\mathcal{L}}_\beta(x^k, y^k, \lambda^k; 1, \cdot) - \widehat{\mathcal{L}}_\beta(x^{k+1}, y^k, \lambda^k; 1, \cdot) \ge \frac{\beta\mu - L_3(y^k) - \omega}{2}\|x^{k+1} - x^k\|^2. \qquad (3.7)$$

Similarly,

$$\begin{aligned}
&\widehat{\mathcal{L}}_\beta(x^{k+1}, y^k, \lambda^k; \alpha, y^k) - \widehat{\mathcal{L}}_\beta(x^{k+1}, y^{k+1}, \lambda^k; \alpha, y^k) \\
&= f(x^{k+1}) + g(y^k) + H(x^{k+1}, y^k) - \langle \lambda^k, Ax^{k+1} + y^k - b\rangle + \frac{\beta}{2}\|\alpha(Ax^{k+1} + y^k - b)\|^2 - \\
&\quad \big\{f(x^{k+1}) + g(y^{k+1}) + H(x^{k+1}, y^{k+1}) - \langle \lambda^k, Ax^{k+1} + y^{k+1} - b\rangle + \\
&\quad \frac{\beta}{2}\|\alpha(Ax^{k+1} + y^k - b) + (y^{k+1} - y^k)\|^2\big\} \\
&= g(y^k) - g(y^{k+1}) + H(x^{k+1}, y^k) - H(x^{k+1}, y^{k+1}) + \langle \lambda^k, y^{k+1} - y^k\rangle - \\
&\quad \alpha\beta\langle Ax^{k+1} + y^k - b, y^{k+1} - y^k\rangle - \frac{\beta}{2}\|y^{k+1} - y^k\|^2 \\
&= g(y^k) - g(y^{k+1}) + H(x^{k+1}, y^k) - H(x^{k+1}, y^{k+1}) + \\
&\quad \langle \lambda^{k+1}, y^{k+1} - y^k\rangle + \frac{\beta}{2}\|y^{k+1} - y^k\|^2, \qquad (3.8)
\end{aligned}$$

where the last equality follows from $-\alpha\beta(Ax^{k+1} + y^k - b) = \lambda^{k+1} - \lambda^k + \beta(y^{k+1} - y^k)$ which is based on the third equality of (3.2). By the Lipschitz continuity of ∇g, it follows from Lemma 2.10 and second equality of (3.2) that

$$g(y^k) - g(y^{k+1}) \ge \langle \lambda^{k+1} - \nabla_y H(x^{k+1}, y^{k+1}), y^k - y^{k+1}\rangle - \frac{L_1}{2}\|y^{k+1} - y^k\|^2.$$

Since $\nabla_y H(x^{k+1}, \cdot)$ is Lipschitz with constant $L_2(x^{k+1})$, we know from Lemma 2.10 that

$$H(x^{k+1}, y^k) - H(x^{k+1}, y^{k+1}) \ge \langle \nabla_y H(x^{k+1}, y^{k+1}), y^k - y^{k+1}\rangle - \frac{L_2(x^{k+1})}{2}\|y^{k+1} - y^k\|^2.$$

Thus, combining the above two inequalities with (3.8), we obtain

$$\widehat{\mathcal{L}}_\beta(x^{k+1},y^k,\lambda^k;\alpha,y^k)-\widehat{\mathcal{L}}_\beta(x^{k+1},y^{k+1},\lambda^k;\alpha,y^k)\geq\frac{\beta-L_1-L_2(x^{k+1})}{2}\|y^{k+1}-y^k\|^2. \tag{3.9}$$

Next, we estimate the remaining terms

$$(\widehat{\mathcal{L}}_\beta(x^{k+1},y^k,\lambda^k;1,\cdot)-\widehat{\mathcal{L}}_\beta(x^{k+1},y^k,\lambda^k;\alpha,y^k))+$$
$$(\widehat{\mathcal{L}}_\beta(x^{k+1},y^{k+1},\lambda^k;\alpha,y^k)-\widehat{\mathcal{L}}_\beta(x^{k+1},y^{k+1},\lambda^{k+1};1,\cdot)).$$

Indeed,

$$\begin{aligned}&\widehat{\mathcal{L}}_\beta(x^{k+1},y^k,\lambda^k;1,\cdot)-\widehat{\mathcal{L}}_\beta(x^{k+1},y^k,\lambda^k;\alpha,y^k)\\&\quad=f(x^{k+1})+g(y^k)+H(x^{k+1},y^k)-\langle\lambda^k,Ax^{k+1}+y^k-b\rangle+\frac{\beta}{2}\|Ax^{k+1}+y^k-b\|^2-\\&\qquad\big\{f(x^{k+1})+g(y^k)+H(x^{k+1},y^k)-\langle\lambda^k,Ax^{k+1}+y^k-b\rangle+\frac{\beta}{2}\|\alpha(Ax^{k+1}+y^k-b)\|^2\big\}\\&\quad=\frac{\beta}{2}\|Ax^{k+1}+y^k-b\|^2-\frac{\beta}{2}\|\alpha(Ax^{k+1}+y^k-b)\|^2\end{aligned}\tag{3.10}$$

and

$$\begin{aligned}&\widehat{\mathcal{L}}_\beta(x^{k+1},y^{k+1},\lambda^k;\alpha,y^k)-\widehat{\mathcal{L}}_\beta(x^{k+1},y^{k+1},\lambda^{k+1};1,\cdot)\\&\quad=f(x^{k+1})+g(y^{k+1})+H(x^{k+1},y^{k+1})-\langle\lambda^k,Ax^{k+1}+y^{k+1}-b\rangle+\\&\qquad\frac{\beta}{2}\|\alpha(Ax^{k+1}+y^k-b)+(y^{k+1}-y^k)\|^2-\big\{f(x^{k+1})+g(y^{k+1})+H(x^{k+1},y^{k+1})-\\&\qquad\langle\lambda^{k+1},Ax^{k+1}+y^{k+1}-b\rangle+\frac{\beta}{2}\|Ax^{k+1}+y^{k+1}-b\|^2\big\}\\&\quad=\langle\lambda^{k+1}-\lambda^k,Ax^{k+1}+y^{k+1}-b\rangle+\frac{\beta}{2}\|\alpha(Ax^{k+1}+y^k-b)+(y^{k+1}-y^k)\|^2-\\&\qquad\frac{\beta}{2}\|Ax^{k+1}+y^{k+1}-b\|^2\\&\quad=\langle\lambda^{k+1}-\lambda^k,Ax^{k+1}+y^{k+1}-b\rangle+\frac{\beta}{2}\|\alpha(Ax^{k+1}+y^k-b)\|^2+\\&\qquad\alpha\beta\langle Ax^{k+1}+y^k-b,y^{k+1}-y^k\rangle-\frac{\beta}{2}\|Ax^{k+1}+y^k-b\|^2-\beta\langle Ax^{k+1}+y^k-b,y^{k+1}-y^k\rangle.\end{aligned}\tag{3.11}$$

Adding (3.10) and (3.11), we have

$$\begin{aligned}&(\widehat{\mathcal{L}}_\beta(x^{k+1},y^k,\lambda^k;1,\cdot)-\widehat{\mathcal{L}}_\beta(x^{k+1},y^k,\lambda^k;\alpha,y^k))+(\widehat{\mathcal{L}}_\beta(x^{k+1},y^{k+1},\lambda^k;\alpha,y^k)-\\&\qquad\widehat{\mathcal{L}}_\beta(x^{k+1},y^{k+1},\lambda^{k+1};1,\cdot))\\&\quad=\langle\lambda^{k+1}-\lambda^k,-\frac{1}{\alpha\beta}(\lambda^{k+1}-\lambda^k)-\frac{1-\alpha}{\alpha}(y^{k+1}-y^k)\rangle-\\&\qquad(1-\alpha)\beta\langle-\frac{1}{\alpha\beta}(\lambda^{k+1}-\lambda^k)-\frac{1}{\alpha}(y^{k+1}-y^k),y^{k+1}-y^k\rangle\\&\quad=-\frac{1}{\alpha\beta}\|\lambda^{k+1}-\lambda^k\|^2+\frac{(1-\alpha)\beta}{\alpha}\|y^{k+1}-y^k\|^2.\end{aligned}\tag{3.12}$$

Because ∇H is Lipschitz continuous on bounded subsets and $\{(x^k,y^k)\}_{k\in N}$ is bounded, we know

$$\|\lambda^{k+1}-\lambda^k\|^2$$

$$
\begin{aligned}
&= \|\nabla g(y^{k+1}) + \nabla_y H(x^{k+1}, y^{k+1}) - \nabla g(y^k) - \nabla_y H(x^k, y^k)\|^2 \\
&\le 2\|\nabla g(y^{k+1}) - \nabla g(y^k)\|^2 + 2\|\nabla_y H(x^{k+1}, y^{k+1}) - \nabla_y H(x^k, y^k)\|^2 \\
&\le 2L_1^2\|y^{k+1} - y^k\|^2 + 2M^2\|x^{k+1} - x^k\|^2 + 2M^2\|y^{k+1} - y^k\|^2 \\
&= (2L_1^2 + 2M^2)\|y^{k+1} - y^k\|^2 + 2M^2\|x^{k+1} - x^k\|^2, \qquad (3.13)
\end{aligned}
$$

where the second inequality follows from the Lipschitz continuity of ∇g and (iii) of Assumption 3.1. Substituting (3.13) into (3.12), we can get

$$
\begin{aligned}
&(\widehat{\mathcal{L}}_\beta(x^{k+1}, y^k, \lambda^k; 1, \cdot) - \widehat{\mathcal{L}}_\beta(x^{k+1}, y^k, \lambda^k; \alpha, y^k)) + (\widehat{\mathcal{L}}_\beta(x^{k+1}, y^{k+1}, \lambda^k; \alpha, y^k) - \\
&\quad \widehat{\mathcal{L}}_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}; 1, \cdot)) \\
&\ge -\frac{2M^2}{\alpha\beta}\|x^{k+1} - x^k\|^2 + (\frac{(1-\alpha)\beta}{\alpha} - \frac{2L_1^2 + 2M^2}{\alpha\beta})\|y^{k+1} - y^k\|^2. \qquad (3.14)
\end{aligned}
$$

Observe that

$$
\begin{aligned}
&\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^{k+1}) = \mathcal{L}_\beta(x^k, y^k, \lambda^k) - \mathcal{L}_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}) \\
&\quad = \widehat{\mathcal{L}}_\beta(x^k, y^k, \lambda^k; 1, \cdot) - \widehat{\mathcal{L}}_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}; 1, \cdot) \\
&\quad = (\widehat{\mathcal{L}}_\beta(x^k, y^k, \lambda^k; 1, \cdot) - \widehat{\mathcal{L}}_\beta(x^{k+1}, y^k, \lambda^k; 1, \cdot)) + (\widehat{\mathcal{L}}_\beta(x^{k+1}, y^k, \lambda^k; 1, \cdot) - \\
&\qquad \widehat{\mathcal{L}}_\beta(x^{k+1}, y^k, \lambda^k; \alpha, y^k)) + (\widehat{\mathcal{L}}_\beta(x^{k+1}, y^k, \lambda^k; \alpha, y^k) - \widehat{\mathcal{L}}_\beta(x^{k+1}, y^{k+1}, \lambda^k; \alpha, y^k)) + \\
&\qquad (\widehat{\mathcal{L}}_\beta(x^{k+1}, y^{k+1}, \lambda^k; \alpha, y^k) - \widehat{\mathcal{L}}_\beta(x^{k+1}, y^{k+1}, \lambda^{k+1}; 1, \cdot)). \qquad (3.15)
\end{aligned}
$$

Thus, substituting (3.7), (3.9) and (3.14) into (3.15), we obtain

$$
\begin{aligned}
&\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^{k+1}) \ge (\frac{\beta\mu - L_3(y^k) - \omega}{2} - \frac{2M^2}{\alpha\beta})\|x^{k+1} - x^k\|^2 + \\
&\quad (\frac{\beta - L_1 - L_2(x^{k+1})}{2} - \frac{2L_1^2 + 2M^2}{\alpha\beta} + \frac{(1-\alpha)\beta}{\alpha})\|y^{k+1} - y^k\|^2 \\
&\ge (\frac{\beta\mu - L_3 - \omega}{2} - \frac{2M^2}{\alpha\beta})\|x^{k+1} - x^k\|^2 + (\frac{\beta - L_1 - L_2}{2} - \frac{2L_1^2 + 2M^2}{\alpha\beta} + \frac{(1-\alpha)\beta}{\alpha})\|y^{k+1} - y^k\|^2 \\
&\ge \delta\|v^{k+1} - v^k\|^2,
\end{aligned}
$$

where the second inequality follows from (v) of Assumption 3.1 and the last inequality follows from (vi) of Assumption 3.1. The proof is completed. □

Lemma 3.3 *Let $\{\omega^k\}_{k\in N}$ be the sequence generated by the GADMM (1.5) which is assumed to be bounded. Then the following holds:*

$$
\sum_{k=0}^{+\infty} \|\omega^{k+1} - \omega^k\|^2 < +\infty. \qquad (3.16)
$$

Proof Since $\{\omega^k\}_{k\in N}$ is bounded, there exists a subsequence $\{\omega^{k_j}\}_{j\in N}$ such that $\omega^{k_j} \to \omega^*$. Firstly, we know $\mathcal{L}_\beta(\cdot)$ is lower semicontinuous due to the continuity of g and H and the closedness of f, that means

$$
\mathcal{L}_\beta(\omega^*) \le \liminf_{j\to\infty} \mathcal{L}_\beta(\omega^{k_j}).
$$

Consequently, $\{\mathcal{L}_\beta(\omega^{k_j})\}_{j\in N}$ is bounded from below. Note that, (3.5) implies that $\{\mathcal{L}_\beta(\omega^k)\}_{k\in N}$

is nonincreasing and thus $\{\mathcal{L}_\beta(\omega^{k_j})\}_{j\in N}$ is convergent. Moreover, we have $\{\mathcal{L}_\beta(\omega^k)\}_{k\in N}$ is convergent and $\mathcal{L}_\beta(\omega^k)\ge\mathcal{L}_\beta(\omega^*)$. Rearranging terms of (3.5) leads to

$$\delta\|v^{k+1}-v^k\|^2\le\mathcal{L}_\beta(\omega^k)-\mathcal{L}_\beta(\omega^{k+1}).$$

Adding the above inequality from $k=0$ to $k=m$

$$\sum_{k=0}^{m}\delta\|v^{k+1}-v^k\|^2\le\mathcal{L}_\beta(\omega^0)-\mathcal{L}_\beta(\omega^{m+1})\le\mathcal{L}_\beta(\omega^0)-\mathcal{L}_\beta(\omega^*).$$

Thus, letting $m\to+\infty$, we get

$$\sum_{k=0}^{+\infty}\delta\|v^{k+1}-v^k\|^2\le\mathcal{L}_\beta(\omega^0)-\mathcal{L}_\beta(\omega^*)<+\infty.$$

In view of $\delta>0$, the above inequality yields

$$\sum_{k=0}^{+\infty}\|v^{k+1}-v^k\|^2\le+\infty.$$

Hence, we obtain

$$\sum_{k=0}^{+\infty}\|x^{k+1}-x^k\|^2<+\infty,\quad \sum_{k=0}^{+\infty}\|y^{k+1}-y^k\|^2<+\infty. \tag{3.17}$$

Moreover, it follows from (3.13) and (3.17) that

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

Therefore,

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

The proof is completed. □

Lemma 3.4 *Let $\{\omega^k\}_{k\in N}$ be the sequence generated by the GADMM (1.5) which is assumed to be bounded. Then there exists $\eta>0$ such that*

$$d(0,\partial\mathcal{L}_\beta(\omega^{k+1}))\le\eta\|v^{k+1}-v^k\|. \tag{3.18}$$

Proof By the definition of the augmented Lagrangian function $\mathcal{L}_\beta(\cdot)$ and (iv) of Remark 2.2, we have

$$\begin{cases}\partial_x\mathcal{L}_\beta(\omega^{k+1})=\partial f(x^{k+1})+\nabla_x H(x^{k+1},y^{k+1})-A^T\lambda^{k+1}+\beta A^T(Ax^{k+1}+y^{k+1}-b),\\ \partial_y\mathcal{L}_\beta(\omega^{k+1})=\nabla g(y^{k+1})+\nabla_y H(x^{k+1},y^{k+1})-\lambda^{k+1}+\beta(Ax^{k+1}+y^{k+1}-b),\\ \partial_\lambda\mathcal{L}_\beta(\omega^{k+1})=-(Ax^{k+1}+y^{k+1}-b).\end{cases} \tag{3.19}$$

Substituting (3.2) into (3.19) leads to

$$\begin{cases}A^T(\lambda^k-\lambda^{k+1})+\beta A^T(y^{k+1}-y^k)+\nabla_x H(x^{k+1},y^{k+1})-\nabla_x H(x^{k+1},y^k)\in\partial_x\mathcal{L}_\beta(\omega^{k+1}),\\ \frac{1}{\alpha}(\lambda^k-\lambda^{k+1})+\frac{(1-\alpha)\beta}{\alpha}(y^k-y^{k+1})\in\partial_y\mathcal{L}_\beta(\omega^{k+1}),\\ \frac{1}{\alpha\beta}(\lambda^{k+1}-\lambda^k)+\frac{1-\alpha}{\alpha}(y^{k+1}-y^k)\in\partial_\lambda\mathcal{L}_\beta(\omega^{k+1}).\end{cases} \tag{3.20}$$

Thus, if we set

$$
\begin{aligned}
(\xi_1^{k+1}, \xi_2^{k+1}, \xi_3^{k+1}) :=&(A^T(\lambda^k - \lambda^{k+1}) + \beta A^T(y^{k+1} - y^k) + \\
&\nabla_x H(x^{k+1}, y^{k+1}) - \nabla_x H(x^{k+1}, y^k), \\
\frac{1}{\alpha}(\lambda^k - \lambda^{k+1}) + \frac{(1-\alpha)\beta}{\alpha}&(y^k - y^{k+1}), \frac{1}{\alpha\beta}(\lambda^{k+1} - \lambda^k) + \frac{1-\alpha}{\alpha}(y^{k+1} - y^k)),
\end{aligned}
$$

then it follows from Lemma 2.3 that $(\xi_1^{k+1}, \xi_2^{k+1}, \xi_3^{k+1}) \in \partial \mathcal{L}_\beta(\omega^{k+1})$. Moreover, there exist $\eta_1, \eta_2, \eta_3 > 0$ such that

$$
\begin{aligned}
&\|(\xi_1^{k+1}, \xi_2^{k+1}, \xi_3^{k+1})\| \\
&\le \eta_1 \|y^{k+1} - y^k\| + \eta_2 \|\lambda^{k+1} - \lambda^k\| + \eta_3 \|\nabla_x H(x^{k+1}, y^{k+1}) - \nabla_x H(x^{k+1}, y^k)\|. \qquad (3.21)
\end{aligned}
$$

Since ∇H is Lipschitz continuous on bounded subsets and $\{(x^k, y^k)\}_{k\in N}$ is bounded, by (iii) of Assumption 3.1 there exists $M > 0$ such that

$$
\|\nabla_x H(x^{k+1}, y^{k+1}) - \nabla_x H(x^{k+1}, y^k)\| \le M \|y^{k+1} - y^k\|. \qquad (3.22)
$$

Notice that, we can deduce from (3.13) that

$$
\|\lambda^{k+1} - \lambda^k\| \le \sqrt{2L_1^2 + 2M^2} \cdot \|y^{k+1} - y^k\| + \sqrt{2}M \cdot \|x^{k+1} - x^k\|. \qquad (3.23)
$$

By setting $\eta := \sqrt{(\eta_1 + \sqrt{2L_1^2 + 2M^2}\eta_2 + M\eta_3)^2 + (\sqrt{2}M\eta_2)^2}$, it follows from (3.21), (3.22) and (3.23) that

$$
\begin{aligned}
d(0, \partial \mathcal{L}_\beta(\omega^{k+1})) &\le \|(\xi_1^{k+1}, \xi_2^{k+1}, \xi_3^{k+1})\| \\
&\le (\eta_1 + \sqrt{2L_1^2 + 2M^2}\eta_2 + M\eta_3) \cdot \|y^{k+1} - y^k\| + \sqrt{2}M\eta_2 \cdot \|x^{k+1} - x^k\| \\
&\le \eta \|v^{k+1} - v^k\|
\end{aligned}
$$

where the third inequality follows from the Cauchy inequality. The proof is completed. □

Let $\{\omega^k\}_{k\in N}$ be the sequence generated by the GADMM (1.5) from a starting point ω^0. The set of all limit points is denoted by $S(\omega^0)$, i.e., $S(\omega^0) := \{\omega^* : \exists$ a subsequence $\{\omega^{k_j}\}_{j\in N}$ of $\{\omega^k\}_{k\in N}$ converges to $\omega^*\}$. In the following, we summarize several properties of the limit point set.

Lemma 3.5 *Let $\{\omega^k\}_{k\in N}$ be the sequence generated by the GADMM (1.5) which is assumed to be bounded. Let $S(\omega^0)$ denote the set of its limit points. Then*

(i) $S(\omega^0)$ is a nonempty compact set, and $d(\omega^k, S(\omega^0)) \to 0$, as $k \to +\infty$;

(ii) $S(\omega^0) \subset$ crit $\mathcal{L}_\beta$;

(iii) $\mathcal{L}_\beta(\cdot)$ is finite and constant on $S(\omega^0)$, equal to

$$
\inf_{k\in N} \mathcal{L}_\beta(\omega^k) = \lim_{k\to+\infty} \mathcal{L}_\beta(\omega^k).
$$

Proof We prove the results item by item.

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

(ii) For any fixed $(x^*, y^*, \lambda^*) \in S(\omega^0)$, there exists a subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}_{j\in N}$ converging to (x^*, y^*, λ^*). By the definition of the augmented Lagrangian function $\mathcal{L}_\beta(\cdot)$ (3.4), the

x-subproblem of (1.5) is equivalent to

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

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

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

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

$$\limsup_{j\to+\infty} \mathcal{L}_\beta(x^{k_j+1}, y^{k_j}, \lambda^{k_j}) = \limsup_{j\to+\infty} \mathcal{L}_\beta(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1}) \le \mathcal{L}_\beta(x^*, y^*, \lambda^*). \tag{3.24}$$

On the other hand, (3.16) implies $\|\omega^{k+1} - \omega^k\| \to 0$, which means that the subsequence

$$\{(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1})\}_{j\in N}$$

also converges to (x^*, y^*, λ^*). From the lower semicontinuity of $\mathcal{L}_\beta(\cdot)$, we have

$$\liminf_{j\to+\infty} \mathcal{L}_\beta(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1}) \ge \mathcal{L}_\beta(x^*, y^*, \lambda^*). \tag{3.25}$$

Then by combining (3.24) and (3.25) together we can get

$$\lim_{j\to+\infty} \mathcal{L}_\beta(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1}) = \mathcal{L}_\beta(x^*, y^*, \lambda^*),$$

which implies

$$\lim_{j\to+\infty} f(x^{k_j+1}) = f(x^*). \tag{3.26}$$

Passing to the limit in (3.2) along the subsequence $\{(x^{k_j+1}, y^{k_j+1}, \lambda^{k_j+1})\}_{j\in N}$ and invoking (3.26) and the continuity of $\nabla g, \nabla_x H(\cdot,\cdot), \nabla_y H(\cdot,\cdot)$, it follows that

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

Thus, $(x^*, y^*, \lambda^*) \in \text{crit}\ \mathcal{L}_\beta$.

(iii) For any point $(x^*, y^*, \lambda^*) \in S(\omega^0)$, there exists a subsequence $\{(x^{k_j}, y^{k_j}, \lambda^{k_j})\}_{j\in N}$ converging to (x^*, y^*, λ^*). Since $\mathcal{L}_\beta(\omega^k)$ is nonincreasing, combining (3.24) and (3.25) together we can get

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

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

$$\inf_{k\in N} \mathcal{L}_\beta(\omega^k) = \lim_{k\to+\infty} \mathcal{L}_\beta(\omega^k).$$

The proof is completed. □

Theorem 3.6 *Let $\{\omega^k\}_{k\in N}$ be the sequence generated by the GADMM (1.5) which is assumed to be bounded. Suppose that $\mathcal{L}_\beta(\cdot)$ is a KL function. Then $\{\omega^k\}_{k\in N}$ has finite length, that is*

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

and as a consequence, $\{\omega^k\}_{k\in N}$ *converges to a critical point of* $\mathcal{L}_\beta(\cdot)$.

Proof Since from the proof of Lemma 3.5, it follows that $\mathcal{L}_\beta(\omega^k) \to \mathcal{L}_\beta(\omega^*)$ for all $\omega^* \in S(\omega^0)$. We consider two cases.

Case 1 If there exists an integer k_0 for which $\mathcal{L}_\beta(\omega^{k_0}) = \mathcal{L}_\beta(\omega^*)$. Rearranging the terms of (3.5), we have that for any $k > k_0$,

$$\delta\|v^{k+1} - v^k\|^2 \le \mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^{k+1}) \le \mathcal{L}_\beta(\omega^{k_0}) - \mathcal{L}_\beta(\omega^*) = 0,$$

which implies that $v^{k+1} = v^k$ for any $k > k_0$. Associated with (3.13), for any $k > k_0 + 1$, it follows that $\omega^{k+1} = \omega^k$ and the assertion holds.

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

$$\delta\|v^{k+1} - v^k\|^2 \le \eta \cdot \|v^k - v^{k-1}\| \cdot \Delta_{k,k+1}, \tag{3.27}$$

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

$$d(\omega^k, S(\omega^0)) < \epsilon, \quad \mathcal{L}_\beta(\omega^*) < \mathcal{L}_\beta(\omega^k) < \mathcal{L}_\beta(\omega^*) + \kappa.$$

Since $S(\omega^0)$ is nonempty compact set and $\mathcal{L}_\beta(\cdot)$ is constant on $S(\omega^0)$, applying Lemma 2.9 with $\Omega := S(\omega^0)$, we deduce that for all $k > \tilde{k}$

$$\varphi'(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*)) d(0, \partial\mathcal{L}_\beta(\omega^k)) \ge 1.$$

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

$$\varphi(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*)) - \varphi(\mathcal{L}_\beta(\omega^{k+1}) - \mathcal{L}_\beta(\omega^*)) \ge \varphi'(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*))(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^{k+1})).$$

Thus, using the inequalities $d(0, \partial\mathcal{L}_\beta(\omega^k)) \le \eta\|v^{k+1} - v^k\|$ and $\varphi'(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*)) > 0$, we obtain

$$\begin{aligned}
&\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^{k+1}) \\
&\le \frac{\varphi(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*)) - \varphi(\mathcal{L}_\beta(\omega^{k+1}) - \mathcal{L}_\beta(\omega^*))}{\varphi'(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*))} \\
&\le \eta\|v^k - v^{k-1}\|[\varphi(\mathcal{L}_\beta(\omega^k) - \mathcal{L}_\beta(\omega^*)) - \varphi(\mathcal{L}_\beta(\omega^{k+1}) - \mathcal{L}_\beta(\omega^*))].
\end{aligned}$$

Combining Lemma 3.2 with the above relation gives (3.27) as desired. Moreover, (3.27) implies

$$\|v^{k+1} - v^k\| \le \sqrt{\frac{\eta}{\delta}\Delta_{k,k+1}} \cdot \|v^k - v^{k-1}\|^{\frac{1}{2}}.$$

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

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

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

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

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

$$\sum_{k=\tilde{k}+1}^{+\infty}\|v^{k+1}-v^k\|\le\|v^{\tilde{k}+1}-v^{\tilde{k}}\|+\frac{\eta}{\delta}\varphi(\mathcal{L}_\beta(\omega^{\tilde{k}+1})-\mathcal{L}_\beta(\omega^*)), \tag{3.28}$$

which implies $\sum_{k=0}^{+\infty}\|v^{k+1}-v^k\|<+\infty$. Thus, it follows that

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

From these together with (3.23), we obtain

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

Moreover, note that

$$\begin{aligned}\|\omega^{k+1}-\omega^k\|&=(\|x^{k+1}-x^k\|^2+\|y^{k+1}-y^k\|^2+\|\lambda^{k+1}-\lambda^k\|^2)^{\frac{1}{2}}\\&\le\|x^{k+1}-x^k\|+\|y^{k+1}-y^k\|+\|\lambda^{k+1}-\lambda^k\|.\end{aligned}$$

Therefore,

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

and $\{\omega^k\}_{k\in N}$ is a Cauchy sequence which converges. The proof is completed. $\square$

Next, we give some sufficient conditions to guarantee the sequence $\{\omega^k\}_{k\in N}$ generated by the GADMM (1.5) is bounded.

Lemma 3.7 *Let $\{\omega^k\}_{k\in N}$ be the sequence generated by the GADMM (1.5). Suppose that ∇H is global Lipschitz continuous and*

$$H_g:=\inf_{x\in\mathcal{R}^n,y\in\mathcal{R}^m}\Big\{g(y)+H(x,y)-\frac{1}{2\widetilde{\beta}}\|\nabla g(y)+\nabla_y H(x,y)\|^2\Big\}>-\infty,$$

where $\widetilde{\beta}$ is defined in (3.3). If $\liminf_{\|x\|\to+\infty}f(x)=+\infty$, then $\{\omega^k\}_{k\in N}$ is bounded.

Proof Since ∇H is global Lipschitz continuous, following the same proof line of Lemma 3.2, we can show that

$$\mathcal{L}_\beta(x^{k+1},y^{k+1},\lambda^{k+1})\le\mathcal{L}_\beta(x^k,y^k,\lambda^k).$$

Then, combining with $\lambda^k=\nabla g(y^k)+\nabla_y H(x^k,y^k)$, we get

$$\begin{aligned}&\mathcal{L}_\beta(x^1,y^1,\lambda^1)\\&\quad\ge f(x^k)+g(y^k)+H(x^k,y^k)-\langle\lambda^k,Ax^k+y^k-b\rangle+\frac{\beta}{2}\|Ax^k+y^k-b\|^2\\&\quad=f(x^k)+g(y^k)+H(x^k,y^k)-\frac{1}{2\beta}\|\lambda^k\|^2+\frac{\beta}{2}\|Ax^k+y^k-b-\frac{1}{\beta}\lambda^k\|^2\\&\quad=f(x^k)+(g(y^k)+H(x^k,y^k)-\frac{1}{2\widetilde{\beta}}\|\lambda^k\|^2)+(\frac{1}{2\widetilde{\beta}}-\frac{1}{2\beta})\|\lambda^k\|^2+\frac{\beta}{2}\|Ax^k+y^k-b-\frac{1}{\beta}\lambda^k\|^2\\&\quad\ge f(x^k)+H_g+(\frac{1}{2\widetilde{\beta}}-\frac{1}{2\beta})\|\lambda^k\|^2+\frac{\beta}{2}\|Ax^k+y^k-b-\frac{1}{\beta}\lambda^k\|^2.\end{aligned}$$

Note that, $\liminf_{\|x\|\to+\infty} f(x) = +\infty$ implies that $\inf_x f(x) > -\infty$. By means of this and $\beta > \widetilde{\beta}$, we deduce that the sequences $\{x^k\}_{k\in N}$, $\{\lambda^k\}_{k\in N}$ and $\{\frac{\beta}{2}\|Ax^k + y^k - b - \frac{1}{\beta}\lambda^k\|^2\}_{k\in N}$ are bounded. Therefore, $\{y^k\}_{k\in N}$ is also bounded and hence $\{\omega^k\}$ is bounded. The proof is completed. □

Theorem 3.8 *Let $\{\omega^k\}_{k\in N}$ be the sequence generated by the GADMM (1.5) that converges to $\{\omega^* := (x^*, y^*, \lambda^*)\}$. Assume that $\mathcal{L}_\beta(\cdot)$ has the KL property at (x^*, y^*, λ^*) with $\varphi(s) := cs^{1-\theta}, \theta \in [0, 1), c > 0$. Then the following estimations hold:*

(i) If $\theta = 0$, then the sequence converges in a finite number of steps;

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

$$\|(x^k, y^k, \lambda^k) - (x^*, y^*, \lambda^*)\| \le c\tau^k;$$

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

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

Proof For the case that $\theta = 0$, we have $\varphi(s) = cs$ and $\varphi'(s) = c$. If $\{\omega^k\}_{k\in N}$ does not converge in a finite number of steps, then the KL property at (x^*, y^*, λ^*) yields for any k sufficiently large, $c \cdot d(0, \partial\mathcal{L}_\beta(\omega^k)) \ge 1$, a contradiction to (3.18).

Now, suppose that $\theta > 0$ and set

$$\Delta_k := \sum_{i=k}^{+\infty} \|v^{i+1} - v^i\|, \quad k \ge 0.$$

The triangle inequality yields $\Delta_k \ge \|v^k - v^*\|$, and it is therefore sufficient to estimate Δ_k. With these notations, it follows from (3.28) that

$$\Delta_{\tilde{k}+1} \le (\Delta_{\tilde{k}} - \Delta_{\tilde{k}+1}) + \frac{\eta}{\delta}\varphi(\mathcal{L}_\beta(\omega^{\tilde{k}+1}) - \mathcal{L}_\beta(\omega^*)).$$

Because $\mathcal{L}_\beta$ has the KL property at (x^*, y^*, λ^*), we have

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

Due to $\varphi(s) = cs^{1-\theta}$, the above inequality is equivalent to

$$(\mathcal{L}_\beta(\omega^{\tilde{k}+1}) - \mathcal{L}_\beta(\omega^*))^\theta \le c \cdot (1-\theta)d(0, \partial\mathcal{L}_\beta(\omega^{\tilde{k}+1})). \tag{3.29}$$

By means of (3.18), we can get

$$d(0, \partial\mathcal{L}_\beta(\omega^{\tilde{k}+1})) \le \eta\|v^{\tilde{k}+1} - v^{\tilde{k}}\| = \eta(\Delta_{\tilde{k}} - \Delta_{\tilde{k}+1}). \tag{3.30}$$

Combining (3.29) and (3.30), we obtain that there exists $\gamma > 0$ such that

$$\varphi(\mathcal{L}_\beta(\omega^{\tilde{k}+1}) - \mathcal{L}_\beta(\omega^*)) = c \cdot (\mathcal{L}_\beta(\omega^{\tilde{k}+1}) - \mathcal{L}_\beta(\omega^*))^{1-\theta} \le \gamma(\Delta_{\tilde{k}} - \Delta_{\tilde{k}+1})^{\frac{1-\theta}{\theta}},$$

and hence

$$\Delta_{\tilde{k}+1} \le (\Delta_{\tilde{k}} - \Delta_{\tilde{k}+1}) + \frac{\eta}{\delta}\gamma(\Delta_{\tilde{k}} - \Delta_{\tilde{k}+1})^{\frac{1-\theta}{\theta}}.$$

Sequences satisfying such inequalities have been studied in Attouch and Bolte [32]. It follows that

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

$$\|v^k - v^*\| \le c_1 \tau^k; \tag{3.31}$$

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

$$\|v^k - v^*\| \le c_2 k^{\frac{\theta-1}{2\theta-1}}, \tag{3.32}$$

which implies that

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

$$\|x^k - x^*\| \le c_1 \tau^k, \quad \|y^k - y^*\| \le c_1 \tau^k; \tag{3.33}$$

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

$$\|x^k - x^*\| \le c_2 k^{\frac{\theta-1}{2\theta-1}}, \quad \|y^k - y^*\| \le c_2 k^{\frac{\theta-1}{2\theta-1}}. \tag{3.34}$$

Note that

$$\begin{aligned} \|\lambda^k - \lambda^*\| &= \|\nabla g(y^k) + \nabla_y H(x^k, y^k) - \nabla g(y^*) - \nabla_y H(x^*, y^*)\| \\ &\le \|\nabla g(y^k) - \nabla g(y^*)\| + \|\nabla_y H(x^k, y^k) - \nabla_y H(x^*, y^*)\| \\ &\le M\|x^k - x^*\| + (L_1 + M)\|y^k - y^*\|, \end{aligned} \tag{3.35}$$

where the inequality follows from the Lipschitz continuity of ∇g and (iii) of Assumption 3.1. Substituting (3.33) and (3.34) into (3.35), we get that

(i) If $\theta \in (0, \frac{1}{2}]$, then there exist $c_3 := c_1(L_1 + 2M)$ and $\tau \in [0, 1)$, such that

$$\|\lambda^k - \lambda^*\| \le c_3 \tau^k; \tag{3.36}$$

(ii) If $\theta \in (\frac{1}{2}, 1)$, then there exists $c_4 := c_2(L_1 + 2M)$, such that

$$\|\lambda^k - \lambda^*\| \le c_4 k^{\frac{\theta-1}{2\theta-1}}. \tag{3.37}$$

Combining (3.36) and (3.37), we get the desired inequalities immediately from (3.31) and (3.32). The proof is completed. □

4. Conclusions

In this paper, we analyzed the convergence of the generalized alternating direction method of multipliers (GADMM) for solving linearly constrained nonconvex minimization problem whose objective contains coupled functions. Under the assumption that the augmented Lagrangian function satisfies the Kurdyka-Łojasiewicz inequality, we proved that the iterate sequence generated by the GADMM converges to a critical point of the augmented Lagrangian function, provided that the penalty parameter in the augmented Lagrangian function is larger than a threshold. Under some further conditions on the problem's data, the convergence rate of the algorithm was also established.

References

[1] R. GLOWINSKI, A. MARROCCO. *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 Fr. Autom. Inform. Rech. Opér. Anal. Numér., 1975, **2**: 41–76.

[2] D. BOLEY. *Local linear convergence of ADMM on quadratic or linear programs*. SIAM J. Optim., 2013, **23**: 2183–2207.

[3] Deren HAN, Xiaoming YUAN. *Local linear convergence of the alternating direction method of multipliers for quadratic programs*. SIAM J. Numer. Anal., 2013, **51**(6): 3446–3457.

[4] Bingsheng HE, Xiaoming YUAN. *On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method*. SIAM J. Numer. Anal., 2012, **50**(2): 700–709.

[5] Weihong YANG, Deren HAN. *Linear convergence of alternating direction method of multipliers for a class of convex optimization problems*. SIAM J. Numer. Anal., 2016, **54**(2): 625–640.

[6] Ke GUO, Deren HAN, D. Z. W. WANG, et al. *Convergence of ADMM for multi-block nonconvex separable optimization models*. Front. Math. China., 2017, **12**(5): 1139–1162.

[7] Ke GUO, Deren HAN, Xiaoming YUAN. *Convergence analysis of Douglas-Rachford splitting method for "strongly+weakly" convex programming*. SIAM J. Numer. Anal., 2017, **55**(4): 1549–1577.

[8] Ke GUO, Deren HAN, Tingting WU. *Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints*. Int. J. Comput. Math., 2017, **94**(8): 1653–1669.

[9] Guoyin LI, T. K. PONG. *Global convergence of splitting methods for nonconvex composite optimization*. SIAM J. Optim., 2015, **25**(4): 2434–2460.

[10] Fenghui WANG, Zongben XU, Hongkun XU. *Convergence of Bregman alternating direction method with multipliers for nonconvex composite problems*. arXiv:1410.8625v3.

[11] Xiang GAO, Shuzhong ZHANG. *First-order algorithms for convex optimization with nonseparable objective and coupled constraints*. J. Oper. Res. Soc. China, 2017, **5**(2): 131–159.

[12] Caihua CHEN, Min LI, Xin LIU, et al. *Extended ADMM and BCD for nonseparable convex minimization models with quadratic coupling terms: convergence analysis and insights*. Math. Program., 2017, doi: 10.1007/s10107-017-1205-9.

[13] Ke GUO, Deren HAN, Tingting WU. *Convergence analysis for optimization problems with nonseparable nonconvex objective and linear constraints*. Pac. J. Optim., Accepted.

[14] H. ATTOUCH, J. BOLTE, P. REDONT, et al. *Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality*. Math. Oper. Res, 2010, **35**(2): 438–457.

[15] H. ATTOUCH, J. BOLTE, B. F. SVAITER. *Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods*. Math. Program., 2013, **137**(1-2): 91–129.

[16] J. BOLTE, A. DANIILIDIS, A. LEWIS. *The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems*. SIAM J. Optim., 2006, **17**(4): 1205–1223.

[17] J. BOLTE, A. DANIILIDIS, A. LEWIS, et al. *Clarke subgradients of stratifiable functions*. SIAM J. Optim., 2007, **18**(2): 556–572.

[18] D. GABAY. *Applications of the Method of Multipliers to Variational Inequalities*. North-Holland, Amsterdam, 1983.

[19] P. L. LIONS, B. MERCIER. *Splitting algorithms for the sum of two nonlinear operators*. SIAM J. Numer. Anal., 1979, **16**(6): 964–979.

[20] J. ECKSTEIN, D. BERTSEKAS. *On the Douglas-Rachford splitting method and proximal point algorithm for maximal monotone operators*. Math. Programming, 1992, **55**(3): 293–318.

[21] B. MARTINET. *Regularisation d'inéquations variations par approximations succesives*. Rev. Francaise d'Inform. Recherche Oper., 1970, **4**: 154–159.

[22] E. G. GOL'SHTEIN, N. V. TRET'YAKOV. *Modified Lagrangian in convex programming and their generalizations*. Math. Programming Stud., 1979, **10**: 86–97.

[23] D. P. BERTSEKAS. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York, 1982.

[24] Xingju CAI, Yannan CHEN, Deren HAN. *Nonnegative tensor factorizations using an alternating direction method*. Front. Math. China., 2013, **8**(1): 3–18.

[25] J. ECKSTEIN. *Parallel alternating direction multiplier decomposition of convex programs*. J. Optim. Theory Appl., 1994, **80**(1): 39–62.

[26] J. BOLTE, A. DANIILIDIS, O. LEY, et al. *Characterizations of Łojasiewicz inequalities: Subgradient flows, talweg, convexity*. Trans. Amer. Math. Soc., 2010, **362**(6): 3319–3363.

[27] M. DEGIOVANNI, A. MARINO, M. TOSQUES. *Evolution equations with lack of convexity*. Nonlinear Anal., 1985, **9**(12): 1401–1443.

[28] S. MARCELLIN, L. THIBAULT. *Evolution problems associated with primal lower nice functions*. J. Convex Anal., 2006, **13**(2): 385–421.

[29] Xianfu WANG. *On Chebyshev functions and Klee functions*. J. Math. Anal. Appl., 2010, **368**(1): 293–310.

[30] J. BOLTE, S. SABACH, M. TEBOULLE. *Proximal alternating linearized minimization for nonconvex and nonsmooth problem*. Math. Program., 2014, **146**(1-2): 459–494.

[31] Y. NESTEROV. *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, Boston, 2004.

[32] H. ATTOUCH, J. BOLTE. *On the convergence of the proximal algorithm for nonsmooth functions involving analytic features*. Math. Program., 2009, **116**(1-2): 5–16.